# Learning more from complex psychological and social interventions in mental health

A thesis submitted to the University of Manchester

for the degree of Doctor of Philosophy

in the Faculty of Medical and Human Sciences

2014

Clare A M Flach

School of Medicine

University of Manchester, UK

**Table of Contents**

Word count main body text: 64,707

**List of Tables**

**List of Figures**

University of Manchester

Abstract of thesis submitted by Clare Flach

For the degree of Doctor of Philosophy (PhD)

Learning more from complex psychological and social interventions in mental health

2014

Complex interventions by definition consist of many parts. Once it has been established that a complex intervention is effective the next step is to determine how it is effective and what the active ingredients are. In a randomised controlled trial the causal effect of the intervention can be determined simply but when the mechanism of the intervention is investigated the exposure is no longer randomised. Instrumental variable methods have been developed to overcome problems of unmeasured confounding which result from the loss of randomisation but they require additional assumptions.

When applying instrumental variable techniques in real data sets with finite samples the identification of effective instruments and the use of weak instruments can cause bias in estimation. This thesis compares methods for the identification of instruments and estimation in the presence of many weak instruments. Shrinkage techniques, the LASSO (Least Absolute Shrinkage and Selection Operator) and Elastic Net, utilised in data mining are applied to the context of instrumental variable selection and compared to a single instrument, all instruments and backward stepwise selection. The commonly used two stage least squares estimator is compared to the limited information maximum likelihood estimator with and without Fuller's adjustment in the presence of many weak instruments. The selection and estimation methods are compared in simulated data replicating the design of a clinical trial of a complex intervention with varying levels of instrument strength and number.

The simulation results indicate that when there are multiple true instruments using multiple instruments is preferred to a single instrument. The benefit of multiple instruments increases as the individual instruments become weaker. Selection by the LASSO increases the first stage F-statistic and reduces bias but precision can suffer in the more parsimonious models. Estimation by two-stage least squares is preferred over limited information maximum likelihood in terms of the mean-squared error in the presence of many weak instruments but the maximum likelihood estimators perform better in terms of median bias. When the process variable is categorical the two-stage least squares is preferred in terms of bias and precision.

The statistical methods identified to be effective in the simulated data are applied to clinical trial datasets to answer substantive questions regarding the important components of cognitive behavioural therapy and to determine if the therapy works through the expected processes. The results indicate that formulation is a key component of CBT therapy for the prevention of psychosis and suggests that homework and active change strategies are also important. However due to the high correlation between these factors it is not possible to distinguish the importance of one aspect over another.

**Declaration**

I declare that no portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

**Copyright Statement**

i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the "Copyright") and she has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

iii. The ownership of certain Copyright, patents, designs, trade marks and other intellectual property (the "Intellectual Property") and any reproductions of copyright works in the thesis, for example graphs and tables ("Reproductions"), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://www.campus.manchester.ac.uk/medialibrary/policies/intellectual-property.pdf), in any relevant Thesis restriction declarations deposited in the University Library, The University Library's regulations (see http://www.manchester.ac.uk/library/aboutus/regulations) and in The University's policy on presentation of Theses

**Acknowledgements**

I have had a fantastic experience researching and writing this thesis. I have learnt a great deal and have had the opportunity to work with some incredible people. I would like to thank some of those people here.

My greatest thanks go to my supervisors Graham Dunn and Tony Morrison who have been brilliant throughout. They have offered their knowledge, experience and advice with such patience (and far more laughs than I ever expected in supervisory meetings) that the whole experience has been very enjoyable. I can only hope that some of their wisdom has rubbed off.

I would like to extend my gratitude to the rest of the biostatistics department, particularly to my advisor Chris Roberts, Richard Emsley and Wendy Lamb for their advice and help throughout my time in the team.

I thank the Medical Research Council and the Economic and Social Research Council, without funding from these bodies I would not have been able to carry out this work.

I have been fortunate to have made some great friends who have been an enormous help through my research, listening to my gripes, checking code and acting as my personal thesaurus. My massive thanks go to them, my fellow PhD sufferers and office buddies - Ian Jacob, Kim Hannam, Matt Gittins, Lesley-Anne Carter and Fiona Holland, without whom I'm not sure that I would have made it to the end and definitely wouldn't have had so much fun in the process. They have offered incredible support – tea, biscuits and a surprising number of statistics related jokes!

Finally I would like to thank my family and friends who have put up with me being a student ...again. Particularly to Mum and Dad who are a constant support and have always encouraged me to take on any new challenge and experience, a sentiment for which I am very grateful.

**List of Abbreviations**

2SLS - Two Stage Least Squares

ARMS – At Risk Mental State

ATE – Average Treatment Effect

ATT – Average Treatment Effect on the Treated

B&K – Baron and Kenny method

BAPS – Beliefs About Paranoia Scale

BCSS – Brief Core Schema Scale

BDI – Beck Depression Inventory

CAARMS – Comprehensive Assessment of At-Risk Mental State

CACE – Complier Average Causal Effect

CBT – Cognitive Behavioural Therapy

CI – Confidence Interval

Coef. - Coefficient

CTCH – Cognitive Therapy for Command Hallucinations

EDIE-II – Early Detection and Intervention Evaluation 2

EN – Elastic Net

EQ5D – European Quality of life- 5 dimensions

Fuller's – Fuller's adjustment to Limited Information Maximum Likelihood

GAF – Global Assessment of Functioning

GMM – Generalised Method of Moments

IPTW – Inverse Probability of Treatment Weighting

ITT – Intention To Treat

IV – Instrumental Variables

LASSO – Least Absolute Shrinkage and Selection Operator

LCGA – Latent Class Growth Analysis

LIML – Limited Information Maximum Likelihood

MANSA – Manchester Short Assessment of quality of life

MAR – Missing At Random

MCAR – Missing Completely At Random

MCQ – Meta-Cognitive Questionnaire

MSE – Mean Squared Error

NAE – Negative Appraisals of Experience

OLS – Ordinary Least Squares

PBEQ – Personal Beliefs about Experiences Questionnaire

PS – Principal Stratification

RCT – Randomised Controlled Trial

SAE – Social Acceptance of Experiences

SIAS – Social Interaction Anxiety Scale

SMM – Structural Mean Model

Std. Err. – Standard Error

SUTVA – Stable Unit Treatment Value Assumption

TAU – Treatment As Usual

UHR – Ultra High Risk

VPD – Voice Power Differential scale

# 1 Introduction and motivation

The introduction to this thesis begins with a brief background to the mental health disorder schizophrenia and the cognitive therapy treatment which motivates the substantive questions of this study. The paper then moves on to introduce the statistical concept of causation applied to mediation analysis and current statistical analysis techniques. Finally the overall aims and objectives of the remainder of the thesis are established.

## 1.1 Introduction to substantive question

### 1.1.1 What is psychosis?

Psychosis is a broad term for mental disorders which include the main and most recognised diagnosis, schizophrenia. Psychotic disorders are characterised by a loss of contact with reality and symptoms often include hallucinations and delusions, negative symptoms or behavioural disorganisation. Delusions are false beliefs, for example the patient may believe that they are someone they are not or that people are trying to harm them, and these beliefs affect the individual's behaviour and functioning. Hallucinations can affect any of the senses and manifest in the form of false visions, hearing, smell, touch or taste. Negative symptoms are characterised by a flattening effect on behaviour and mood; for example, an inability to experience pleasure, spontaneous speech or willpower. The clinical signs under the umbrella of negative symptoms include alogia, affective flattening, avolition and anhedonia. Behavioural disorganisation describes a disconnect between thought, affect and behaviour for example rambling disconnected speech, lack of logical thought or inappropriate behaviour. The clinical signs of behavioural disorganisation are termed positive formal thought disorder, inappropriate affect or bizarre behaviour.

There are two main classification systems for diagnosis of psychosis, the DSM[1] (Diagnostic and Statistical Manual of Mental Disorders) and the ICD[2] (International Classification of Disease). The DSM recently released the 5$^{th}$ edition of the manual, though the 4$^{th}$ edition was in use for the studies analysed in this thesis, the ICD is currently on the 10$^{th}$ edition (ICD-10). The DSM tends to be used more in North America and the ICD elsewhere. The criteria for diagnosis of schizophrenia are very similar under the two classification systems in terms of the symptoms that must be present; the main difference is that the DSM requires symptoms to have lasted for 6 months whereas the ICD is only for one month. This means that incidence and prevalence rates are lower with the DSM but prognosis is worse. The broad diagnosis of psychosis is split further into specific disorders;

schizophrenia - the main form of psychosis; schizophreniform – typically used as a preliminary diagnosis of schizophrenia; schizoaffective – a combination of schizophrenia and affective disorder (also known as mood disorder e.g. depression, anxiety); delusional disorder – which can be further defined as bizarre and non-bizarre delusions, non-bizarre delusions are those that could possibly occur in real life situations e.g. being followed or poisoned and bizarre delusions are those that are impossible for example being tracked by aliens; brief psychotic disorder is a short term disorder where symptoms have not lasted for more than a month; shared psychotic disorders are rare and describe the situation where an otherwise non-psychotic person takes on the delusional beliefs of a diagnosed psychotic person.

Psychosis onset is usually in the late teens or early 20s, and symptoms develop gradually over time though the speed of development varies. The incidence of schizophrenia is estimated at around 1.5-2.5 cases per 10,000 population per year[3,4], however due to the early onset and persistent or fluctuating symptoms prevalence can be much higher. The DSM-IV publication estimates the lifetime prevalence rate to be 0.5% - 1%[5] though a recent review of prevalence estimates suggests that it may be lower at around 4 per 1,000 (0.4%)[6]. The UK Office for National Statistics (ONS) in 1998 estimated a treated schizophrenia rate amongst registered general practice patients of around 2 per 1000[7]. We would expect this estimate to be lower than the overall prevalence since it is treated cases only. However, this suggests that less than half of cases were treated which is much lower than expected.

These incidence and prevalence studies have also indicated certain demographic risk factors associated with psychosis. Males tend to have a higher risk of psychosis and to have an earlier onset than females; peak onset in males is around 21-26 years compared to 25-32 years in females. Migrant populations are at higher risk of schizophrenia than their indigenous counterparts, and rates are also higher in urban compared to mixed urban/rural areas[4,6]. Other studies have found genetic/family history associations with psychosis as well as obstetric complications, early separation from parents[8] and more immediate risk factors such as drug use[9]. There is evidence of indicators of psychosis early in childhood, with birth cohort studies suggesting that those who develop schizophrenia are slower to reach developmental milestones, have lower educational test scores and lower levels of motor cognition, e.g. walking[8,10,11].

### 1.1.2 Diagnosing psychosis

Symptoms of psychosis include but are not restricted to: illusions, mood symptoms such as anxiety, depression and mood swings, cognitive symptoms, being overly distracted and having difficulty with concentration[12]. In the early stages symptoms intensify over time and may gradually become more noticeable to the patient or their family and friends. Symptoms must continue at a significant level of severity to affect functioning for a period of time before a diagnosis of psychosis is given[13,14]. This period before a diagnosis is made but when symptoms are appearing is now termed the pre-psychotic or prodromal period. At this stage the initial changes in character and behaviour occur but are not yet deemed severe enough to be fully psychotic. This leaves a grey area between what is considered different or strange behaviour and actual psychosis making a clear diagnosis difficult[14]. This prodromal period cannot be identified at the time but can only be determined in retrospect when the individual has developed psychosis.

There is evidence that although the path of psychosis for individuals is heterogeneous there is a tendency for a deterioration in symptoms over the initial period after diagnosis which stabilises after the first years and then may or may not improve again[15]. Research has indicated that it is the first stage that is important in determining future prognosis as the longer the duration of untreated psychosis (DUP) the worse outcomes are likely to be[15]. However, the evidence is not completely clear as there is some overlap with early-onset which has also been shown to result in worse outcomes[16] making it difficult to distinguish the individual impact of each.

Detection and treatment of those at high risk of developing psychosis has the potential to both delay onset of psychosis and reduce the time that patients are left untreated and so hopefully lead to improved outcomes. Steps have therefore been taken to identify people that are at risk of developing full psychosis prospectively and intervene at this early stage. Since this is a prospective diagnosis that does not necessarily mean the patient will become psychotic the terms ‘clinical high-risk’, ‘ultra-high risk’ or having an ‘at risk mental state’[13,14] are used instead of prodromal or pre-psychotic.

#### *1.1.2.1 Identifying individuals at high-risk*

Early studies following cohorts of participants thought to be at high risk of psychosis were based on defining high-risk purely by family history. However, although there is evidence of some genetic risk most diagnosed psychosis cases do not have a first degree relative

with psychosis. In addition the follow-up time for such a study is prolonged, making them unpopular. The more recent approach is to select cohorts based on factors that have been shown by prediction studies to put a person at increased risk of psychosis and to consider more immediate indicators so that the transition to psychosis is expected in a shorter space of time, making research studies more feasible[17]. Several studies have investigated those at risk of psychosis and developed measures to formally identify the group. The first of these came from Yung and colleagues[18] who studied young people accessing services at the PACE (Personal Assistance and Crisis Evaluation) and EPPIC (Early Psychosis Prevention and Intervention Centre) clinics both of which are based in Australia. The PACE clinic was specifically established to monitor young people at high risk of psychosis. People at high-risk were defined as having at least one of the following three criteria: a family history of psychosis and a recent change in mental state from DSM-III criteria for prodrome; at least one positive prodromal element of the DSM-III criteria; a history of psychotic episodes that recovered within a week (brief limited intermittent psychotic symptoms - BLIPS). This team then went on to develop the CAARMS – Comprehensive Assessment of At-Risk Mental States[19], the measure used in the EDIE-II trial, one of the example datasets in this study. This measure accounts separately for the frequency and severity of symptoms and so can be sensitive to small changes in experiences. Other measures include the Bonn Scale for Assessment of Basic Symptoms (BSABS), the Structured Interview for Prodromal Syndromes (SIPS)[20] and the Scale of Prodromal Symptoms (SOPS)[21]. SOPS is the tool of choice in North America and like the CAARMS is a semi-structured interview covering positive, negative and general symptoms.

The risk indicators used in these measures are based on the results of studies assessing predictors of psychosis and there is inevitably some uncertainty and debate around the importance and strength of these leading to different measures and definitions. It is important to realise that these tools are not providing an early diagnosis of psychosis but, they are highlighting people at higher risk of developing the disorder[22].

The main outcome in studies related to the at-risk population, be it validation of the at-risk measure or effectiveness of a treatment, is the transition rate to psychosis. A recent review by Ruhrmann et al.[22] has summarised the results of studies assessing the rate of transition in patients deemed at-risk. They found 17 studies assessing conversion rates with results ranging from 10% - 70%. The authors highlight the heterogeneity between studies in criteria for being at-risk, the follow-up times and the definition of transition to psychosis;

even studies using the same criteria implement it in different ways leading to large variation in transition rates. Included in the review are the results from Yung and colleagues, who first tried to define the at-risk group. They found in their first study a transition rate of 21% (of 33 patients) over the first 20 months with transition to full psychosis defined as having at least one of hallucinations, delusions or unusual thought by the DSM-III, BPRS or CASH measures[18]. The highest conversion rate reported in the Ruhrmann review was found in a study by Klosterkotter et al[23]. In this study an historical cohort of patients referred to psychiatric outpatient departments was used. After seven to ten years the overall rate of transition to psychosis was 49%, rising to 70% in those that had prodromal symptoms in their first examination. Since this was an historical cohort where individuals were re-contacted after a large time gap there was a great deal of loss to follow-up which may create bias in the transition rate estimates. Haroun et al.[24] reported on the Cognitive Assessment and Risk Evaluation (CARE) program, another small study with only 40 participants at 1-year follow-up, of which 15% had developed psychosis as defined by the SIPS. The EPOS (European Prediction of Psychosis Study) is a much larger study (N=245) designed to identify predictors of psychosis in an ultra high risk group and define more specific high risk groups. They found a psychosis incidence rate at 18 months follow-up of 19%[25]. The largest study to date recruited 370 high-risk participants for 30 months as part of the North American Prodrome Longitudinal Study. Transition to psychosis was diagnosed in 82 of the 291 participants with follow-up giving a rate of 28%[26]. The EDIE trial, which is the precursor to the EDIE-II trial analysed in this thesis, randomised 58 high-risk patients to cognitive therapy intervention or treatment as usual. A transition rate of 22% in the 6-12 month follow-up period was found in the 23 participants randomised to the treatment as usual group who did not receive cognitive therapy[27].

Yung et al[28] continued to monitor the rates of transition in the PACE clinic and suggest that the rate of transition is decreasing. It is difficult to know why this is the case. It could be because of more effective treatment or the inclusion of more patients that are not high-risk. However, the transition rates are low, with the majority of those deemed at high-risk never actually developing full psychosis. The treatment for these high-risk individuals is controversial as is the possibility of introducing an at risk category to the DSM since many will not become psychotic, there is no definitive treatment and a diagnosis can be stigmatising. Ruhrmann[22] argues that even though these at-risk patients are not and may

never become psychotic, they are ill and should receive treatment or help. In fact most studies are in help-seeking individuals.

With heterogeneity in the definition of transition and diagnoses between studies and the use of arbitrary cut-points in measures of diagnosis the usefulness of defining and studying transition rates is questionable. It may be of greater use to look at outcome measures related to the intentions of the therapy, such as emotional dysfunction or distress[29]. To this end the importance of service user involvement in research has now been recognised and a number of small qualitative studies have investigated the view of the patient in terms of recovery measures. These studies use psychotic rather than pre-psychotic samples but highlight the importance to the patient of a reduction in symptoms and an improvement in social functioning which could be incorporated as measures of success[30,31].

#### *1.1.2.2 Predictors of transition to psychosis*

The EPOS study in Europe[25], the CARE program in the US[24], the North American Prodromal Longitudinal study[26] and the PACE clinic in Australia[32] have all followed individuals deemed at high-risk of psychosis and investigated predictors of transition to full psychosis. Each of the studies has used different measurement methods and predictors and are therefore difficult to compare. However, they all suggest that greater levels of initial symptoms (the specific symptoms included vary) and lower functioning are strong predictors of transition. There is also a suggestion that the duration of symptoms, a family history of psychosis and substance abuse may be important factors.

### 1.1.3 Command hallucinations in psychosis

Auditory hallucinations (voices) are reported to be the most common symptom in schizophrenia with a prevalence estimated at over 60% in psychosis patients[33]. The effect of hearing voices on the individual can vary; some finding them stressful and distressing, whereas for others they can be a comfort. The effect is dependent on the type, content and appraisal of the voice. Shawyer et al[34] summarise eight studies that have assessed the prevalence of command hallucinations amongst those hearing voices and report a median rate of 53% (range 18-89%). The range of prevalence rate estimates is wide, demonstrating heterogeneity in the studies and the difficulty in estimation, namely that incidence of command hallucinations relies on self reports, which also introduces the possibility that prevalence is underestimated.

Command hallucinations are voices that tell the person to do something. These can range from harmless orders such as "make a cup of tea" to anti-social commands like shouting at someone to committing serious criminal offences. It has generally been assumed that individuals experiencing command hallucinations feel compelled to obey the commands and are, therefore, a danger to both themselves and others. However, the evidence for this is unclear. The rate of compliance with commands has been estimated in several studies, the compliance rates summarised by Braham et al[35] ranging from 40% to 88%. However, compliance itself is an ambiguous term since individuals will experience multiple commands, some of which they will comply with and others that they will resist. They may also appease the voice by carrying out a less severe act if they deem the consequences of the command to be too great. Several authors have stated that it is important to distinguish commands by their severity when considering compliance since people are less likely to comply with severe commands whereas mild commands will be obeyed regularly[36,37]. Other predictors of compliance have been summarised in reviews by Braham et al.[35], Barrowcliff and Haddock[38] and Shawyer et al.[34] The predictors established as being of importance are the perceived malevolence or benevolence of voice; the perceived power/omnipotence of voice; recognition of voice; and content of the command. Voices that are considered to be kind are more likely to be obeyed and malevolent voices resisted. If the individual believes that the voice is powerful and controlling and they are weak then they are more likely to carry out the commands. There is also evidence that if the individual recognises the voice and can identify it then they will be more likely to trust and therefore comply with the voice. Finally, the actual content of the command has a strong affect on whether it is obeyed. Compliance is low if the command is considered to be dangerous but mild commands will regularly be carried out.

Cognitive models of compliance with command hallucinations have drawn on Social Rank Theory[39]; the application of this theory to command hallucinations is described by Singer and Addington[40] and Braham et al[35]. Social Rank Theory comes from the hierarchies seen in the animal kingdom that stronger more skilful individuals elicit power over weaker individuals by intimidation or threat. The weaker individuals will obey the dominant individual or appease them when it would be dangerous to obey. This can be applied to hearing voices ordering actions from an individual that may or may not have negative consequences. Within this theoretical model it is clear that an individual's perceived power of the voice, as with the perceived power of any other individual they come in contact

with, will influence their compliance with the orders. This theory also supports the evidence of reaction to perceived malevolent and benevolent voices stated earlier. If a voice is deemed to be evil then its commands will be resisted or appeased whereas commands from voices that are deemed to be kind are more likely to be obeyed.

### 1.1.4 Psychological intervention

Treatments for mental health disorders often fall under either pharmacological or psychological interventions, though other alternative therapies are available. Antipsychotic medications are known to have severe adverse effects particularly extrapyramidal symptoms (movement disorders), loss of motivation, weight gain and sexual dysfunction[41]. Treatment with medication that can have severe side effects for people that may not become ill raises ethical questions regarding the principle of non-maleficence in medical practice[42,43]. Although most medications have some risk of side-effects the choice of treatment must be based on a balance of likely benefit to harm. Psychological interventions are advocated as an alternative treatment to medication as they do not suffer the same side effects. Efficacy of these interventions has therefore been a focus of recent studies[41]

Cognitive Behavioural Therapy (CBT) is a form of psychotherapy first designed by Beck in the 1960s for the treatment of depression and has since been developed and expanded to other illnesses. He gives a review of the development of CBT since then in his 2005 paper[44]. The therapy model was based on the theory that people read and react to situations or events in a biased way depending on previous experiences. For example, people with depression have negative perceptions of themselves and so their appraisal of and reactions to life events are negative, which in turn enhances the negative views of themselves. Patients with panic disorder have a belief that certain events are far more important than they are and so their interpretation of and reaction to the event is exaggerated. This can then build to the extent of inducing a panic attack and they are unable to realistically appraise the event. The therapy therefore seeks to identify how the patient thinks about themselves, the world and other people, how these thoughts affect their behaviour and vice versa so that unhelpful thoughts or behaviours can be highlighted, realistically evaluated and changed if necessary. Rather than focussing on the past causes of symptoms, CBT concentrates on the current situation and how to change it for the better.

The theory behind the cognitive model has evolved since the Beck model of 50 years ago. Garety and colleagues[45] and Morrison[46] have developed cognitive models of psychosis,

which suggest that rather than a patient's belief about themselves and others, it is their appraisal of the event which is central to the development and maintenance of psychosis. For example, a non-psychotic person may experience a hallucination or hear voices but would not react to them in a psychotic manner. Instead, they may dismiss it as tiredness or stress. The reaction in a psychotic patient may be to interpret the experience as the devil putting thoughts into their head or being chased by the police. So, rather than necessarily the hallucinations or delusions themselves being indicative of psychosis, it is the patients appraisal and their subsequent behaviour that is psychotic[46].

The cognitive model for psychosis is formed from the established belief that some people have a vulnerability to psychosis, and if they then experience some kind of trigger, for example, a traumatic life event, substance use or a hallucination/delusion this can lead to emotional and cognitive changes, disturbances in perception and judgement resulting in psychotic behaviour and a negative impact on functioning. Based on this theory, cognitive therapy then focuses on evaluating patients beliefs about their symptoms and experiences: what they mean, why they have occurred; exploring alternative explanations for these symptoms and questioning their beliefs; then addressing their subsequent behaviour and reaction to the event and suggesting alternative ways of coping. The therapeutic model targets patients' beliefs in their symptoms and formulates tasks for the patient to think about these symptoms in a more helpful way with the intention that their behaviour and reaction to the symptoms will improve. The general model can therefore be applied to the treatment of the specific symptoms of an individual. Current NICE (National Institute for Health and Clinical Excellence) guidelines recommend CBT use for depression, obsessive compulsive disorder, post-traumatic stress disorder, body dysmorphic disorder, anxiety and smoking cessation[47] as well as psychosis[48].

A Delphi study was carried out in 2009 to obtain a consensus from experts on the important components of CBT for psychosis[49]. This produced 77 statements describing the components grouped under seven broad headings by the authors: engagement with the client, structure and principles, formulation, assessment and model, homework, change strategies and therapist assumptions. The Delphi process extracts opinions from a large group of experts in a structured way, so although they are based on expert knowledge the results are still opinion. The components highlighted in the study that are expected to be important in the efficacy of CBT have not been tested empirically. The analyses in this thesis seek to provide evidence to test these opinions.

#### *1.1.4.1 CBT for the prevention of psychosis*

The CBT model developed by French and Morrison has been used in several trials[50-52], including the EDIE-II[53] trial for the prevention of psychosis that will be analysed in this study. The therapy is based on a specific formulation covering elements highlighted in the Delphi study. The treatment manual for the therapy details the components involved, which are summarised below[54]:

1. Agreement of problems and goals – this task should provide a shared list of problems and goals for the client and therapist to inform formulation of strategies for change and targets to measure improvement. It should include client history, beliefs about themselves, the world and others, drug use and risk of harm to themselves and others.
2. Formulation – case formulation or case conceptualisation is a detailed account of the client's difficulties, and development of a strategy based on the specific cognitive model of the clients disorder.
3. Homework – setting and completing homework tasks is integral for clients to maintain change strategies outside of the therapy sessions. Types of homework are behavioural experiment (changing reactive behaviours), monitoring (e.g. levels of anxiety or frequency of events) and education (reading relevant information).
4. Active change strategies – this term covers a range of specific strategies, any one or combination of which should be used as appropriate
    i. Normalisation – this is a process by which an individual is encouraged to consider their symptoms in a less catastrophic manner, to understand that other people experience similar feelings and events;
    ii. Generation of alternatives – question explanations for symptoms and explore alternative explanations in an ordered manner to reduce distress;
    iii. Manipulation of safety behaviours – safety behaviours are developed in order to avoid or prevent some feared event occurring, some behaviours can be debilitating and distressing in themselves and encourage unusual thoughts. The safety behaviours adopted must be fully assessed and tested to encourage the client to challenge the success of the behaviours and alter them accordingly.
    iv. Evaluation of metacognitive beliefs or responses – assessment of a clients beliefs about their illness and symptoms and how they react to unusual experiences to formulate strategies to question and alter distressing beliefs.

v. Evaluation of beliefs about self and others - identify and challenge beliefs about self that cause negative reaction and unusual behaviour.
vi. Reducing social isolation – unusual thoughts or experiences can be evaluated and challenged when discussed with others. CBT seeks to encourage contact with family and friends or development of social support networks.
vii. Preventing relapse – create a blueprint summarising work carried out during the course of therapy, determine warning signs of possible relapse and identify early interventions to initiate in the event of deterioration.

Some of these tasks are designed to target particular beliefs that the patient holds about themselves, others and the world in order to change their beliefs about and reactions to events. The CBT described above is specifically targeted at patient's beliefs about their illness, beliefs about paranoia, core beliefs and their meta-cognitive appraisal of their illness.

Metacognition is described as thoughts about thoughts, these can be positive or negative. Positive beliefs such as finding comfort from hallucinations or believing that paranoid thoughts are useful in order to keep safe, may not cause distress to the individual but having a positive attitude to these unusual experiences may be an indicator of increased risk of progression to psychosis. If beliefs about thoughts are negative, for example that the voices are uncontrollable or dangerous and have to be obeyed then they can be distressing for the individual. CBT therefore seeks to reduce distress from these beliefs by questioning the individual's thoughts about their experiences and beliefs and seeking alternatives. An individual's beliefs about themselves can also influence how they view themselves and others. It is found that people at risk of psychosis hold beliefs about themselves that cause them to behave and react to situations in a negative way. A common belief is that they are different from other people. CBT seeks to identify and question these beliefs and so we expect that changes in these beliefs will help to improve symptoms[54].

#### *1.1.4.2 CBT for command hallucinations*

The cognitive therapy model for the treatment of command hallucinations is similar to that for prevention. Recent work is being carried out on the use of CBT for this symptom under the Social Rank Theory model. Under this cognitive model it is expected that an individual's perceptions and beliefs about the voice influence their behaviour and response to the commands. The first study to consider CBT specifically for the treatment of

command hallucinations was a randomised controlled trial (RCT) by Trower et al[55]. In this study the cognitive therapy treatment follows the model described previously with an assessment of problems, formulation and intervention. In terms of command hallucinations the therapy focuses on four core beliefs, the client-voice power relationship, the punishment associated with non-compliance/appeasement, the identity of the voice and the meaning attached to the voice e.g. punishment for previous behaviour. The individual's beliefs about the voice are identified and these beliefs, as well as the orders that are given, are challenged. This therapeutic model was then developed further and formed into a treatment manual by Byrne et al[56] specifying three therapy stages: assessment, intervention and reformulation. At the assessment stage the therapist determines the client-voice power relationship and the client's beliefs about the voice power, compliance, resistance, appeasement and meaning. They distinguish with the client between the voice and the client's interpretation of it, identify coping strategies and set goals for therapy. The intervention stage seeks to question and challenge the client's beliefs about the voice, for example, by not complying with or appeasing the voice to see if the consequences of disobeying that the client thinks will happen do in fact occur, or by answering back to the voice and increasing the clients control to increase their social rank relative to the voice. Reformulation is concerned with the client's broader beliefs about themselves and others and seeks to question and challenge these in order to improve the client's perceptions about themselves, others and the voice.

### 1.1.5 Effectiveness of CBT in psychosis

There has been a great deal of research into CBT for various mental health conditions, some of which have been in the treatment of psychosis. A review by Rector and Beck[57] published in 2001 found six randomised studies of CBT in schizophrenia and concluded a positive effect of CBT compared to routine care. Five studies of positive symptoms gave a combined standardised effect size of 1.31 (sd=0.71) for CBT versus routine care and three studies analysing negative symptoms gave a treatment effect size of 1.08 (sd=0.83) in favour of therapy. These are standardised effect measures and are presented in units of standard deviations. The effect of 1.31 indicates that on average those receiving CBT are expected to score 1.31 standard deviations higher than those receiving routine care. This is generally considered to be a large effect. A popular rule of thumb is that a standardised effect of less than 0.3 is small and greater that 0.8 is large[58].

CBT was also reported to show a positive effect when compared to supportive therapy, with a combined effect size across five studies of 0.91 (sd=0.14). A more recent review published in 2008 by Wykes et al[59] found 34 studies, an additional 29 studies in the intervening years, which tested the impact of CBT in schizophrenia. The authors conclude that there is an overall positive effect of CBT for psychosis in terms of positive and negative symptoms, functioning, mood and anxiety with standardised effect sizes of 0.37, 0.44, 0.38 and 0.36 respectively. They found no benefit for hopelessness. As would be expected these are smaller effects than those found when comparing CBT to routine care but remain statistically significant. Based on the target symptom of the trial the pooled standardised effect of CBT was estimated at 0.4 (95% CI 0.25 – 0.55). The authors also assessed the methodological rigor of the trials, rating each trial using the Clinical Trial Assessment Measure (CTAM). The CTAM rates the quality of the trial covering the following areas: sample characteristics, treatment allocation, outcome assessment, control group, description of the treatments and analysis carried out. They report that the quality of the trial appears to have an impact on the effect sizes, with lower quality trials finding larger effects, specifically that they were larger when assessors were unmasked to the patients treatment allocation, though the association for this is weak. The authors report that over half of the studies included were judged to have unsatisfactory statistical analysis to account for loss to follow-up. However, encouragingly they also note that the quality of the trials tends to be higher in more recent studies.

A review of CBT for first episode and early psychosis by Morrison[60] discussed the findings of 12 studies in the area and found that overall there is not a great deal of evidence that CBT is better than treatment as usual for early or first episode psychosis in terms of relapse or readmission. There is a benefit of CBT for other outcomes; improvement in symptoms, rate of recovery and quality of life. The review highlights problems with trials in this area, one of which is that first-episode patients tend to improve quickly anyway and so testing CBT in those that are unlikely to respond quickly may be more beneficial. It also means that the length of treatment and timing of treatment is all the more important. The studies use the same outcome measures as those used for medication interventions but it may be that other, therapy specific, outcomes are more appropriate.

These reviews consider the effect sizes under an intention-to-treat methodology. The individual trials however, report heterogeneity in the treatment received by participants,

this may have an effect on observed treatment effects and future attempts to refine and optimise treatment strategies.

#### *1.1.5.1 Effectiveness of CBT in high-risk samples*

Very few trials have evaluated the effects of cognitive behavioural therapy for those who are not currently psychotic but are at high-risk of psychosis. This may be because treatment intervention in this area is controversial due to low transition rates to psychosis even in the high-risk group. Two systematic reviews published in 2013 have consolidated the evidence for early interventions in the prevention of psychosis[61] and specifically CBT for prevention[62]. The primary outcome in the trials is transition to psychosis and secondary outcomes of symptom severity and functioning are reported.

Stafford and colleagues report on 11 randomised controlled trials that tested the effects of early interventions on the prevention of psychosis. In addition to psychotherapies the authors found studies testing pharmacological (risperidone and olanzapine) and nutritional interventions (omega 3 fatty acids). Of the 11 RCTs, five specifically allowed for the testing of CBT against a control group which, in all cases, was supportive counselling. The authors pooled the results in a meta-analysis. The second review by Hutton and Taylor reports on six trials comparing CBT to a control group for the prevention of psychosis. They include a trial[63] in the comparison of CBT to supportive counselling that was not included in the Stafford review because the intervention was not just CBT but an integrated psychological intervention involving CBT along with group skills training, cognitive remediation and psychoeducational multifamily group sessions. The trials ranged in size from 29 to 288 participants with information on the primary outcome, the largest of these being the EDIE-II trial. Rates of transition to psychosis were low, with a total of only 69 events in 645 participants across all studies at 12 months. The rate of transition in the supportive care group ranged from 7% to 22% and in the CBT group from 0 to 16%

The meta-analyses conducted by the two sets of authors differ in their results due to the methods of analysis. The Stafford review pools transition rates in only those participants who complete each study and found that all five favoured CBT over supportive counselling for reducing transition to psychosis at 12 months follow-up. However, none were individually significant at the 5% level. When combined in a meta-analysis using a Mantel-Haenszel random effects model, CBT was found to reduce the risk of transition to psychosis by almost half compared to supportive counselling (pooled risk ratio = 0.54,

95% CI 0.34 to 0.86). The Hutton and Taylor review not only includes an additional study but also pools rates with the denominator being based on all participants randomised, whether they completed follow-up or not. The pooled risk ratio of transition to psychosis at 12 months again using a random effects model is reported as 0.45 (95% CI 0.28 to 0.73). The impact of analysing the data under complete case or intention to treat is unknown. If the drop outs were missing at random then the pooled treatment effect would be unbiased. However, if the reason participants left the study was connected with their illness the results may be biased. Fortunately, in this case the two reviews report pooled effects of similar magnitude and come to the same conclusion that CBT has a beneficial effect over supportive counselling in reducing transition to psychosis in those at high risk.

In terms of secondary outcomes, Stafford et al report a significant improvement in positive symptoms at 12 months in the CBT groups compared to the control group (pooled standardised mean difference= -0.17, 95% CI -0.35 to -0.01) though only one study, the large EDIE-II trial shows a significant positive effect.

Several of these studies[50-52,64] report heterogeneity in the therapy received by participants in terms of the number of therapy sessions attended, which varied greatly within each study. The Addington[51] study also collected data on the content and focus of the therapies. Unfortunately analyses were limited to intention-to-treat methods and did not consider further investigation of heterogeneity of therapy received on treatment effect.

#### *1.1.5.2 Effectiveness of CBT for command hallucinations*

There is currently very little evidence for treatment specifically targeted at command hallucinations rather than general symptoms of psychosis. Chadwick and Birchwood[36] describe the CBT model and give a qualitative description of four case studies where they have applied CBT techniques to patients suffering from commanding voices. An improvement is seen in all four patients; this does not prove the benefits of CBT but provided the first evidence of its plausibility as a therapy and led to larger scale studies. The first RCT study to test the effectiveness of CBT for command hallucinations was that of Trower et al.[55], the cognitive model used was as described earlier. The therapy does not seek to stop the hallucinations but to prevent compliance and power of the voices. The study randomised 38 participants to Cognitive Therapy for Command Hallucinations (CTCH) or treatment as usual (TAU) and found a significant reduction in compliance as well as a reduction in the perceived power of the voice, the belief in the voices

omnipotence and an increase in the patients control over the voice at six and twelve months follow-up. The participants were selected as high-risk individuals since they had previously complied with serious commands to harm themselves, others or commit other 'serious social transgressions'. Though the study is small and in a high-risk sample it gives a promising outlook for therapy.

This study has been followed up with a larger RCT of CTCH called the COMMAND trial (see Birchwood et al. for study protocol[65]), a secondary analysis of which will be reported in the present thesis. The cognitive model was again as described earlier. The primary results of the larger study which have not yet been published[66] show a significant reduction in the odds of full-compliance in the group randomised to receive CTCH compared to TAU, odds ratio 0.57, (95% confidence interval 0.33 to 0.98, p=0.042). A further detailed analysis of this study is provided in the results section of this thesis.

A recent Australian RCT[67] for command hallucinations TORCH (Treatment Of Resistant Command Hallucinations) compared acceptance based CBT to befriending, an active control. The trial also had a waiting list control and members of this group were subsequently randomised to one of the two treatment conditions and included in the overall analysis as well as separately. The study of 44 participants did not find any significant difference in the primary outcome or confidence to resist harmful commands (compliance with harmful commands was deemed an unsatisfactory outcome in this group as few had complied with harmful commands in the 4-6 months pre-baseline). The authors suggest that the difference in results between this and the Trower et al study is due to differences in the characteristics of the sample population, namely the severity and frequency of experiencing harmful command hallucinations. They also compare the cognitive therapy to befriending which may be considered an active control and could reduce the effect sizes. In terms of size, both studies are comparable in the number of participants and both are small, making it difficult to uncover significant treatment effects and limiting the generalisability of findings.

### 1.1.6 How does CBT work and who can benefit?

Once it is established that an intervention is effective the next question is how is it effective? CBT therapy is designed from a clear theoretical model and so there are strong hypotheses as to the specific effects on the individual and how it will adjust the course of their illness. If the therapy works in the way indicated by the therapeutic model then we

expect changes in the hypothesised beliefs will mediate the effect of CBT on outcome. It is important to establish if the therapy is working as intended or via alternative pathways in order to apply it appropriately and improve it.

The therapy employed in the COMMAND trial seeks to alter the client-voice power differential in order to give the client more control over the voice and reduce compliance with the commands. It is therefore expected that power of the voice will mediate the treatment effect on compliance. The components of the therapeutic model used in the EDIE-II trial target certain psychological processes: beliefs about self, illness, paranoia and beliefs about beliefs (metacognition) in order to reduce the risk of transition to psychosis and improve symptoms these are therefore expected to act as mediators on the causal pathway.

Many studies carry out secondary analyses to get a better understanding of the benefits of CBT but few consider mediation. Instead research has often concentrated on identifying predictors of outcome with CBT. These explorations tend to be an analysis of both randomisation arms together and therefore provide general prognostic information rather than specific predictors of improvement for those receiving CBT. If the impact of interactions between treatment group and covariates on outcome were considered we would know prognostic factors specifically in CBT and may be able to find treatment effect moderators. However, the information can be helpful to establish possible confounders or mediators of the treatment process. Predictors can be grouped in terms of characteristics of the patient, of the therapist, the relationship between patient and therapist, the therapy received and the impact on mediators of the treatment process, for example, appraisal of and beliefs about illness. Unfortunately, few studies investigating predictors are in psychotic patients and even less specifically in the pre-psychotic, high-risk group or investigating issues related to command hallucinations. A short summary of studies exploring predictors of outcome is given below.

#### *1.1.6.1 Predictors of outcome*

An analysis of the London East Anglia Trial of CBT for psychosis determined predictors of outcome that are specific to those receiving CBT (i.e. treatment effect moderators) by testing for interactions between baseline covariates and randomisation group on outcome. The analysis provides evidence that greater severity of illness (indicated by number of recent admissions) is associated with a greater effect of CBT[68]. A secondary analysis of

the SOCRATES trial (Study of Cognitive Reality Alignment Therapy in Early Schizophrenia)[69,70], a study with two active interventions and a control condition, also considered a treatment effect moderator, age, and found that younger patients (under 21 years) benefit more from supportive counselling whereas older patients (over 21 years) benefit more from CBT in terms of symptom reduction[71]

An analysis by Naeem et al[72] of the Insight trial[73] of CBT for psychotic patients found that a higher level of insight and greater psychopathology (symptoms) resulted in better outcome. They attribute this specifically to a positive response to CBT though it appears to be more generally predictive of improved outcome across the CBT and control arms. Insight was also found to be a significant predictor of outcome in a further analysis of the SOCRATES trial where higher insight significantly reduced risk of relapse and readmission over the 18 month follow-up period[74]. Secondary analysis of trials by Morrison et al.[75], Drury et al.[76] and Tarrier et al.[77] support the hypothesis that a longer duration of illness is associated with worse outcomes, again the analyses are carried out in all participants rather than being attributed to any particular course of therapy. The Drury study also found that a shorter period of untreated psychosis and being female predicted faster recovery time. Gender was found to be the only predictor of outcome in the CBT group of a trial by Brabban, Tai and Turkington[78] after adjustment for previous diagnosis, affective blunting (lack of emotional response), alogia (lack of speech) and insight. These analyses provide information on characteristics that are associated with a participant outcome. They do not provide proof of causal pathways as there are likely to be other factors confounding the association that are not accounted for in the basic regression analyses used.

The literature on command hallucinations uses the term mediators for predictors of compliance and has indicated several important predictors; malevolence/benevolence of the voice; content of the command; omnipotence and power of the voice; identification of the voice.

#### *1.1.6.2 Therapy content*

The cognitive therapy used in both the COMMAND and EDIE-II trials follow specific protocols designed to target aspects of the patient's belief system in order to improve outcomes, for example formulation and active change strategies are expected to be necessary to make the therapy effective. Unfortunately it is not always possible to carry out

all of these parts of therapy so the treatment may not be effective. To know the mechanism by which the treatment works and in whom it is most effective it may be possible to improve, refine and tailor the treatment to be more effective.

A study exploring the heterogeneity in CBT treatment received on treatment benefit in patients with psychosis has been published by Dunn et al[79]. This is a secondary analysis of the Psychological Prevention of Relapse in Psychosis (PRP) trial comparing CBT to a family intervention and TAU to prevent relapse and improve symptoms in people with psychosis. The primary analysis did not find a significant benefit of CBT for improving relapse or symptoms but did find a reduction in depression. However, the amount and content of therapy received varied between participants introducing the possibility that the benefits of receiving therapy as prescribed could be diluted. The authors used information on therapy adherence to group participants randomised to CBT as receiving no therapy, partial therapy or full therapy. Using principal stratification techniques they predicted which of these three strata those allocated to TAU would have received had they been allocated to CBT and calculated within stratum treatment effects. The authors concluded that CBT is only effective when full therapy is received, no benefit is observed if only partial therapy is given and there may even be evidence of a negative effect of having partial therapy. This type of analysis accounts for selection effects associated with receiving different quality of therapy that cannot be accounted for by adjustment of confounders. The analysis provides solid evidence that adherence to the cognitive model of therapy is essential to provide a benefit for people with psychosis. Freeman et al[80] follow on from this analysis by investigating predictors of the three levels of treatment received defined in the previous study as none, partial or full-therapy. They find that the participants perception about their illness in terms of cause of illness and the control that they have over the illness are associated with the treatment received. Specifically they find that participants who agree that personality and state of mind are causes of their illness are more likely to receive full therapy. Those who received partial therapy were more likely than those receiving no therapy or full therapy to agree that their illness was caused by pollution. Those in the no therapy group were less likely to think that their illness was long lasting than those who received partial or full therapy.

The PRP study data provided information comparing full to partial or no therapy but did not allow for the effect of particular aspects of therapy to be assessed. A study by Chadwick et al[81] aimed to determine if case formulation is necessary for CBT to be

effective in the treatment of psychotic patients. The authors carried out two studies following a small sample of psychosis patients, thirteen and four for the first and second experiments respectively, to determine formulation effects on therapeutic alliance and distress. In the first study baseline measures were taken followed by three follow-up measures, the first two directly after case formulation was carried out. Overall trends and pairwise comparison of scores for alliance and distress were compared between the time points. The results did not indicate that case formulation had a direct impact on either alliance or depression. In the second experiment only four patients were followed but this time the number of data points increased. Baseline measures were taken over the course of five time points and then different key aspects of CBT therapy were introduced starting with case formulation then restructuring of negative self beliefs and finally restructuring of secondary delusions, each of these was implemented for at least four sessions. Outcome measures of depression and symptoms were taken regularly. The authors reported no significant improvement in either depression or symptoms that could be attributed specifically to the formulation component of CBT therapy. The sample sizes of these two experiments were very small even with multiple measures on the participants so lack power to determine any change that is not due to chance. However, even if a significant change in outcome over time had been found this would not provide evidence of the effect of formulation as there is no control group to compare to. It may be that symptoms and depression change over time anyway or that it is just the regular contact that leads to improvements. Therapeutic alliance may also improve naturally over time as the client and therapist have more contact rather than being specifically attributed to the use of formulation. In order to determine the impact of CBT and specific components of therapy a control group not receiving the aspect in question must be followed as well. A qualitative aspect of the study asking participants opinions of formulation gave mixed results, nine of the eleven agreeing to interview said that the formulation helped, six reported positive emotions and six reported negative reactions to the experience of formulation, though four of the six reporting negative feelings also reported positive reactions.

An aspect of CBT thought to be key to effectiveness is homework. Types of homework (information gathering, hypothesis testing, practice of change strategies) and their role in cognitive therapy are described by Dunn and Morrison[82] and Rector[83]. A meta-analysis found 23 studies investigating the impact of homework within CBT, the studies were mainly for the treatment of anxiety and depression with only two applied to a

schizophrenia sample[84]. The authors conducted subgroup analyses and conclude that the source (client/therapist) and time (during/post treatment) of assessment affected the relationship between use of homework and outcome. The studies assessed the effect of homework within the CBT framework, participants were not randomly assigned to receive homework and so only associations can be inferred. The meta-analysis combined correlation coefficients from the studies; the analyses these were taken from ranged from adjustment for baseline measures, change scores or summary statistics. The quality of the results is therefore questionable. The authors report a positive association between homework assignment or compliance and outcome estimating the correlation between homework compliance and treatment outcome to be approximately 0.27 (95% CI 0.19 – 0.33). This is a pooled estimate over all studies rather than the effect of homework in a psychosis sample, which was not reported specifically.

The two studies investigating homework in psychosis patients consisted of a trial of Cognitive Behavioural Social Skills Training (CBSST) in 76 older people with schizophrenia (age range of sample 42-74 years)[85] and a sub-sample of 29 patient-therapist dyads that were part of a larger effectiveness study of CBT in psychosis[86]. The trial of CBSST investigated homework as a process variable by looking at correlation between proportion of homework assignments completed and outcome measures within those assigned CBSST, they found no significant effect on symptoms, depression or insight but some correlation with skills acquisition. The analysis was conducted only in those assigned to the intervention and so no control group was provided and there did not appear to be any adjustment for potential confounders. The analysis of patient-therapist pairs also found no impact of homework compliance on symptoms at follow-up. This was a small sub-sample of a larger trial consisting only of patients receiving therapy. Homework was assessed at therapy session three of a possible 35, (mean number of session =18 range 4-35) raising the possibility that homework compliance may change over the course of therapy. Additionally, no adjustment made for potential confounders, though with such a small sample this would be difficult.

An observational study of homework assignment and quality of homework completion followed 129 patients being treated for severe mental illness, predominately schizophrenia, within a national study evaluating recovery-orientated case management[87]. The study found an association between amount of homework given and functioning, but no association with symptom distress or recovery. The analysis did not seem to adjust for any

confounding factors, which is of particular importance in an observational study where there are likely to be underlying factors effecting both engagement with homework tasks and measures of outcome. As such, only associations can be inferred.

#### *1.1.6.3 Therapeutic alliance*

There has been some interest in the role of the therapist-patient relationship on the success of the therapeutic process with the hypothesis that the better the relationship the better the treatment will work. The Dunn, Morrison and Bentall[86] analysis of patient and therapist dyads that was mentioned previously in terms of homework compliance also considered the relationship between therapeutic alliance and symptoms. They found that patients and therapists generally agreed on the level of alliance and that over the course of therapy the correlation between the patient and therapist ratings increased in magnitude. There was no significant increase in alliance over the period and they found no association between alliance and change in symptoms. The sample was small however, with only 20 client therapist pairs completing follow-up. Another study of 26 mental health patients (20 schizophrenic) undergoing cognitive therapy found no relation between initial patient reported alliance on change in outcome, but found a significant positive correlation with therapist rated alliance and change in GAF (functioning)[88]. This study involved collecting patient and therapist reports of alliance throughout the therapy, giving an average of 12 (range 3-28) assessments. The authors averaged these assessments for the initial phase (1$^{st}$ two assessments), final phase (last two assessments) and the working phase (middle assessments). These averages were then categorised as low, fair and good alliance. Associations between alliance and outcome were limited to correlations and did not take into account any confounding between them so no causal inferences can be derived. Another analysis of the SOCRATES study investigated the effect of attendance and alliance as an example of the statistical methods used later in this thesis. The analysis showed an improvement in symptoms with increased attendance at therapy and a multiplicative effect of attendance and increased therapeutic alliance[89]. Steel et al[90] added to the meta-analysis of Wykes et al[59] and considered therapist influences on treatment effects in RCT studies. They found that treatment effects were larger when the proportion of therapists dedicated entirely to the trial was greater and when there was more supervision of therapists during the trial.

These studies investigating the impact of potential mediators such as therapeutic alliance or homework on success of therapy have indicated associations at most. Participants are not

(and often cannot be for ethical reasons) randomly allocated to have low or high alliance with their therapist or adhere to homework and so any association seen may be due to other factors. There may be a host of factors influencing the alliance between a participant and therapist or whether somebody does their homework, for example severity of illness or specific symptoms experienced by the participant, these in turn may affect their outcome therefore confounding any association seen. Since these mediators are not randomised, special statistical methods must be used for causal links to be inferred.

### 1.1.7 Substantive aims

This brief summary of evidence for the effectiveness of CBT in the prevention and treatment of psychosis as well as reasons for effectiveness highlights that there is a lack of research in the area. There is particularly a lack of information as to the treatment effects related to heterogeneity in treatment received. Although authors acknowledge that heterogeneity is present no further steps are taken to delve deeper into the results. Where mediation analysis has been carried out the statistical methods are often unsatisfactory. Analyses conducted within the treatment arm only or mediation analyses that do not account for confounders do not allow for any meaningful conclusions to be drawn. Without further analysis as to the effectiveness of particular aspects of complex interventions alterations and improvements the therapy models are merely theory rather than evidence based. Additionally if the therapy does not work in the way that is expected by changing behaviours that it is designed to target then this may indicate that the intervention should be changed or that the previously hypothesised behaviours are not as important as some other factor that is propelling change in outcomes.

The work reported in the present thesis aims to answer specific substantive questions relating to the EDIE-II trial of CBT for the prevention of psychosis and the COMMAND trial of CBT for compliance with command hallucinations:

1. Estimate the effect of dose of therapy on symptom reduction in those at high risk of psychosis
2. Estimate the effectiveness of including the following aspects of therapy on reduction of symptoms in those at high risk of psychosis; agreement of problems and goals, formulation, homework and active change strategies
3. Determine if the effectiveness of CBT for high-risk individuals is mediated by changes in beliefs

4. Determine if there is a causal pathway from the amount and content of therapy received through changes in beliefs to a reduction in symptom severity in those at high risk
5. Determine if the effectiveness of CBT on compliance with command hallucinations is mediated by a change in the power of the voice

In order to answer these questions statistical methods for causal inference are applied. The background to these methods and aims for the thesis will be discussed in the following section.

## 1.2 Introduction to statistical methods for causation and mediation

This thesis is focussed on the application of statistical methods to answer the substantive questions detailed above with emphasis on improving estimation in the face of problems found in the analysis of real data sets. To begin, the statistical methods of causal inference are introduced, and their application to the analysis of mediation and process evaluation explained.

### 1.2.1 Notation

Participants are randomised to receive either the treatment (e.g. CBT+mental state monitoring) or control intervention (e.g. mental state monitoring alone). Participant characteristics are measured at baseline, their outcome is recorded at follow-up and a post-randomisation process variable or mediator is measured at a mid-point between randomisation and follow-up. The following notation describes these measures for the i-th subject:

$Z_i$ – randomisation group indicator: $Z_i = 1$ if randomised to treatment, $Z_i = 0$ if randomised to control

$D_i(z) = D_i(Z_i = z)$ – treatment received indicator: $D_i(Z_i = 1) = D_i(1) =$ treatment received when allocated to treatment condition , $D_i(Z_i = 0) = D_i(0) =$ treatment received if assigned to control condition.

$X_{1i}, X_{2i}$, ... - baseline covariates
$M_i(z, d)$ – value of the mediator when allocated to treatment $z$ and receive treatment $d$
$Y_i(z, d, m)$ – outcome when allocated to treatment $z$ receive treatment $d$ and have mediator at level $m$

These definitions introduce the concept of potential outcomes. Prior to treatment allocation (randomisation, for example) there are two sets of potential outcomes associated with the i-th participant. In the control condition $Z_i = 0$, we have $Y_i(0, d, m)$, $D_i(0)$ and $M_i(0, d)$ and under the treatment condition we have $Y_i(1, d, m)$, $D_i(1)$ and $M_i(1, d)$. Following treatment allocation we only observe one of these two sets. In the treatment group we observe $Y_i(1, d, m)$, $D_i(1)$ and $M_i(1, d)$ but not the outcomes associated with the control condition (the latter frequently being labelled as 'counterfactuals') and in the control group the opposite will be seen. Baseline covariates are available for all participants. If there is perfect compliance with treatment allocation then $D_i(z) = z_i$. When $M_i$ is a post-randomisation process variable then it is only available for those who are allocated to treatment and by convention is zero if randomised to the control group, $M_i(0) = 0$. The rationale for this convention will be explained later.

### 1.2.2 Associative models

Analysis of clinical datasets is often focussed around determining statistical associations between variables: do older patients have worse symptoms than younger patients? Do women have fewer symptoms then men? Statistical analyses test whether the associations seen in the data are more than we would expect to see by chance and we use regression models to quantify the associations. In these questions symptoms are the outcome recorded for each individual and in statistical notation referred to as $Y_i$. The explanatory variables are age and gender which again are recorded for each individual and labelled $X_i$. To quantify the association of each of these explanatory variables with the continuous outcome we can use a simple linear regression where we model.

$$Y_i = \alpha + \beta X_i + \varepsilon_i$$

Where the coefficient $\alpha$ is the mean value of Y when X is zero, $\beta$ is the average increase in Y for each one unit increase in X and $\varepsilon_i$ is an error term which is assumed to describe a normal distribution with mean zero (with $X_i$ and $\varepsilon_i$ assumed to be uncorrelated). This model can be extended to include additional explanatory variables as well as non-linear effects. When the outcome is a continuous measure the coefficient of association $\beta$ can be estimated using ordinary least squares (OLS). Ordinary least squares is a method used to calculate parameter coefficients in a linear regression by minimising the sum of the squared differences between observed and fitted values (i.e. sum of square residuals). The sum of the squared value of the residuals rather than the sum of the actual residuals must

be minimised since positive and negative error terms can cancel each other out if not squared first. It can be shown that the coefficients can be calculated as follows:

$$\hat{\beta}_{OLS} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})\,(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

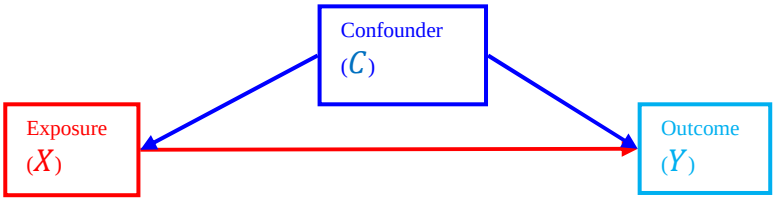
$$\hat{\alpha}_{OLS} = \overline{Y} - \hat{\beta}\overline{X}$$

Where $\overline{X}$ and $\overline{Y}$ are the mean of the observed X and Y values respectively. The estimate of the parameters will be unbiased if, as assumed above, there is no covariance (correlation) between the independent variable and the error term if $\mathrm{cov}(X_i, \varepsilon_i) = 0, E(\varepsilon_i|X_i) = 0$. However, if there is a factor (common cause) that influences both the explanatory variable and the outcome (or equivalently the variation not accounted for by the explanatory variable in the outcome) then the estimate will be biased[91]. This is known as confounding.

### 1.2.3 Confounding

Finding that Y is associated (correlated) with X does not necessarily imply that there is a causal effect of X on Y. X might influence Y; the effect might be in the reverse direction; or there might be other variable(s) that are a common cause of both. Note that these three explanations are not mutually exclusive. A variable is a confounder of the relationship between an exposure and outcome (or any two variables of interest) if it is associated with both the outcome and the exposure but is not a result of either the outcome or exposure[92]. Typically it is a common cause of both X and Y. An estimate of the causal effect of X on Y from the simple linear regression equation (structural model – see below) $Y_i = \alpha + \beta X_i + \varepsilon_i$ will be biased because of the effect that C has on both variables (Figure 1.1). This means that part of the unexplained variability in the outcome $\varepsilon_i$, is due to C and C is also associated with X so $\mathrm{cov}(X_i, \varepsilon_i) \neq 0$ and the estimate of the causal effect β will be biased. We account for confounding statistically by adjustment for this variable so that the effect of X on Y is estimated by $Y_i = \alpha + \beta_1 X_i + \beta_2 C_i + e_i$ . Now $e_i$ does not contain this confounding factor and if all confounders are accounted for we can assume that $\mathrm{cov}(X_i, e_i) = 0$. Rules and methods for determining appropriate adjustments are detailed by Shrier[93], Greenland[94] and Shipley[95].

**Figure 1.1: A graphical representation of confounding**




If the confounder is known and measured then it is easy to adjust for this in the analysis and therefore achieve an unbiased estimate of the effect of X on Y. Unfortunately in practice it is very unlikely that all possible confounders will be measured and it is likely that there will be some residual confounding that cannot be accounted for, therefore giving biased estimates. For this reason even when adjusted for all of the possible confounders that are measured the coefficient is described as associative rather than causative.

### 1.2.4 Path diagrams

A path diagram is a representation of the hypothesised relationships between variables and can be useful when considering the causal relationships and effects of confounding as in Figure 1.1[96]. A path diagram is expected to describe all of the relationships between variables and the nature of the relationships. An expected causal relationship is indicated by a single headed arrow and an association with a double headed arrow; with the omission of an arrow signifying no direct link between variables. Since all relationships are described the interpretation of causal associations can be hypothesised. In Figure 1.1 for example a causal effect of X on Y is hypothesised though some of the association is expected to be explained by C. If it was expected that C completely confounded the relationship between X and Y the red arrow connecting them would be removed.

A simple path diagram of the relationships between receiving treatment (D), a mediator (M) and outcome (Y) would be shown as

$$D \rightarrow M \rightarrow Y$$

In the above diagram we expect that Y is directly influenced by M but is only influenced by D through M. This and the graph in Figure 1.1 is an example of a directed acyclic graph (DAG) [91]. A DAG is a path diagram which consists of directed edges that are not cyclical, i.e. starting from any point it is not possible to follow the directed paths back to that starting point.

Graphical diagrams are non-parametric models and make no assumptions about the distribution of the variables. Terms used in graphical modelling come from the descriptions of family relations. A parent has a causal effect on a child, the child is a descendant of the parent and an ancestor of a child would be a further step away from the child e.g. the parent's parent. In the simple path diagram above D is the parent of M and ancestor of Y, M is the child of D and parent of Y and Y is the child of M and descendent of D. An endogenous variable is one that has parents within the stated causal model; it is determined by other variables within the model, in this example both M and Y are endogenous. Alternatively, an exogenous variable is one that has no parents in the model, it is not caused by other variables in the model, D in this example. A variable can be exogenous in a particular model of interest if it is caused only by factors outside the system in question.

The path diagram is a useful tool to determine and describe all possible pathways of a treatment mechanism but on its own provides no quantifiable measure of the associations. To do this the relationships shown by the arrows in the path diagram can be estimated using structural equation models of which correctly-specified regression equations are an example.

### 1.2.5 Structural models

The relationships shown by the arrows in the path diagram can be estimated using a set of regression equations. The path diagram above can be described by the following:

$$M = \alpha D$$
$$Y = \beta M$$

The path coefficients α and β can be estimated using regression techniques to quantify the causal effects of D on M and M on Y respectively. A key conceptual rather than practical difference between structural equations and standard regression models is described by Pearl[91]. He states that they describe the effect of manipulating elements in the model rather than from purely observing the elements at different values. The interpretation of the coefficients above would therefore for be that α is the expected change in M if the value of D is increased by one unit from d to d+1, rather than the observational concept which would define α as the expected difference in M if D was observed at the level of d+1 rather than d. Another cornerstone of the structural equation that differs from standard regression equations is the direction of the equality. Standard regression equations can be read in

either direction and allow for rearrangement so that setting values of the variables on either the left-hand or right-hand side of the equation allows assumptions to be made about the other. In structural equations the equality is asymmetrical and could be replaced by an arrow indicating the direction so $m = \beta d$ is equivalent to $m \leftarrow \beta d$. From this, the effect on M when we manipulate D can be determined but there is no information about D if we manipulate M.

The equations describe every hypothesised direct causal association between variables. As with path diagrams the exclusion of a variable in a structural equation indicates that there is no direct causal link from that variable. The statements above declare that M is the only immediate cause of Y. Although D affects Y though M it does not have a direct effect, manipulation of D does not affect Y if M is held constant. The coefficient β can therefore be interpreted as the increase in Y per unit increase in M regardless of the value of any other variable in the model (D).

### 1.2.6 Causation

If we wish to know if receiving a particular treatment, CBT therapy for example, causes an improvement in outcomes we could compare the outcome measure of interest in patients who have received the treatment to those that have not. From this we could not interpret any effect as a causal effect because the model is likely to be described by the diagram in Figure 1.1, there is potential confounding that has not been accounted for between the exposure and outcome. There could be a multitude of reasons why particular patients have undergone CBT therapy which may also be associated with their outcome in some way. In order to infer the causal pathway of CBT to an improvement in outcomes all other possible reasons for the difference in outcome must be eliminated.

In the following section the concepts of causation will be explained in terms of determining a causal effect of a new treatment regime compared to an established treatment (control). The concepts also apply to epidemiological studies where the terms treatment and control can be exchanged for exposed and unexposed.

Rubin[97 98] popularised the concept of ‘potential outcomes’ for causal analysis. Assuming there are two treatment conditions, there are two potential outcomes. The potential outcome that is observed is the actual outcome for the treatment actually received by the individual. The potential outcome under the condition that is not observed is the outcome that would have been observed had they experienced the alternative treatment condition.

Comparing treatment with a control condition, Rubin states that to make causal inferences we must compare the outcome in an individual under the treatment condition $Y_i(D_i = 1)$, to the outcome in that same individual under the control condition $Y_i(D_i = 0)$, giving the causal effect as $Y_i(D_i = 1) - Y_i(D_i = 0)$. The average treatment effect (ATE) across all individuals is then:

$$ATE = E[Y_i(D_i = 1) - Y_i(D_i = 0)]$$

Unfortunately for any one individual it is only possible to observe one of these outcomes as no participant can experience both the treatment and control condition at the same time, this a problem that Holland[99] calls the 'fundamental problem of causation'.

Random allocation of participants to the variable of interest is the standard method of making the groups comparable. Randomised Controlled Trials (RCTs), as used in the EDIE-II and COMMAND trials are designed to randomly assign people to treatment so that an individual is equally as likely to be in either trial arm $Pr(Z_i = 1) = Pr(Z_i = 0) = 0.5$, their allocation is independent of any baseline characteristics (no confounding) and independent of their outcome other than through the receipt of treatment. This means that the average outcome in those assigned to the treatment condition can be assumed to be equivalent to the average outcome of those assigned to the control condition if (counter-to-fact) the latter had been assigned to the treatment. The average outcome in those randomised to treatment is compared to the average outcome of those randomised to the control condition. This is the traditional and gold standard method of analysing an RCT, it is called an Intention To Treat (ITT) analysis.

$$ITT = E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]$$

If all individuals adhere to their allocation and receive the treatment they are randomised to $D_i(Z) = Z_i$ then the ITT is a valid estimate of the ATE. It can be assumed that subjects randomised to the treatment are exchangeable with subjects assigned to the control, potential outcomes are independent of treatment allocation $E[Y_i|Z_i = 1] = E[Y_i|D_i = 1] = E[Y_i(1)]$ and $E[Y_i|Z_i = 0] = E[Y_i|D_i = 0] = E[Y_i(0)]$ then $E[Y_i(1)|Z_i = 1] - E[Y_i(0)|Z_i = 0] = E[Y_i(1)] - E[Y_i(0)]$.

In reality this is not the case and the ITT analysis answers the pragmatic question; what is the effect of allocating treatment? This question is of interest as it estimates the practical

impact of prescribing treatments to patients who will not always adhere; it also maintains the benefits of randomisation. However, if there is variation in the type and amount of treatment that people receive (heterogeneity in treatment received) or there are missing data then the calculated treatment effect may not be answering the question of real interest. Rather than answering 'what is the effect of allocating treatment?' it may be important to answer 'what is the effect of receiving treatment?'

One way to define treatment effects when the treatment received does not reflect the treatment allocation is to look at treatment effects within subgroups of participants, for example only in those that receive the treatment or only in those that comply with their allocation. The average treatment effect on the treated (ATT) estimates the average treatment effect only in those that received treatment, rather than the average treatment effect over all participants whether they receive the treatment or not.

$$\mathrm{ATT} = \mathrm{E}[\mathrm{Y_i}(\mathrm{D_i} = 1) - \mathrm{Y_i}(\mathrm{D_i} = 0) | \mathrm{D_i} = 1]$$

This is a local average treatment effect as it is the average treatment effect in a specific population of participants that receive the treatment. Another local average treatment effect is the complier average causal effect (CACE). This is the treatment effect only in those that comply with their allocated treatment and is defined as:

$$\mathrm{CACE} = \mathrm{E}[\mathrm{Y_i}(\mathrm{D_i} = 1) - \mathrm{Y_i}(\mathrm{D_i} = 0) | \mathrm{D_i}(1) - \mathrm{D_i}(0) = 1]$$

The ATT and CACE are treatment effects within a subgroup of individuals and do not provide a comparison with other groups. As such their usefulness may be limited, especially in more complicated scenarios of treatment heterogeneity.

### 1.2.7 Processes and mediators

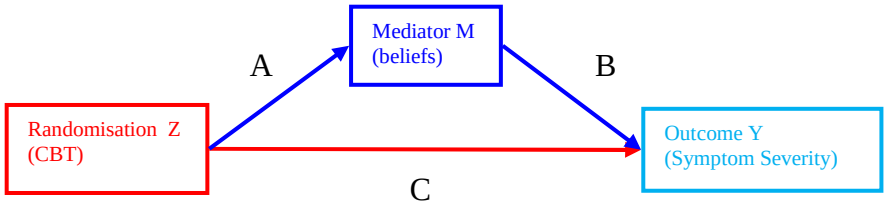
Once a treatment effect has been established the next task is to find out how it works, or more precisely if it is working as intended. Understanding the process by which the therapy is effective can lead to further improvements in its content and implementation. There are two forms that this can take; processes and mediators, though the analyses are similar.

Variables specifying the treatment received, for example, attendance at therapy or inclusion of homework in therapy sessions I will refer to as process variables. The process variables occur after randomisation but the processes are specific to the treatment condition

and so will only be available to those randomised to the intervention. They can describe the reasons for differences in treatment effect and highlight important elements of treatment.

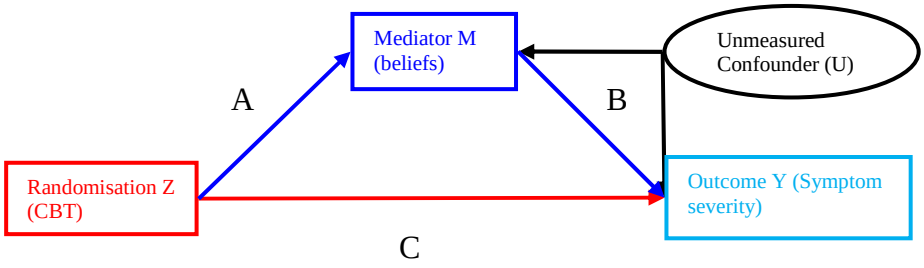
Mediators are factors on the causal pathway which again occur after the treatment allocation. They are hypothesised to be altered by the receipt of treatment and in turn lead to changes in outcome. In CBT trials these would include factors such as a patient's beliefs about their illness. CBT is expected to change the patient's beliefs and this in turn improves their overall symptoms. These factors are measured in both the treatment and control arms. A simple graphical example is given in Figure 1.2 for the situation where therapy influences beliefs which in turn have an effect on symptoms. The treatment may not work entirely through the post randomisation process or mediator, there may still be an effect on outcome even if these factors do not change, this is the direct effect (C) and the indirect effect is that experienced through the mediator (AB).

**Figure 1.2: A graphical representation of a simple mediation**




The direct and indirect effects can be estimated using a least squares regression (Chapter 1.2.2) of Y on Z and M. This is the Baron & Kenny mediation model, described in more detail later in Chapter 1.2.8.1. The model, as shown in Figure 1.2, assumes that there is no unmeasured confounding of the effects of randomisation on the mediator or outcome and that there is no unmeasured confounding of the effect of the mediator on outcome (measured confounders could be included though they are not shown in this graph). When investigating direct and indirect treatment effects the problem arises that the mediator is not randomly assigned; it is an intermediate outcome of treatment. The level of the mediator is likely to be influenced by the individual's characteristics which may also impact their outcome; this is called confounding. A confounder of the mediator and outcome relationship (U in Figure 1.3), means that a participant who improves in the mediator may also improve in their outcome because of other influencing factors for example, higher education, functioning or symptoms. These are examples of measurable and often measured confounders but there are likely to be unmeasured confounders as well that are unknown or cannot be measured.

**Figure 1.3: A graphical representation of a simple mediation with unmeasured confounding**




*1.2.7.1 Mediation – defining direct and indirect treatment effects*

Rubin describes the causal effect of treatment on outcome as $Y_i(1) - Y_i(0)$. When a mediator is introduced, as illustrated in Figure 1.2, this can be broken down into the direct effect, C, and the indirect effect A*B. This process is described by Sobel[100] and Emsley et al.[101] and presented below. For the purposes of defining the direct and indirect terms it is assumed in the following summary that there is perfect adherence to treatment allocation so that treatment allocation is the same as treatment receipt, $D_i(z) = Z_i$ , that is $D_i(1) = 1$, $D_i(0) = 0$ and so $Y_i(z, d, m) = Y_i(z, m)$. The total effect can be written as the difference in outcome if assigned to treatment and receive the mediator at treatment level and outcome when assigned to control and receive the mediator at control level. This is a description of a potential outcomes model since values under both eventualities for the one individual are specified.

$$\text{Total Effect} = Y_i(1) - Y_i(0) = Y_i\big(1, M_i(1)\big) - Y_i\big(0, M_i(0)\big) = \tau_i \qquad \textbf{Eq. 1.1}$$

We wish to break down the total effect in Eq. 1.1 to show direct and indirect effects. To do this we write the total effect as the direct effect of randomisation on outcome holding the mediator constant $Y_i\big(1, M_i(1)\big) - Y_i\big(0, M_i(1)\big)$ plus the effect of a change in the mediator on outcome holding randomisation group constant $Y_i\big(0, M_i(1)\big) - Y_i\big(0, M_i(0)\big)$ (the indirect effect). The $Y_i\big(0, M_i(1)\big)$ terms (in red) cancel out so expressing it in this way does not change the overall model:

$$Y_i\big(1, M_i(1)\big) - Y_i\big(0, M_i(0)\big) = [Y_i\big(1, M_i(1)\big) - Y_i(0, M_i(1))] + [Y_i\big(0, M_i(1)\big) - Y_i(0, M_i(0))] \qquad \textbf{Eq. 1.2}$$

We assume that there is no interaction between mediator and treatment on outcome, that is that the treatment effect is the same at each level of M i.e. $Y_i(1, m) - Y_i(0, m) = Y_i(1, m^*) - Y_i(0, m^*)$. Then the first square brackets of this equation; the direct unmediated effect of randomisation on outcome, C in Figure 1.2, is defined as

$$\text{direct effect of randomisation on outcome} = Y_i(1, m) - Y_i(0, m) = \beta_{zi}$$

This provides expressions for A, C and the total effect C+A*B. The effect of the treatment via the mediator, the second part of Eq. 1.2 and A*B in the diagram, is defined for an individual as:

$$Y_i\big(0, M_i(1)\big) - Y_i\big(0, M_i(0)\big) = \beta_{mi}\big(M_i(1) - M_i(0)\big)$$

Where $\beta_m$is the effect coefficient for the mediator on outcome. The effect of treatment allocation Z on the mediator (A in Figure 1.2) for individual i is:

$$\text{Total effect of randomisation on mediator} = M_i(1) - M_i(0) = \gamma_i$$

The total effect for an individual by potential outcomes can therefore be written as:

$$\tau_i = Y_i(1) - Y_i(0) = \beta_{zi} + \beta_{mi}\big(M_i(1) - M_i(0)\big) = \beta_{zi} + \beta_{mi}\gamma_i \qquad \textbf{Eq. 1.3}$$

Distinguishing the parts of the effect in this way means that the effect attributable to randomisation and the mediator can be calculated separately. Unfortunately, under these definitions the outcomes are observed under both treatment conditions in each individual, indicated by the i subscripts. Instead, the sample is aggregated to give average treatment effects $\beta_z$ and $\beta_m$ which, under certain assumptions, will allow causal interpretation. In the description above these are still dependent on unobserved potential outcomes rather than observations. In the next section we generalise to observed data.

#### *1.2.7.1.1 Definition of direct and indirect effects – generalising for observed data*

The description of direct and indirect effects in the previous section is given in potential outcomes notation. Here it is generalised to describe the breakdown of direct and indirect effects in terms of observed data. First the description in Eq. 1.3 is recalled

$$Y_i\big(1, m_i(1)\big) - Y_i\big(0, m_i(0)\big) = \beta_{zi} + \beta_{mi}\gamma_i$$

Aggregating this to give individual outcomes in terms of average treatment effects plus some individual error we have:

$$Y_i\big(1, m_i(1)\big) - Y_i\big(0, m_i(0)\big) = E[\beta_{zi}] + \varepsilon_{1i} + E[\beta_{mi}\gamma_i] + \varepsilon_{2i}$$
$$= \beta_z + \beta_m\gamma + \varepsilon_i$$
Where $\boldsymbol{\varepsilon_i = \varepsilon_{1i} + \varepsilon_{2i}}$

Values of the outcome and mediator under specific conditions are described for individuals as the average within groups plus individual error:

$$Y_i(z = 0, m = 0) = E[Y_i(0,0)] + \varepsilon_i = \alpha_0 + \varepsilon_i$$
$$M_i(z = 0) = E[M_i(0)] + e_i = \delta_o + e_i$$
$$Y_i\big(0, m_i(0)\big) = Y_i(0,0) + \beta_m * M_i(0)$$
$$= \alpha_0 + \varepsilon_i + \beta_m * (\delta_o + e_i)$$

$$Y_i\big(1, m_i(1)\big) = Y_i\big(0, m_i(0)\big) + \beta_z + \beta_m\gamma$$
$$= \alpha_0 + \varepsilon_i + \beta_m * (\delta_o + e_i) + \beta_z + \beta_m\gamma$$
$$= \alpha_0 + \varepsilon_i + \beta_z + \beta_m(\delta_o + e_i + \gamma)$$

The observed outcome for any participant with observed values $Z_i$ and $M_i$ is defined as:

$$Y_i(Z_i, M_i) = Z_i * Y_i\big(1, m(1)\big) + (1 - Z_i) * Y_i\big(0, m(0)\big)$$

Substituting in the above values this becomes:

$$Y_i(Z_i, M_i) = Z_i * (\alpha_0 + \varepsilon_i + \beta_z + \beta_m(\delta_o + e_i + \gamma)) + (1 - Z_i) * (\alpha_0 + \varepsilon_i + \beta_m * (\delta_o + e_i))$$
$$= \alpha_0 + \varepsilon_i + \beta_m(\delta_o + e_i) + Z_i * \beta_z + Z_i * \beta_m(\delta_o + e_i + \gamma) - Z_i * \beta_m * (\delta_o + e_i)$$
$$= \alpha_0 + \varepsilon_i + Z_i * \beta_z + \beta_m(\delta_o + e_i) + Z_i * \beta_m * \gamma$$
$$= \alpha_0 + Z_i * \beta_z + \beta_m * (\delta_o + e_i + Z_i * \gamma) + \varepsilon_i$$

$$Y_i(Z_i, M_i) = \alpha_0 + \beta_z * Z_i + \beta_m * M_i + \varepsilon_i \qquad \textbf{Eq. 1.4}$$

Where $\alpha_0$ is the average outcome in the control group with the mediator at control group level, $\beta_z$ is the average direct effect of randomisation when there is no change in the mediator and $\beta_m$ is the average additional effect of the post-randomisation mediator on outcome. The aim is to estimate both $\beta_z$ and $\beta_m$. This expression describes the observed outcome and is similar to the definition of the potential outcome for individual i under conditions $Z_i$=z and $M_i$=m:

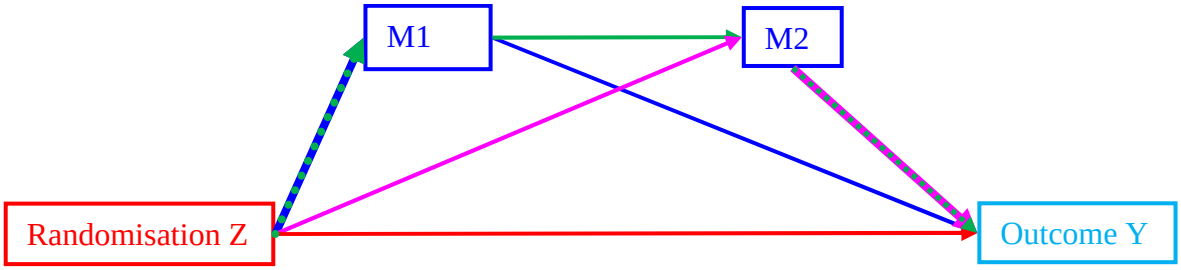
$$Y_i(z, m) = Y_i(0,0) + \beta_{zi} * Z + \beta_{mi} * M$$

Applying the example of treatment adherence in Eq. 1.4 gives a special case of this situation where $M_i$ is treatment adherence and Z remains as a randomisation indicator. In this situation we would assume that $\beta_z = 0$ indicating no affect of randomisation if the treatment is not received and $\beta_m$ provides the effect of attending therapy sessions.

*1.2.7.2 More than one mediator*

The single mediator model is extended to two mediators, one feeding into another, illustrated in the diagram below.

**Figure 1.4: A graphical representation of two mediators**




In this example the notation changes slightly:

$M_{1i}(z)$ – the level of mediator 1 for individual i at randomisation z

$M_{2i}(z, m_1)$ – the level of the mediator 2 for individual i at randomisation z and level of mediator 1 at $m_1$.

$Y_i(z, M_{1i}(z), M_{2i}(z, m_1))$ - the outcome for individual i under randomisation z, with mediator 1 at the level of z and mediator 2 at the level of z and $m_1$.

Extending Eq. 1.1 above the total effect of randomisation is defined as

$$Y_i(1) - Y_i(0) = Y_i\big(1, M_{1i}(1), M_{2i}(1, M_{1i}(1))\big) - Y_i(0, M_{1i}(0), M_{2i}(0, M_{1i}(0)))$$

As previously this is broken up into a sum of the direct and indirect effects. The total effect is therefore a sum of the direct effect of randomisation (red line) plus three indirect effects through mediator 1 only (blue line), mediator 2 only (pink line) and both mediator 1 and 2 (green line).

Each of these can be defined as follows:

Direct effect= $Y_i\big(1, M_{1i}(1), M_{2i}(1, 1)\big) - Y_i\big(0, M_{1i}(1), M_{2i}(1, 1)\big)$

Indirect effect through mediator 1 only

$=Y_i\big(0, M_{1i}(1), M_{2i}(0, 0)\big) - Y_i\big(0, M_{1i}(0), M_{2i}(0,0)\big)$

Indirect effect through mediator 2 only

$=Y_i(0, M_{1i}(0), M_{2i}(1,0)) - Y_i(0, M_{1i}(0), M_{2i}(0,0))$

Indirect effect through mediator 1 and 2

$=Y_i(0, M_{1i}(0), M_{2i}(0,1)) - Y_i(0, M_{1i}(0), M_{2i}(0,0))$

These represent the effects of the variable in question when holding other variables constant at zero. If it is assumed, as previously, that the effect is the same regardless of the level at which the other variables are held then it can be shown that the sum of these gives us the total effect of randomisation.

For example:

Indirect effect through mediator 1 $=Y_i(0, M_{1i}(1), M_{2i}(0,0)) - Y_i(0, M_{1i}(0), M_{2i}(0,0)) = Y_i(0, M_{1i}(1), M_{2i}(1,1)) - Y_i(0, M_{1i}(0), M_{2i}(1,1))$

Indirect effect through mediator 2 $=Y_i(0, M_{1i}(0), M_{2i}(1,0)) - Y_i(0, M_{1i}(0), M_{2i}(0,0)) = Y_i(0, M_{1i}(0), M_{2i}(1,1)) - Y_i(0, M_{1i}(0), M_{2i}(0,1))$

Total effect =sum of direct and indirect effects =

$$\begin{aligned} &Y_i(1, M_{1i}(1), M_{2i}(1,1)) \\ &-Y_i(0, M_{1i}(1), M_{2i}(1,1)) + Y_i(0, M_{1i}(1), M_{2i}(1,1)) \\ &-Y_i(0, M_{1i}(0), M_{2i}(1,1)) + Y_i(0, M_{1i}(0), M_{2i}(1,1)) \\ &-Y_i(0, M_{1i}(0), M_{2i}(0,1)) + Y_i(0, M_{1i}(0), M_{2i}(0,1)) \\ -Y_i(0, M_{1i}(0), M_{2i}(0,0)) &= Y_i(1, M_{1i}(1), M_{2i}(1,1)) - Y_i(0, M_{1i}(0), M_{2i}(0,0)) \end{aligned}$$

The direct effect of randomisation on outcome holding both mediators at some constant level can be taken from the single mediator model defined previously in Chapter1.2.7.1:

$$\begin{aligned} &Y_i(1, M_{1i}(1), M_{2i}(1,1)) - Y_i(0, M_{1i}(1), M_{2i}(1,1)) \\ &= Y_i(1, M_{1i}(m1), M_{2i}(m2)) - Y_i(0, M_{1i}(m1), M_{2i}(m2)) = \beta_{zi} \end{aligned}$$

The indirect effects through each individual mediator (red and blue lines) are similar to the model for a single mediator scenario but with the additional restriction that the remaining mediator is held at a constant level.

Indirect effect through mediator 1 only

$$IE1_i = Y_i(0, M_{1i}(1), M_{2i}(0,0)) - Y_i(0, M_{1i}(0), M_{2i}(0,0)) = \beta_{m1Di}(M_{1i}(1) - M_{1i}(0))$$

Where the effect of randomisation on mediator 1 is defined as:

$$M_{1i}(1) - M_{1i}(0) = \gamma_{m1i}$$

and $\beta_{m1Di}$is the direct effect of mediator 1 on outcome that is not through mediator 2. The total effect of mediator 1 would then be described as:

$$\beta_{m1i} = \beta_{m1Di} + \beta_{m2i}[M_{2i}(0,1) - M_{2i}(0,0)]$$

Similarly, the indirect effect through mediator 2 only

$$IE2_i = Y_i\big(0, M_{1i}(0), M_{2i}(1,0)\big) - Y_i\big(0, M_{1i}(0), M_{2i}(0,0)\big) = \beta_{m2i}(M_{2i}(1,0) - M_{2i}(0,0))$$

Where the effect of randomisation on mediator 2 is:

$$M_{2i}(1,0) - M_{2i}(0,0) = \gamma_{m2i}$$

and $\beta_{m2i}$is the effect of mediator 2 on outcome.

This leaves only the joint indirect effect of both mediators on the outcome which is a product of the effect of randomisation of mediator 1 ($\gamma_{m1i}$), the effect of mediator 1 on mediator 2 $[M_{2i}(0,1) - M_{2i}(0,0)]$ and the effect of mediator 2 on outcome ($\beta_{m2i}$). It is assumed that the effect of a change in mediator 2 on outcome is the same regardless of how that change has come about, i.e. whether it is due to the direct effect of randomisation or the effect through mediator 1, the joint indirect effect is therefore defined as:

$$Y_i\big(0, M_{1i}(0), M_{2i}(0,1)\big) - Y_i\big(0, M_{1i}(0), M_{2i}(0,0)\big)$$

$$= \gamma_{m1i} * \beta_{m2i} * [M_{2i}(0,1) - M_{2i}(0,0)] = \gamma_{m1i} * \beta_{m2i} * \gamma_{m12i}$$

And the average total effect becomes

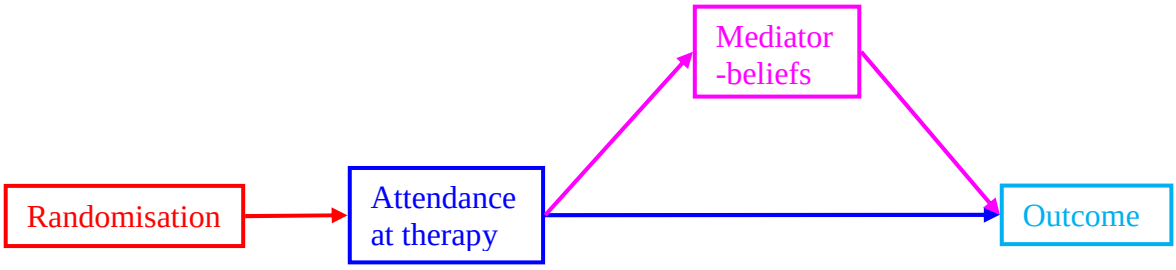
$$E\left[Y_i\left(1, M_{1i}(1), M_{2i}\big(1, M_{1i}(1)\big)\right) - Y_i\left(0, M_{1i}(0), M_{2i}\big(0, M_{1i}(0)\big)\right)\right]$$

$$= \beta_z + \beta_{m1D}\gamma_{m1} + \beta_{m2}\gamma_{m2} + \beta_{m2}\gamma_{m1}\gamma_{m12} \qquad \textbf{Eq. 1.5}$$

#### *1.2.7.2.1 A special case of the two mediator model: sessions and beliefs*

The model is simplified to establish the joint effect of attendance at therapy sessions and changes in beliefs about illness on outcome. In this scenario there is no direct causal effect of randomisation on outcome or on beliefs about illness other than through the effect on attendance at therapy. Additionally attendance at sessions is set to zero for all individuals in the control arm since they are unable to access therapy, the second mediator is allowed to vary in both arms. The diagram in Figure 1.4 can therefore be reduced to that shown in Figure 1.5.

**Figure 1.5: A graphical representation of mediation by sessions and beliefs**




In this example it is expected that $\beta_z = 0$ and $\gamma_{m2} = 0$. The total effect given in Eq. 1.5 above then reduces to

$$Y_i\left(1, M_{1i}(1), M_{2i}\left(1, M_{1i}(1)\right)\right) - Y_i\left(0, M_{1i}(0), M_{2i}\left(0, M_{1i}(0)\right)\right)$$
$$= \gamma_{m1i}\beta_{m1Di} + \gamma_{m1i}\beta_{m2i}[M_{2i}(1, M_{1i}(1)) - M_{2i}(1, M_{1i}(0))]$$

*1.2.7.2.2 Estimation*

As previously, values of the outcome and mediators under specific conditions are described for individuals as the average within groups plus individual error:

$$Y_i(0,0,0) = E[Y_i(0,0,0)] + \varepsilon_i = \alpha_0 + \varepsilon_i$$
$$M_{1i}(0) = 0$$
$$M_{1i}(1) = E[M_{1i}(1)] + e_i = m_1 + e_i$$
$$M_{2i}(0, M_{1i}(0)) = M_{2i}(1, M_{1i}(0)) = E[M_{2i}(0)] + e_{m2i} = m_2 + e_{m2i}$$
$$M_{2i}(1, M_{1i}(1)) = M_{2i}(0,0) + M_{1i}(1) * \gamma_{m12} = m_2 + e_{m2i} + \gamma_{m12}(m_1 + e_i)$$
$$Y_i(1,1,1) = E[Y_i(0,0,0)] + \varepsilon_i + \beta_{m1D}(m_1 + e_i) + \beta_{m2}\gamma_{m1}(m_2 + e_{m2i} + \gamma_{m12}(m_1 + e_i))$$
$$= \alpha_0 + \varepsilon_i + (\beta_{m1D} + \beta_{m2}\gamma_{m1}\gamma_{m12}) * (m_1 + e_i) + \beta_{m2}\gamma_{m1}(m_2 + e_{m2i})$$

The observed outcome for any participant with observed values $Z_1$ , $M_{1i}$and $M_{2i}$is defined as:

$$Y_i(z, m1, m2) = Z_i * Y_i(1, M_1(1), (M_2(1,1)) + (1 - Z_i) * Y_i(0, M_1(0), M_2(0,0))$$
$$= Z_i * [\alpha_0 + \varepsilon_i + (\beta_{m1D} + \beta_{m2}\gamma_{m1}\gamma_{m12}) * (m_1 + e_i) + \beta_{m2}\gamma_{m1}(m_2 + e_{m2i})] + (1 - Z_i) * [\alpha_0 + \varepsilon_i]$$
$$= \alpha_0 + \varepsilon_i + Z_i(\beta_{m1D} + \beta_{m2}\gamma_{m1}\gamma_{m12}) * (m_1 + e_i) + Z_i\beta_{m2}\gamma_{m1}(m_2 + e_{m2i})$$

An estimate of the total effect of mediator 1 $\beta_{m1D} + \beta_{m2}\gamma_{m1}\gamma_{m12}$ and the total effect of mediator 2 $\beta_{m2}\gamma_{m1}$ is required. It is believed that there may be confounding between the two mediators and between these and the outcome that cannot be account for, that the

errors $\varepsilon_i, e_i$ and $e_{m2i}$ are correlated. If this is the case then a standard regression of the two mediators on the outcome will give biased effect estimates.

### 1.2.8 Estimation methods for mediation analysis

Recall the basic mediation model from Eq. 1.4 for a randomised trial with randomised treatment allocation denoted by Z and mediator observed in both arms denoted by M:

$$Y_i(Z_i, M_i) = \alpha_0 + \beta_z * Z_i + \beta_m * M_i + \varepsilon_i$$

We would like to estimate the parameters $\beta_z$ and $\beta_m$.

#### *1.2.8.1 Baron and Kenny's mediation model*

The most cited method for mediation analysis is that of Baron and Kenny[102] who propose a 4-step process to determine if a factor is a mediator and estimate the parameters (see also Judd and Kenny[103]). The parameters $\beta_z$ and $\beta_m$ are estimated using an ordinary least squares regression model (Chapter 1.2.2) of $Y_i(Z_i, M_i) = \alpha_0 + \beta_z * Z_i + \beta_m * M_i + \varepsilon_i$ above.

Baron and Kenny stipulate that four separate criteria must be met in order to determine if mediation is present; that there is a significant association between:

1. the treatment and outcome (regress Y on Z); $Y = \mu_1 + dZ + \varepsilon_1$
2. the treatment and mediator (regress M on Z to estimate A in Figure 1.2); $M = \mu_2 + AZ + \varepsilon_2$
3. the mediator and outcome (regress Y on M to estimate B in Figure 1.2); $Y = \mu_3 + BM + \varepsilon_3$
4. and that the mediator reduces the treatment effect on the outcome when both are included in the regression (regress Y on Z and M); $Y = \mu_0 + \beta_z Z + \beta_m M + \varepsilon_4$

The estimates of $\beta_z$ and $\beta_m$ from this Baron and Kenny model ($\hat{\beta}_z^{OLS}$ and $\hat{\beta}_m^{OLS}$) are therefore taken from the OLS estimation of the model in 4. The direct effect of the treatment (C in Figure 1.2) is estimated by the treatment effect when adjusted for the mediator. The indirect effect is the multiple of the treatment to mediator (A) and mediator to outcome (B) effects. The total treatment effect can be computed as the sum of the direct and indirect effects (A*B+C).

A further criterion to the Baron & Kenny model has been suggested by work from The MacArthur Foundation group in order to determine mediation and moderation. They state

that knowledge of the temporal order of events is necessary; a mediator on the treatment pathway must occur after treatment, a moderator must occur before[104].

This Baron and Kenny model is based on the associative model described in Chapter 1.2.2 and as such the effect estimates $\hat{\beta}_z^{OLS}$ and $\hat{\beta}_m^{OLS}$ can only be interpreted as causal effects if there is no unmeasured confounding present; that there is no characteristic that may influence the value of the covariates M and Z that may also influence the outcome which hasn't been accounted for. This causes no problem in the first two steps where the total effect of randomisation on the outcome and the effect on the mediator is assessed, since randomisation should ensure that no confounding is present. However, when the effect of the mediator on outcome is measured randomisation is no longer present and the estimate may be biased.

Since the value of the mediator has not been assigned randomly it is dependent upon the individual participant. If all factors associated with the mediator status and outcome are accounted for then the estimate will be unbiased. However, if there are factors that are not accounted for then the error term will be correlated with the mediator and both the treatment and mediator effect estimates will be biased. There has therefore been an emphasis on developing estimation methods that do not require this assumption and will account for possible unmeasured confounding in order to give an unbiased estimate of the direct and indirect effects.

#### *1.2.8.2 Instrumental variables (IV)*

The instrumental variable (IV) estimation approach has been a popular method in econometrics to account for associations between the mediator and outcome. Angrist and Krueger[105] explain that the IV answers the problem of unmeasured confounding by only using the variability in the mediator that can be accounted for by observed variables, and not using the part that could be confounded by other factors. The observed variables that are used to explain the variability must not themselves be associated with the outcome. We begin with the mediation model described in Eq. 1.4, extended below to include baseline covariates X:

$$Y_i(Z_i, M_i) = \alpha_0 + \beta_X * X_i + \beta_z * Z_i + \beta_m * M_i + \varepsilon_i$$

Where $\beta_z = E[Y_i(1, m) - Y_i(0, m)|\, X]$ and $\beta_m = E[Y_i(z, M(1)) - Y_i(z, M(0))|\, X]$, $\beta_x$ is effect of baseline covariates X on outcome and $\varepsilon_{zm}$ is a zero mean error term.

In a simple model with no mediator the estimate $\beta_z$ is expected to be unbiased since Z is randomisation status and is assumed to be independent of the error term; $\mathrm{cov}(Z_i, \varepsilon_i) = 0$. When a mediator is considered, that is not randomised, there may be unmeasured confounding present between the mediator and outcome. In this case M and ε are associated, $\mathrm{cov}(M_i, \varepsilon_i) \neq 0$ and the OLS estimator of $\beta_m$will be biased. The bias in the estimate of $\beta_m$ will mean that $\beta_z$is also biased. Instead a new variable V is introduced to estimate M which is strongly associated with the mediator M but not with the outcome Y (except through M) and so is independent of ε, this is called an instrumental variable.

$$M_i = \alpha V_i + e_i$$

In order to act as an instrument the variable V must satisfy two conditions, that it is associated with the mediator M, so $\alpha \neq 0$ and $\mathrm{cov}(M_i, V_i) \neq 0$ and is only associated with Y through M, not in any other way, this is called the 'exclusion restriction' and means that there must be no correlation between the instrument V and any other explanatory variables of Y, $\mathrm{cov}(V_i, \varepsilon_i) = \mathrm{cov}(V_i, Z_i) = \mathrm{cov}(V_i, X_i) = 0$.

To derive the IV estimate from the equation above the covariance with the instrument V is assessed and re-arranged to describe the error

$$\mathrm{cov}(V_i, Y_i) = \beta_x \mathrm{cov}(V_i, X_i) + \beta_z \mathrm{cov}(V_i, Z_i) + \beta_m \mathrm{cov}(V_i, M_i) + \mathrm{cov}(V_i, \varepsilon_i)$$

$$\beta_m \mathrm{cov}(V_i, M_i) = \mathrm{cov}(V_i, Y_i)$$

The instrumental variable estimate of the population parameter $\beta_m$ is therefore:

$$\hat{\beta}_m^{IV} = \frac{\mathrm{cov}(V_i, Y_i)}{\mathrm{cov}(V_i, M_i)} = \frac{E[V_i Y_i] - E[V_i]E[Y_i]}{E[V_i M_i] - E[V_i]E[M_i]}$$

The estimate of the sample parameter

$$\hat{\beta}_m^{IV} = \frac{\frac{1}{n}\sum_{i=1}^{n} V_i Y_i - \frac{1}{n^2}\sum_{i=1}^{n} V_i \sum_{i=1}^{n} Y_i}{\frac{1}{n}\sum_{i=1}^{n}(V_i M_i) - \frac{1}{n^2}\sum_{i=1}^{n}(V_i)\sum_{i=1}^{n} M_i} = \frac{\sum_{i=1}^{n}(V_i - \bar{V})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(V_i - \bar{V})(M_i - \bar{M})}$$

This is reminiscent of the ordinary least squares estimate and it is clear that when V=M, i.e. the mediator rather than the instrument is used, the formula gives the OLS estimate of $\beta_m$.

Consider the special case of the mediator M being attendance at therapy sessions. It is assumed that there is no effect of therapy other than through attendance at therapy sessions, $\beta_Z = 0$. The estimating equation therefore becomes:

$$Y_i(Z_i, M_i) = \mu_0 + \beta_x X_i + \beta_m M_i + \varepsilon_i$$

This means that the randomisation indicator Z can act as an instrument on the mediator sessions if it satisfies the requirements of an instrument.

1. $\text{cov}(M_i, Z_i) \neq 0$, it is assumed that allocation of treatment is associated with the number of therapy sessions attended. This is true of the case where the control group are not able to receive therapy and so the number of sessions attended should be zero in this group. This is an assumption that can be tested in the data.
2. $\text{cov}(Z_i, \varepsilon_i) = \text{cov}(Z_i, X_i) = 0$, due to randomisation it is assumed that pre-randomisation covariates and the error terms are independent of allocation.

In this situation using only randomisation as an instrument for the mediator (sessions) $\hat{\beta}_m^{IV}$ would effectively be the ITT treatment effect divided by the average number of sessions in the treated group[89].

When a direct effect of randomisation group cannot be excluded or multiple mediators are involved additional instruments must be found. Baseline covariates by randomisation interactions may be used as instruments if it is valid to assume that they maintain the requirements of an IV. The model for $M_i$ above would then also include $Z_i$*$X_i$ interaction terms where $X_i$ represents baseline covariates:

$$M_i = \alpha_z Z_i + \alpha_x X_i + \alpha_{xz}(X_i * Z_i) + e_i$$

$$Y_i(Z_i, M_i) = \mu_0 + \beta_x X_i + \beta_z Z_i + \beta_m M_i + \varepsilon_i$$

**Eq. 1.6**

To act as an instrument the interaction, shortened to $X_iZ_i$ for ease, must satisfy the two criteria defined previously:

1. The interaction is associated with the mediator $\alpha_{xz} \neq 0$ and $\text{cov}(M_i, X_iZ_i) \neq 0$
2. There is no correlation between the interaction and the error term of Y, $\text{cov}(X_iZ_i, \varepsilon_i) = 0$

The first requirement is testable in the data by regressing the mediator on randomisation, the covariate and the interaction between randomisation and covariates. If a significant interaction is found then criterion 1 is met. This interaction is interpreted as there being a different effect of the covariate on mediator in the two randomisation groups. In the situation where the effects are allowed to vary between individuals Eq. 1.3 Small[106] demonstrates that the second requirement is valid if the following additional assumptions can be made:

A. The average direct effect of randomisation on outcome and the average effect of the mediator on outcome are independent of covariates; effects are the same at all levels of the covariates $E[\beta_{zi}|X_i = X] = \beta_z$ and $E[\beta_{Mi}|X_i = X] = \beta_M$
B. The effect of the mediator on outcome and the value of the mediator are independent given the randomisation and baseline covariates; effects are the same at all levels of the mediator for people with the same baseline characteristics and treatment $M_i \perp \beta_{Mi}|Z_i, X_i$ .

Small's[106] proof that if assumptions A and B are valid then randomisation by covariate interactions are valid instruments is described below. In order to demonstrate that group by covariate interactions are valid instruments he starts with defining the error as a sum of the differences between the individual and average treatment effects for each component of Eq. 1.6.

$$\varepsilon_i = Y_i(0,0) - E[Y_i(0,0)|X_i] + (\beta_{zi} - \beta_z)Z_i + (\beta_{mi} - \beta_m)M_i$$

**Eq. 1.7**

To test the requirement of no association with the error term of Y the equality in criteria 2 is applied to Eq. 1.7

$$\begin{aligned}\mathrm{cov}(X_iZ_i, \varepsilon_i) = \mathrm{cov}(X_iZ_i, Y_i(0,0) - E[Y_i(0,0)|X_i]) + \mathrm{cov}(X_iZ_i, (\beta_{zi} - \beta_z)Z_i) \\ + \mathrm{cov}(X_iZ_i, (\beta_{mi} - \beta_m)M_i) = 0\end{aligned}$$

To demonstrate that this is true each component is considered separately, starting with the direct effect of randomisation. The second covariance is defined as usual in terms of expectations:

$$\mathrm{cov}(X_iZ_i, (\beta_{zi} - \beta_z)Z_i) = E[X_iZ_i(\beta_{zi} - \beta_z)Z_i] - E[X_iZ_i]E[(\beta_{zi} - \beta_z)Z_i)]$$

Due to randomisation it can be assumed that $E[(\beta_{zi} - \beta_z)Z_i)] = E[Z_i]E[(\beta_{zi} - \beta_z)]$=0, this states that the treatment effect is independent of treatment allocation; the effect of receiving treatment will be the same whether randomised to the treatment or control group. This then leaves the first term $E[X_iZ_i(\beta_{zi} - \beta_z)Z_i] = E[Z_i^2]E[X_i(\beta_{zi} - \beta_z)]$ since $Z_i$ is randomised. Applying assumption A the treatment effect is independent of covariates so $E[Z_i^2]E[X_i(\beta_{zi} - \beta_z)] = E[Z_i^2]E[X_i]E[(\beta_{zi} - \beta_z)] = 0$.

Repeating the same process on the effect of the mediator gives:

$$\text{cov}(X_iZ_i, (\beta_{mi} - \beta_m)M_i) = E[X_iZ_i(\beta_{mi} - \beta_m)M_i] - E[X_iZ_i]E[(\beta_{mi} - \beta_m)M_i]$$

Small first employs a property of conditional expectation[107] that $E\big[E[X\,|Y]\big] = E[X]$ to state the equality $E[(\beta_{mi} - \beta_m)M_i] = E[E[(\beta_{mi} - \beta_m)M_i|Z_i, X_i]]$ applying assumption B that the mediator effect is independent of mediator level conditional on $Z_i$ and $X_i$ allows the terms to be separated out to give $E[(\beta_{mi} - \beta_m)M_i] = E[E[(\beta_{mi} - \beta_m)|Z_i, X_i]E[M_i|Z_i,\ X_i]]$ and since $E[(\beta_{mi} - \beta_m)|Z_i, X_i] = 0$ the whole term is zero.

Repeating the method again on the first term of the equality gives

$$E[X_iZ_i(\beta_{mi} - \beta_m)M_i] = E[E[X_iZ_i(\beta_{mi} - \beta_m)M_i|Z_i, X_i]] = E[X_iZ_iE[(\beta_{mi} - \beta_m)M_i|Z_i, X_i]]$$

The second expectation is now equivalent to the term above which has been shown to be equal to zero if assumption B and randomisation are valid.

The final expression is

$$\begin{aligned}&\text{cov}(X_iZ_i, Y_i(0,0) - E[Y_i(0,0)|X_i])\\ &\qquad = E[X_iZ_i(\,Y_i(0,0) - E[Y_i(0,0)|X_i])] - E[X_iZ_i]E[\,Y_i(0,0) - E[Y_i(0,0)|X_i]]\end{aligned}$$

Since $Z_i$ is randomised and so independent of baseline values of $Y_i$ and $X_i$ we have

$$\begin{aligned}E[X_iZ_i(\,Y_i(0,0) - E[Y_i(0,0)|X_i])] &= E[Z_i]E[X_i(\,Y_i(0,0) - E[Y_i(0,0)|X_i])]\\ &= E[Z_i](E[X_iY_i(0,0)] - E[X_iE[Y_i(0,0)|X_i]])\\ &= E[Z_i](E[X_iY_i(0,0)] - E[X_iY_i(0,0)]) = 0\end{aligned}$$

It is therefore shown using Small's proof that randomisation by covariate interactions are valid instruments if assumptions A and B are considered to be true. This definition has been described in the simplified case of one covariate by randomisation interaction but can

be extended to multiple covariate interactions in which case each would have to satisfy the above criteria.

#### *1.2.8.2.1 IV estimation methods*

##### *1.2.8.2.1.1 Two-stage least squares*

The most widely used estimation method for instrumental variables is two-stage least squares where the instrumental variables analysis specified above is effectively carried out in two stages. In the first stage the mediator is modelled on the IV, that is M is regressed on the instrument in an ordinary least squares estimation. Here the randomisation indicator Z is the instrument:

$$M_i = \alpha Z_i + e_i$$

using the estimate of α we predict fitted values for the mediator, $\widehat{M}_i = \alpha Z_i$ for all observations. These predicted mediator values are then used in the regression on the outcome to estimate $\beta_m$:

$$Y_i = \mu_0 + \beta_m \widehat{M}_i + \varepsilon_i$$

The 2SLS estimate of the mediator $\beta_m$is therefore an OLS estimate with the predicted values $\widehat{M}_i$ in place of X. Details and examples of the process are given by Dunn et al.[89,108] and Maracy and Dunn[109]. The standard errors of the treatment effects will be underestimated if analysed in two steps as described above since it will not account for uncertainty around the predicted values of the mediator. This two-step process is useful to explain the method but in practice it is estimated in one step so that the correct standard errors are given. This has been made possible in several software packages. In the following thesis the ivregress[110] command in the software package Stata[111] will be used.

A variation on this method is the adjusted treatment received IV(ATR)[89] where the residuals of the mediator are predicted and applied as an adjustment in the second part of the estimation model. These would be the distance from the actual to the average value of the mediator for each individual, the treatment effect of the mediator controlling for error. The results are the same as the IV estimation described above.

The method can be extended to introduce more mediators, for example an individual's beliefs about their illness, this would be measurable in both the treatment and control groups. As more mediators are included more instruments must be found, this can be

difficult but the benefit of randomisation is that covariate by randomisation interactions fulfil the criteria of an instrument as has been shown.

#### *1.2.8.2.1.2 G-estimation*

The G-estimation method has come from graphical modelling of causal pathways and the structural equation methods described earlier. The methods were developed by Robins[112] in a survival analysis context and are explained in an example by Fischer-Lapp and Goetghebeur[113]. It is the same as the IV 2SLS estimation when the same covariate by randomisation interactions are used as instruments[109]. The following process describes how G-estimation is applied to solve the mediation model described above:

$$Y_i(Z_i, M_i) = \alpha_0 + \beta_x X_i + \beta_z * Z_i + \beta_m * M_i + \varepsilon_i$$

1. define a linear regression model for response in control group on baseline covariates X.

$$Y_i(0, m) = \alpha_{y0} + \beta_{y0} X_i + \varepsilon_{y0i} \quad \text{for Z=0}$$

Outcome when receiving the control treatment is observed only in those randomised to control. The model is run in this group only to estimate $\alpha_{y0}$ and $\beta_{y0}$. These estimates are used to predict values $\widehat{Y}_i(0)$ for all participants in both the treatment and control groups.

2. define a linear regression model for response in treatment group on baseline covariates X.

$$Y_i(1, m) = \alpha_{y1} + \beta_{y1} X_i + \varepsilon_{y1i} \quad \text{for Z=1}$$

Outcome when receiving treatment is observed only in those randomised to treatment so the model is applied in the treatment group to estimate the parameters $\alpha_{y1}$ and $\beta_{y1}$. The estimates are used to predict values $\widehat{Y}_i(1)$ for participants in both randomisation groups.

3. define the mediator in the control group as a function of the baseline covariates

$$M_i(0) = \alpha_{m0} + \gamma_{mo} X_i + \varepsilon_{m0i} \qquad \text{for Z=0}$$

The same process is repeated for values of the mediator. The mediator is regressed on covariates within those randomised to the control group and predictions of the mediator when under the control condition $\widehat{M}_i(0)$ are made for all.

4. define the mediator in the treatment group as a function of the baseline covariates

$$M_i(1) = \alpha_{m1} + \gamma_{m1}X_i + \varepsilon_{m1i} \qquad \text{for Z=1}$$

The mediator is regressed on covariates within those randomised to the treatment group and predictions of the mediator when under the treatment condition $\widehat{M}_i(1)$ are made for all.

Once these 4 steps have been completed each participant will have a predicted value of the mediator and outcome under both the control and treatment conditions. The treatment effect on the mediator $\widehat{M}_i(1) - \widehat{M}_i(0)$ and the outcome $\widehat{Y}_i(1) - \widehat{Y}_i(0)$ can therefore be calculated for each individual. Regressing the difference in outcome on the difference in mediator we obtain the mediator effect.

$$\widehat{Y}_i(1) - \widehat{Y}_i(0) = \beta_m \left(\widehat{M}_i(1) - \widehat{M}_i(0)\right) + \beta_z$$

As described previously $\beta_z$ is the direct effect of randomisation on the outcome and $\beta_m$is the effect of mediator on outcome.

A special case of the G-estimation model is the example of attendance at therapy sessions as a mediator of the treatment effect, this example is described by Fischer-Lapp and Goetghebeur[113] and reviewed by Maracy and Dunn[109]. In this case the therapy is not available when in the control condition and so the predicted value of the mediator is zero for all participants when assigned to the control $\widehat{M}_i(0) = 0$. In this special case an additional assumption can be made, that there is no direct effect of randomisation other than through the mediator, i.e. there is no treatment effect unless therapy is actually received, this is the exclusion restriction and is applied by stating that $\beta_z = 0$. The restriction is enforced by removing the constant in the final regression and the model for attendance as the mediator therefore becomes

$$\widehat{Y}_i(1) - \widehat{Y}_i(0) = \beta_m \widehat{M}_i(1)$$

The difference with the standard ordinary least squares regression (OLS, Baron & Kenny approach) is that the traditional method requires the assumption that there is no hidden confounding, which is unnecessary by the G-estimation method.

It can be seen that G-estimation can be carried out as a series of simple linear regression models. As with the 2SLS method the standard errors of the effect estimates will be underestimated due to predicted rather than observed values being used. In order to correctly estimate the uncertainty around the estimates the whole process must be bootstrapped. The bootstrapping process is described in detail in Chapter 3.

#### *1.2.8.2.1.3 Generalised method of moments*

The generalised method of moments is an instrumental variables estimator popular in econometrics. GMM bases the estimation of parameters on moment conditions. A moment in mathematics is a quantitative description of a distribution[114]. The $k^{th}$ moment of a variable X, is the expectation of the variable to the $k^{th}$ order, $E[X^k]$, the first order moment of X is therefore $E[X^1]$ which is the mean of X. This can also be described relative to another value, the $k^{th}$ order of X relative to a value a is $E[(X-a)^k]$. In this way the $2^{nd}$ order of X relative to the first order of X is $E[(X-E(X))^2]$, also known as the second-order central moment, or the variance of X.

The model is estimated by forming a series of simultaneous equations based on the assumption that the covariance of the instruments with the error term in the second stage equation is zero. This assumption provides the first-order moment condition $E[Z\varepsilon] = 0$. This is equivalent to the derivation of the 2SLS estimator. Since $\varepsilon = Y - \beta_m M$ the moment condition can be written as:

$$E(Z_i\varepsilon_i) = E\big(Z_i(Y_i - \beta_M M_i)\big) = 0$$

Solving for $\beta_M$ gives:

$$E(\beta_M) = \frac{E(Z_i Y_i)}{E(Z_i M_i)}$$

Which can be expressed in terms of sample estimation as:

$$\hat{\beta}_M = \frac{\frac{1}{n}\sum_{i=1}^{n}(Z_i Y_i)}{\frac{1}{n}\sum_{i=1}^{n}(Z_i M_i)} = \frac{\sum_{i=1}^{n}(Z_i Y_i)}{\sum_{i=1}^{n}(Z_i M_i)}$$

When there is one endogenous variable and one instrument this is equivalent to the 2SLS model. If there are as many endogenous variables as instruments, there are the same amount of moment condition equations as unknown parameters and this can be solved for values of $\beta_M$ that satisfy all of them. However when the system is over-identified, there are more condition equations than unknowns. In this situation it may be difficult to find a single vector for $\beta_M$ that satisfies all of the moment conditions for all Z. Instead the GMM seeks to minimise a quadratic form of the moment conditions to be as close to zero as possible using a weighting matrix giving different weights to each condition. This weighting takes into account the covariance of the residuals and therefore the main benefit

of GMM is in the scenario where the residuals are heteroskedastic. If the residuals have constant variance (homoskedastic) then the GMM estimator is equivalent to the 2SLS estimator[115]. Software packages enable the GMM to be estimated simply in one step, for example GMM estimation is an option in the ivregress command in Stata.

#### *1.2.8.2.2 Principal stratification and CACE analysis*

The use of principal stratification is equivalent to that of instrumental variable methods when a categorical mediator is considered. It was introduced by Frangakis and Rubin[116,117] and has been explored further by Rubin[118], Jin and Rubin[119] and Dunn et al.[108]. Angrist and Imbens[120] describe this under observational study conditions. They estimate the mediator effect on outcome using an IV analysis (using randomisation and randomisation by covariate interactions as possible instruments) but when condensed down to a binary variable it is reduced to a principal stratification analysis with the baseline covariates predicting stratum membership (this will be explained below).

The premise of principal stratification (PS) is that in order to determine a causal effect of a mediator that is not randomised individuals are grouped by their potential mediator levels under the different treatment conditions. Comparisons are made within each group independently of any confounder of the mediator. This has developed from the basic definition of a causal effect calculated as $Y_i(1) - Y_i(0)$ but specifies that $Y_i(1) - Y_i(0) | \, i \in S$ is estimated. Outcomes for subjects within the same group are calculated where the group *S* consists of all individuals who will have the same combination of mediator level under the treatment and control conditions.

In order to understand principal stratification we begin with a special case called the Complier Averaged Causal Effect. In its simplest form this has two randomisation arms and two levels of the post-randomisation process, participants are randomised to a treatment or control and will either comply with the treatment or not (for example, attend at least four therapy sessions or not). These provide four principal strata, they have been named in many different ways but the following will be used for this example;
Compliers – receive what they are allocated $\mathrm{M_i}(0) = 0 \; \mathrm{M_i}(1) = 1$;
Never takers – never receive the treatment regardless of allocation $\mathrm{M_i}(0) = \mathrm{M_i}(1) = 0$;
Always takers – will always receive the treatment regardless of allocation $\mathrm{M_i}(0) = \mathrm{M_i}(1) = 1$;

Defiers - will always receive the opposite treatment to their allocation $M_i(0) = 1\ M_i(1) = 0$.

In an RCT these four groups cannot be distinguished, for example, those that are randomised to treatment and take it could be either compliers or always takers and if they do not take the treatment they could be defiers or never takers (Table 1.1). Additional assumptions must be made in order to be able to identify individuals under these classifications. It seems sensible for example, to assume that there will be no defiers (that people will not purposefully do the opposite of what they are asked), this is also called the monotonicity assumption ($M_i(1) \geq M_i(0)$). We can additionally assume in many RCTs that there are no always-takers since the treatment is only available to those allocated to treatment.

**Table 1.1: Categorisation of the four observed groups for the binary attendance mediator**

| | **<4 sessions** | **>=4 sessions** |
|---|---|---|
| **Treatment** | M(1)=0<br>Never-taker/Defier | M(1)=1<br>Complier/ always-taker |
| **Control** | M(0)=0<br>Complier/ never-taker | M(0)=1<br>Always-taker/defier<br>(Assume not possible in RCT) |

This leaves only two possible strata. It is clear now that a person allocated to therapy who does not receive it must be a never-taker and a person allocated to therapy who takes it must be a complier. However, a participant allocated to the control condition and does not receive the treatment could be a complier or a never-taker. In order to estimate the average treatment effect in the compliers two further assumptions must be made.

1. The proportions of the three classes are (on average) the same in the two arms of the trial. This is valid due to randomisation.
2. The effect of allocation on outcome in the never-takers is zero (the so-called exclusion restriction). Since there is no effect of randomisation on treatment received $M_i(1) - M_i(0) = 0$ then any observed effect must be direct and $Y_i(z, M_i(1)) - Y_i(z, M_i(0)) = 0$.

The overall treatment effect (intention to treat effect) can be expressed as a sum of the intention to treat effects within the two strata weighted by the proportion of individuals in each stratum:

$$ITT_{all} = p_{never} * ITT_{never} + p_{complier} * ITT_{complier}$$

Which can be written as:

$$ITT_{all} = (1 - P_{comply}) * ITT_{non-comply} + P_{comply} * ITT_{comply}$$

Where $P_{comply}$ is the proportion of the treated group who comply and $ITT_{all}$, $ITT_{non-comply}$ and $ITT_{comply}$ are the intention to treat effects of randomisation on outcome overall, in non-compliers and compliers respectively.

Applying the first assumption the proportion of compliers is determined by the proportion of the treated group that receive therapy. Since by assumption two, the exclusion restriction $ITT_{never} = 0$, it is clear that the treatment effect in compliers can be calculated as the overall treatment effect divided by the proportion of compliers.

The CACE can be estimated using the two-stage least squares instrumental variables model described in Chapter 1.2.8.2.1.1 with a binary endogenous variable. It is valid to use a binary variable in the model as the 2SLS method is not dependent on the endogenous variable being normally distributed. A binary indicator of compliance, for example attendance at a minimum of four therapy sessions, is included as the instrumented variable and randomisation as the instrumental variable.

“Compliance” can be defined in any way, for example inclusion of formulation in therapy sessions or not, homework received or not. These binary indicators can be analysed as a CACE estimator using 2SLS regression. The CACE analysis however makes the assumption that there is no direct effect of randomisation in the non-compliers, those that do not receive homework for example. This is the exclusion restriction and it may not always be a valid assumption. In this example it may be that an individual will attend several therapy sessions and will see some benefit of these even if homework is not part of the therapy given. In order to relax the exclusion restriction we can think of it in terms of a principal stratification problem. In the example of homework received versus homework not received we observe in those allocated to treatment, participants that do not receive anything (never-takers), participants that receive treatment but no homework (never-

takers), those that receive treatment with homework (always-takers or compliers) and the control group that do not receive treatment with or without homework (never-takers or compliers). If we assume that everyone allocated to treatment receives treatment either with or without homework then this is reduced to two groups within the treatment arm and the overall ITT effect can be described by two strata as:

$$\mathrm{ITT}_{\mathrm{all}} = (1 - \mathrm{P}_{\mathrm{homework}}) * \mathrm{ITT}_{\mathrm{no-homework}} + \mathrm{P}_{\mathrm{homework}} * \mathrm{ITT}_{\mathrm{homework}}$$

Where $\mathrm{P}_{\mathrm{homework}}$ is the proportion of individuals that receive homework in the treatment group and $\mathrm{ITT}_{\mathrm{all}}$, $\mathrm{ITT}_{\mathrm{no-homework}}$ and $\mathrm{ITT}_{\mathrm{homework}}$ are the ITT effects of randomisation on outcome overall, in those not receiving homework and those receiving homework respectively. Relaxing the exclusion restriction means that in those that do not receive homework we may still expect to see an ITT effect of randomisation, this is that $\mathrm{ITT}_{\mathrm{never}} \neq 0$. The total effect is therefore a combination of two ITT effects and it is not possible to determine these effects within each strata by simply dividing the total by the proportion in the strata. Instead we use baseline covariates to predict stratum membership; that is we use covariates that predict whether homework is received within the treatment group and apply this to the control group. The treatment effect is then estimated within each stratum. The use of baseline covariates to predict stratum membership is equivalent to using baseline covariates by treatment group interactions in an instrumental variables analysis.

This example of principal stratification with a binary process measure can be extended to a measure with several categories, for example no components of therapy, 1-3 components of therapy or 4 components of therapy. Baseline covariates are again used to predict stratum membership in order to estimate treatment effects within strata.

#### *1.2.8.3 Causal estimation model assumptions*

The estimation models described above, 2SLS, G-estimation, GMM and PS all make the following assumptions:

1. Whichever outcome is unobserved is in fact possible; the individual could potentially receive either the experimental or alternative treatment.

2. Stable Unit Treatment Value Assumption (SUTVA): this consists of two assumptions, firstly that there is independence between individuals and secondly that there is consistency, that the treatment effect or mediator effect is the same regardless of how

treatment is allocated/administered i.e. the treatment effect is the same if the person chooses the treatment or it is given to them by random allocation.

3. Monotonicity – mediator level is not lower when randomised to treatment rather than control $M_i(1) \geq M_i(0)$. In the context of a CACE analysis this means that there are no defiers.

4. Randomisation: there is no mediator by randomisation interaction on outcome. So for example, if attendance is a mediator of the effect of CBT on outcome it is assumed that the effect of attendance in the treatment arm would have the same effect in those assigned to the control arm were they to receive treatment. It does not however, assume that there is no imbalance in the mediator across intervention and control group just that the mediator effect on the outcome is no different in the two groups.

Angrist, Imbens and Rubin[121] consider problems with violations of these assumptions in principal stratification and assess the level of bias in the treatment effect when these assumptions are not true. They specifically concentrate on the monotonicity assumption. The bias associated with violation of the monotonicity assumption is shown to increase as the difference in treatment effect between compliers and defiers increases and the proportion of defiers increases. The authors do not delve into assessing bias due to failure of the SUTVA assumption and possible clustering effects. In each model, the stronger the instrument is (the better it fits the instrumented variable) the less bias incurred if the assumptions are flawed. It is therefore important to ensure that strong instruments are found to reduce the risk of biased estimates.

These methods seek to improve upon the OLS method to infer causation when unmeasured confounding cannot be ignored. These methods are very similar; they have been derived from the concept of potential outcomes, share the same basic assumptions and can be shown to give the same results in particular situations. The instrumental variables approach allows for both continuous and categorical mediators the IV approach will therefore be applied to answer the mediation questions of the EDIE-II and COMMAND trials.

### 1.3 Aims and objectives

The substantive aims of this thesis described in chapter 1.1.7 are reiterated here:

1. Estimate the effect of dose of therapy on symptom reduction in those at high risk of psychosis

2. Estimate the effectiveness of including the following aspects of therapy on reduction of symptoms in those at high risk of psychosis; agreement of problems and goals, formulation, homework and active change strategies
3. Determine if the effectiveness of CBT for high-risk individuals is mediated by changes in beliefs
4. Determine if there is a causal pathway from the amount and content of therapy received through changes in beliefs to a reduction in symptom severity in those at high risk
5. Determine if the effectiveness of CBT on compliance with command hallucinations is mediated by the power of voice

In this chapter the statistical concept of causal inference in mediation analysis and the estimation methods for mediation have been introduced. It has been demonstrated that instrumental variable methods can be used to give unbiased estimates of the treatment and mediator effect in the presence of unmeasured confounding. It has also been shown that randomisation group is a valid instrument for the process variable or mediator if it can be assumed that there is no direct effect of randomisation on the outcome. If this cannot be assumed then baseline covariate by randomisation group interactions can be used as instruments. What has not been tackled so far is the choice of instruments. In practice the interactions are weaker instruments than randomisation and the effect estimates can suffer from large standard errors. Additionally there may be many baseline variables that could act as an instrument when used in an interaction with randomisation. The choice of instruments is an important factor for instrumental variables analysis especially in the absence of prior knowledge and with a potentially weak set of valid instruments to select from. In order to answer the substantive questions this thesis first investigates methods for the selection of instrumental variables and estimation methods to tackle problems associated with bias due to weak instruments and multiple instruments, with the following aims:

1. Determine the best method to select instrumental variables when there is no prior hypothesis and many potential candidates
2. Determine the preferred estimation method in the presence of many, potentially weak instruments
3. Apply these methods to estimate mediation effects in the EDIE-II trial data

**Objectives**

- Summarise problems of bias in instrumental variables
- Compare methods for instrument selection in simulated data.
- Compare methods for instrumental variable estimation in simulated data
- Apply methods for instrumental variable selection and estimation to the EDIE-II and COMMAND trial data to answer the substantive questions.

# 2 Bias of two-stage least squares

## 2.1 Introduction

The 2SLS instrumental variable method is designed to overcome bias that is present in OLS due to unmeasured confounding between post-randomisation variables and outcome. However, there still remains a possibility of bias, especially in finite samples. In addition to bias the variance of the estimator must be taken into account. If the estimator is unbiased but has a large variance then the chance of producing a misleading estimate is increased. Alternatively if the estimator has a small amount of bias but also small variance then the probability of obtaining an estimate close to the true estimate may be improved compared to an unbiased but imprecise estimator. In this chapter the bias present in 2SLS estimates is described and the factors that influence bias are highlighted. It is shown that the strength of the instrument and the number of instruments used are crucial to reducing bias in a post-hoc analysis. Methods for the selection of instruments are considered and estimation methods expected to be robust to the choice of instruments are summarised. Finally, methods for model selection and estimation are compared in simulated datasets, in order to determine the most effective way to analyse clinical trial data relating to the substantive questions of this thesis.

## 2.2 Defining and calculating bias

When using instrumental variables analysis to replace the observed values of an endogenous variable there must be at least as many instruments as endogenous variables. If there are fewer instruments then the model is under-identified and cannot be estimated. If the model is exactly-identified then the effect estimates will be valid if the sample size is fairly large and the number of instruments is less than the sample size. In an over-identified model, where there are more instruments than there are endogenous variables and the instruments are strongly associated with the instrumented variable then the bias is small and the errors are approximately normal. However, if the instruments are weak or the sample is small then bias may be present. In addition to the detectable effect size, variation and power which is necessary in determining sample size in a standard analysis in an instrumental variables analysis, the number of observations needed is dependent on the strength of the instruments[122]. Instrumental variables analysis has been demonstrated in large population studies for example Angrist and Krueger's investigation of education on subsequent salary[123] and Angrist's analysis of military service on civilian earnings[124]. These were both observational studies involving large samples of tens of thousands of

individuals. Despite these large samples Bound et al[125] have suggested that their findings may still be biased due to weak instruments. Weak instrument literature is summarised and illustrated with econometric examples by Murray[126].

Hahn and Hausman[127] give an approximation for the bias of the 2SLS estimator in a simplified example where the outcome is explained by one endogenous variable and the errors of the first and second stage regressions are standardised so that their variances are one. Using the same notation as in previous chapters:

$$Y = \beta_m * M + \varepsilon$$

$$M = \beta_z * Z + \nu$$

The bias of the 2SLS estimator when there are more instruments than endogenous variables is described in terms of the number of instruments z, the covariance between the error terms of the mediator and outcome which in this situation becomes the correlation between the errors $\rho = \mathrm{corr}(\varepsilon, \nu)$, the proportion of variance in the mediator described by the instruments $R^2$ in the first stage of the 2SLS model and the sample size *n*. They give the second-order approximation as:

$$\mathrm{Bias}(\beta_m^{2sls}) = E\left[\beta_m^{2sls}\right] - \beta_m \approx \frac{z\rho}{nR^2 var(M)}$$

Where $\beta_m^{2sls}$is the estimate of the coefficient of the instrumented variable (mediator) in the second stage regression and $\beta_m$ is the true coefficient of the instrumented variable.

The bias of the over-identified 2SLS estimate is also described relative to the bias present in the equivalent OLS model. The 2SLS method is preferred if it reduces bias compared to the OLS method even if there is still some bias present. The bias of the OLS estimate of $\beta_m$ is approximately:

$$\mathrm{Bias}(\beta_m^{ols}) = E\left[\beta_m^{ols}\right] - \beta_m \approx \frac{\mathrm{cov}(M, \varepsilon)}{\mathrm{var}(M)}$$

The endogenous variable M can be described by the instruments and an error term so the bias of the OLS estimate can be written as $\mathrm{Bias}(\beta_m^{ols}) \approx \frac{\mathrm{cov}(\beta_z * Z + \nu, \varepsilon)}{\mathrm{var}(M)}$. Since Z is an instrument $\mathrm{cov}(Z, \varepsilon) = 0$ and $\mathrm{cov}(\beta_z * Z + \nu, \varepsilon) = \mathrm{cov}(\nu, \varepsilon)$. Hahn and Hausman[127]

therefore show that in this simplified and over-identified model the relative bias of the 2SLS compared to the OLS estimator is:

$$\frac{\text{Bias}(\beta_m^{2sls})}{\text{Bias}(\beta_m^{ols})} \approx \frac{z}{nR^2}$$

The relative bias of 2SLS compared to OLS increases as the number of instruments used in the 2SLS model increase but decreases with larger samples and as the amount of variation in the mediator explained by the instrument(s) in the 2SLS model increases. Instrumental variables analysis will therefore have less bias than the OLS method if $z/n > R^2$. Once the trial has been conducted and all measured confounders are accounted for the only values that can be manipulated to reduce bias are the $R^2$ and $z$. The $R^2$ can be increased to reduce bias by introducing more instruments ($z$) but more instruments also increase bias and so adding instruments that only explain a small amount of variation can increase bias.

### 2.3 Identifying weak instruments

Instruments that explain only a small part of the variation in the mediator (have a low first stage $R^2$ value) introduce more bias into the estimates and have been termed 'weak instruments'. When instruments are weak not only can they give biased estimates but the error in the 2SLS model is under-estimated leading to over-confidence in results and potentially unwarranted positive results from significance tests[128,129]. However, a larger variance increases the chances of producing incorrect estimates even if the estimator is unbiased. An estimator that has a small amount of bias and less variation may be preferred.

Work by Stock, Wright and Yogo[130], Staiger and Stock[128] and Stock and Yogo[131] has sought to define weak instruments in two ways; firstly based on bias relative to the bias of the OLS estimate and secondly based on the size of the Wald test of significance of the estimate. The authors define the relative bias of the IV estimator to the OLS estimator $B_{rel} = \frac{E[\beta_{IV}]-\beta}{E[\beta_{OLS}]-\beta}$ and show that asymptotically this is approximately equivalent to the inverse of the F-statistic of the instruments on the endogenous variable. This is the F-statistic from the test of the null hypothesis that the coefficient of association between the instrument and endogenous variable in the first-stage regression is zero. This finding is also demonstrated in simulations by Burgess and Thompson[132]. The F-statistic can be expressed in terms of the variance in the endogenous variable explained by the instruments ($R^2$), the sample size n and the number of instruments z, $F = \left(\frac{n-z-1}{z}\right)\left(\frac{R^2}{1-R^2}\right)$ [128,131,133].

Staiger and Stock[128] and Stock and Yogo[131] propose that weak instruments can be defined by stating that the relative bias of the 2SLS estimator should not be greater than some specified value, which they suggest to be 0.1; that is 10% of the bias of the OLS estimator. Alternatively they propose that a 5% hypothesis test should not be rejected more than a specified proportion of the time (i.e. 10% or 15%). From these definitions the authors provide critical values of the F-statistic (i.e. strength of the instruments) for different numbers of instruments, number of endogenous variables and for various levels of acceptable bias[131]. A rule of thumb for weak instruments has developed from these critical values, the Staiger-Stock rule of thumb is that a first stage F-statistic<10 is indicative of a weak instrument problem. This is based on testing at the 5% level of significance that the bias is at least 10% of (or at least 90% less than) the bias of the OLS estimate i.e. $B_{rel} \approx \frac{1}{F} < 0.1$. A significant finding means that the bias is more than 10% of the OLS estimate and therefore the instruments are not effective and are deemed weak.

### 2.3.1 Methods to improve estimation in the presence of weak instruments

Several estimators have been suggested to combat the problem of weak instruments in two-stage least squares. They are closely associated to 2SLS which in general is a special case of these alternative estimators. The Limited Information Maximum Likelihood (LIML) and Fuller's adjustment are described below as alternatives to the least squares formulation.

#### *2.3.1.1 Limited information maximum likelihood (LIML) and Fuller's adjustment*

The Limited Information Maximum Likelihood (LIML) is similar to 2SLS in that they are both single equation methods, hence the term 'limited information'. The equations of the model can be estimated one at a time rather than together as with 'full information' or system methods. The LIML estimator can be described as minimising the ratio of residuals from regressions of the endogenous variables (outcome and mediator) on the specific exogenous variables in that equation to the residuals from the regressions of the endogenous variables on all exogenous variables (covariates and instruments) in the entire model. If the system is exactly identified so that there are only as many instruments as endogenous variables then the LIML estimator will be the same as the 2SLS estimator[134,135].

Stock and Yogo[131] show that in the many weak instruments case the LIML (and Fuller's adjustment described below) are asymptotically unbiased but the 2SLS estimator is not. Many authors therefore use the median when comparing results from this estimator and

advocate LIML since it is shown to be median unbiased even in the presence of weak instruments[131 136 129]. However, the LIML estimator does not have any finite sample moments. This means that mathematically the estimator does not have a sample mean and in practice the distribution of the LIML has thicker tails and is more dispersed than the 2SLS[137]. Hahn, Hausmann and Kuersteiner[137] do not recommend the use of the LIML since the wide dispersion makes results unreliable especially when weakly identified. Burgess and Thompson[132] dismiss this characteristic of the LIML estimator as a theoretical rather than a practical problem since extreme values of the estimate would generally be ignored as implausible and they recommend using the estimator as a sensitivity tool in the presence of weak instruments. However because the LIML estimator is much less precise than the 2SLS and so gives widely dispersed distribution and much larger mean-squared errors [136,137], the chance of a biased estimate in a sample can be higher even if the estimator is unbiased on average.

The Fuller estimator is a modification of the LIML designed to have finite sample moments and as such is preferred by Hahn et al[137]. It is similar to the LIML when the number of exogenous variables is small compared to the sample size. As the number of exogenous variables approaches the sample size the difference between the estimators will increase. The Fuller estimate is shown to perform better than the 2SLS in terms of median bias and is comparable in terms of the mean–squared error[132 137].

The LIML estimator has been shown to be less precise than the 2SLS estimator in the presence of weak instruments[137]. As described previously this can be considered beneficial as it indicates uncertainty in the estimates but also increases the chance of incorrect estimates. The 2SLS may have greater bias but with smaller variance the estimates may be preferred. The imprecision of the LIML may be improved with the Fuller adjustment to the estimator providing a reduction in bias without forfeiting precision. This highlights that when choosing the best method for IV estimation a balance must be achieved between bias and precision.

### 2.4 Instrument selection

The second hurdle in instrumental variables is the choice of instruments and number of instruments to ensure that the instrument set is strong. In his summary of weak instruments Murray[126] states that instruments must be found to increase the amount of variation explained without over-fitting by including too many instruments. A balance must be

achieved between the number of instruments used and the amount of variation explained. In an RCT setting randomisation is a strong instrument for post-randomisation mediators such as attendance at therapy when we can assume that there is no direct effect of randomisation. However, when a direct effect of randomisation that is not through the post-randomisation mediator cannot be discounted, randomisation is no longer a valid instrument and another must be found. Ideally there would exist a pre-specified instrument based on clinical knowledge which is also strongly statistically associated with the post-randomsation mediator. It would be necessary to incorporate this into the design of the study from the beginning to ensure that the potential instrument is measured. Unfortunately prior beliefs do not necessarily ensure that instruments are statistically valid and in real datasets like the EDIE-II trial, that are not specifically designed to include an instrument there may be several possibilities. It is therefore important to select appropriate instruments with the strength to give unbiased results.

It is established that the larger the first stage F-statistic the smaller the relative bias of the IV estimate. This suggests that the selection of instruments should focus on increasing the F-statistic and from this Staiger-Stock have developed a rule of thumb suggesting that instruments should be selected so that F>10 ensuring that the relative bias is <10%. This may be an oversimplification and the authors note that the rule of thumb is not as accurate when there are a small number of instruments, for example with one instrument the mean of the IV estimator is not mathematically defined. To overcome this problem Burgess and Thompson use the relative median bias. They find that relative median bias is also approximately equal to the inverse of the first stage F-statistic when only a small number of instruments are used[132]. However, they warn that in simulations the F-statistic varies greatly and therefore may not be a reliable measure of the true mean F-statistic in a real study. An additional problem with using the F-statistic is that it is not appropriate when there are multiple endogenous variables[136]. If there are two endogenous variables then 2 instruments are needed, one may be a strong predictor of both endogenous variables and the other weak, the first stage F-statistic will still be high for both but the model will be weak as 2 strong instruments are needed. Additionally when categorical endogenous variables are considered the F-statistic is not an appropriate measure of model fit.

A focus on increasing the F-statistic only takes into account bias in the estimate and not precision. Work by Burgess and Thompson[132] finds that the inclusion of multiple instruments increases bias in the estimate of the endogenous variable on outcome even

when the instruments are valid but decreases variation in the estimator, overall, by considering the median absolute bias the authors advocate the use of multiple valid instruments. These results indicate that by sacrificing some bias the inclusion of additional instruments can improve the precision of the estimate. The best choice of instruments may therefore not be those that maximise the F-statistic but achieve a balance between reducing bias and reducing the variance of the estimator. So how can we choose the best set of instruments from a pool of potentially valid instruments in order to achieve a balance between bias and precision?

Univariate analysis looking at significant associations between possible predictors and the outcome is often a first stage and those with a significant result may be included in a model. However, if 20 variables are considered for inclusion in a prediction model, setting the inclusion criterion at the usual 5% significance level on average one of the 20 variables will be significantly associated even when none of them truly are. This one predictor would be included in the model even though it is not truly associated. Equally if some of these 20 are real predictors then the magnitude of their effect will vary. If only those with a large effect are included their importance can be overestimated.

Stepwise regression, another popular data reduction method is relatively simple to apply and does cut down the dataset, however there are certain problems: the selection of predictors is unstable and can be influenced by the sample size, the number of possible predictors considered and the correlation between them; the $R^2$ value of variation explained is likely to be overoptimistic and p-values exaggerated; the coefficient estimates are biased and predictions can be worse than the full model[138].

The question of instrument selection has seen a focus on automated methods to increase efficiency by selecting a parsimonious model in order to improve prediction without increasing bias. Variable selection techniques utilised in prediction modelling improve on basic variable selection methods such as univariate analyses or stepwise selection.

Donald and Newey[139] propose instrument selection to minimise the mean-squared error, a measure of both bias and precision, they compare results of this selection to using all simulated instruments. The authors advocate their method when some prior information is known about the importance of the potential instruments. However when there is no prior knowledge, as in the application of variable selection in this thesis, the median bias and precision of the 2SLS estimator appear to be worse when this method is used.

Several authors[140-142] have considered shrinkage methods as a way of selecting a smaller set of instruments from a large potential set. Shrinkage methods have been developed to combat problems with stepwise selection. The method adds a penalty to the ordinary least squares estimator that results in some coefficients reducing to zero, therefore excluding them from the model. This penalty restricts the sum of the coefficient values to be less than a certain value known as the shrinkage factor, in this way shrinkage can also be used to select predictors. Their main application previously has been in prediction models as shrinkage to the mean increases the validity of predictions in new samples.

Ng and Bai[140] consider boosting for instrumental variable selection as well as selection of instruments derived from principal components. Boosting as an iterative procedure which fits least squares regressions to the residuals of the previous model. In each iteration an additional predictor that minimises the square error is selected. One predictor is selected at each iteration but a predictor can be selected more than once so that lots of small adjustments are made[143]. The authors simulate data in which all potential instruments are associated with the endogenous variable, either being equally associated or reducing in importance. The simulation results indicate that the boosting method is comparable if not a little better than selection by t-test or BIC (Bayesian Information Criterion) in terms of bias and root mean-squared error. However, their results indicate that the boosting always selects the largest instrument set in this scenario and when the instruments explain only a small part of the endogenous variable the boosting method uses all instruments available.

There has been some research into the Least Absolute Shrinkage and Selection Operator (LASSO) in terms of variable selection in causal inference models. The LASSO selection model applies the specific shrinkage factor of the form $\sum_{j=1}^{p}|\beta_j| \leq s$. This restriction means that as the number of predictors, p, increases the value of their coefficients must decrease. The limit forces some coefficients to zero and therefore excludes these as predictors. Belloni and colleagues[142] apply this LASSO selection technique to a scenario where there is a large set of potential instruments relative to the sample size. They use the LASSO coefficient estimates as well as discarding the coefficients and using only the variables selected by the procedure; they term this the post-LASSO. Their simulations indicate that when the instruments are strong and there is a small subgroup of relevant instruments or the instrument relevance varies then selection of variables by the LASSO is preferable to using all instruments. When there are many instruments (50 in these simulations) which are

equally associated with the endogenous variable the LASSO selection does not perform as well as including all variables. When the instruments are very weak the LASSO does not find relevant instruments.

Odondi and McNamee[141] apply the LASSO in the analysis of an RCT using principal stratification. The authors compare the LASSO model selection to backward stepwise selection using the Akaike Information Criterion (AIC) and all nine potential predictors to predict compliance to treatment in both arms of an RCT prior to applying a principal stratification for causal inference. The authors find that when applied to the real dataset the LASSO performed better than the backward AIC selection in the validation statistics used. When the predictors selected by the AIC were included as if pre-specified (i.e. applying the selected variables directly in a regression model rather than using the coefficients generated by the selection process, like the post-LASSO used by Belloni et al.) the performance of this model, though comparable to that of the LASSO, was preferred.

The elastic net, another selection procedure, has a very similar form to the LASSO but with a different penalty applied, in this case the $\beta_j$ values are constrained by a shrinkage factor defined as $\lambda(1-\alpha)\sum_{j=1}^{p}\left|\beta_j\right|^2 + \lambda\alpha\sum_{j=1}^{p}\left|\beta_j\right| \leq t$. This is equivalent to the LASSO when $\alpha = 1$. We specify the Elastic Net penalty at $\alpha = 0.5$ for comparison. This difference in the penalty applied means that if there are a group of variables that are correlated with each other and are associated with the dependent variable then the elastic net will select the group whereas the LASSO will select only one[144]. It is likely that in an RCT context there will be groups of variables that are correlated, for example measures of symptoms and functioning. The elastic net may therefore be a better selection method, though if more instruments are included bias may increase.

The optimum penalty, $\lambda$, applied in the LASSO and Elastic Net estimators is determined by cross-validation. The glmnet[145] software package in R[146] involves running the LASSO and Elastic Net models over a range of values of $\lambda$, giving estimates that apply $\lambda$ to minimise the mean square error ($\lambda_{min}$) of the predicted outcome and the largest lambda that gives an error within one standard error of the minimum ($\lambda_{1se}$). A larger $\lambda$ will force fewer predictors to be included in the model therefore $\lambda_{1se}$ will give a more restricted set of predictors than $\lambda_{min}$. The two outputs from these selections therefore provide a model that gives the best fit in terms of minimising the mean squared error and another that gives the most parsimonious model that still gives a reasonable mean-squared error. Applying this to

an instrumental variables context, we are reminded of the trade off between increasing the amount of variation explained in the first stage regression and limiting the number of instruments to reduce bias. Using the shrinkage methods to select instruments should therefore provided choices of instruments that will give either the best fit in the first stage regression or a compromise between fit and parsimony. Applying the instruments that are selected by these methods in an instrumental variables analysis allows for the overall impact on the estimates of the direct and indirect effects to be assessed.

Most work in the area of instrument selection has been within the economics sector where they have large numbers of potential instruments, possibly more instruments than sample size and the methods have been tested in observational data scenarios. In an RCT setting it is unlikely that such a large pool of potential instruments will be available and so the methods may not be as useful. The methods suggested for these more extreme circumstances are considered and applied in an RCT context. In the application of the methods to real datasets we do not have any prior knowledge of the importance of the instruments considered, the Donald and Newey method is therefore not compared here. In the following simulation studies the LASSO and Elastic net shrinkage methods will be investigated for selection. The LASSO has been shown to be effective in both simulated and real datasets similar to those that will be analysed in this thesis. The Elastic net though it has not been tested in this situation may prove to be effective since the instruments that will be considered are likely to be correlated with each other. Additionally by altering the penalty applied these two methods allow for a direct comparison between selecting the most parsimonious model and the model that reduces the mean-squared error. These methods will be compared to using all potential variables and simple stepwise selection since it is probably the most popular and well understood method for variable selection. Data will be simulated to replicate an RCT of a complex intervention under differing levels of instrument strength.

## 2.5 Simulation Study

In the following section several simulation studies are carried out. These simulations will compare methods described above for:

- The selection of instruments to use in a two-stage least squares instrumental variables analysis of a post-randomisation process variable.
- Instrumental variables estimation of a post-randomisation process variable.

- Instrumental variables estimation of a post-randomisation process variable and a mediator available in both treatment arms.

The simulations will consider the performance of the methods with varying levels of unmeasured variation in the process variable/mediator and unmeasured confounding between the process variable/mediator and outcome. The details of the simulation studies are given below.

### 2.5.1 Monte Carlo simulation

Data simulations are the mathematical equivalent of a scientific experiment. It is possible to compare statistical methods and their properties in a theoretical sense by formulae but these are often derived for large (infinite) samples and may behave differently in small samples that are encountered in real data. They are also subject to assumptions, for example normality of data distribution or independent observations. Performance of the methods when these assumptions are violated is not clear. There are various questions that may be asked of a statistical method. Does it provide consistent and unbiased estimates? How does it compare to other methods in terms of bias and precision? What sample size is needed to provide a certain level of power for hypothesis testing?

It is not possible to answer these questions of statistical methods applied to real data as the true answer is not known. Instead, advances in computational ability mean that data can be created under specific conditions to be tested, for example non-normal distribution, correlation between variables/observations, so that the true values of the parameters of interest are known. It is then possible to apply the methods to the simulated data and generate the parameter/test-statistic and assess accuracy and precision.

Monte Carlo simulation is probably the most widely used simulation method in statistics and the one used in the following analyses. The Monte Carlo method generates random samples from specified distributions. This is usually done by drawing a random value from a uniform distribution and then applying a transformation to replicate another distribution. In practice, computer generated random numbers are not entirely random but ‘pseudo-random’ as they use an algorithm to create a sequence of numbers from a set starting point, known as the seed. The seed can be selected by the computer, usually based on the time it is run or can be specified by the programmer allowing the exact ‘random’ number sequence to be replicated. Although the number generation is not completely random it has been shown to produce values that behave like random numbers for the purposes of

statistical analysis. Using this method a simulated dataset can be created containing independent observations of variables from particular distributions with specified relationships between the variables so that it is similar to a real dataset but the relationships between the variables are known.

As with any scientific experiment it is important to plan the analysis effectively and report the results clearly to enable replication. A guide to the design and reporting of simulation studies is provided by Burton et al[147]. They stress the importance of detailed planning of the study setting aims and objectives, designing the study to control for any event outside the scenario under investigation and specifying the parameters of interest to be saved from each simulation. The number of replications, seed and random number generator should all be reported. The results should include the summary measures of interest as well as some indication of their bias, accuracy or coverage.

### 2.5.2 Simulation study 1: one post randomisation process variable present in the intervention arm only

#### *2.5.2.1 Simulation study 1a: comparison of selection methods*

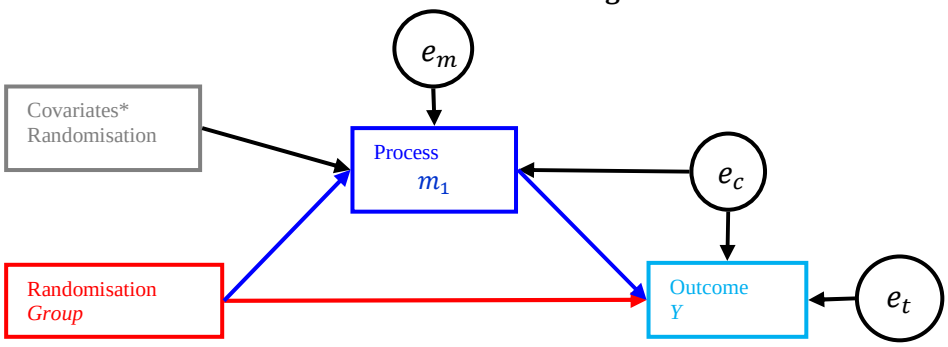
**Aim:**

To compare methods of instrument selection for an instrumental variable analysis when the continuous or categorical process variable is available only to those assigned to the intervention and determine the method which minimises the mean squared error and bias of the effect estimate.

**Method:**

The data are constructed with similar attributes to a real randomised trial dataset and based on the post-randomisation process model described in Figure 2.1. A dataset of N observations was created with what are considered to be 20 baseline variables (independent of randomisation), a randomisation group indicator (50:50 randomisation in all simulations), a post-randomisation process variable and an outcome.

**Figure 2.1: A graphical representation of the simulation model with one post-randomisation mediator and unmeasured confounding**




The post-randomisation process variable is described by the sum of either five or ten standard normal variables and two error terms. One of these error terms ($e_c$) is also associated with the outcome and therefore introduces a level of confounding between the post-randomisation process variable and outcome. The other error term ($e_m$) is specific to the post-randomisation process.

When five variables are used to describe $m_1$:

$$m_1 = 0.6 + x_1 + x_2 + x_3 + x_4 + x_5 + e_c + e_m$$

When ten variables are used to describe $m_1$:

$$m_1 = 0.6 + x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 + x_8 + x_9 + x_{10} + e_c + e_m$$

In both cases a categorical process variable is formed by dichotomising the linear process variable at the mean. To replicate the information available in a post-randomisation process variable $m_1$ is set to zero in the control group.

The outcome in the control condition is a combination of a baseline variable and the error term also associated with the process variable.

$$y_c = x_{11} + e_c$$

The outcome in the treated group is that of the control group plus a treatment group effect (-10) and a process effect (50).

$$y_{t=}y_c - 10 + 50 * m_1 + e_t$$

Outcome under the control condition $y_c$ and the treatment condition $y_t$ is simulated for every individual and their observed outcome calculated according to their group allocation as follows:

$$Y = y_t * \text{group} + y_c * (1 - \text{group})$$

The error terms $e_c, e_m$ and $e_t$are normally distributed with zero mean and standard deviation one. In the simulations, the variation in the process variable explained by the covariates is reduced whilst keeping the overall variation in the process variable and the level of confounding the same. This is achieved by altering the variation due to the error terms $e_c$ and $e_m$ and the covariates $x_1 - x_5$ (or $x_1 - x_{10}$). The standard deviations of each of these variables for the simulations are detailed below and in Table 2.1 and Table 2.2. The 20 baseline variables consist of either 5 (or 10) standard normal variables $x_1 - x_5$ ($x_1 - x_{10}$) associated with the post-randomisation process $m_1$ which are the explanatory variables, 5 standard normal variables not associated with $m_1$ (noise) and 10 noise variables correlated with the explanatory variables to different degrees e.g.

Explanatory variables – $x_1 \sim N(0,1), \ldots, x_5 \sim N(0,1)$
The variance of the explanatory variables $x_1 - x_5$ or $x_1 - x_{10}$ is reduced by multiplying each by a factor less than one, as the variance in the error term is increased by multiplying by a factor greater than one. In this way the overall variation in $m_1$ is kept the same. The standard deviations of the variables in each simulation are detailed below.

Noise variables– $x_6 \sim N(0,1), \ldots, x_{10} \sim N(0,1)$ (these are explanatory variables in the 10 explanatory variables model)

Correlated noise variables: $x_{11} = x_1 + N(0,1), \ldots, x_{15} = x_5 + N(0,1)$

Correlated noise variables: $x_{16} = x_1 + N(0,5), \ldots, x_{20} = x_5 + N(0,5)$

In this set up the post-randomisation variable is explained by the five or ten explanatory variables in the treatment group and is zero in the control group as well as some error that is associated with the outcome (unmeasured confounding). An interaction between group and covariates therefore explains this relationship. Since the process variable in the treatment group arm is explained by several explanatory variables each individual interaction will only be weakly associated with the process. The strength of the instruments

is also altered by the amount of additional unmeasured variation present in the mediator $e_m$.

In order to investigate the selection methods thoroughly the following are altered:

1. The variation in the process variable accounted for by the explanatory variables. Increasing the amount of unexplained variation means that the instruments are more weakly associated with the process variable allowing us to compare methods at different levels of instrument strength.

   The variance in the error term $e_m$ is increased whilst keeping the unmeasured confounding and the total amount of variation in the process variable at the same level. Unmeasured confounding is measured as the correlation between the error terms of the process variable and outcome ie. $\mathrm{corr}(e_c + e_m, e_c)$ and so the variation in $e_c$ is also increased to maintain the ratio between the error terms in the process variable. The variation due to the covariates must be reduced to keep the same overall variation in the process variable. The error terms and explanatory variables are generated as standard normal variables with mean zero and standard deviation one. In order to alter the standard deviation of the error and explanatory variables they are multiplied by a factor, the standard deviations for each simulation are detailed in Table 2.1 and Table 2.2.

**Table 2.1: Standard deviations for variables in simulated models with 5 explanatory variables**

| Simulation model | $e_m$ | $e_c$ | $x_1 - x_5$ |
|---|---|---|---|
| 1 | 1 | 0.1 | 2.65 |
| 2 | 2 | 0.2 | 2.41 |
| 3 | 3 | 0.3 | 1.95 |
| 4 | 4 | 0.4 | 1.00 |

**Table 2.2: Standard deviations for variables in simulated models with 10 explanatory variables**

| Simulation model | $e_m$ | $e_c$ | $x_1 - x_{10}$ |
|---|---|---|---|
| 1 | 1 | 0.1 | 2.00 |
| 2 | 2 | 0.2 | 1.84 |
| 3 | 3 | 0.3 | 1.55 |
| 4 | 4 | 0.4 | 1.00 |

2. The number of true instruments associated with the process variable.
   Including more true variables that are associated with the process variable compares the selection methods in the situation where there are more but weaker instruments to choose from. This is set as either five or ten variables. The F-statistic for individual instruments in the first stage regression is lower with more instruments.

Number of simulations

Systematic difference between the estimates can be tested using a paired t-test. The number of simulations necessary is therefore based on detecting a difference between estimation methods in the bias of the effect estimates. The calculation of the number of simulations given by Burton et al[147] requires an estimate of the variance of the bias across simulations and the detectable difference in the bias that is required. In order to inform the sample size calculation the magnitude and variance of the bias is estimated by running the simulation study described above for just 100 simulations. This is carried out for the categorical measures and run on the scenario with the highest amount of unexplained variation as this is expected to generate the largest bias with the most variation in estimates. Belloni et al[142] report the root mean squared error (RMSE) as a comparison of methods. The percentage reduction in RMSE between selection by the LASSO and using all variables ranges from around 20%-30%. The reduction in bias between the LIML and 2SLS reported by Burgess and Thompson[129] ranged from around 7% difference. The number of simulations here will be calculated based on detecting at least a 10% difference between methods in the bias. The bias for a categorical process variable with weak instruments is estimated to be around 2.7 with standard deviation 4. According to Burton et al[147] in order to detect a difference in bias of at least 10% at the 5% level of significance 843 simulations are needed. It is expected that 1000 simulations will be enough to detect a difference in methods of at least 10% for all scenarios.

**Analysis:**

To compare instrument selection methods several analyses are run on the simulated data and summary measures extracted for comparison. The traditional Baron & Kenny OLS mediation is compared to instrumental variables regression under the different methods for obtaining instruments. An OLS regression with randomisation group and the post-randomisation mediator will give a biased estimate of the mediator effect due to

confounding. Interactions between randomisation group and each of the explanatory variables are instruments for the post-randomisation process variable as they are associated with the process variable but not directly with outcome. The direct effect of group and the effect of the mediator on outcome are estimated using the two-stage least squares method to remove bias due to unmeasured confounding.

To select instruments the selection methods LASSO, Elastic Net and backward stepwise are applied to the regression of covariates $x_1 - x_{20}$ in the treatment arm only. The selected variables by group interactions are then used as instruments in the 2SLS analysis. The IV analysis is calculated in two stages, regression of instruments on $m_1$ to make predictions $\hat{m}_1$ which are then regressed with group and covariates on Y. Results for the following models are therefore compared:

- Ordinary least squares regression of treatment group and $\hat{m}_1$ on Y adjusting for all 20 baseline variables $x_1 - x_{20}$ (this is repeated for each of the following selection methods, the results are not shown here but are given in the full tables in Appendix 1).
- IV 2SLS using all 20 baseline variable by randomisation arm interactions as instruments i.e. no selection of instruments
- IV 2SLS using the variable by randomisation interactions selected via:
    - LASSO$\lambda_{\min}$
    - LASSO $\lambda_{1se}$
    - Elastic Net $\lambda_{\min}$
    - Elastic Net $\lambda_{1se}$
    - Backward stepwise selection
- IV 2SLS using only the relevant variable by randomisation interactions specified as instruments in the simulation set-up. This is for comparison only; this would not be possible in an analysis of real data as it is an unknown.

The parameters of interest are those associated with the coefficients of the randomisation group and the process variable in the second stage regression. The mean parameter estimate and associated standard deviation across the simulations, the bias or mean absolute difference (AD) and the mean square error (MSE) from the true parameter value are reported along with the F-statistic for the first stage of the 2SLS regressions. The F-

statistic specifically associated with the instruments cannot be taken directly from the first stage regression as this includes other variables namely group and the covariates selected. To calculate the F-statistic specifically associated with the instruments the first-stage regression is run with and without the instruments keeping everything else the same; an ANOVA of the nested models is conducted and the F-statistic produced is that associated with the instruments.

The analyses are repeated simulating continuous and categorical process variables on a continuous outcome. The impact of sample size is also considered by running simulations on samples of 200 and 400 observations. All simulations were carried out 1000 times setting the seed as 2012 for replication. The estimated coefficient of group and process variable effect on outcome is saved for each model within each simulation. The bias and mean squared error of the coefficients are compared across models to assess performance. These performance measures are calculated as follows:

1. $\text{Mean overall bias} = \bar{\beta} - \beta = \sum_{i=1}^{nsim} \hat{\beta}_i / nsim - \beta$
2. $\text{Mean absolute difference} = \left(\sum_{i=1}^{nsim} |\hat{\beta}_i - \beta|\right) / nsim$
3. $\text{Mean squared error} = \left(\sum_{i=1}^{nsim} \left(\hat{\beta}_i - \beta\right)^2\right) \Big/ nsim = \text{bias}^2 + (\text{s.e.}(\hat{\beta}))^2$
4. $\text{Median overall bias} = \text{median}(\hat{\beta}_i) - \beta$

Where $\beta$ is the true value of the parameter of interest, $\bar{\beta}$ is the average estimate of the parameter over all simulations, $\hat{\beta}_i$ is the parameter estimate for simulation i, $\text{s.e.}(\hat{\beta})$ is the standard error of the parameter estimate over all simulation and nsim is the number of simulations.

As discussed earlier in the chapter it is important to consider both bias and precision of the model estimates. Differences in bias between the selection methods will be tested using a paired t-test of the absolute bias of the group and process effect estimates within simulations. The difference in precision between effect estimates from the same simulation will be tested using the Pitman-Morgan test for the difference between two correlated variances[148-150]. The mean-squared error is a summary of both bias and precision. The mean-squared error incorporates both of these and therefore is the preferred measure of performance though differences in MSE between methods are not tested. The statistical

methods considered here will be compared based primarily on the bias and precision. The preferred method is that which gives the smallest mean-squared error.

**Simulation results:**

The graphs in Figure 2.2 and Figure 2.3 provide an example of the cross-validation process for the Elastic Net and LASSO variable selection of instruments on one simulated dataset. The graphs illustrate the relationship between the mean-squared error of the predicted mediator variable and number of variables included in the model. Figure 2.2 shows the cross-validation of the LASSO and Elastic Net methods on a simulated dataset where five variables are associated with the post-randomisation process variable. The top two graphs show the results for a single simulated dataset in which the five explanatory variables explain a large amount of the variation in the process variable. In terms of instrumental variables this simulates a scenario of strong instruments (F-statistic>10). In the bottom two graphs the amount of unexplained variation is greater, the instruments are weaker (F-statistic<10). The graphs show the trade off between the reducing the mean-squared error of the predicted mediator and reducing the number of variables selected. This is not as important when the variables explain a large amount of variation but can make a large difference to the mean-squared error when the variables are more weakly associated. When the variables are weakly associated with the process variable the more parsimonious models do not find any variables. This may be appropriate and indicative that the variables are too weakly associated to be used effectively. However, the best fitting model does select variables and for the purposes of an instrumental variable analysis may still provide an effective instrument. The number of variables selected is similar between the Elastic Net and LASSO in this example, if anything the Elastic Net tends to select more variables as expected. Figure 2.3 replicates the plots given in Figure 2.2 but for a simulated data set in which the post-randomisation process variable is constructed from ten baseline covariates in the treatment arm (the value of the process variable is still set to zero for all observations in the control arm). The same patterns are seen when comparing the Elastic Net selection to the LASSO and at different levels of lambda in the five versus the ten explanatory variables set up. Interestingly, when only five variables were designed to be associated with the process variables both methods selected many more than five variables for the best model in both the weakly and strongly associated scenarios. In the simulation where ten variables were associated with the post-randomisation process the number of variables selected is similar to that of the five variable scenario but is closer to the number

we would expect. The graphs suggest the selection methods may work better when the instruments are weaker. The graphs illustrate in one simulated dataset the selection of variables in the treated arm to be used as instruments when interacted with group. The impact of these methods of treatment selection on an instrumental variables analysis are summarised in the results of the simulation study.

**Figure 2.2: Simulation example of cross-validation indicating number of variables selected by selection method; 5 true variables associated with continuous post-randomisation process variable**

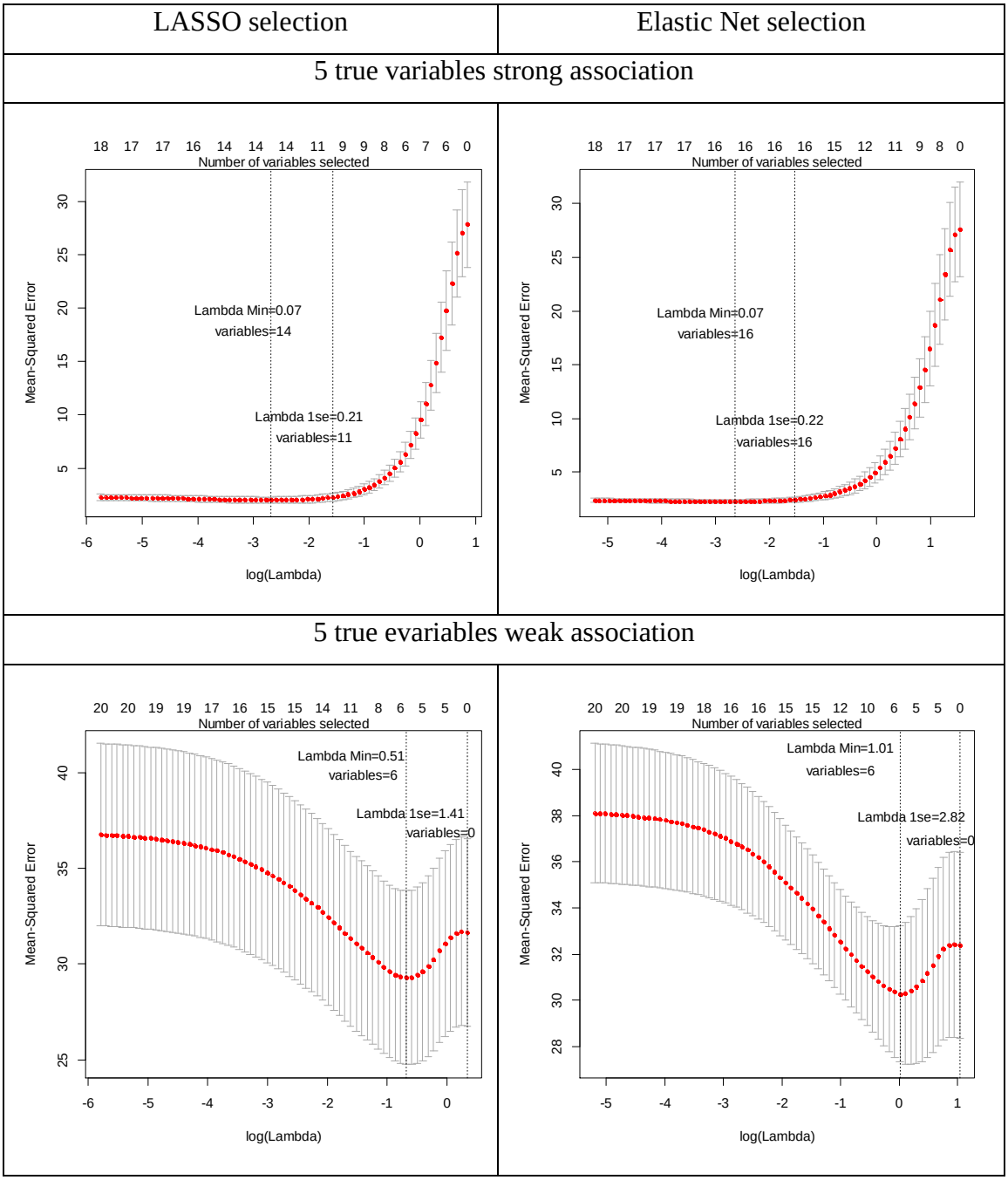
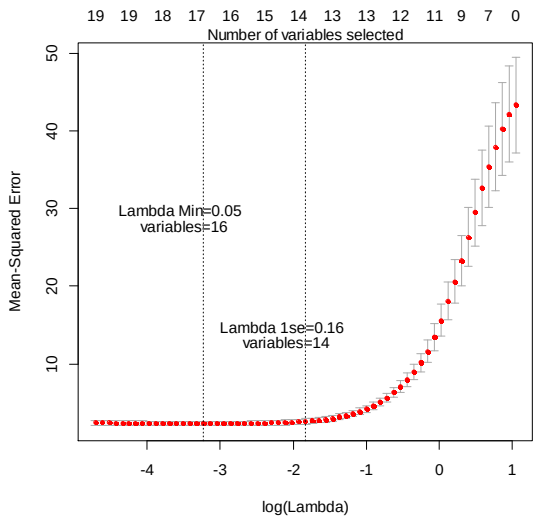
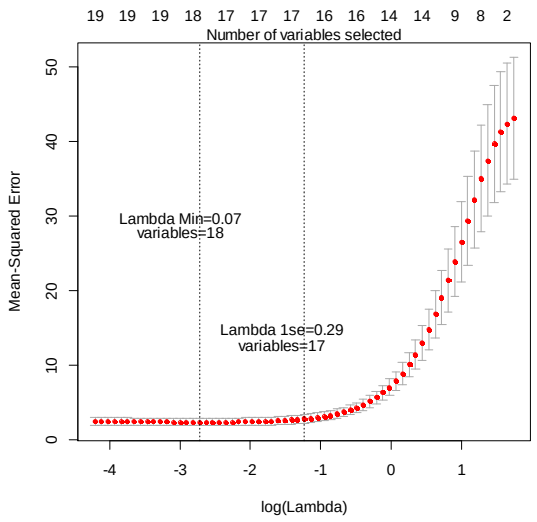
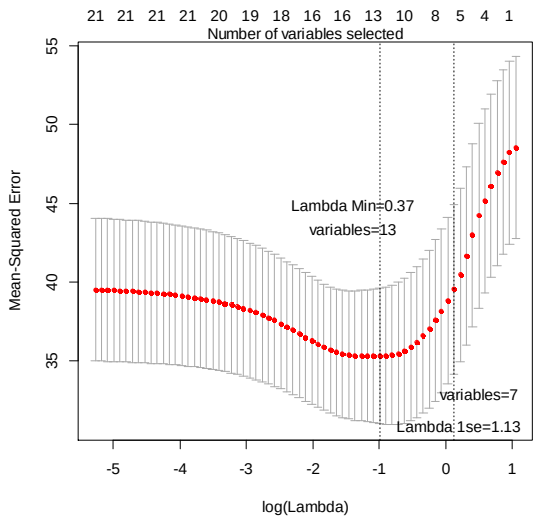
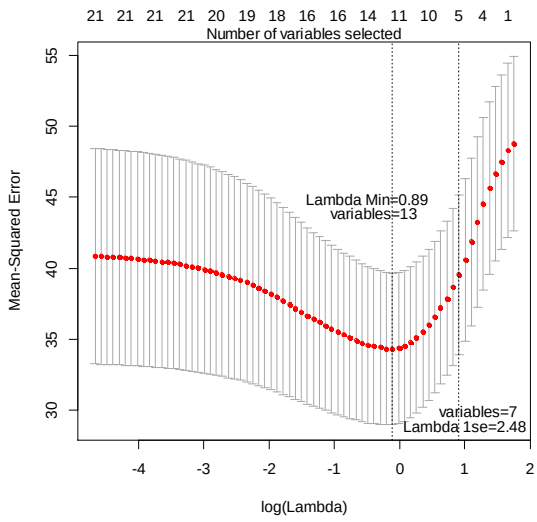


**Figure 2.3: Simulation example of cross-validation indicating number of variables selected by shrinkage method; 10 true variables associated with continuous post-randomisation process variable**

| LASSO selection | Elastic Net selection |
|---|---|
| 10 true variables strong association | |
|  |  |
| 10 true variables weak association | |
|  |  |

A summary of results for the simulations with a sample size of 200 are presented below with full results given in Appendix 1. Table 2.3 to Table 2.6 show the results of altering the amount of variation in the process variable explained and the number of explanatory variables associated with the process. Instrument selection methods are compared by

applying the instruments in a two-stage least squares instrumental variables analysis and extracting the coefficient estimates. In this shortened summary only the mean-squared error and mean bias are shown though the mean absolute difference along with the coefficient estimate and standard error are given in the tables of Appendix 1. The smallest value of the mean-squared error and bias indicates the most accurate method reading across each row of the table and in each case is highlighted in bold and blue. The first two tables (Table 2.3 and Table 2.4) refer to a continuous post-randomisation mediator and include the average first stage F-statistic in brackets to serve as a measure of strength of instruments in the first stage prediction model and as a comparison to the literature regarding weak instruments. The following two tables (Table 2.5 and Table 2.6) give the results of a binary post-randomisation mediator, the first stage F-statistic is not provided in these tables as it is not applicable to a binary outcome.

**Continuous post-randomisation process variable**

The results in Table 2.3 and Table 2.4 show that although there is bias present in both the instrumental variables and ordinary least squares (Baron and Kenny) estimate the IV estimation outperforms OLS method for all levels of unmeasured variation when considering a continuous post-randomisation process variable. This is based on the mean-squared error and mean bias being smaller for values estimated by IV versus OLS regression. Work by Staiger and Stock[128] and Hausman[127] quantify the benefit of IV analysis in terms of the relative bias of the IV estimate to the OLS estimate. This is the basis of the Staiger Stock rule of thumb that a first stage F-statistic greater than ten provides an IV model with bias less than 10% that of the OLS estimator. In these simulations the bias of the IV estimates remain on average less than 10% of the bias of the OLS estimates whilst the first stage F-statistic stays above ten as expected. In the ten explanatory variable example when the instruments are strong the percent bias of the IV relative to the OLS ranges from around 2% when only the relevant instruments are selected to 10% when all variables are used and the selection methods provide average relative bias of between 3% and 9%. In the same scenario when the instruments are weak the bias in the estimate of the process variable for the IV is 25% of the bias of the OLS estimate with all relevant instruments, 48% with all instruments and under the various selection methods ranges from 40% to 56%. These are termed weak instruments by Staiger and Stock though the use of the IV model still halves the bias in the estimate of the post-randomisation process variable relative to the OLS model.

Bias of estimates

There is greater discrepancy in the mean bias between selection methods than in the mean-squared error. Paired t-tests indicate that there is no significant difference in bias of the group effect between methods of instrumental variable selection at any level of unmeasured confounding when ten variables truly explain the post-randomisation process. The pattern is not as clear in the simulations with five explanatory variables. Overall in the five variable model when comparing the LASSO results to the stepwise the most parsimonious LASSO (LASSO$\lambda_{1se}$) tends to have greatest absolute bias and the LASSO with the smallest mean-squared error has the least (LASSO$\lambda_{min}$). However, there is no significant difference in absolute bias in the process variable between an IV analysis with instruments selected by stepwise or LASSO$\lambda_{min}$. The most parsimonious LASSO selection is significantly less bias than these two in the estimation of the process variable effect apart from the scenario with the most unmeasured variation. When the instruments are very weak the parsimonious LASSO model is more biased than models with instruments selected by the stepwise or LASSO$\lambda_{min}$.

Variance of estimates

In order to compare the precision between methods we test for a difference in the variance of the effect estimates. There is no significant difference in the variance of the group effect in the IV analysis between different instrument selection methods when unmeasured variation in process variable is low. The LASSO that minimises the MSE (LASSO$\lambda_{min}$) has significantly less variance in it's estimate of the process variable than when instruments are selected by stepwise. Shrinkage methods that minimise the mean-squared error of the prediction model are preferred over those that reduce the number of predictors selected (instruments) in terms of minimising the variance of the process variable estimate.

Mean-squared error

The mean squared error combines the bias and precision of estimates and is the preferred measure of estimator performance. Table 2.3 shows that when the instruments explain a large amount of the variation in the process-variable there is little to distinguish between the methods in terms of MSE but as the instruments explain less of the process variable the choice of instruments become more important. When the instruments are weak so that the F-statistic associated with including all of the known relevant instruments is less than 10

the effect estimate of the process variable remains quite accurate regardless of the instruments used whereas the direct effect of randomisation suffers a lack of precision.

When there is a high level of unmeasured variation the LASSO and Elastic Net methods that give the most parsimonious model struggle to find any instruments at all. This may be considered to be a good thing if the weak instruments actually increase bias and would not meet the requirements of the Stock and Staiger rule of thumb, but the 2SLS estimate is still better than the OLS estimates even when all variable by group interactions are included as instruments.

It seems that the F-statistic of the instruments in the first stage regression is not a good indicator of the quality of the model since the highest average F-statistic is not always associated with the most precise model. In only a small portion of the 1000 simulated datasets does the best model defined to be the model providing the lowest mean-squared error of the effect estimate of the post-randomisation process variable also have the highest first-stage F-statistic. Within each scenario less than 10% of the 1000 simulations gave results in which these two criteria were met by the same model.

As a comparison the explanatory variable by group interactions are used individually as instruments in each simulation, the results provided in the set of tables in Appendix 1. When one instrument is used the bias remains low even when instruments are weak but the mean-squared error is larger than when multiple instruments are used and this becomes more extreme as the instruments become weaker. In the ten explanatory variable simulation when the variables explain a large amount of the variation in the process variable the MSE of the process variable when only one instrument is used ranges from 0.9 to 34 (average=7), higher than the MSE for any of the selection models which are around 0.06. In the weak instrument case where there is a large amount of unexplained variation in the process variable the MSE for one instrument is very large at over 100 in every case compared to around 0.5 by the selection methods.

**Table 2.3: MEAN-SQUARED ERROR (first-stage f-statistic in brackets) of estimates by variable selection method with increasing uncertainty in post-randomisation process variable: 1000 simulations, CONTINUOUS process variable, correlation of error terms =0.69, sample size=200**

| Standard deviation of mediator error | Average first stage f-statistic | | | OLS | Two-stage Least Squares | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | All relevant variables | Individual variables | Effect | All variables | All variables | LASSO$\lambda_{1se}$ | LASSO$\lambda_{min}$ | Elastic Net $\lambda_{1se}$ | Elastic Net $\lambda_{min}$ | Stepwise | Relevant only |
| 5 true explanatory variables | | | | | | | | | | | |
| 1 | 341 | 24.3 | Group | 2.115 | 2.114 | 2.176 | **2.101** | 2.174 | 2.120 | 2.165 | 2.144 |
| | | | Process | 0.316 | 0.066 (78) | **0.064 (252)** | 0.065 (161) | 0.064 (186) | 0.066 (129) | 0.064 (223) | 0.063 (341) |
| 2 | 71.3 | 19.6 | Group | 2.155 | 2.082 | 2.147 | 2.071 | 2.146 | **2.070** | 2.109 | 2.123 |
| | | | Process | 0.831 | 0.087 (17) | 0.076 (53) | 0.082 (34) | 0.077 (42) | 0.083 (29) | **0.076 (47)** | 0.075 (71) |
| 3 | 21.3 | 12.4 | Group | 2.143 | 2.015 | 14.810 | **1.997** | 2.062 | 2.018 | 2.008 | 2.088 |
| | | | Process | 1.218 | 0.162 (5.7) | **0.119 (17)** | 0.141 (11) | 0.124 (15) | 0.143 (10) | 0.122 (15) | 0.111 (21) |
| 3.5 | 10.8 | 8.2 | Group | 2.110 | 1.965 | 173.379 | 12.264 | 179.027 | 2.204 | **1.949** | 2.071 |
| | | | Process | 1.319 | 0.280 (3.3) | **0.223 (11)** | 0.235 (6.5) | 0.217 (10) | 0.238 (6.0) | 0.224 (8.7) | 0.165 (11) |
| 4 | 3.94 | 3.7 | Group | 2.063 | 1.957 | 1492.959 | 250.772 | 1493.701 | 253.604 | **1.892** | 2.099 |
| | | | Process | 1.364 | 0.675 (1.8) | **0.638 (11)** | 0.687 (5.0) | 0.665 (9.8) | 0.644 (4.7) | 0.659 (4.8) | 0.398 (3.9) |
| 10 true explanatory variables | | | | | | | | | | | |
| 1 | 191 | 11.5 | Group | 2.547 | 2.562 | 2.492 | 2.495 | **2.490** | 2.523 | 2.490 | 2.485 |
| | | | Process | 0.257 | 0.063 (84) | **0.060 (150)** | 0.062 (118) | 0.061 (135) | 0.063 (110) | 0.064 (138) | 0.061 (191) |
| 2 | 42.3 | 9.78 | Group | 2.524 | 2.521 | **2.452** | 2.469 | 2.457 | 2.482 | 2.455 | 2.461 |
| | | | Process | 0.678 | 0.080 (19) | **0.073 (33)** | 0.078 (26) | 0.074 (30) | 0.078 (25) | 0.078 (32) | 0.072 (41) |
| 3 | 14.0 | 7.02 | Group | 2.461 | 2.437 | 10.016 | 2.381 | 10.028 | 2.396 | **2.372** | 2.415 |
| | | | Process | 1.022 | 0.133 (6.5) | 0.119 (12) | 0.126 (9.1) | **0.119 (11)** | 0.127 (8.8) | 0.130 (12) | 0.102 (14) |
| 3.5 | 7.95 | 5.3 | Group | 2.410 | 2.371 | 97.95 | 2.299 | 137.8 | 11.45 | **2.274** | 2.376 |
| | | | Process | 1.123 | 0.204 (4.0) | 0.202 (8.6) | **0.191 (5.9)** | 0.200 (8.0) | 0.191 (5.6) | 0.199 (7.7) | 0.141 (7.7) |
| 4 | 4.06 | 3.4 | Group | 2.347 | 2.295 | 1163.5 | **2.206** | 1156.3 | 160.9 | 2.207 | 2.323 |
| | | | Process | 1.179 | 0.389 (2.5) | 0.465 (5.0) | **0.337 (4.6)** | 0.440 (4.9) | 0.380 (4.3) | 0.387 (5.6) | 0.250 (4.1) |

Note: $\lambda_{1se}$- penalty lambda that gives an error within one standard error of the minimum, $\lambda_{min}$ - penalty lambda that minimises the mean square error of the predicted outcome

**Table 2.4: BIAS of estimates by variable selection method with increasing uncertainty in post-randomisation process variable: 1000 simulations, CONTINUOUS process variable, correlation of error terms =0.69, sample size=200**

| Standard deviation of mediator error | Average first stage f-statistic of instruments | | | OLS | Two-stage Least Squares | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | All relevant variables | Individual variables | Effect | All variables | All variables | LASSO$\lambda_{1se}$ | LASSO$\lambda_{min}$ | Elastic Net $\lambda_{1se}$ | Elastic Net$\lambda_{min}$ | Stepwise | Relevant only |
| 5 true explanatory variables | | | | | | | | | | | |
| 1 | 341 | 24.3 | Group | 0.199 | **-0.072** | -0.073 | -0.088 | -0.082 | -0.076 | -0.079 | -0.075 |
| | | | Process | -0.507 | -0.049 | -0.011 | -0.031 | -0.019 | -0.035 | **0.008** | -0.002 |
| 2 | 71.3 | 19.6 | Group | 0.426 | **-0.024** | -0.048 | -0.046 | -0.051 | -0.040 | -0.048 | -0.065 |
| | | | Process | -0.886 | -0.126 | -0.049 | -0.093 | -0.062 | -0.100 | **-0.026** | -0.014 |
| 3 | 21.3 | 12.4 | Group | 0.552 | 0.063 | -0.108 | 0.028 | **0.009** | 0.036 | 0.018 | -0.046 |
| | | | Process | -1.090 | -0.267 | **-0.134** | -0.211 | -0.146 | -0.216 | -0.140 | -0.040 |
| 3.5 | 10.8 | 8.2 | Group | 0.584 | 0.153 | -2.968 | **0.009** | -3.078 | 0.098 | 0.097 | -0.022 |
| | | | Process | -1.138 | -0.410 | -0.288 | -0.332 | **-0.278** | -0.338 | -0.281 | -0.075 |
| 4 | 3.94 | 3.7 | Group | 0.601 | 0.346 | -24.763 | -4.095 | -24.977 | -3.928 | **0.306** | 0.065 |
| | | | Process | -1.159 | -0.718 | **-0.621** | -0.655 | -0.659 | -0.640 | -0.628 | -0.215 |
| 10 true explanatory variables | | | | | | | | | | | |
| 1 | 191 | 11.5 | Group | 0.192 | -0.050 | -0.055 | -0.050 | -0.062 | -0.055 | **-0.044** | -0.047 |
| | | | Process | -0.448 | -0.045 | -0.023 | -0.037 | -0.026 | -0.039 | **-0.012** | -0.010 |
| 2 | 42.3 | 9.78 | Group | 0.405 | -0.006 | -0.021 | -0.011 | -0.026 | -0.013 | **-0.006** | -0.029 |
| | | | Process | -0.795 | -0.114 | -0.074 | -0.101 | -0.079 | -0.103 | **-0.062** | -0.038 |
| 3 | 14.0 | 7.02 | Group | 0.532 | 0.067 | **-0.032** | 0.062 | -0.039 | 0.061 | 0.055 | 0.003 |
| | | | Process | -0.995 | -0.229 | **-0.180** | -0.210 | -0.180 | -0.212 | -0.180 | -0.090 |
| 3.5 | 7.95 | 5.3 | Group | 0.567 | 0.133 | -1.193 | 0.122 | -1.782 | **0.044** | 0.121 | 0.037 |
| | | | Process | -1.048 | -0.334 | -0.295 | -0.297 | -0.301 | -0.305 | **-0.287** | -0.144 |
| 4 | 4.06 | 3.4 | Group | 0.588 | 0.255 | -15.803 | **0.200** | -15.468 | -2.002 | 0.219 | 0.112 |
| | | | Process | -1.077 | -0.524 | -0.557 | **-0.425** | -0.536 | -0.486 | -0.473 | -0.270 |

Note: $\lambda_{1se}$- penalty lambda that gives an error within one standard error of the minimum, $\lambda_{min}$ - penalty lambda that minimises the mean square error of the predicted outcome

**Categorical post-randomisation process variable**

The results of simulations for a categorical process variable are presented in Table 2.5, comparing estimation models by the mean-squared error and Table 2.6 by mean bias. When the process variable is categorical the first stage F-statistic is not an appropriate measure of model fit and so this summary measure is not provided in the tables. The errors associated with the estimates of the categorical process variable are much higher than that of a continuous process measure but the overall conclusions remain the same. The results of the categorical process variable also show that instrumental variable methods give more accurate estimates than the ordinary least squares method even when the instruments are weak. In the simulations of five explanatory variables the IV estimate of the process when the instruments are strong and only the relevant instruments are used is 2% of the bias of the OLS estimator, this ranges from 7% to 15% for the different selection methods and 18% when all variables are used as instruments. The equivalent results when a large amount of variation in the process variable is unexplained showed that the IV estimate for all relevant instruments was 8% of the bias of the OLS estimate, across the variable selection methods this value ranged from 25% to 30% and was 37% when all variables were used as instruments. When ten variables are simulated to explain the process measure the IV estimator is not as effective when considering the relative bias. When the variables explain a large portion of variation in the mediator the relative bias when only the relevant instruments are used is 8% and this value ranges from 15% to 18% depending on the selection method used. When there is large unexplained variation in the process variable IV using all relevant variables resulted in bias in the process estimate that was 17% of the bias of the OLS estimate, the bias of the selection methods ranged from 33% to 35% of the bias of the OLS estimate.

Bias of estimates

Bias of the estimates are presented in Table 2.6. Testing for differences between the methods we find that there is no significant difference in absolute bias of the group effect when instruments are strong (F-statistic>10). When the instruments are weak the LASSO that minimises the MSE has significantly less bias in the group effect estimate than the stepwise selection procedure. The bias in the process effect is greater when instruments are selected by stepwise than by the LASSO$\lambda_{min}$. When there are ten explanatory variables the

parsimonious LASSO is also more bias than LASSO$\lambda_{min}$ but this is not always a significant difference in the five explanatory variable simulations.

Variance of estimates

When comparing the variance of the estimates the stepwise selection results in significantly greater variance of estimates compared to both LASSO selection procedures. The parsimonious LASSO$\lambda_{1se}$ has greater variance of estimates than the LASSO$\lambda_{min}$ selection.

Mean-squared error of estimates

Considering both the bias and precision the LASSO that minimises the mean-squared error of the first stage model is preferred for selecting instruments. This can be seen by the mean-squared error shown in Table 2.5 which is lowest for the LASSO$\lambda_{min}$ selected instruments when the instruments are weaker.

**Conclusion:**

In the presence of weak instruments using one instrument may provide unbiased estimates but with such uncertainty that there is a risk that the estimate will be far from the true value. Increasing the number of instruments will improve precision but at a cost to bias. In these simulations methods for the selection of instruments have been compared in order to achieve a balance between bias and precision for effective estimation. The simulations indicate that the LASSO variable selection maximises the F-statistic for instruments in the first stage regression (average F-statistic for each selection method given in brackets in Table 2.3), however this does not always indicate the best model. The model with the highest first stage F-statistic is also the model with the lowest mean-squared error in only a small proportion of simulations regardless of the strength of the instruments indicating that using the F-statistic to judge model quality may not be effective. There is not a distinct method for instrument selection that has proven to be the best under all scenarios. However, the LASSO that minimises the mean-squared error of the first stage model tends to produce estimates with lower variance than the stepwise selection method for a continuous process variable and is comparable in terms of absolute bias. The parsimonious selection models can reduce bias but when instruments are very weak struggle to find any instruments.

When a categorical post-randomisation process variable is used the LASSO that minimises the mean-squared error is preferred in terms of reducing the mean-squared error when instruments are weak. This selection method is significantly better than the stepwise and parsimonious LASSO in terms of reducing variance and bias.

This analysis compared selection methods within the context of an IV analysis and compared to an ordinary least squares regression adjusting for all variables and showed them to be effective. When instruments are very weak the parsimonious models of the LASSO and Elastic Net have trouble finding any instruments which may be indicative that instrumental variable analysis is not appropriate. In the next simulations estimation methods to improve precision in the presence of weak instruments are considered.

**Table 2.5: MEAN-SQUARED ERROR of estimates by variable selection method with increasing uncertainty in post-randomisation process variable: 1000 simulations, CATEGORICAL process variable, correlation of error terms =0.69, sample size=200**

| | | OLS | Two-stage Least Squares | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Standard deviation of mediator error | Effect | All variables | All variables | LASSO$\lambda_{1se}$ | LASSO$\lambda_{min}$ | Elastic Net $\lambda_{1se}$ | Elastic Net $\lambda_{min}$ | Stepwise | Relevant only |
| 5 explanatory variables | | | | | | | | | |
| 1 | Group | 6.82 | **5.43** | 5.55 | 5.44 | 5.57 | 5.44 | 5.61 | 5.62 |
| | Process | 20.20 | 13.69 | 13.96 | **13.49** | 13.94 | 13.57 | 15.03 | 14.07 |
| 2 | Group | 14.54 | 6.22 | **6.15** | 6.28 | 6.26 | 6.25 | 6.44 | 6.25 |
| | Process | 53.21 | 17.62 | **16.94** | 17.19 | 17.18 | 17.21 | 19.35 | 16.57 |
| 3 | Group | 23.94 | 9.17 | 18.95 | 9.62 | 20.58 | **8.77** | 9.03 | 8.36 |
| | Process | 92.65 | 30.07 | 28.95 | **27.93** | 28.34 | 28.23 | 31.24 | 25.10 |
| 10 explanatory variables | | | | | | | | | |
| 1 | Group | 6.85 | 5.96 | 6.01 | 5.86 | 5.92 | **5.84** | 6.52 | 6.20 |
| | Process | 19.33 | 14.07 | 13.98 | **13.59** | 13.82 | 13.62 | 17.93 | 14.81 |
| 2 | Group | 14.13 | 6.93 | 6.90 | **6.76** | 6.87 | 6.77 | 7.73 | 6.82 |
| | Process | 49.38 | 18.19 | 17.95 | **17.48** | 17.74 | 17.58 | 22.75 | 17.53 |
| 3 | Group | 22.96 | 9.49 | 35.18 | **9.11** | 26.28 | 10.68 | 10.20 | 8.45 |
| | Process | 85.63 | 28.59 | 30.92 | **27.33** | 30.47 | 27.91 | 34.00 | 24.16 |

Note: $\lambda_{1se}$- penalty lambda that gives an error within one standard error of the minimum, $\lambda_{min}$ - penalty lambda that minimises the mean square error of the predicted outcome

**Table 2.6: BIAS of estimates by variable selection method with increasing uncertainty in post-randomisation process variable: 1000 simulations, CATEGORICAL process variable, correlation of error terms =0.69, sample size=200**

| | | OLS | Two-stage Least Squares | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Standard deviation of mediator error | Effect | All variables | All variables | LASSO$\lambda_{1se}$ | LASSO$\lambda_{min}$ | Elastic Net $\lambda_{1se}$ | Elastic Net $\lambda_{min}$ | Stepwise | Relevant only |
| 5 explanatory variables | | | | | | | | | |
| 1 | Group | 1.74 | 0.22 | **0.05** | 0.13 | 0.08 | 0.17 | 0.10 | -0.05 |
| | Process | -3.70 | -0.66 | **-0.27** | -0.44 | -0.31 | -0.53 | -0.32 | -0.06 |
| 2 | Group | 3.35 | 0.73 | **0.38** | 0.57 | 0.42 | 0.59 | 0.45 | 0.08 |
| | Process | -6.93 | -1.70 | **-0.94** | -1.32 | -1.01 | -1.35 | -1.03 | -0.32 |
| 3 | Group | 4.58 | 1.60 | 0.66 | 1.27 | **0.66** | 1.33 | 1.16 | 0.29 |
| | Process | -9.40 | -3.43 | **-2.33** | -2.78 | -2.43 | -2.86 | -2.45 | -0.75 |
| 10 explanatory variables | | | | | | | | | |
| 1 | Group | 1.67 | 0.29 | **0.20** | 0.23 | 0.20 | 0.24 | 0.24 | 0.08 |
| | Process | -3.51 | -0.73 | **-0.53** | -0.62 | -0.54 | -0.63 | -0.60 | -0.27 |
| 2 | Group | 3.21 | 0.80 | 0.66 | 0.71 | **0.66** | 0.73 | 0.68 | 0.31 |
| | Process | -6.59 | -1.75 | -1.44 | -1.54 | **-1.43** | -1.57 | -1.50 | -0.73 |
| 3 | Group | 4.42 | 1.57 | **0.35** | 1.39 | 0.71 | 1.36 | 1.38 | 0.71 |
| | Process | -8.99 | -3.29 | -2.99 | **-2.90** | -3.00 | -2.97 | -2.93 | -1.53 |

Note: $\lambda_{1se}$- penalty lambda that gives an error within one standard error of the minimum, $\lambda_{min}$ - penalty lambda that minimises the mean square error of the predicted outcome

#### *2.5.2.2 Simulation study 1b: comparison of estimation methods*

**Aim:**

To determine the estimation technique for instrumental variable analysis of post-randomisation process variables with multiple and potentially weak instruments that gives the smallest mean-squared error and bias of the effect estimate.

**Method:**

The data sets are created in exactly the same way as in simulation study 1a: a group indicator, 20 baseline covariates either 5 or 10 of which explain the post-randomisation process variable which is set to the value of zero for all observations in the control arm, a continuous outcome calculated by a group effect and an effect of the post-randomisation process variable. Specifically, when five variables are used to describe $m_1$:

$$m_1 = 0.6 + x_1 + x_2 + x_3 + x_4 + x_5 + e_c + e_m$$

When ten variables are used to describe $m_1$:

$$m_1 = 0.6 + x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 + x_8 + x_9 + x_{10} + e_c + e_m$$

The outcome for an individual is calculated as:

$$Y = y_t * group + y_c * (1 - group)$$

Where the outcome under the control condition, $y_c$ is:

$$y_c = x_{11} + e_c$$

The outcome under the treatment condition, $y_t$ is:

$$y_{t=}y_c - 10 + 50 * m_1 + e_t$$

The covariates $x_1 - x_5$ ($x_{10}$) and the error terms $e_c$, $e_m$ and $e_t$are normally distributed with zero mean and standard deviation one. The estimation methods will be tested by altering the amount of variation in the post-randomisation process that can be explained by the variables whilst keeping the amount of unmeasured confounding the same. In order to alter the standard deviation of the error and explanatory variables they are multiplied by a factor, the standard deviations for each simulation are detailed in Table 2.1 and Table 2.2

The details of the set up are provided in Chapter 2.5.2 and in Table 2.1 and Table 2.2.

The Two Stage Least Squares, Limited Information Maximum Likelihood and Fuller's estimates of the group and post-randomisation process variable are compared in terms of bias, variance and mean squared error. Since the LIML estimator does not have finite sample moments the median bias is often used to assess this estimator. For consistency with previous simulations the mean bias is presented and differences tested using a parametric paired t-test but median bias is given in Appendix 1. The methods are compared when all variables are included as instruments and when only those selected via the LASSO variable selection method are used.

**Simulation Results:**

Table 2.7 to Table 2.10 compare two-stage least squares estimation to limited information maximum likelihood and Fuller's using the Stata[111] ivregress and ivreg2 commands using both the mean-squared error and bias. Results of simulations with a continuous process variable are presented in Table 2.7 and Table 2.8 and simulations with a categorical process variable are provided in Table 2.9 and Table 2.10.

**Continuous post-randomisation process variable**

Differences in bias of the estimates are tested using paired t-tests. The overall results indicate that 2SLS is significantly less biased than both the LIML and Fuller in estimation of the group effect but significantly more biased in estimation of the process variable when instruments are weaker. The Fuller estimator shows the least bias in the process variable until the simulations with the most severely weak instruments. Comparison of estimator variances indicate that the 2SLS has significantly less variance than LIML and Fuller in estimation of both the group and process effects at all strengths of instruments.

Comparing the mean-squared errors of the estimates presented in Table 2.7 the results are similar across all methods. As expected from tests of bias and precision when the instruments are selected by the LASSO the 2SLS estimator has the lowest MSE for the group effect and Fuller adjustment for the process effect.

**Categorical post-randomisation process variable**

As has been shown in simulation exercise 1a the estimates when the post-randomisation process variable is categorical are much less precise. When a categorical process variable is simulated the 2SLS provides estimates with the smallest bias and variance for both the

group and process effect at all strengths of instruments. The significant test results for bias and variance are summarised by the mean-squared error in Table 2.9.

Some authors use the median bias when presenting results for the LIML estimator. The results for this outcome measure are provided in Appendix 1 and indicate that the LIML and Fuller methods perform better than the 2SLS in terms of median bias.

**Conclusion**

When measured variables explain a large amount of the variation in a continuous process variable there is little to distinguish between the estimators regardless of the number of instruments used. This is true for both continuous and categorical process variables. The 2SLS is more precise than both the LIML and Fuller under all simulated scenarios. When a continuous process variable is modelled the 2SLS is more biased than both the LIML and Fuller in estimation of the process effect especially when instruments are weak. When a categorical process variable is used the 2SLS estimator is less bias than the LIML and Fuller. The bias and precision is summarised by the mean-squared error which is lowest for the 2SLS estimator for a categorical process variable. The median bias has been used by other authors to compare estimation methods but measures bias only and does not account for the precision of the estimates, whereas the mean-squared error summarises both bias and precision. It is therefore preferable to minimise the mean-squared error.

**Table 2.7: MEAN-SQUARED ERROR of estimates by estimation method with increasing uncertainty in post-randomisation process variable: 1000 simulations, CONTINUOUS process variable, correlation of error terms =0.69, sample size=200**

| Standard deviation of mediator error | Average first stage F-statistic | | | OLS | All variables | | | LASSO selected | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | All relevant variables | ALL variables | Effect | All | 2SLS | LIML | Fuller | 2SLS | LIML | Fuller |
| 5 explanatory variables | | | | | | | | | | |
| 1 | 343 | 80.1 | Group | 2.331 | **2.274** | 2.284 | 2.284 | **2.269** | 2.275 | 2.274 |
| | | | Mediator | 0.315 | 0.063 | 0.063 | **0.063** | **0.063** | 0.063 | 0.063 |
| 2 | 71.8 | 17.4 | Group | 2.390 | **2.251** | 2.301 | 2.297 | **2.237** | 2.266 | 2.263 |
| | | | Mediator | 0.830 | 0.083 | 0.079 | **0.078** | 0.079 | 0.077 | **0.076** |
| 3 | 21.6 | 5.8 | Group | 2.358 | **2.195** | 2.360 | 2.343 | **2.157** | 2.237 | 2.226 |
| | | | Mediator | 1.213 | 0.156 | 0.143 | **0.136** | 0.134 | 0.119 | **0.118** |
| 4 | 4.2 | 1.8 | Group | 2.279 | **2.154** | 100.694 | 2.778 | **2.125** | 2.215 | 4.960 |
| | | | Mediator | 1.352 | **0.660** | 114.144 | 1.062 | 0.610 | **0.566** | 13.030 |
| 10 explanatory variables | | | | | | | | | | |
| 1 | 161.4 | 87.4 | Group | 2.203 | **2.151** | 2.161 | 2.160 | **2.147** | 2.154 | 2.154 |
| | | | Mediator | 0.252 | 0.058 | 0.058 | **0.058** | 0.058 | 0.058 | **0.058** |
| 2 | 35.0 | 19.3 | Group | 2.268 | **2.129** | 2.174 | 2.171 | **2.120** | 2.154 | 2.151 |
| | | | Mediator | 0.672 | 0.074 | 0.071 | **0.070** | 0.073 | 0.070 | **0.069** |
| 3 | 11.7 | 6.7 | Group | 2.280 | **2.082** | 2.217 | 2.206 | **2.065** | 2.154 | 2.145 |
| | | | Mediator | 1.022 | 0.126 | 0.115 | **0.111** | 0.120 | 0.106 | **0.104** |
| 4 | 3.5 | 2.4 | Group | 2.218 | **2.007** | 3.025 | 2.519 | **1.968** | 2.066 | 7.181 |
| | | | Mediator | 1.187 | **0.377** | 1.211 | 0.479 | 0.367 | **0.308** | 9.860 |

**Table 2.8: BIAS of estimates by estimation method with increasing uncertainty in post-randomisation process variable: 1000 simulations, CONTINUOUS process variable, correlation of error terms =0.69, sample size=200**

| Standard deviation of mediator error | Average first stage f-statistic | | | OLS | All variables | | | LASSO selected | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | All relevant variable | ALL variables | Effect | All | 2SLS | LIML | Fuller | 2SLS | LIML | Fuller |
| 5 explanatory variables | | | | | | | | | | |
| 1 | 343 | 80.1 | Group | -0.332 | -0.055 | **-0.019** | -0.021 | -0.044 | **-0.024** | -0.026 |
| | | | Mediator | 0.508 | 0.048 | -0.013 | **-0.010** | 0.031 | -0.003 | **0.000** |
| 2 | 71.8 | 17.4 | Group | 0.559 | 0.102 | **0.016** | 0.021 | 0.081 | **0.034** | 0.038 |
| | | | Mediator | -0.885 | -0.123 | 0.020 | **0.012** | -0.090 | **-0.010** | -0.018 |
| 3 | 21.6 | 5.8 | Group | -0.681 | -0.187 | **-0.008** | -0.020 | -0.162 | **-0.070** | -0.080 |
| | | | Mediator | 1.087 | 0.261 | -0.039 | **-0.019** | 0.198 | **0.044** | 0.060 |
| 4 | 4.2 | 1.8 | Group | 0.739 | 0.478 | -0.189 | **0.126** | 0.406 | 0.316 | **0.287** |
| | | | Mediator | -1.155 | -0.704 | 0.459 | **-0.064** | -0.603 | -0.443 | **-0.195** |
| 10 explanatory variables | | | | | | | | | | |
| 1 | 161.4 | 87.4 | Group | -0.267 | -0.023 | 0.011 | **0.010** | -0.023 | 0.002 | **0.000** |
| | | | Mediator | 0.446 | 0.045 | -0.012 | **-0.009** | 0.038 | -0.004 | **-0.001** |
| 2 | 35.0 | 19.3 | Group | 0.482 | 0.064 | -0.015 | **-0.011** | 0.059 | **0.001** | 0.005 |
| | | | Mediator | -0.793 | -0.112 | 0.017 | **0.010** | -0.097 | **-0.002** | -0.009 |
| 3 | 11.7 | 6.7 | Group | 0.611 | 0.134 | -0.025 | **-0.016** | 0.126 | **0.017** | 0.025 |
| | | | Mediator | -0.996 | -0.226 | 0.029 | **0.013** | -0.199 | **-0.024** | -0.037 |
| 4 | 3.5 | 2.4 | Group | 0.669 | 0.312 | -0.123 | **-0.047** | 0.278 | 0.155 | **0.056** |
| | | | Mediator | -1.081 | -0.516 | 0.145 | **0.036** | -0.472 | -0.275 | **-0.152** |

**Table 2.9: MEAN-SQUARED ERROR of estimates by estimation method with increasing uncertainty in post-randomisation process variable: 1000 simulations, CATEGORICAL process variable, correlation of error terms =0.69, sample size=200**

| | | OLS | All variables | | | LASSO selected | | |
|---|---|---|---|---|---|---|---|---|
| Standard deviation of mediator error | Effect | All | 2SLS | LIML | Fuller | 2SLS | LIML | Fuller |
| 5 explanatory variables | | | | | | | | |
| 1 | Group | 7.56 | **5.59** | 6.55 | 6.47 | **5.83** | 6.43 | 6.36 |
| | Mediator | 20.37 | **12.84** | 16.80 | 16.48 | **13.89** | 16.39 | 16.12 |
| 2 | Group | 15.98 | **6.45** | 7.66 | 7.49 | **6.64** | 7.36 | 7.25 |
| | Mediator | 54.38 | **16.49** | 21.35 | 20.69 | **17.13** | 19.98 | 19.52 |
| 3 | Group | 25.66 | **9.70** | 13.68 | 12.60 | **9.42** | 10.43 | 10.09 |
| | Mediator | 93.04 | **29.48** | 45.64 | 41.35 | **28.27** | 32.35 | 30.97 |
| 10 explanatory variables | | | | | | | | |
| 1 | Group | 6.83 | **5.55** | 6.53 | 6.46 | **5.59** | 6.35 | 6.29 |
| | Mediator | 19.18 | **13.67** | 17.48 | 17.19 | **13.83** | 16.77 | 16.52 |
| 2 | Group | 14.37 | **6.50** | 7.87 | 7.71 | **6.53** | 7.52 | 7.40 |
| | Mediator | 49.78 | **17.86** | 23.08 | 22.46 | **17.89** | 21.60 | 21.13 |
| 3 | Group | 23.49 | **8.99** | 12.35 | 11.65 | **8.89** | 10.74 | 10.31 |
| | Mediator | 86.46 | **28.08** | 41.33 | 38.53 | **27.43** | 34.61 | 32.88 |

**Table 2.10: BIAS of estimates by estimation method with increasing uncertainty in post-randomisation process variable: 1000 simulations, CATEGORICAL process variable, correlation of error terms =0.69, sample size=200**

| Standard deviation of mediator error | Effect | OLS | All variables | | | LASSO selected | | |
|---|---|---|---|---|---|---|---|---|
| | | All | 2SLS | LIML | Fuller | 2SLS | LIML | Fuller |
| 5 explanatory variables | | | | | | | | |
| 1 | Group | -1.934 | -0.385 | 0.052 | **0.025** | -0.281 | **-0.032** | -0.056 |
| | Mediator | 3.795 | 0.703 | -0.171 | **-0.117** | 0.489 | **-0.010** | 0.036 |
| 2 | Group | -3.555 | -0.885 | 0.097 | **0.032** | -0.685 | **-0.127** | -0.182 |
| | Mediator | 7.043 | 1.711 | -0.251 | **-0.121** | 1.305 | **0.190** | 0.300 |
| 3 | Group | -4.748 | -1.732 | 0.214 | **0.045** | -1.426 | **-0.421** | -0.543 |
| | Mediator | 9.423 | 3.392 | -0.494 | **-0.156** | 2.750 | **0.740** | 0.984 |
| 10 explanatory variables | | | | | | | | |
| 1 | Group | 1.735 | 0.337 | -0.077 | **-0.053** | 0.299 | -0.013 | **0.009** |
| | Mediator | -3.485 | -0.694 | 0.134 | **0.086** | -0.605 | **0.017** | -0.027 |
| 2 | Group | 3.286 | 0.816 | -0.116 | **-0.058** | 0.734 | **0.040** | 0.092 |
| | Mediator | -6.602 | -1.658 | 0.206 | **0.090** | -1.485 | **-0.096** | -0.201 |
| 3 | Group | 4.514 | 1.580 | -0.220 | **-0.081** | 1.454 | 0.193 | **0.308** |
| | Mediator | -9.047 | -3.182 | 0.421 | **0.143** | -2.914 | **-0.395** | -0.627 |

### 2.5.3 Simulation study 2: validating the estimation of the post-randomisation process and mediator model

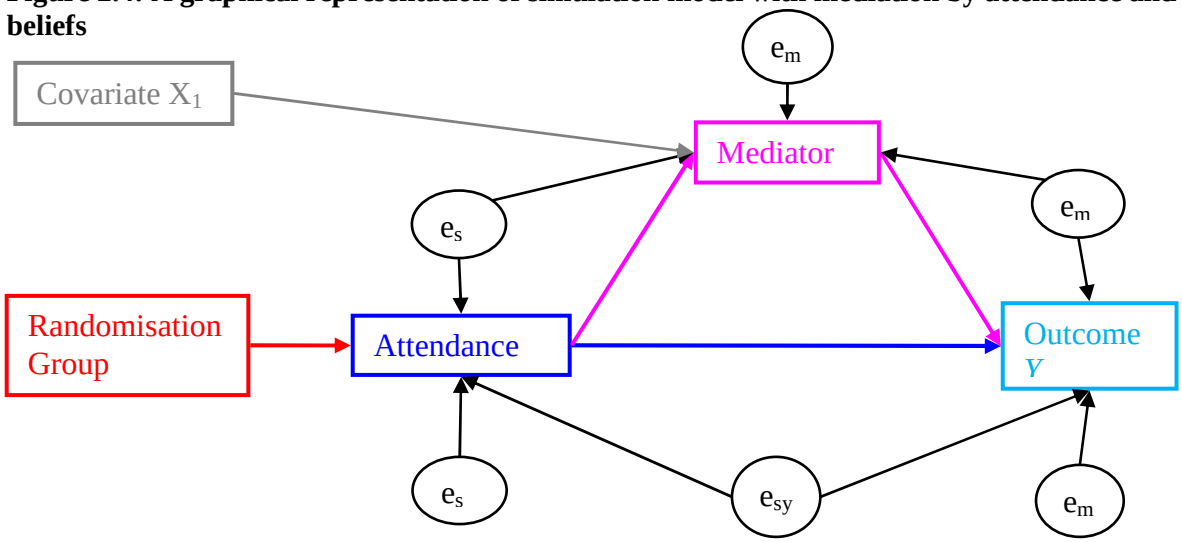
**Aim:**

To establish the accuracy of two-stage least squares estimation for a process variable and mediator model of attendance at therapy sessions and beliefs.

**Method:**

The simulation model with all error terms included is presented below (Figure 2.4). This is the situation where a process variable, attendance at therapy, causes a change in a mediator but where confounding is present and was described earlier. This situation is simulated to establish the performance of the instrumental variable estimator against the ordinary least squares estimator. Data are simulated to replicate that of a randomised trial with an indicator for group allocation, a post randomisation process variable, attendance, available only in the treated arm, a mediator recorded in both arms, an outcome and baseline covariates recorded for all.

**Figure 2.4: A graphical representation of simulation model with mediation by attendance and beliefs**




The variable that is considered to represent the number of sessions attended has a value of zero in the control group and is a positive continuous variable in the treatment arm normally distributed around a mean of ten. The error in the attendance variable consists of three parts, one that is also associated with the mediator, one that is also associated with the outcome thereby introducing confounding of these two relationships and a third which is just on attendance.

The mediator is a continuous variable calculated as a function of number of sessions attended, a baseline covariate $x_1$ and error. The baseline covariate $x_1$ is normally distributed with mean zero and a standard deviation which is altered to balance the variation in error so that the total standard deviation of the mediator variable remains approximately the same between simulation models. As with the attendance the error term for the mediator consists of three parts, one that is associated with the error in the process variable, one that is associated with the error in the outcome variable and one that is just on the mediator. The outcome is also a continuous variable calculated as a function of attendance and mediator and three error terms. The error term includes a part that is associated with attendance, a part that is associated with the mediator and a part that is only on the outcome.

Explanatory variable – $x_1 \sim N(0, sd_x)$

Attendance variable $m_1 = \begin{cases} 0 \text{ if group}=0 \\ 10+(a*e_{sy}+b*e_{sm}+c*e_s) \text{ if group}=1 \end{cases}$

Mediator $m_2 = 3 + x_1 + 2 * m_1 + (d * e_{my} + e * e_{sm} + f * e_m)$

Outcome $y = 50 - 10 * m_1 + 50 * m_2 + (g * e_{sy} + h * e_{my} + i * e_y)$

Where $e_m, e_s, e_y, e_{sm}, e_{sy}, e_{my}$ are standard normal variables. The total error in each variable is calculated as:

Total error in sessions attended: $\varepsilon_s = a * e_{sy} + b * e_{sm} + c * e_s$

Total error in mediator: $\varepsilon_m = d * e_{my} + e * e_{sm} + f * e_m$

Total error in outcome: $\varepsilon_y = g * e_{sy} + h * e_{my} + i * e_y$

It is expected that an ordinary least squares regression including both attendance and mediator as that described by Baron and Kenny will be biased if there is confounding present that cannot be accounted for. Instrumental variable estimation described earlier is expected to give unbiased estimates in this situation by using instruments to predict attendance and mediator values and so remove correlations with the outcome. To estimate the model there must be at least as many instruments as mediators. If a process variable and a mediator are considered then an instrument must be found for each one. In this

situation the instrument must be associated with the mediator in question but not with the outcome other than through the mediator.

The performance of the instrumental variables model against the ordinary least squares estimation is tested in relation to

1. Bias (unmeasured confounding)
   Assess the impact of increasing unmeasured confounding between: process and outcome, process and mediator, mediator and outcome. The level of unmeasured confounding is indicated by measuring the correlation in the errors of the variables in question. Two pairs of error correlations are held constant whilst altering the correlation in the errors of the other. In order for each model to be comparable the the total error in each variable is kept the same. This is achieved by altering the values of a-i in the equations for attendance, mediator and outcome stated above. The values given to a-i in each simulation model and the associated correlations of the error terms in the model are given in Table 2.11. In these simulations the standard error of the covariate $x_1$ is kept the same at a value of one.

**Table 2.11: Details of the simulation models to assess the impact of altering the level of unmeasured confounding between variables**

| | attendance error values | | | mediator error values | | | outcome error values | | | Correlations | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Model | a | b | c | d | e | f | g | h | i | $\varepsilon_s, \varepsilon_m$ | $\varepsilon_s, \varepsilon_y$ | $\varepsilon_m, \varepsilon_y$ |
| | vary confounding between attendance and outcome | | | | | | | | | | | |
| 1 | 1 | 1 | 1 | 2 | 1 | 1 | 1.00 | 1 | 1.00 | 0.237 | 0.331 | 0.468 |
| 2 | 1 | 1 | 1 | 2 | 1 | 1 | 1.39 | 1 | 0.28 | 0.237 | 0.462 | 0.468 |
| 3 | 1 | 1 | 1 | 2 | 1.39 | 0.28 | 4.47 | 1 | 0.45 | 0.235 | 0.647 | 0.468 |
| 4 | 1 | 1 | 1 | 2 | 4.47 | 0.45 | 1.41 | 1 | 0.07 | 0.234 | 0.659 | 0.468 |
| 5 | 1 | 1 | 1 | 2 | 1.41 | 0.07 | 1.41 | 1 | 0.05 | 0.234 | 0.662 | 0.468 |
| | vary confounding between mediator and outcome | | | | | | | | | | | |
| 1 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1.00 | 1.00 | 0.236 | 0.468 | 0.328 |
| 2 | 2 | 1 | 1 | 1.39 | 1 | 0.28 | 1 | 1.00 | 1.00 | 0.238 | 0.468 | 0.460 |
| 3 | 2 | 1 | 1 | 4.47 | 1 | 0.45 | 1 | 1.39 | 0.28 | 0.238 | 0.470 | 0.648 |
| 4 | 2 | 1 | 1 | 1.41 | 1 | 0.07 | 1 | 4.47 | 0.45 | 0.239 | 0.471 | 0.661 |
| 5 | 2 | 1 | 1 | 1.41 | 1 | 0.05 | 1 | 1.41 | 0.07 | 0.239 | 0.471 | 0.664 |
| | vary confounding between sessions and mediator | | | | | | | | | | | |
| 1 | 1 | 1 | 1 | 1 | 1.00 | 1.00 | 2 | 1 | 1 | 0.334 | 0.470 | 0.231 |
| 2 | 1 | 1 | 1 | 1 | 1.39 | 0.28 | 2 | 1 | 1 | 0.464 | 0.470 | 0.234 |
| 3 | 1 | 1.39 | 0.28 | 1 | 4.47 | 0.45 | 2 | 1 | 1 | 0.650 | 0.468 | 0.234 |
| 4 | 1 | 4.47 | 0.45 | 1 | 1.41 | 0.07 | 2 | 1 | 1 | 0.661 | 0.468 | 0.235 |
| 5 | 1 | 1.41 | 0.07 | 1 | 1.41 | 0.05 | 2 | 1 | 1 | 0.664 | 0.468 | 0.235 |

Note: $\varepsilon_s$ is total error in sessions attended, $\varepsilon_m$is total error in mediator, $\varepsilon_y$ is total error in outcome

2. Unmeasured variation in the mediator

   To assess the impact of increasing the proportion of unmeasured variation in the mediator the amount of bias is kept constant. This is achieved by keeping the same level of correlation between the process, mediator and outcome. The total variation in the mediator is kept the same but the amount of unknown variation is increased by trading off variation in known covariate $x_1$ ($sd_x$) and that attributable to error in the mediator only ($f * e_m$). In these simulations the values of a- e and g-i and the correlation between the errors of the process, mediator and outcome are kept the

same at a=2, b=1, c=0, f=0, g=1, h=1.5 and i=0; $\text{cor}(\varepsilon_s, \varepsilon_m)$=0.26, $\text{cor}(\varepsilon_s, \varepsilon_y)$=0.39 and $\text{cor}(\varepsilon_m, \varepsilon_y)$=0.77.

**Table 2.12: Details of the simulation models to assess the impact of altering the amount of explained variation in the mediator**

| Values | | Standard deviations | | |
|---|---|---|---|---|
| c | d | x | attendance | Y |
| 2.50 | 1.00 | 10.48 | 14.9 | 713.3 |
| 5.00 | 2.00 | 9.38 | 15.0 | 716.5 |
| 7.50 | 3.00 | 6.96 | 15.0 | 713.7 |
| 8.75 | 3.50 | 5.31 | 15.1 | 721.9 |
| 9.38 | 3.75 | 3.88 | 15.1 | 722.7 |
| 10.00 | 4.00 | 1.00 | 15.1 | 721.8 |

Two analyses are conducted on the simulation model to estimate the attendance and mediator effects on outcome. The ordinary least squares estimation is a regression of outcome on attendance and the mediator. The instrumental variable estimation uses the most appropriate instrument. Randomisation group is used as an instrument for attendance as it is strongly associated with number of sessions but is not associated with either the mediator or outcome other than through attendance. Covariate $x_1$ is used as an instrument for the mediator as it is associated with the mediator but not directly associated with either the process or outcome. The instruments are analysed in a structural equation model using the sem command in Stata[111] with attendance explained by randomisation group, mediator explained by attendance and $x_1$, and outcome explained by attendance and mediator. The model is defined within Stata allowing for correlations between attendance, mediator and outcome.

Estimated coefficients for the attendance effect and mediator effect on outcome are saved for each simulation and models are compared by calculation of the absolute difference (bias) and root mean squared error (precision) from the true coefficients of attendance and mediator on outcome. Differences in bias between the selection methods will be tested as in the previous simulations using a paired t-test of the absolute bias and Pitman-Morgan test for the difference between two correlated variances. 1000 simulations were conducted in each instance.

#### *2.5.3.1 Simulation results*

**Effect of bias on model estimates**

The effect of bias in each pathway of the model is investigated by holding bias constant in two pathways and varying the bias in the third. To be able to compare models fairly the overall variation in the attendance, mediator and outcome variables is kept the same through the simulations.

Results are provided in Table 2.13 giving the mean squared error for both the process variable and mediator variable under the ordinary least squares and instrumental variable estimations. The unmeasured confounding in each model is described using the correlation in errors between the variables. The three overall models hold two correlations constant and allow the third to vary. The relationships that are held constant are indicated at the top of the set of columns and the value of the correlation that varies is given for each model across the rows.

In each simulation model the absolute bias of both the attendance and mediator in the OLS model is significantly greater than the absolute bias in the instrumental variable model, measured by a paired t-test. However, the variation of the attendance and mediator estimates is significantly greater in the instrumental variables model than the OLS model. These two measures are summarised by the mean squared error which is shown in Table 2.13 to be lower for the instrumental variables analysis than the OLS estimation in all cases.

The results show that as the correlation in errors between the mediator and outcome increases the performance of the OLS estimate declines whereas the instrumental variable estimates remain approximately the same. When the correlation in errors between attendance and mediator or attendance and outcome are allowed to vary keeping the others constant the OLS estimates of the attendance and mediator effect remain approximately the same.

There is very little difference in the performance of the instrumental variable as bias is increased in any path of the model, this is indicated by similar results for the IV estimates when reading down the columns of Table 2.13. The IV estimates are calculated with a known and strong instrument so the simulations show that the method works under these conditions when there is unmeasured confounding present.

**Effect of unmeasured variation in the mediator**

The impact of higher levels of unmeasured variation in the mediator on effect estimates is investigated whilst keeping the level of confounding the same. The correlations between each pair of variables is set as follows, correlation of attendance and mediator is 0.26, correlation of attendance and outcome is 0.39, correlation of mediator and outcome is 0.77. The level of explained variation in attendance is kept the same, summarised by an average $R^2$ value of 0.961 in each of the simulation specifications. The simulations indicate as expected that the estimates from the IV analysis have less bias and better precision than the OLS estimator when the instrument explains a large amount of the variation in the mediator. Once the level of unmeasured variation is too great the IV analysis loses precision, indicated by a large mean-squared error though the mean bias remains smaller than the bias in the OLS estimate.

**Table 2.13: Comparison of estimation method performance as the level of unmeasured confounding varies, MEAN-SQUARED ERROR reported, sample size N=1000, simulations=1000**

| | Vary confounding between attendance and outcome $cor(\varepsilon_s, \varepsilon_m) = 0.24, cor(\varepsilon_m, \varepsilon_y) = 0.47$ | | | | Vary confounding between mediator and outcome $cor(\varepsilon_s, \varepsilon_m) = 0.24, cor(\varepsilon_s, \varepsilon_y) = 0.47$ | | | | Vary confounding between attendance and mediator $cor(\varepsilon_s, \varepsilon_y) = 0.47, cor(\varepsilon_m, \varepsilon_y) = 0.23$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Model correlations | | Attendance | Mediator | Model correlations | | Attendance | Mediator | Model correlations | | Attendance | Mediator |
| 1 | $cor(\varepsilon_s, \varepsilon_y)$ | 0.331 | OLS | 0.316 | 0.082 | $cor(\varepsilon_m, \varepsilon_y)$ | 0.328 | OLS | 0.221 | 0.062 | $cor(\varepsilon_s, \varepsilon_m)$ | 0.334 | OLS | 0.234 | 0.065 |
| | | | IV | 0.072 | 0.018 | | | IV | 0.062 | 0.015 | | | IV | 0.117 | 0.029 |
| 2 | $cor(\varepsilon_s, \varepsilon_y)$ | 0.462 | OLS | 0.305 | 0.082 | $cor(\varepsilon_m, \varepsilon_y)$ | 0.460 | OLS | 0.445 | 0.121 | $cor(\varepsilon_s, \varepsilon_m)$ | 0.464 | OLS | 0.238 | 0.066 |
| | | | IV | 0.064 | 0.016 | | | IV | 0.062 | 0.015 | | | IV | 0.117 | 0.029 |
| 3 | $cor(\varepsilon_s, \varepsilon_y)$ | 0.647 | OLS | 0.292 | 0.081 | $cor(\varepsilon_m, \varepsilon_y)$ | 0.648 | OLS | 0.897 | 0.237 | $cor(\varepsilon_s, \varepsilon_m)$ | 0.650 | OLS | 0.239 | 0.066 |
| | | | IV | 0.057 | 0.014 | | | IV | 0.061 | 0.015 | | | IV | 0.115 | 0.029 |
| 4 | $cor(\varepsilon_s, \varepsilon_y)$ | 0.659 | OLS | 0.291 | 0.081 | $cor(\varepsilon_m, \varepsilon_y)$ | 0.661 | OLS | 0.932 | 0.246 | $cor(\varepsilon_s, \varepsilon_m)$ | 0.661 | OLS | 0.240 | 0.066 |
| | | | IV | 0.056 | 0.014 | | | IV | 0.061 | 0.015 | | | IV | 0.115 | 0.029 |
| 5 | $cor(\varepsilon_s, \varepsilon_y)$ | 0.662 | OLS | 0.291 | 0.081 | $cor(\varepsilon_m, \varepsilon_y)$ | 0.664 | OLS | 0.941 | 0.248 | $cor(\varepsilon_s, \varepsilon_m)$ | 0.664 | OLS | 0.242 | 0.070 |
| | | | IV | 0.056 | 0.014 | | | IV | 0.061 | 0.015 | | | IV | 0.203 | 0.050 |

Note: $cor(\varepsilon_a, \varepsilon_m)$=correlation of error in attendance and mediator, $cor(\varepsilon_a, \varepsilon_y)$=correlation in errors of attendance and outcome, $cor(\varepsilon_m, \varepsilon_y)$=correlation in errors of mediator and outcome

**Table 2.14: Comparison of estimation methods with a process and mediator variable as strength of the instrument for the mediator reduces, n=200, simulations=1000**

| Simulation summary statistics | | Estimation | Attendance | | Mediator | |
|---|---|---|---|---|---|---|
| | | | Coefficient | MSE | Coefficient | MSE |
| attendance $R^2$ value | 0.962 | OLS BK | -10.042 | 0.013 | 50.031 | 0.002 |
| mediator $R^2$ value | 0.941 | SEM | -9.996 | 0.011 | 49.999 | 0.001 |
| sd of mediator | 14.946 | | | | | |
| attendance $R^2$ value | 0.962 | OLS BK | -10.108 | 0.022 | 50.065 | 0.005 |
| mediator $R^2$ value | 0.837 | SEM | -9.996 | 0.012 | 50.001 | 0.002 |
| sd of mediator | 15.007 | | | | | |
| attendance $R^2$ value | 0.962 | OLS BK | -10.181 | 0.043 | 50.100 | 0.011 |
| mediator $R^2$ value | 0.667 | SEM | -9.996 | 0.018 | 50.000 | 0.003 |
| sd of mediator | 14.955 | | | | | |
| attendance $R^2$ value | 0.962 | OLS BK | -10.194 | 0.048 | 50.104 | 0.012 |
| mediator $R^2$ value | 0.566 | SEM | -9.991 | 0.022 | 49.997 | 0.004 |
| sd of mediator | 15.114 | | | | | |
| attendance $R^2$ value | 0.962 | OLS BK | -10.227 | 0.062 | 50.120 | 0.015 |
| mediator $R^2$ value | 0.508 | SEM | -9.983 | 0.048 | 49.993 | 0.010 |
| sd of mediator | 15.129 | | | | | |
| attendance $R^2$ value | 0.962 | OLS BK | -10.244 | 0.070 | 50.130 | 0.018 |
| mediator $R^2$ value | 0.446 | SEM | -9.909 | 16.981 | 49.958 | 4.460 |
| sd of mediator | 15.110 | | | | | |

## 2.6 Summary and conclusions

In this chapter it was shown that the 2SLS IV estimator can be biased in finite samples and that this is dependent upon the amount of unmeasured confounding, the number of instruments used, the strength of those instruments and the sample size. Once a trial has been conducted the sample size and level of unmeasured confounding can no longer be influenced and the analysis can only control the number and strength of the instruments. When an instrument has not been included in the design of the trial an appropriate instrument must be selected. In practice there may be several potential instruments none of which may be very strong alone. Bias in the 2SLS estimates will be low if only one instrument is used but if weak then the estimates will be very imprecise and could give misleading results. A balance between bias and precision is required. The rule of thumb to

select instruments providing a first stage F-statistic>10 has been criticised by several authors and some authors have found that it is beneficial to use multiple instruments when individual instruments are weak. This chapter focussed on the selection of instruments and estimation of the IV model.

Simulations designed to represent an RCT of a complex intervention compared methods for instrument selection and estimation methods in the presence of weak instruments. These simulations have shown that when instruments are strongly associated with the process variable the choice of instrument selection method is not very important as they are comparable in terms of bias and precision of estimates. The LASSO that minimises the mean-squared error of the first stage model is preferred over the stepwise selection procedure as it tends to produce estimates with lower variance but is comparable in terms of absolute bias. When instruments are only weakly associated with the process variable the mean squared error is lower for the LASSO. The parsimonious selection models can reduce bias but when instruments are very weak struggle to find any instruments. When a categorical post-randomisation process variable is used the LASSO that minimises the mean-squared error is preferred in terms of reducing the mean-squared error when instruments are weak. This selection method is significantly better than the stepwise and parsimonious LASSO in terms of reducing variance and bias.

The second set of simulations compared estimation methods in the same scenario. The results indicate that the 2SLS estimator is preferred over LIML and Fuller in terms of reducing the variance of the estimates. Although the 2SLS estimate of a continuous process variable is more biased than the LIML and Fuller when instruments are weak the mean-squared error is smaller. The 2SLS is less biased and more precise than the LIML and Fuller's when the process variable is binary. LIML and Fuller's adjusted estimators are preferred in terms of the median bias but this does not take into account the precision of the estimates.

Finally a two mediator process was simulated and the instrumental variable results compared to ordinary least squares regression for different types of confounding and instrument strength. If no unmeasured confounding is expected between the mediator and outcome then an OLS model will give estimates comparable to the IV. Instrumental variables regression provides better estimates than OLS when compared by bias and mean-squared error in the presence of unmeasured confounding between mediator and outcome.

# 3 Additional statistical methods

## 3.1 Introduction

The main focus of this thesis is in the application of instrumental variable methods in mediation analyses of complex intervention trials. In order to carry out the specific analyses in the EDIE-II and COMMAND trial datasets thoroughly some additional methods are applied. These are established statistical methods summarised here briefly to give a full presentation of the analyses that have been undertaken. Methods for longitudinal analysis, missing data, bootstrapping and instrumental variables regression of binary outcomes are detailed.

## 3.2 Longitudinal analysis

The EDIE-II trial benefits from having multiple measurement points. Outcome and possible mediator measurements have been recorded at monthly intervals for the first 6 months and then every 3 months until the end of follow-up. The therapy is given over the course of the first six months so during this time the amount of the therapy received will be different at each measurement point. No more therapy is given after six months and so for the rest of the follow-up points the amount of therapy received is unchanging. This type of data provides a wealth of information and the ability to look at temporal changes necessary for causal inference.

In the situation when the exposure of interest, for example treatment received, does not change over the course of the measurements analysed, which in this example would be if looking only at the outcome measures post-treatment, the analysis is relatively straight forward and established techniques can be used. The basic issue when measures are taken on the same individuals at multiple time points is that the observations within one person are no longer independent; which is an assumption of simple regression methods. Standard techniques for analysing repeated measures data where the exposure does not change over time are time-series models with a random effect, repeated measures analysis of variance or growth curve analysis. These methods are general regression methods that account for variation within as well as between individuals. In the following analyses of the EDIE-II and COMMAND data treatment effects are estimated from outcomes measured over time in an intention-to-treat analysis using time-series models. These models are implemented using the xtreg or xtlogit[110] commands in Stata[111] which are random effects models accounting for repeated measures within an individual. The outcome time points are

modelled from the first follow-up point and adjusted for the baseline measure of the outcome. All models include a time component which is a measure of the average change in outcome over time across the treatment groups. Treatment group by time interactions are investigated in the models, the interaction shows how the treatment effect differs over the follow-up time points. If the interaction is not significant this indicates that the treatment effect does not differ over time. In this case the interaction may be removed and the average treatment effect over the follow-up time period is estimated.

When considering a post-randomisation process or mediation analysis we are interested in knowing how the outcome differs over time at different levels of the process/mediator. To do this we can simply include the mediator/process as a covariate in the analysis model. However, this is analogous to using a Baron and Kenny mediation model on an outcome with a single time point and assumes that there is no unmeasured confounding. The mediation methods described previously can be applied to longitudinal data if the mediator or post-randomisation process occurs at a time prior to the outcome measures. In this way the standard IV regression will still predict values of the mediator or process variable based on the instrument and apply these in the longitudinal model rather than the observed values.

A more recent method is the Latent Class Growth Analysis (LCGA) or group-based trajectory modelling which has developed from growth curve modelling to analyse patterns in longitudinal data[151]. Growth curve models describe longitudinal data by the path trajectory of an individual's outcome over time. Each individual has their own trajectory (plot of their outcome over time) and an average trajectory over the sample can be described by the average intercept and slope. The intercept is the average outcome at baseline and the slope is the change in outcome over time for individuals in the latent class. Rather than one overall average trajectory the latent class model splits observations into groups with similar outcome trajectories, these groups of trajectories are the latent classes. They are called latent classes as the similarity between observations within classes is inferred rather than observed. Once classified into trajectory groups associations between class membership and covariates can be investigated and characteristics of the classes described. This method can be used to determine factors accounting for the variance in patterns of outcome between participants.

Principal stratification methods (as a cross-section or longitudinal analysis) can be applied to the LCGA by defining the classes rather than allowing them to be determined by the data. Principal strata are defined as described in Chapter 1.2.8.2.2, so for example we can investigate treatment effects within strata of compliers and non-compliers. Probability of latent class membership is estimated by randomisation or randomisation by baseline covariate interactions. The effect of treatment is then estimated within each class over time. The intercept is interpreted as the treatment effect when time is coded as zero (see below) and the slope as the change in the treatment effect over time. LCGA can be estimated in the Mplus software package[152]. Binary indicators of compliance class are defined in the treatment arm and baseline covariates used to predict probability of class membership in the treatment and control arms (equivalent to the use of the baseline covariate by treatment interactions as instrumental variables in a conventional instrumental variables regression). The intercept and slope is calculated to describe the outcome over time within the compliance classes. In the analysis of the EDIE-II trial CACE analysis methods will be applied to growth curve analyses to estimate the trajectory of outcome measures over time within people that would be compliers. Estimating the treatment effect at baseline is not of interest (it would be expected to be very close to zero) and so time is centred at the primary outcome point 12 months. The intercept is therefore the treatment effect at 12 months within compliers and the slope is still the change in treatment effect over time.

When exposures change over the time that the outcome is measured the analysis is more complicated, for example, outcome measures that are taken during the treatment window when the amount of treatment received will change. There are additional statistical considerations to take into account in these situations as the multiple measurements will be interrelated; exposure level at time 1 is expected to influence outcome at time 1 and time 2 but outcome at time 1 may also influence exposure at time 2. In addition there is likely to be confounding of these associations at each time point. This means that a simple association between an exposure and outcome over time will give biased effect estimates[153].

The complexity of causal inference analysis in data with time-varying confounding is not the subject of this thesis and so the analysis of the trial data will only consider participant outcomes after the completion of therapy.

## 3.3 Missing data

Unfortunately missing information is a problem in all studies including both EDIE-II and COMMAND. No study provides perfect data. There are two main missing data problems, missing baseline covariate data and missing outcome data. They are likely to have different reasons for being missing and are therefore dealt with in different ways.

If data are missing completely at random (MCAR) it is assumed that there is no pattern to the missing data at all. If this is the case then there would be no bias associated with losing information and it could be ignored. Missing at random (MAR) assumes that the missing data can be predicted by observed covariates but that it is not dependent on the value of the data that is missing. Many statistical packages will automatically remove observations that have missing information on the variables included in the analysis. This complete case analysis may not bias the results if they are truly MCAR but it does mean that the sample is smaller and will have less statistical power. It is likely that missing data in the baseline questionnaire is an accidental oversight and can be considered missing completely at random, as long as there is no systematic error causing the missing data.

Missing follow-up data is a little different as this is likely to be a conscious decision on the participant's part not to continue or an inability to follow-up a participant for reasons that may be due to characteristics of that participant. It is likely that the people that are not included because they have refused to continue or cannot be located are different in some way to those that have taken part in the full study and so results only of those that have completed will not be representative of the whole. If all variables that are associated with missing outcome are included as covariates then the analysis will not be biased under the assumption that data are missing at random dependent on the covariates. Alternatively they can be used to estimate each individual participant's probability of completion and used as an inverse probability weight to give greater weight to individuals who are similar to those that have not completed the study[154]. In the following analyses of the trial data baseline covariates that are found to be associated with missing follow-up will be included as an adjustment in the regression models. Missing outcome is then assumed to be missing at random dependent on included covariates.

When baseline data is missing at random imputation methods can be used to fill in the covariate data. In a univariate imputation a single variable with missing data is regressed on other complete variables to give a prediction model for it. Given a respondents observed

data, the value of the missing items is then predicted from this model. Multiple imputation by chained equations (MICE)[155] extends the univariate imputation method to fill in the missing information in all variables. Each variable is regressed on the other variables including both observed and imputed information to predict the missing values of the dependent variable, this process continues through all variables with missing data. Several complete datasets are produced. The analysis is then carried out separately in each complete dataset and the results combined using Rubin's rules[156] to produce average effect estimates and appropriate standard errors which account for the uncertainty in the estimates as well as that of the imputations. If the covariate data is complete the imputation of missing outcome values by regression of the observed outcome values on the covariates adds no additional information in a likelihood-based analysis to the regression of outcome on covariates that is of interest[157,158]. The imputed outcome values contain no information of the regression of the outcome on covariates but the inclusion of the outcome is essential in the imputation of covariates to ensure that all of the associations are represented in the imputed data[154]. Von Hippel therefore recommends a method that he calls multiple imputation then deletion (MID) where the outcome variable is included in the imputation and therefore values of the outcome are imputed along with those of the covariates but only the observed outcomes are used to determine the likelihood of the model.

### 3.3.1 Multiple imputation of trial datasets

In the COMMAND data 19 (10%) of participants had missing baseline data on age at onset and missing covariate data was less than 3% on the other five measures that had missing data. Schafer suggests that a low rate of missing data, <5% may not have a great impact[159] though it is likely that the mechanism of the missing data is of greater importance[160]. It is assumed that the missing covariate data is missing at random and with such low levels of missingness multiple imputation is not deemed necessary. Since age at onset has the largest amount of missing data it will not be included as a potential instrument. As a sensitivity analysis the procedure will be repeated on the smaller set of subjects with complete data including the variable as a possible instrument.

In the EDIE-II data several clinical measures had data missing in more than 10% of cases with 19% missing an anxiety score. Removing participants with missing covariate data would result in a large reduction in the sample. To improve efficiency, missing covariate values were imputed using multiple imputation by chained equations implemented with the 'ice' command in Stata[155]. This was carried out on the full dataset, including all variables

used in the analysis as well as follow-up data from all time points and variables thought to be related to non-response. Five imputations were created using all variables. Variables were imputed using the appropriate model: linear, logistic or multinomial logistic. The matching method was used for variables with skewed distributions to maintain the shape and range of the distributions. The matching method predicts the missing value and then allocates the closest observed value to the prediction as the imputed value. Missing values of the outcome variables are imputed together with the covariates; however, analyses only use imputed information for baseline covariates and do not included respondents with no outcome reported. Analyses are run on the imputed datasets separately and pooled using Rubin's rules either by the mim command in Stata where possible or in excel to give appropriate estimates and standard errors.

### 3.4 The bootstrap

A 2SLS instrumental variables analysis can be carried out in two ways, as either a one-step process or in two separate stages. The one-step process provides a valid estimate of the parameter and it's associated standard error but only participants with data on both their process/mediator and outcome will be used in the estimation. The two-stage process benefits from using all participants with data available on the process/mediator in the first stage though by the second stage only those with an outcome will be used. However, the standard errors associated with the parameter estimates when the model is estimated in two stages will be too small. The second stage regression does not account for uncertainty in the values of the mediator because they are predicted rather than observed. The bootstrap can be used to obtain valid estimates of the accuracy of parameter estimates when the IV is conducted in two stages by bootstrapping the entire two-stage process.

Bootstrapping is a resampling method used to quantify the accuracy of a sample parameter estimate. The usual way of assessing accuracy of an estimate is by the standard error. For example, a set of observations $x_1, x_2, \ldots x_n$ may be summarised by the sample mean as $\bar{x} = \sum_1^n x_i/n$ and the standard error of the sample mean given by, $s.e. = \sqrt{s^2/n}$ where $s^2 = \sum_1^n (x_i - \bar{x})^2/(n-1)$. This is standard statistical practice and provides a valid measure of the mean and standard error in many cases. The bootstrap is an alternative method to determine the standard error and is described in detail by Efron and Tibshirani[161]. The authors describe the bootstrap in terms of the simple example above, a summary of which is given here.

The bootstrap uses sampling with replacement from the original data to obtain a set of estimates of a summary statistic of interest. The standard deviation of this summary measure is then calculated, and this provides an estimate of the required standard error. Specifically for the example given above, we have a set of n observations $x=(x_1, x_2, \dots x_n)$ from which we obtain a summary statistic c(x), in this case the sample mean $c(x) = \bar{x}$. We now draw a sample which again is a set of n observations $x^* = (x_1^*, x_2^*, \dots, x_n^*)$ that are randomly selected with replacement from the original x observations to provide another sample, the bootstrap sample. Since they are sampled with replacement some observations will appear multiple times in the new set and others will not be selected at all. The same summary statistic is taken of the bootstrap sample to give $c(x^*)$, in this case the sample mean $c(x^*) = \bar{x}^*$. This is repeated many times providing a set of summary statistics from the B bootstrap samples $c(x_1^*), c(x_2^*), \dots, c(x_b^*)$ . The bootstrapped standard error is then calculated as $s.e._{boot} = \sqrt{\sum_1^B \left[c(x_b^*) - \overline{c(x_b^*)}\right]^2/(B-1)}$ where $\overline{c(x_b^*)} = \sum_1^B c(x_b^*)/B$. The calculations are similar to the usual calculation of the standard error of the mean but instead applied to a sample of the summary statistics. The sample mean is described here as a summary statistic but it could be any other measure, for example the median or a regression coefficient. The benefit of the bootstrap is that no assumptions are made of the distribution of the statistic as the distribution is created empirically it is therefore useful if parametric assumptions cannot be met, for example to obtain a standard error of a sample median or if the sample size is small.

This is a simple example of one vector of observations and a simple summary statistic but the principle can be extended to multiple measures and complex statistical analyses. In the trial datasets analysed in this thesis the bootstrap selects random samples of observations from the datasets with replacement. Within each selected sample the specified analyses are conducted and the parameter estimates of interest within each sample are saved. The bootstrapped standard error for each parameter estimate can then be calculated as above.

The bootstrapped standard error can be used to create confidence intervals in the usual way under the assumption of a normal distribution of the parameter estimates. A (1- α)% confidence interval for the sample statistic c(x) would therefore be $c(x) \pm z^{(\alpha/2)} * s.e_{boot}$ where $z^{(\alpha/2)}$ is the $(\alpha/2)^{th}$ percentile of the standard normal distribution. This is an approximate confidence interval and several other ways of calculating bootstrapped confidence intervals have been developed to improve it.

The percentile interval method gives lower and upper values of the confidence interval as the $\alpha/2$ and $1-\alpha/2$ percentiles of the bootstrap sample distribution. If 1000 bootstraps are carried out and so there are 1000 estimates of the sample statistic the 95% confidence interval by the percentile method will give the $25^{th}$ and the $975^{th}$ value of the ordered sample statistic estimates. This method does not make any assumptions as to the distribution of the estimates. If they are normally distributed then the confidence interval will be similar to the normal confidence interval. A similar method to this is the bootstrap-t but this is not as consistent as the percentile interval method. The bias-corrected and accelerated ($BC_a$) interval improves on the percentile and bootstrap-t methods. The $BC_a$ is similar to the percentile interval in that the lower and upper values are taken at percentiles of the observed distribution of values. The $\alpha$ values that define the percentiles to use are not taken as the usual 5% or 10% but are adjusted to account for bias and skewness in the data.

The larger the number of bootstrap samples taken the better the bootstrap estimate of the standard error will be. If the distribution of the bootstrapped estimates is normal or the sample is large then these different methods to derive the confidence interval of the sample statistic will be approximately the same. The boot.ci command in the R program allows for the calculation of confidence intervals under each method and so they can be compared as a form of sensitivity analysis. The normal approximation of the confidence intervals will be used for consistency with other estimates unless there is large discrepancy with other methods, which will be highlighted.

Efron and Tibshirani[161] state that 50 to 200 bootstraps is usually enough to provide a good estimate of the standard error but a larger number of bootstraps, around 1000 is needed for confidence intervals. The analyses in this thesis use 1000.

### 3.5 Binary Outcomes

So far all models that have been described in this thesis have been of a continuous outcome measure, however the outcome in the COMPLIANCE example dataset is a binary indicator of compliance with the voice. It is important to consider the implications to the interpretation of results and assumptions that are made when analysing a binary outcome. If we begin with a simple model where we wish to estimate the effect of an exposure X on outcome Y we have:

$$Y_i(1) - Y_i(0) = \beta X_i + \varepsilon_i$$

This can be generalised to:

$$h\big(Y_i(1)\big) - h\big(Y_i(0)\big)\big) = \beta X_i + \varepsilon_i \qquad \textbf{Eq. 3.1}$$

where h() is a function known as a link function in generalised linear models. When Y is a continuous measure β is estimated using a linear regression model with an identity link and the error $\varepsilon_i$is assumed to follow a normal distribution with zero mean $\varepsilon_i \sim N(0, \sigma^2)$.

If Y is a binary outcome but it's binary nature is ignored and the effect estimated using linear regression (linear probability model), the outcome is an estimate of the risk difference. When the outcome is binary $E[Y_i]$ becomes $\Pr(Y = 1)$. The treatment effect $E[Y_i(1)] - E[Y_i(0)]$ is therefore the difference in probabilities and the β values are interpreted as the increase in probability of the outcome when X increases by 1 unit. This estimation is exactly the same as specified for the linear model and as such the assumptions of a linear model apply. This means that there are certain problems, the errors will be non-normal as they can only take on two values, they are heteroskedastic and although $0 \leq \Pr(Y = 1) \leq 1$ the predicted values are unbounded and so can take values outside of these limits. Although these problems do not cause bias in the point estimates the probit link function has been suggested to improve upon the linear probability model.

The probit link function models the inverse normal distribution of the probability so Eq. 3.1 becomes $\Phi^{-1}\big(Y_i(1)\big) - \Phi^{-1}\big(Y_i(0)\big)\big) = \alpha + \beta X_i + \varepsilon_i$. The framework of the probit model is based on the assumption that the binary outcome Y is an indicator of whether a normally distributed latent variable $Y^*$ is positive. Where:

$$Y_i^* = \alpha + \beta X_i + \varepsilon_i \quad \text{and so } Y_i = \begin{cases} 1 \text{ if } Y_i^* > 0 \\ 0 \text{ if } Y_i^* \leq 0 \end{cases}$$

This means that the outcome of interest $\Pr(Y = 1) = \Pr(\varepsilon_i < \alpha + \beta X_i) = \Phi(\alpha + \beta X_i)$ where Φ() is the cumulative distribution function of the normal distribution. The probit function is not easily interpretable, it models the probability of the outcome transformed to the inverse normal distribution. It is therefore not as intuitive in its interpretation as the risk difference. The coefficient is the estimated change in z-score of the outcome that a one unit change in the predictor brings about. Average marginal probabilities can be used to aid interpretation. The average marginal probability is the average change in probability of the outcome associated with a one point change in the variable of interest.

### 3.5.1 Instrumental variables

The effect measures and their assumptions for binary outcomes have been described here in terms of associative regression models but apply also to the use of instrumental variables when analysing binary outcomes. Referring to Eq. 3.1 above we expect that there is unmeasured confounding present so $\text{cov}(x, \varepsilon_i) \neq 0$ but an instrument is available Z which is associated with x but not with Y except through x.

When applying an instrumental variables model the modelling assumptions of the outcome and instrument must be considered. An instrumental variables model applying a linear probability model to the binary outcome can be applied using the standard IV 2SLS estimator described previously treating the outcome as a continuous measure. In this case the effect estimates are interpreted as risk differences. An instrumental variables probit model begins with the same first stage as the 2SLS estimator fitting the model for the endogenous variable and predicting values. The second stage regresses these predicted rather than observed values of the endogenous variable on the outcome using a probit model. This estimator assumes that the first stage model of the mediator is linear with normally distributed error terms. It can therefore only be used when the mediator is continuous as is the case in the COMMAND dataset presented in this thesis.

The analysis of the COMMAND dataset in Chapter 5 will therefore apply the instrumental variables probit model since the endogenous mediator of interest (power of voice) is a continuous measure and the outcome is binary (compliance with voice).

## 3.6 Summary

In this chapter details of the additional statistical methods that will be applied in the analysis of EDIE-II and COMMAND have been summarised. Missing baseline data in the EDIE-II trial will be imputed using multiple imputation by chained equations in order to maximise the information available. Due to the low level of missing baseline data in the COMMAND trial this is not deemed necessary. Analyses of both the EDIE-II and COMMAND data will be adjusted for any baseline covariates that are found to be associated with missing follow-up. 2SLS instrumental variables analyses will be conducted in two stages to increase the participant information used and standard errors will be estimated by bootstrapping the whole process. The longitudinal aspect of the EDIE-II data will be utilised using latent class growth models.

# 4 EDIE-II trial

## 4.1 Trial design

The following analyses are based on the EDIE-II trial, a follow-on from the preliminary EDIE trial of CBT in those deemed at high risk of developing psychosis. The EDIE-II trial was a randomised controlled trial of cognitive behavioural therapy with mental state monitoring versus mental state monitoring alone for the prevention of psychosis in ultra high-risk (UHR) individuals. The trial took place across 5 sites; Manchester, Birmingham/Worcester, Glasgow, Cambridge and Norfolk. The goal was to recruit 320 participants to the study, 160 per randomisation arm. To be included in the trial participants had to be aged between 14 and 35 years, seeking help for symptoms and had to satisfy the CAARMS criteria for at-risk mental state, without having a diagnosis of psychosis. Exclusions were previous or current anti-psychotic medication, moderate to severe learning difficulties and insufficient English. A double baseline assessment over two to four weeks was employed to ensure that no individual was currently experiencing psychosis. Participants were randomised to their trial arm using computerised varying block randomisation stratified by site and gender. All participants received monthly monitoring for the first six months and then assessments every three months, the follow-up period ranges from 12-24 months depending on when the participant was recruited. Those recruited early have a full follow-up period but due to time constraints those starting later have a shorter follow-up.

All participants received mental state monitoring which involves frequent one-to-one meetings to assess mental state, crisis cards and signposting to other services if necessary. The therapeutic intervention was individual CBT in addition to the mental-state monitoring. This was offered on a weekly basis for up to 25 sessions plus up to 4 booster sessions in the following 6 months. The process is a problem-orientated approach based on determining the individual's goals and formulating ways to achieve them through exercises or tasks with assessments of progress throughout. Permissible interventions are set out in the manual to ensure some consistency between therapists.

The primary outcome measures were time to psychosis transition, reduction in symptom severity of At Risk Mental State (ARMS) and reduction in distress caused by ARMS. Transition to psychosis and severity/distress are measured using the Comprehensive Assessment of At-Risk Mental State (CAARMS)[19]. The CAARMS is a semi-structured

interview to discuss recent experiences followed by a clinician rated scale of symptoms, frequency and distress under the following subheadings: disorders of thought content, perceptual abnormalities, conceptual disorganisation, motor changes, concentration and attention, emotion and affect, subjectively impaired energy and impaired tolerance to normal stress. A study of the measures reliability and validity by Yung et al[19] showed it to have good inter-rater reliability with an intraclass correlation (ICC) of at least 0.62 on each subscale and an overall ICC of 0.85 based on seven raters of 34 patients. The scale was able to discriminate between ultra-high risk and non-patients and was a strong predictor of onset of psychosis with the BPRS/CASH defined ultra high-risk group.

The EDIE-II trial used the short version of the CAARMS[19] including only the following subscales: unusual thought content, non-bizarre ideas, perceptual abnormalities, disorganised speech, aggression/dangerous behaviour, suicidality and self-harm. To be included in the study patients had to be considered at risk, the criteria being in one or more of the following groups, as defined in the trial protocol paper and replicated here[53]:

**Group 1: Attenuated psychosis group**

(i) Subthreshold intensity:

This group was characterised by a severity scale score of 3–5 on disorders of thought content subscale, 3–4 on perceptual abnormalities subscale and/or 4–5 on disorganized speech subscale of the CAARMS;

frequency scale score of 3–6 on disorders of thought content, perceptual abnormalities and/or disorganized speech subscale of the CAARMS for at least 1 week; OR

frequency scale score of 2 on disorders of thought content, perceptual abnormalities and disorganized speech subscale of the CAARMS on more than two occasions.

(ii) Subthreshold frequency:

Characterised by a severity scale score of 6 on disorders of thought content subscale, 5–6 on perceptual abnormalities subscale and/or 6 on disorganized speech subscale of the CAARMS;

frequency scale score of 3 on disorders of thought content, perceptual abnormalities and/or disorganized speech subscale of the CAARMS; (for both categories)

symptoms present in past year and for not longer than 5 years.

**Group 2: BLIPS group**

This group had a severity scale score of 6 on disorders of thought content subscale, 5 or 6

on perceptual abnormalities subscale and/or 6 on disorganized speech subscale of the CAARMS;
frequency scale score of 4–6 on disorders of thought content, perceptual abnormalities and/or disorganized speech subscale;
each episode of symptoms present for less than 1 week and symptoms spontaneously remit on every occasion;
symptoms occurred during last year and for not longer than 5 years.

**Group 3: Vulnerable group:**
These had a family history of psychosis in first degree relative OR schizotypal personality disorder in identified patient;
30% drop in GAF score from premorbid level, sustained for 1 month;
a change in functioning occurred within last year and maintained at least 1 month.

Psychotic disorder threshold:
Severity scale score of 6 on disorders of thought content subscale, 5 or 6 on perceptual abnormalities subscale and/or 6 on disorganized speech subscale of the CAARMS;
Frequency scale score of greater than or equal to 4 on disorders of thought content, perceptual abnormalities and/or disorganized speech subscale;
Psychotic symptoms present for longer than 1 week.

### 4.1.1 Primary Outcome

The primary outcomes were (1) transition to psychosis (2) severity (a product of frequency and duration) of symptoms and (3) distress of symptoms across the four subscales of unusual thought content, non-bizarre ideas, perceptual abnormalities and disorganised speech. These were recorded every month for the first 6 months and then every 3 months until the end of follow-up.

#### *4.1.1.1 Primary analysis results*

The primary analysis of the EDIE-II trial data conducted by Morrison et al (statistical analysis by Dunn)[64] reported that 10 of 144 (6.9%) who received CBT were clinically diagnosed as having transitioned to psychosis whereas 13 of 144 (9%) who did not receive CBT were defined as having transitioned to psychosis. Analysing the odds of transition at each time interval using a logistic regression model gave a non-significant treatment effect (proportional odds ratio 0.73, 95% CI 0.32 to 1.68).

Measures of symptom severity and distress from symptoms were compared between randomisation groups at each time interval using a cross-sectional time series model with time centred at 12 months adjusting for site and baseline of the outcome measure. Squared and cubic time adjustments and a time by treatment group interaction were fitted for both severity and distress measures separately. A significant improvement was found in the severity of symptoms between those allocated to the treatment compared to the control (estimated difference in severity of symptoms=-5.12, 95% CI -8.60 to -1.64, p=0.004) but there was no treatment effect found in terms of distress (estimated difference in distress from symptoms =-3.00, 95% CI -6.95 to 0.94, p=0.136). There was no difference in the treatment effect over time from 6 months onwards for either outcome (no interaction).

### 4.1.2 Secondary outcomes and other measures

Secondary outcomes were depression -Beck Depression Inventory (BDI)[162], anxiety - Social Interactions and Anxiety Scale (SIAS)[163] and quality of life – the EQ5D[164,165] recorded at the same monthly intervals as the CAARMS, additionally the MANSA[166] quality of life questionnaire was administered at 6-monthly intervals.

In addition to these measures demographic details for example age, ethnicity, education, occupation, smoking and alcohol use were collected before randomisation.

### 4.1.3 Post–randomisation process variables

Details of the treatment received in the therapeutic arm were collected in order to determine the quality of the therapy received; these are referred to as post-randomisation process variables. Therapist notes of CBT sessions carried out as part of the EDIE-II trial were evaluated by trial clinicians for evidence of particular aspects of therapy that should be present. This included recording for each session if an agenda was set, if homework was given, if there was formulation as an intervention, if there was work on problems and goals and if there were other interventions used, each was rated as either not present, present but not a full dose or a full dose. This is dependent upon detailed and accurate notes and so provides a conservative estimate of the occurrence of the practices, if it was not stated in the notes that a particular aspect of therapy was conducted then it was not recorded as present. Presence of an aspect of therapy was graded as present but not a full dose or a full dose. Present but not a full dose indicates that there was evidence of some use of the particular intervention but no indication that every aspect was carried out. Full dose indicates that the particular component of therapy was conducted as per the therapy

manual. Present but not a full dose and a full dose are both considered as indication that the practice occurred.

Problem agreement is considered present if in any session there was a discussion of the participant's problems and goals. Formulation or case conceptualisation is a detailed account of the patient's history including diagnosis, triggers, cycle of events and treatment plan. Problem agreement and formulation are both considered as binary measures. Problem agreement is likely to only occur once in the course of therapy and although formulation may occur many times for the purposes of measuring fidelity at least one occurrence is required.

Homework was recorded as present at each session if a review of homework was made. Additional detail on the type of homework given was also recorded under the groupings of behavioural experiment (changing reactive behaviours), monitoring (e.g. levels of anxiety or frequency of events) and education (reading relevant information). The involvement of other interventions was also recorded for each session if any of the following change strategies were mentioned in the notes: provision of normalising information, generating alternative explanations for problematic appraisals, manipulation of safety behaviours, evaluation of metacognitive beliefs or responses, evaluation of beliefs about self and others, efforts to reduce social isolation and attempts to promote relapse prevention.

Presence of homework and other interventions recorded for each session were analysed in two ways: (a) the proportion of sessions in which these were involved, and (b) a binary measure for each of presence in more than half of sessions.

Two composite measures are calculated from the four binary component measures. The first composite measure indicates a 'some versus all' comparison dichotomising participants as receiving: 0-3 components (none/some) or 4 components (all). The second measure splits this further into three groups: 0 components (none), 1-3 components (some), 4 components (all). In these composite measures the dichotomised versions of homework and active change strategies are used rather than the percentage of sessions involving these interventions.

### 4.1.4 Mediator variables

A range of psychological measures were taken at 1 month and 6 months after randomisation. Treatment is expected to cause an improvement in these measures which

will in turn improve symptoms. They are expected to mediate the treatment effect. The measures were recorded for both the treatment and control group participants.

The Meta-Cognitions Questionnaire (MCQ) revised [167] - 30-item measure generating 5 subscales: positive beliefs about worry i.e worrying is helpful; uncontrollability and danger i.e. worry must be controlled; cognitive confidence i.e. lack of confidence in memory and ability to concentrate; negative beliefs about thoughts e.g. punishment and responsibility for not controlling thoughts; cognitive self-consciousness i.e. I think a lot about my thoughts. Mediation effects are not hypothesised for the positive beliefs about worry scale but CBT is expected to reduce scores on the other four scales.

Brief Core Schema Scale (BCSS)[168] – 24-item scale of questions with a 0-4 rating, grouped into four subscales of six items each giving total possible scores in the range 0-24 for each subscale. The subscales evaluate perceptions of self and others under the four dimensions: negative-self, positive-self, negative-other, positive-other. The measure is shown to have good stability (test-retest correlation>=0.7 for all subscales) and internal consistency (Cronbach's alpha range 0.78-0.86 for all scales). CBT treatment is expected to increase scores on the positive perceptions of self and decrease scores on negative perceptions of self. Treatment may also increase positive perceptions of others and decrease negative perceptions of others but this is not a specific target of the therapy.

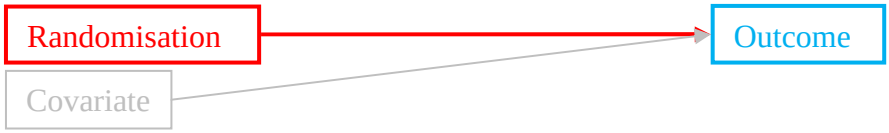
Belief About Paranoia Scale (BAPS)[169,170] –18 item scale of questions with a 1-4 rating generating 3 subscales: negative beliefs, positive beliefs/survival, normal beliefs. CBT is expected to decrease negative beliefs about paranoia and increase normal beliefs about paranoia. There is no expectation of an effect on positive beliefs/survival.

Personal Beliefs about Experiences Questionnaire (PBEQ)[171] – 13 item scale of questions with a 1-4 rating generating 2 subscales: negative appraisals of experiences (NAE) and social acceptance of experiences (SAE). Reliability of subscales has been shown to be good (alpha=0.74) and acceptable (alpha=0.52) for the subscales respectively. Psychotherapy is expected to increase scores on the social acceptance of experiences scale and decrease scores on the negative appraisals of experiences scale

## 4.2 Building the hypothesised mediation process in EDIE

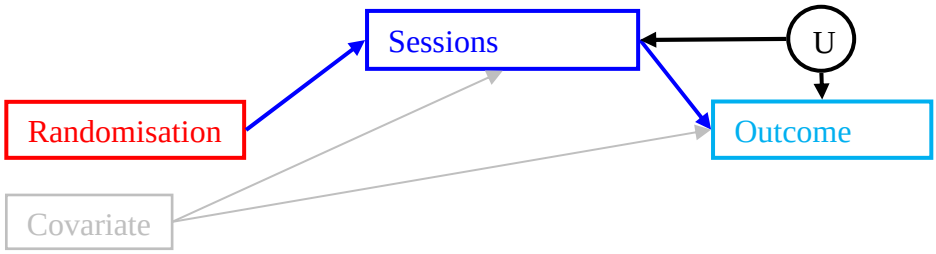
Graphical methods are used to illustrate the mediation process hypothesised in the EDIE trial. The full model is built in stages the results of which will be provided in the following chapters.

**Figure 4.1: Stage 1: intention to treat model**




The diagram for Stage 1 (Figure 4.1) represents the gold standard model for analysis of randomised trials and the primary EDIE-II analysis. The effect of randomising a participant to receive the treatment compared to the control group is estimated. This does not include any investigation of how the treatment works. In this model it is assumed, due to randomisation, that no confounding is present. Covariates are included in the model to increase the precision of the treatment-effect estimates.

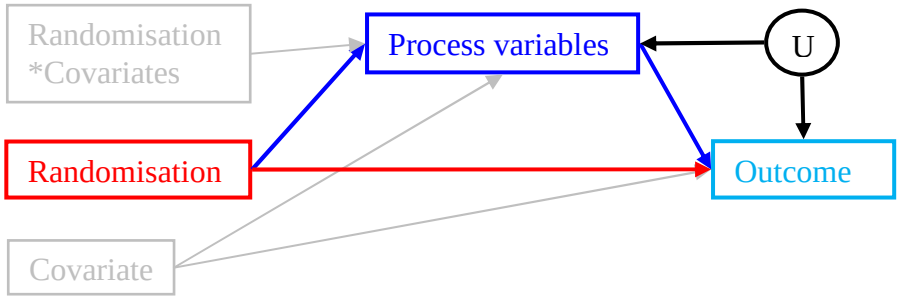
**Figure 4.2: Stage 2: mediation by attendance**




The next step, illustrated in Figure 4.2, which is becoming more popular as an addition to the standard ITT analysis, is to investigate the effect of adherence to treatment on outcome which in this example is measured by attendance at therapy sessions. The hypothesis in this situation is quite straightforward; the more treatment received the more effective it will be and those who do not attend any therapy will see no benefit of it. The second statement implies the exclusion restriction on randomisation, assuming that the only effect of being randomised to treatment will be through an increase in the number of therapy sessions attended, hence there is no direct arrow from randomisation to outcome (that is, randomisation is considered to be an instrumental variable). It is expected that there may be unmeasured confounding between attendance and outcome (some of the confounding will be allowed for by including the baseline covariates in the model and inclusion of these covariates will also help to improve the precision of the estimates of the effect of sessions

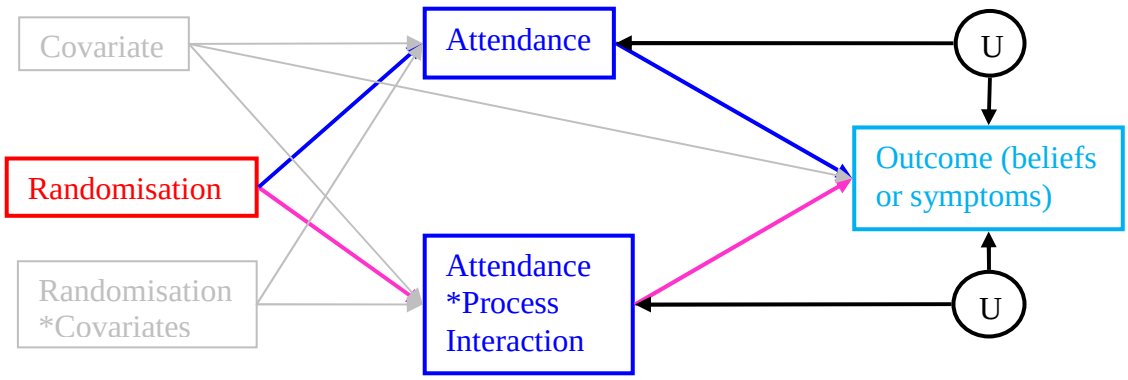
on outcome). Further improvement in efficiency may be obtained by including the randomisation by covariate interactions as instruments.

**Figure 4.3: Stage 3: mediation by post-randomisation process variables**




Step 3 (see Figure 4.3) looks specifically at the mechanisms of CBT. The impact on outcome of receiving particular components of therapy is modelled. In this situation a direct effect of randomisation is included as well as an additional effect of receiving the specific component of therapy in question. Note that the covariate by randomisation interactions are assumed only to act on the therapeutic process and not have a direct effect on outcome (these interactions, but not randomisation itself, are assumed to be instrumental variables)

**Figure 4.4: Stage 4: mediation by attendance and post-randomisation process variables**




Combining steps 2 and 3 (Figure 4.4) the multiplicative effect of attending more therapy sessions which also contain the effective components of therapy is investigated. There are three main aspects to this model, it is expected that:

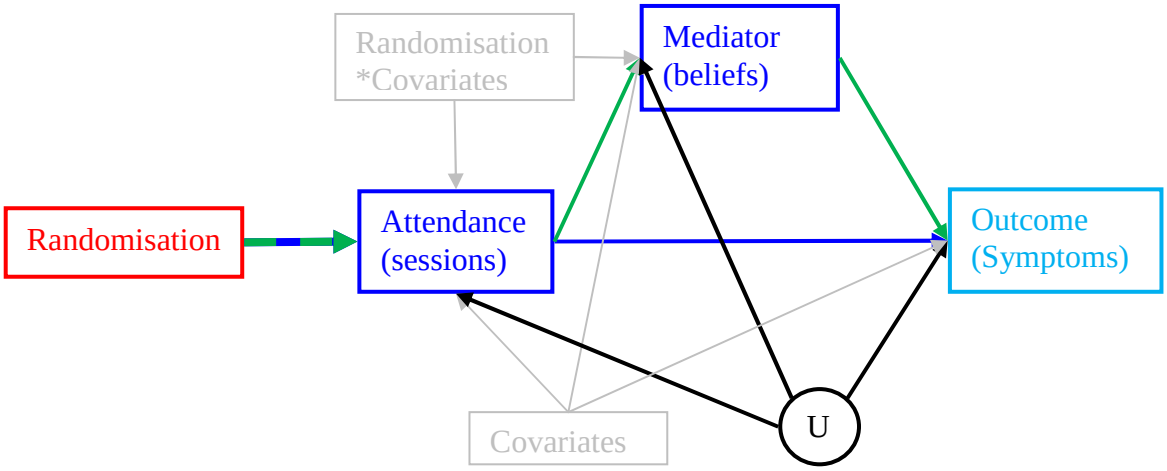
1. attending more therapy will improve outcome even if the expected active ingredients are not included.

2. the impact of attending more therapy will be increased if that therapy includes the active components of CBT
3. the impact of receiving particular components of CBT can only occur if at least one session of therapy is attended.

There is no direct effect of randomisation on outcome; that is, there is no effect of randomisation on outcome other than that through the process variables, specifically through attendance at therapy sessions. There is also no direct effects of the covariate by randomisation interactions on outcome. Here, both randomisation and its interaction with baseline covariates are assumed to be instrumental variables. Although it is likely that attendance at more sessions will be associated with an increase in the chance of those sessions containing the aspects of therapy of interest and the aspects of therapy can only be observed if therapy has been attended it is not a causal relationship. Content of therapy is modelled as an interaction with attendance as a multiplicative effect, the improvement in outcome due to receiving a particular aspect of therapy is expected to increase as more of it is received i.e. a higher dose.

This model can be used to describe the effect on the main outcome or an intermediary outcome which will later be investigated as a mediator.

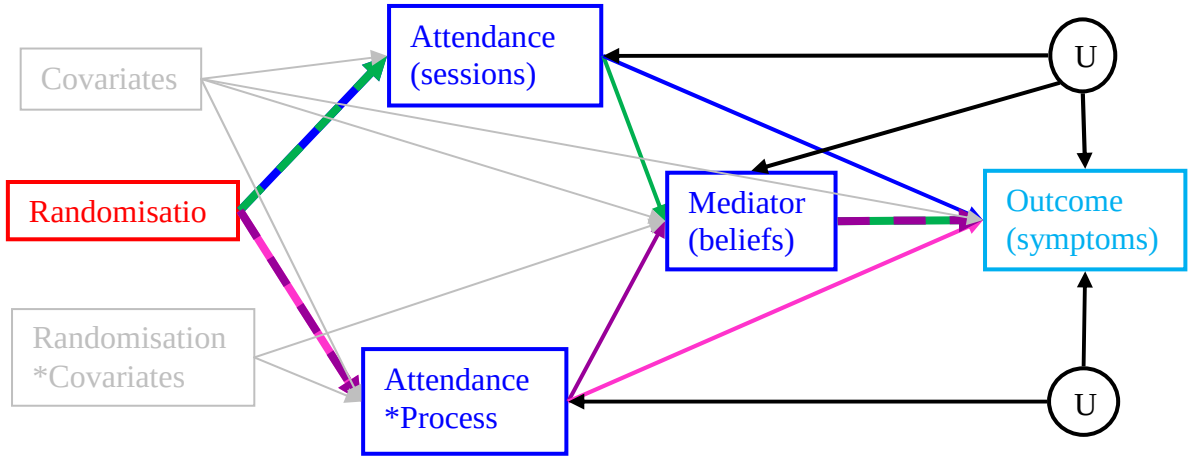
**Figure 4.5: Stage 5: mediation by attendance and changes in beliefs**




In Figure 4.5 belief measures are incorporated into the model as a mediator on the treatment pathway. This model differs from the model of attendance and process variables as a causal relationship is expected between attending therapy and changes in beliefs. The beliefs are a mediator on the causal pathway, measured in both the treatment and control arms rather than a mechanism of the treatment only available to the treated. Confounding

is expected between the two mediators and outcome. Both randomisation and its interaction with baseline covariates are assumed to be instrumental variables for the effects of attendance on beliefs and outcome. For the effects of beliefs on outcome, the interactions are the only instruments (not randomisation itself).

**Figure 4.6: Stage 6: mediation by attendance, process and belief change**




The full model (Figure 4.6) assumes that there is no direct effect of randomisation, that any treatment effect works through attendance at therapy sessions. There is an increased effect if certain components of therapy are included and the therapy may alter patients beliefs which in turn alter their outcome. Both randomisation and randomisation by baseline covariate interactions are instruments for the estimation of the effects of attendance and attendance by process interaction; only the randomisation by covariate interactions are instruments for the estimation of the effect of beliefs on outcome.

## 4.3 Results

### 4.3.1 EDIE-II trial sample description

All analyses were initially carried out on the 12 month follow-up point. Models were fitted in the statistical packages Stata12[111], R[146] and Mplus6[152].

The sample is described in terms of demographic characteristics and baseline symptoms using frequencies, percentages, means and standard deviations as a whole and by randomisation group. Primary outcomes of severity and distress were summarised by treatment group with differences assessed using t-tests. Associations between baseline characteristics and outcome were explored using chi-square or ANOVA tests as appropriate.

The trial recruited 288 participants at high risk of psychosis who were randomised to mental state monitoring only (control=144) or CBT plus mental state monitoring (treatment=144). The average age of participants was 21 years (sd=4.2), they were predominantly white (n=252, 90%) and male (n=180, 63%). Full characteristics by treatment allocation can be seen in Table 4.1.

Baseline predictors of missing follow-up are summarised in Table 4.2; only lower anxiety (SIAS) is associated with missing outcome measure. However, 19% of SIAS scores are missing at baseline which may itself be a source of bias if introduced as a covariate. When baseline associations with missing outcome are considered in the imputed dataset SIAS score is no longer associated with missing follow-up and there are no other variables associated. Therefore, no further adjustment is made for variables associated with missing outcome, it is assumed to be missing at random.

**Table 4.1: Baseline characteristics by treatment allocation**

| | Control N=144 | CBT N=144 | All N=288 | Missing |
|---|---|---|---|---|
| Age | 20.8 (4.5) | 20.7 (4.2) | 20.7 (4.3) | 0 |
| Gender (Male) | 91 (63%) | 89 (62%) | 180 (63%) | 3 (1%) |
| Ethnicity (white) | 124 (90%) | 128 (90%) | 252 (90%) | 8 (3%) |
| Education | 12.9 (2.2) | 13.2 (2.2) | 13.0 (2.2) | 23 (8%) |
| Occupation: Seeking/other | 48 (40%) | 46 (37%) | 94 (39%) | 40 (14%) |
| Employed/Hswk/student | 71 (60%) | 79 (63%) | 150 (62%) | |
| Site | | | | |
| Manchester | 40 (28%) | 40 (28%) | 80 (28%) | 0 |
| Birmingham | 38 (26%) | 39 (27%) | 77 (27%) | |
| Cambridge | 16 (11%) | 14 (10%) | 30 (10%) | |
| Norfolk | 19 (13%) | 21 (15%) | 40 (14%) | |
| Glasgow | 31 (22%) | 30 (21%) | 61 (21%) | |
| | | | | |
| CAARMS Severity | 38.2 (17.8) | 38.7 (16.8) | 38.4 (17.3) | 2 (1%) |
| CAARMS Distress | 42.5 (19.6) | 42.8 (20.5) | 42.6 (20.0) | 24 (8%) |
| BDI | 9.0 (4.7) | 10.4 (4.1) | 9.7 (4.5) | 30 (10%) |
| SIAS | 39.4 (16.9) | 42.9 (16.9) | 41.2 (17.0) | 54 (19%) |
| GAF | 51.1 (10.3) | 51.0 (11.0) | 51.1 (10.6) | 0 |
| MANSA | 4.04 (0.96) | 3.85 (0.78) | 3.94 (0.88) | 41 (14%) |
| EQ5D | 0.63 (0.30) | 0.59 (0.28) | 0.61 (0.29) | 43 (15%) |

**Table 4.2: Baseline associations with missing severity of symptoms at 12 months follow-up**

| | Complete at 12 months n=188 | Missing at 12 months n=100 | p-value complete covariate | p-value imputed covariate |
|---|---|---|---|---|
| Age | 20.4 (4.2) | 21.3 (4.6) | 0.104 | 0.105 |
| Gender (Male) | 117 (62%) | 63 (63%) | 0.898 | 0.898 |
| Ethnicity (white) | 171 (92%) | 81 (86%) | 0.129 | 0.052 |
| Education | 12.9 (2.1) | 13.2 (2.5) | 0.419 | 0.308 |
| Occupation: | | | | |
| Seeking/other | 62 (37%) | 32 (42%) | | |
| Employed/Hswk/student | 105 (63%) | 45 (58%) | 0.508 | 0.408 |
| Site | | | | |
| Manchester | 48 (26%) | 32 (32%) | | |
| Birmingham | 55 (29%) | 22 (22%) | | |
| Cambridge | 19 (10%) | 11 (11%) | | |
| Norfolk | 29 (15%) | 11 (11%) | | |
| Glasgow | 37 (20%) | 24 (24%) | 0.432 | 0.432 |
| | | | | |
| CAARMS Severity | 38.2 (16.1) | 38.9 (19.4) | 0.738 | 0.734 |
| CAARMS Distress | 42.0 (19.4) | 43.7 (21.3) | 0.534 | 0.443 |
| BDI | 10.1 (4.4) | 9.0 (4.5) | 0.053 | 0.056 |
| SIAS | 42.8 (17.1) | 37.5 (16.2) | 0.028 | 0.054 |
| GAF | 50.3 (10.7) | 52.6 (10.3) | 0.077 | 0.078 |
| MANSA | 3.9 (0.91) | 4.0 (0.82) | 0.257 | 0.257 |
| EQ 5D | 0.59 (0.29) | 0.65 (0.28) | 0.185 | 0.514 |

### 4.3.2 Stage 1: intention to treat analysis

A cross-sectional time series model with time centred at 12 months adjusting for site and baseline measure of the outcome with both squared and cubic time adjustments and a time by treatment group interaction was fitted for both severity and distress measures separately. This analysis was undertaken in the primary analysis of the trial[64], and the authors analysis is replicated here to provide a start to a comprehensive mediation analysis. A significant improvement (estimated by the effect of the intervention at 12 months) was found in the severity of symptoms between those allocated to the treatment compared to the control (coefficient=-5.12, 95% CI -8.60 to -1.64, p=0.004) but there was no statistically significant treatment effect found in terms of distress (coefficient=-3.00, 95% CI -6.95 to 0.94, p=0.136), see Table 4.3. There was no significant difference in the treatment effect over the follow-up time points for either severity or distress (no

statistically significant interaction between treatment time and follow-up). This indicates that after adjustment for baseline level of the outcome the treatment effect of CBT compared to TAU is estimated to be no different at each point from 6 to 24 months. When the interaction between randomisation group and time is removed from the model the treatment effect common to all follow-up time points for severity is estimated at -3.67 (95% CI -6.71 to -0.64, p=0.018) and on distress as -3.01 (95% CI -6.95 to 0.94, p=0.14).

**Table 4.3: Treatment effect - ITT analysis**

| | Treatment effect at 12 months | 95% CI | p-value | Treatment by month interaction | 95% CI | p-value |
|---|---|---|---|---|---|---|
| Severity | -5.112 | -8.586 to -1.638 | 0.004 | -0.273 | -0.611 to 0.064 | 0.113 |
| Distress | -3.003 | -6.949 to 0.943 | 0.136 | -0.213 | -0.579 to 0.153 | 0.254 |

### 4.3.3 Stage 2: attendance at therapy as a mediator of treatment

#### *4.3.3.1 Statistical methods*

The role of attendance was explored as a mediator of the treatment effect on outcome, no direct effect of randomisation is allowed as illustrated in Figure 4.2. This was analysed as a continuous variable testing for quadratic as well as linear effects and categorised to determine optimum levels. A dichotomous split with 4 or more sessions attended considered as compliance was generated as well as a 3-category split with groupings of <4, 4-12 and 13+ sessions. These cut-points were provided by clinical experts as hypothesised important groupings. Categorical predictors of sessions attended as both a continuous and categorical measure were explored. For the initial exploration analyses were carried out using 2SLS for the continuous measure of attendance, CACE for the binary indicator and principal stratification for the categorical measure (see Chapter 1).

##### *4.3.3.1.1 Instrumental variables*

The instrumental variable regression uses treatment group and the treatment group by baseline score interaction as the instrument. The first stage is therefore a regression of number of sessions attended on treatment group, group by baseline score interaction and baseline score. The second stage regression is of outcome on the fitted values from the first stage regression and baseline score. No adjustment for treatment group is made in the second stage so that no direct effect of randomisation is allowed. Although described in two steps this is run in one step to ensure the correct standard errors using the Stata

package ivregress and the Mplus software package. The Stata ivregress package uses ordinary least squares estimation and requires complete information on all observations deleting observations with missing data; I have labelled these results ‘complete case’. The Mplus estimation uses all available data from each observation in a maximum likelihood model; I have labelled these results ‘all data’

#### *4.3.3.1.2 Complier average causal effect (CACE)*

Compliance with therapy was defined as having attended at least four sessions. Within the treatment arm compliers (those receiving at least four sessions) and never-takers (those receiving less than four sessions) can be identified. The analysis was set up in the same way as the instrumental variables regression described above but with a binary mediator. The first stage is therefore a regression of compliance with therapy on treatment group, group by baseline score interaction and baseline score. The second stage regression is of outcome on the fitted values from the first stage regression and baseline score.

#### *4.3.3.1.3 Principal stratification / latent class analysis*

The CACE analysis was extended to more than two groups of the mediator. The treatment effect calculated within each stratum/ category. The number of sessions attended in the treatment group is categorised as <4, 4-12 and 13+ sessions. Stratum membership for all participants was predicted based on those in the treatment group using only the baseline measure of the outcome and treatment centre.

#### *4.3.3.1.4 Longitudinal analysis*

Follow-up data recorded after all therapy had been completed i.e. from 6 months onwards was incorporated so that the mediator was not time dependent. Outcome at 6, 9, 12, 15, 18, 21 and 24 months was fitted as a random effects linear regression model with both a random intercept and random slope. Time points were centred at 12 months so the intercept indicates the average level of the outcome at 12 months (in line with cross-sectional analyses) and the slope is the average change over time. The methods above can be extended to multiple time points using both the continuous and categorical measure of sessions of therapy. Extending the CACE analysis, strata were defined as attending more or less than four sessions of therapy. As with the one-time point outcome class membership (complier/non-complier) was predicted using only the baseline measure of the outcome. It is assumed that there is no treatment effect in non-compliers and the average treatment effect is then estimated within the compliers.

#### *4.3.3.2 Results*

##### *4.3.3.2.1 Predictors of outcome and mediator*

Bivariate associations with severity at 12 month follow-up adjusting only for the baseline measure of severity indicate no strong predictors of severity at follow-up. A higher level of depression (BDI) at baseline and site are associated with distress at follow-up (Table 4.4) with higher levels of distress seen in Manchester and Glasgow and the lowest in Cambridge.

Older age, more social anxiety (SIAS), lower functioning (GAF), having a healthworker, and being of non-white ethnicity were associated with attending more sessions. Participants in Manchester and Birmingham attended the most sessions on average with the least on average in Norfolk.

**Table 4.4: Baseline covariate associations with outcome – all participants**

| | Severity | | Distress | |
|---|---|---|---|---|
| | Partial correlations* | p-value* | Partial correlations* | p-value* |
| Age | -0.088 | 0.230 | 0.018 | 0.821 |
| Education | 0.074 | 0.338 | 0.023 | 0.780 |
| Severity | - | - | -0.043 | 0.581 |
| Distress | 0.119 | 0.122 | - | - |
| BDI | 0.124 | 0.106 | 0.163 | 0.043 |
| SIAS | 0.046 | 0.562 | 0.095 | 0.257 |
| GAF | -0.124 | 0.092 | -0.124 | 0.113 |
| MANSA | -0.210 | 0.007 | -0.232 | 0.004 |
| EQ5D | -0.034 | 0.660 | -0.229 | 0.005 |
| | mean (sd) | p-value* | mean (sd) | p-value* |
| Gender: Male | 17.9 (16.8) | 0.911 | 15.4 (17.5) | 0.174 |
| Female | 17.8 (17.1) | | 19.9 (17.8) | |
| Ethnicity: white | 18.0 (17.2) | 0.474 | 16.8 (17.2) | 0.963 |
| non-white | 16.1 (13.7) | | 17.3 (22.0) | |
| Occupation: Seeking/other | 18.3 (16.2) | 0.344 | 19.1 (17.3) | 0.120 |
| Employed/Hswk/student | 17.1 (17.1) | | 15.0 (16.6) | |
| Site: Manchester | 17.5 (13.4) | 0.746 | 22.3 (19.3) | 0.045 |
| Birmingham | 19.7 (17.7) | | 14.1 (16.4) | |
| Cambridge | 12.8 (17.8) | | 7.1 (8.8) | |
| Norfolk | 16.5 (14.9) | | 17.2 (17.9) | |
| Glasgow | 19.0 (20.6) | | 19.9 (18.4) | |
| Healthworker: No | 17.1 (16.5) | 0.835 | 15.6 (16.5) | 0.850 |
| Yes | 18.8 (16.2) | | 18.4 (15.6) | |

*adjusted for baseline of outcome

**Table 4.5: Baseline covariate associations with number of sessions attended - treatment group only**

| | Sessions - treatment group only | |
|---|---|---|
| | Correlation | p-value |
| Age | 0.224 | 0.007 |
| Education | 0.115 | 0.182 |
| Severity | 0.180 | 0.031 |
| Distress | 0.125 | 0.157 |
| BDI | 0.036 | 0.685 |
| SIAS | 0.326 | <0.001 |
| GAF | -0.217 | 0.009 |
| MANSA | -0.097 | 0.280 |
| EQ5D | -0.086 | 0.343 |
| | mean (sd) | p-value |
| Gender: Male | 8.9 (6.8) | 0.920 |
| Female | 9.2(6.5) | |
| Ethnicity: white | 8.6 (6.6) | 0.047 |
| non-white | 12.4 (6.9) | |
| Occupation: | | |
| Seeking/other | 10.1 (7.4) | 0.151 |
| Employed/Hswk/student | 8.3 (6.3) | |
| Site: Manchester | 10.8 (7.4) | <0.001 |
| Birmingham | 10.8 (6.2) | |
| Cambridge | 8.6 (6.3) | |
| Norfolk | 4.0 (2.2) | |
| Glasgow | 7.9 (6.8) | |
| Healthworker: No | 8.6 (6.5) | 0.149 |
| Yes | 11.6 (8.9) | |

#### *4.3.3.2.2 Attendance at sessions as a continuous measure*

The continuous measure of attendance at therapy sessions was tested as a post-randomisation process variable using 2SLS (repeating the 2SLS analysis using Stata ivregress and Mplus software) for severity and distress outcomes at 12 months follow-up. In all of the analyses presented in Table 4.6 baseline severity/distress was the only predictor used and the instruments in IV analysis were randomisation group and the baseline measure by randomisation group interaction. It is assumed that randomisation only effects outcome through the mediator i.e. if no therapy is attended there can be no treatment effect.

**Table 4.6: Sessions as a mediator of treatment effect**

| | | Severity at 12 months | | | Distress at 12 months | | | |
|---|---|---|---|---|---|---|---|---|
| Method | N | Coefficient of Sessions | Std.Err | p-value | N | Coefficient of Sessions | Std.Err | p-value |
| 2SLS Stata – complete cases | 187 | -0.692 | 0.223 | 0.002 | 166 | -0.353 | 0.243 | 0.147 |
| 2SLS Mplus – complete cases | 187 | -0.696 | 0.224 | 0.002 | 166 | -0.353 | 0.243 | 0.147 |
| 2SLS Mplus – All data | 286 | -0.837 | 0.276 | 0.002 | 264 | -0.398 | 0.312 | 0.202 |

The results in Table 4.6 estimate that every additional session attended reduces symptom severity by 0.7 points (95% CI -1.1 to -0.3, p=0.002, complete cases) but that there will be no significant effect on distress (coefficient= -0.35 95% CI -0.8 to 0.1, p=0.147). When all participants with any data were included ('all data') the effect of sessions on severity of illness was estimated to be slightly greater at around -0.8 (95% CI -1.4 to -0.3, p=0.002) and although it also had a greater effect on distress this is still non-significant (coef=-0.4, 95% CI -1.1 to 0.2, p=0.202). A non-linear effect of sessions attended was investigated adding quadratic and cubic terms to the IV model but were found to be non-significant and so only the linear model is reported.

These analyses were repeated for the secondary outcomes of depression (BDI), anxiety (SIAS) and quality of life (MANSA); no significant association of number of sessions was found for any of these outcomes, full results are provided in Appendix 2.

#### *4.3.3.2.3 Attendance at sessions as a categorical measure*

As a further investigation to check the linearity of the mediator effect and to determine if there exists a minimum or optimum number of sessions the number of sessions attended was categorised, initially into two groups and then three.

The 2-group scenario was analysed using the principal stratification CACE model described previously with compliance defined as attending at least four sessions of therapy. The CACE model determines the probability of class membership for participants in the control group by applying associations seen in the treated group between covariates and number of sessions attended. Within each group the effect of being randomised to treatment is calculated. The interpretation is the treatment effect within people that would

have attended the same number of sessions if they had been offered. By applying the exclusion restriction and forcing no treatment effect in the non-compliers the within class treatment effect then becomes a comparison to the non-compliant group. This analysis and interpretation can be extended to more than two classes under the general term of principal stratification or latent class analysis. The analyses were conducted in Mplus6 using latent class analysis; all analyses adjust for baseline measure of outcome and trial site.

**Table 4.7: Analysis of sessions as a post-randomisation process effect, CACE model**

| | Severity at 12 months | | | | Distress at 12 months | | | |
|---|---|---|---|---|---|---|---|---|
| Predictors of class membership included | N | Effect in compliers | Std. Err | p-value | N | Effect in compliers | Std. Err | p-value |
| Baseline of outcome, trial site | 286 | -10.460 | 4.716 | 0.027 | 264 | -7.085 | 4.516 | 0.117 |

A positive impact of attending at least four sessions is seen on severity of symptoms with a smaller and non-significant effect on distress from symptoms (Table 4.7). Attending at least four sessions of therapy is expected to reduce symptoms by approximately 10.5 points compared to attending less than four sessions (std. err =4.7, p=0.027). The analysis was extended to three mediator levels defined as <4, 4-12 and 13+ sessions.

When splitting attendance into three groups the effect of attending 4-12 sessions has no significant effect on either symptoms or distress. A large and statistically significant effect is seen on both symptoms and distress when 13 or more sessions are attended (Table 4.8). This is a surprising result since no quadratic effect was found for the continuous measure of attendance on either symptoms or distress. The magnitude of the effect is large as is the standard error. Introducing additional covariates as predictors of class membership in order to improve estimates results in large changes in the estimated effects giving a lack of confidence in the models.

**Table 4.8: Principal stratification of sessions as a post-randomisation effect, three-levels of attendance**

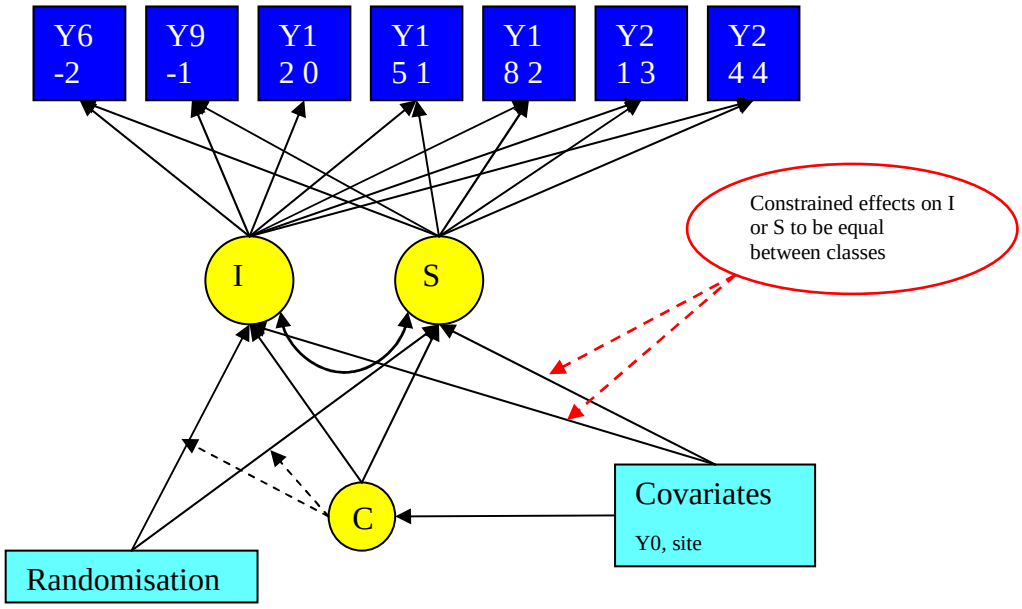
| | | Severity at 12 months | | | | Distress at 12 months | | | |
|---|---|---|---|---|---|---|---|---|---|
| Predictors of class membership included | Stratum | N | Treatment effect in strata | Std. Err | p-value | N | Treatment effect in strata | Std. Err | p-value |
| Baseline of outcome, trial site | 4-12 sessions | 286 | -1.011 | 4.096 | 0.805 | 264 | 1.325 | 8.013 | 0.869 |
| | 13+ sessions | | -27.526 | 8.689 | 0.002 | | -20.507 | 12.596 | 0.104 |

#### *4.3.3.2.4 Longitudinal analysis*

All previous analyses look only at one follow-up time point 12 months after randomisation. A benefit of the EDIE-II trial is that data has been collected at multiple time points, monthly during the treatment period and 3-monthly thereafter. Only the follow-up times that occur after all sessions would have been completed i.e. from 6 months onwards are included in the analysis, this means that the mediator is not time dependent. A CACE analysis is conducted in Mplus with number of sessions attended dichotomised at four sessions to indicate compliance. Class membership is defined using the threshold method with the dichotomous sessions variable. The diagram below (Figure 4.7) shows the model graphically where I is the intercept, months of follow-up is centred on 12 months so that the intercept is interpreted as the treatment effect at 12 months follow-up within each strata and S the slope of the line is the change in treatment effect over time within each strata. C indicates latent class/stratum.

In the Mplus models the covariates are constrained to have the same effect on I and S in each class i.e. association between a demographic variable and outcome is the same whether you comply with treatment or not, only the treatment effect is allowed to vary. The intercept and slope is free to vary across the classes. The exclusion restriction is enforced by setting the treatment effect in those attending <4 sessions at zero and allowing the model to estimate the treatment effect in the 'compliers' (this assumption can be tested by removing the zero effect restriction).

**Figure 4.7: A graphical representation of longitudinal mediation class analysis**




Modelling the categorical measure of attendance dichotomised at four sessions on severity with severity at baseline as a covariate the intercept is estimated at -7.95 (se=2.9, p=0.006) with a slope of -0.25 (se=1.1 p=0.814) and for the distress outcome intercept=-5.15 (se=3.6, p=0.150) with slope=-0.62 (se=1.0, p=0.549). There is an impact of attendance on symptom severity at 12 months but this effect does not change over time. There is no impact on distress (Table 4.9).

The CACE analysis using longitudinal follow-up indicates as with the continuous measure of sessions that there is an association with attending at least four sessions of therapy at the 12 month follow-up point (the intercept, see Table 4.9) which is significant for severity but not distress but there is no change in this over time (the slope) for either. The magnitude of the treatment effect at 12 months is similar to that found in the earlier analysis (see Table 4.7) as expected.

**Table 4.9: Longitudinal CACE analysis of sessions as a mediator of treatment outcome**

| | Severity | | | | Distress | | | |
|---|---|---|---|---|---|---|---|---|
| Comparison | N | Effect in compliers | Std. Err | p-value | N | Effect in compliers | Std. Err | p-value |
| Intercept | 286 | -7.953 | 2.890 | 0.006 | 264 | -5.147 | 3.578 | 0.150 |
| Slope | | -0.253 | 1.074 | 0.814 | | -0.622 | 1.038 | 0.549 |

### 4.3.4 Stage 3: content of therapy as a post-randomisation process variable

The EDIE-II trial investigators consider four components of therapy: problem agreement, formulation, homework and other interventions. Problem agreement and formulation were binary measures considered present if, in any session, there was evidence of their presence. Homework is believed to be a key part of CBT and in this trial several aspects were recorded: if there was a review, monitoring and feedback given from the homework and behavioural experiment. Other interventions used in the course of therapy were also reported and we consider if any of the following change strategies were present: normalising, generating alternatives, safety behaviours, metacognition, "I am different", social isolation and relapse prevention. Details of these interventions are described in Chapter 1. The level of homework and other interventions incorporated into therapy are measured as a proportion of sessions in which they were involved. They are also analysed as binary measures of presence in more than half of sessions or not. Finally an overall measure of having all four components present versus some or none present is calculated as a binary process variable. Since CBT is not found to be effective in reducing distress the post-randomisation and mediator models will only be conducted on the severity of symptoms outcome

#### *4.3.4.1 Statistical Methods*

The inclusion of specific aspects of therapy during CBT sessions on symptom severity is assessed using instrumental variables analysis. The model assumes that there is a direct effect of randomisation to CBT even if the component of therapy is not received (see Figure 4.3). Randomisation cannot be used as an instrument and alternative instruments are found from interactions of randomisation group with each of the following baseline covariates: site, age, gender, education, ethnicity, have a degree, occupation, GAF, BDI, SIAS, CAARMS severity, CAARMS distress, Eq5D, MANSA. It is expected that interactions of baseline covariates with randomisation arm will be valid instruments because, from the properties of randomisation, any association between a covariate and outcome would not be expected to be any different in one randomisation group to another other than through the treatment received. Additionally, there is no evidence in the current literature to suggest that any of these variables are treatment effect modifiers. Variables will therefore be selected that are associated with the process variable in the treatment group in order to satisfy the requirement that an instrument must be associated with the

endogenous variable. Simulations have indicated that the LASSO with the penalty lambda applied which minimises the MSE of the first stage prediction is effective in selecting instruments when instruments are strongly associated with the process variable and suggest that the LASSO may be preferred when instruments are weak. Instruments will therefore be selected using the LASSO with the penalty lambda which minimises the MSE. Instrument selection and instrumental variables analyses are carried using the statistical package R version 2.15. Each process variable is considered separately in a two-stage least squares (2SLS) regression model with randomisation and the covariates selected included in the second stage regression. Results for the simple Baron and Kenny mediation model regressing symptom severity on randomisation, the process variable and covariates are also calculated. This model assumes no unmeasured confounding which we do not believe to be appropriate but present for comparison.

The binary composite measure of none/some components received versus all components received is analysed in the same way as the individual components allowing for a direct effect of randomisation. The analysis of the 3-category indicator assumes that there is no effect of the treatment if none of the four components have been involved (exclusion restriction on randomisation). Variables associated with the two additional levels (some components and all components) are determined separately and all associated variables are used as instruments. The two stage analysis predicts values for receipt of some components and all components categories separately before using them both in the regression on outcome.

All analyses are adjusted for site and baseline severity. The IV regression is estimated in 2 parts and the whole process from variable selection to regression analysis is bootstrapped 1000 times to ensure correct standard errors. The analysis is run on a subset of participants with complete observed data on all variables and also on the imputed datasets where missing baseline covariates have been imputed. Outcome and process variables are included in the imputation process but imputed values for the outcome and process variables are not included in the analysis (details of the imputation model are described in Chapter 3.3.1). There are no missing data in the process variables, the components of therapy received. Missing outcome data is assumed to be missing at random dependent on site, baseline severity and covariates included as instruments. The IV 2SLS regression with bootstrap is carried out on each of the five imputed datasets separately and the results combined using Rubin's Rules by the user made excel macro[172]. This is to ensure that the

standard errors and p-values are correctly adjusted for the uncertainty in the imputed data. Results are provided for both the complete case (only participants with data observed for outcome, process variable/mediator and baseline covariates) and imputed datasets. The results for the imputed data are preferred but complete case results are given for comparison.

Simulations indicate that the 2SLS estimation method is preferred over the LIML and Fullers to reduce bias with binary process variables and reduce variation of estimates for both binary and continuous process variables. Although the LIML and Fuller reduce bias when a continuous process variable is used the 2SLS is reported here and the LIML and Fuller results reported in Appendix 2.

#### *4.3.4.2 Results*

Heterogeneity was present in the treatment actually received by participants in the intervention arm. 100 (69%) had an agreement of problems and goals, 79 (55%) had received formulation at some point in their therapy. On average 39% (sd=32) of sessions involved homework and 43% (sd=34) involved other interventions. 36 (25%) of participants received all 4 components of therapy and 28 (19%) did not receive any. Participants who attended more therapy sessions were more likely to receive any of these components and there is a high correlation between receipt of different components (Table 4.12).

Individual associations between the characteristics of patients and the treatment received within those in the intervention arm are detailed in Table 4.10 and Table 4.11. The tables show observed values including only those who have complete data for the comparison shown. The results indicate that there are measured characteristics that influence the type of therapy received by patients and different characteristics appear to be bring about different components of therapy. The location of the participant seems to be associated with most measures of therapy content with higher rates of compliance in Manchester and lower rates in Norfolk. Higher levels of anxiety (SIAS) and lower functioning (GAF) are indicated in participation in more components of therapy.

**Table 4.10: Demographic associations with content of therapy received - CBT arm only, complete case analysis**

| | Problem Agreement | | | Formulation | | | Homework | | | Other Interventions | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | No | Yes | | No | Yes | | No | Yes | | No | Yes | |
| Gender: | | | | | | | | | | | | |
| Male | 33 (37%) | 56 (63%) | | 41 (46%) | 48 (54%) | | 48 (54%) | 41 (46%) | | 45 (51%) | 44 (49%) | |
| Female | 11 (20%) | 44 (80%) | 0.03 | 24 (44%) | 31 (56%) | 0.78 | 33 (60%) | 22 (40%) | 0.48 | 27 (49%) | 28 (51%) | 0.86 |
| Age | 20.2 (3.9) | 21.0 (4.3) | 0.28 | 20.4 (4.1) | 21.0 (4.3) | 0.46 | 20.2 (3.7) | 21.4 (4.7) | 0.08 | 20.1 (3.9) | 21.4 (4.4) | 0.06 |
| Ethnicity: | | | | | | | | | | | | |
| non-white | 3 (21%) | 11 (79%) | | 4 (29%) | 10 (71%) | | 8 (57%) | 6 (43%) | | 7 (50%) | 7 (50%) | |
| White | 40 (31%) | 88 (69%) | 0.45 | 61 (48%) | 67 (52%) | 0.17 | 72 (56%) | 56 (44%) | 0.95 | 65 (51%) | 63 (49%) | 0.96 |
| Site: | | | | | | | | | | | | |
| Manchester | 4 (10%) | 36 (90%) | | 12 (30%) | 28 (70%) | | 17 (43%) | 23 (58%) | | 16 (40%) | 24 (60%) | |
| Birm'ham | 14 (36%) | 25 (64%) | | 14 (36%) | 25 (64%) | | 26 (67%) | 13 (33%) | | 21 (54%) | 18 (46%) | |
| Cambridge | 3 (21%) | 11 (79%) | | 5 (36%) | 9 (64%) | | 6 (43%) | 8 (57%) | | 3 (21%) | 11 (79%) | |
| Norfolk | 16 (76%) | 5 (24%) | | 18 (86%) | 3 (14%) | | 13 (62%) | 8 (38%) | | 16 (76%) | 5 (24%) | |
| Glasgow | 7 (23%) | 23 (77%) | <0.01 | 16 (53%) | 14 (47%) | <0.01 | 19 (63%) | 11 (37%) | 0.15 | 16 (53%) | 14 (47%) | 0.02 |
| Years of education | 12.8 (2.1) | 13.3 (2.3) | 0.27 | 12.9 (2.3) | 13.4 (2.2) | 0.16 | 12.9 (2.0) | 13.5 (2.5) | 0.08 | 12.9 (2.1) | 13.4 (2.3) | 0.12 |
| Continuing Education: | | | | | | | | | | | | |
| No | 16 (43%) | 21 (57%) | | 21 (57%) | 16 (43%) | | 22 (59%) | 15 (41%) | | 22 (59%) | 15 (41%) | |
| Yes | 22 (26%) | 62 (74%) | 0.06 | 35 (42%) | 49 (58%) | 0.13 | 42 (50%) | 42 (50%) | 0.34 | 35 (42%) | 49 (58%) | 0.07 |
| Degree : | | | | | | | | | | | | |
| No | 33 (32%) | 71 (68%) | | 49 (47%) | 55 (53%) | | 55 (54%) | 49 (46%) | | 49 (47%) | 55 (53%) | |
| Yes | 5 (26%) | 14 (74%) | 0.64 | 8 (42%) | 11 (58%) | 0.69 | 11 (54%) | 8 (46%) | 0.69 | 9 (47%) | 10 (53%) | 0.98 |
| Occupation : | | | | | | | | | | | | |
| Seeking work/other | 12 (26%) | 34 (74%) | | 21 (46%) | 25 (54%) | | 25 (53%) | 21 (47%) | | 21 (46%) | 25 (54%) | |
| Employed/ Student/ Housework | 27 (34%) | 52 (66%) | 0.35 | 37 (47%) | 42 (53%) | 0.90 | 43 (58%) | 36 (42%) | 0.99 | 39 (49%) | 40 (51%) | 0.69 |

**Table 4.11: Associations between baseline health measures and content of therapy received - CBT arm only, complete case**

| | Problem Agreement | | | Formulation | | | Homework | | | Other Interventions | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | No | Yes | p-value | No | Yes | p-value | No | Yes | p-value | No | Yes | p-value |
| SIAS | 37.2 (15.3) | 45.6 (17.1) | <0.01 | 38.7 (16.4) | 46.7 (16.6) | <0.01 | 38.5 (17.0) | 48.1 (15.4) | <0.01 | 36.3 (16.1) | 49.4 (15.2) | <0.01 |
| BDI | 9.6 (3.7) | 10.8 (4.3) | 0.12 | 10.3 (3.8) | 10.5 (4.4) | 0.82 | 10.5 (4.3) | 10.4 (4.0) | 0.89 | 10.4 (4.3) | 10.4 (4.0) | 0.93 |
| Severity | 37.0 (16.2) | 39.5 (17.1) | 0.43 | 36.4 (16.8) | 40.7 (16.8) | 0.13 | 39.1 (17.4) | 38.8 (16.2) | 0.79 | 37.1 (18.3) | 40.4 (15.1) | 0.24 |
| Distress | 42.2 (20.9) | 43.0 (20.5) | 0.82 | 40.8 (20.2) | 44.3 (20.8) | 0.34 | 41.0 (20.5) | 44.9 (20.5) | 0.28 | 41.2 (22.3) | 44.2 (18.8) | 0.41 |
| GAF | 54.2 (10.9) | 49.6 (10.8) | 0.02 | 52.5 (11.5) | 49.7 (10.4) | 0.13 | 52.4 (11.0) | 49.2 (10.7) | 0.08 | 52.8 (12.2) | 49.1 (9.3) | 0.04 |
| EQ5 | 0.65 (0.25) | 0.56 (0.29) | 0.08 | 0.62 (0.28) | 0.56 (0.29) | 0.28 | 0.58 (0.29) | 0.60 (0.28) | 0.71 | 0.61 (0.29) | 0.56 (0.28) | 0.35 |
| MANSA | 4.01 (0.75) | 3.77 (0.79) | 0.11 | 3.86 (0.80) | 3.83 (0.77) | 0.88 | 3.94 (0.82) | 3.73 (0.72) | 0.14 | 3.90 (0.79) | 3.79 (0.78) | 0.46 |

**Table 4.12: Associations between components of therapy**

| | | # sessions | Problem agreement | | Formulation | | >50% homework | |
|---|---|---|---|---|---|---|---|---|
| | | mean (sd) | No | Yes | No | Yes | No | Yes |
| Problem Agreement | No | 4.5 (5.2) | | | | | | |
| | Yes | 10.9 (6.3) | | | | | | |
| Formulation | No | 4.8 (4.9) | 37 (84%) | 28 (28%) | | | | |
| | Yes | 12.4 (6.0) | 7 (16%) | 72 (72%) | | | | |
| >50% sessions with homework | No | 7.1 (6.4) | 34 (77%) | 47 (47%) | 47 (72%) | 34 (43%) | | |
| | Yes | 11.3 (6.7) | 10 (23%) | 53 (53%) | 18 (28%) | 45 (57%) | | |
| >50% sessions with active | No | 5.6 (5.0) | 35 (80%) | 37 (37%) | 51 (78%) | 21 (27%) | 58 (72%) | 14 (22%) |
| change strategy | Yes | 12.3 (6.5) | 9 (20%) | 63 (63%) | 24 (22%) | 58 (73%) | 23 (28%) | 49 (78%) |

Note: all significance tests (t-tests and chi-squared) give p-value<=0.001

Agreement of problems and goals

The magnitude of the randomisation effect on severity at 12 months is much larger in the analysis of participants with complete data compared to the analysis on the imputed datasets (Table 4.13). However the estimate of the effect of problem agreement is similar. We expect that the imputed data will give a better estimate of the effects as the sample is larger and bias due to missing information is reduced. The results are very different for the ordinary least squares model that assumes no unmeasured confounding compared to that of the IV model that does not make this assumption. There is no significant effect of therapy on severity of symptoms when a problem agreement is not given (effect estimate of randomisation, LASSO selected instruments=-0.41, standard error=6.70, p=0.951). Receiving a problem agreement as part of therapy is estimated to reduce symptom severity by over 10 points but there is great uncertainty around the estimates and the effect is not found to be a significant (effect estimate for problem agreement, LASSO selected instruments -11.4, 95% CI -29.8 to 7.1, p=0.226).

**Table 4.13: The effect of randomisation and agreement of problems and goals on symptom severity at 12 months for complete case and imputed datasets**

| | Complete Case N=126 | | | | Imputed N=188 | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | Bootstrap normal 95% | | | | Bootstrap normal 95% | | p- |
| Effect | Coef. | S.E. | CI | | Coef. | S.E. | CI | | value |
| All instruments – OLS | | | | | | | | | |
| Randomisation | -7.46 | 4.48 | -16.42 | 1.16 | -5.79 | 3.54 | -12.73 | 1.15 | 0.102 |
| Problem Agreement | 1.53 | 5.47 | -8.91 | 12.53 | -2.64 | 3.99 | -10.47 | 5.19 | 0.508 |
| LASSO selection of instruments - 2SLS | | | | | | | | | |
| Randomisation | 4.57 | 32.04 | -50.08 | 75.52 | -0.41 | 6.70 | -13.54 | 12.72 | 0.951 |
| Problem Agreement | -13.05 | 40.23 | -102.4 | 55.31 | -11.38 | 9.40 | -29.81 | 7.05 | 0.226 |

Formulation

There is a large discrepancy in the IV effect estimates of formulation when analysing only participants with complete data compared to those with missing covariates that have been imputed (Table 4.14). There is no significant effect of either randomisation or formulation when analysing only complete cases. When covariates are imputed it is estimated that the use of formulation significantly improves symptom severity reducing the CAARMS

symptom score by approximately 20 points by the 12 month follow-up. There is a great deal of uncertainty associated with these estimates indicated by wide confidence intervals (formulation effect estimate for LASSO selected instruments -22, 95% CI -44 to -0.4, p=0.048). The effect of randomisation when no formulation is given is not statistically significant (randomisation effect estimate, LASSO selected instruments=4.3 95% CI=-8.3 to 17.0, p=0.502).

**Table 4.14: The effect of randomisation and formulation on symptom severity at 12 months outcome for complete case and imputed datasets**

| | Complete Case N=126 | | | | Imputed N=188 | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | Bootstrap normal 95% | | | | Bootstrap normal 95% | | p- |
| Effect | Coef. | S.E. | CI | | Coef. | S.E. | CI | | value |
| All instruments - OLS | | | | | | | | | |
| Randomisation | -9.65 | 3.99 | -17.56 | -1.92 | -9.89 | 2.87 | -15.51 | -4.26 | 0.001 |
| Formulation | 5.37 | 5.45 | -5.05 | 16.32 | 3.59 | 3.65 | -3.56 | 10.75 | 0.325 |
| LASSO selection of instruments - 2SLS | | | | | | | | | |
| Randomisation | -3.97 | 7.24 | -18.64 | 9.72 | 4.34 | 6.45 | -8.31 | 16.99 | 0.502 |
| Formulation | -1.73 | 10.55 | -21.18 | 20.18 | -22.15 | 11.08 | -43.87 | -0.43 | 0.048 |

Homework

Table 4.15 and Table 4.16 provide the results for the effect of having homework as part of therapy using continuous and binary measures of homework participation respectively. The magnitudes of the effect estimates for receiving homework are similar between complete case and imputed data when the confidence intervals around the estimates are also taken into account. In both the complete case and imputed data when the instruments are selected by the LASSO the effect estimate is not significant (imputed effect estimate for homework, LASSO selected instruments=-0.20, standard error=0.15, p=0.182). The direct effect of therapy if homework is not involved in any sessions is not statistically significant in any of the analyses. The average first-stage F-statistic of the instruments selected by the LASSO method is 3.5, indicating the possibility of weak instrument bias.

**Table 4.15: The effect of randomisation and proportion of sessions involving homework on symptom severity at 12 months outcome for complete case and imputed datasets**

| | Complete Case N=126 | | | | Imputed N=188 | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | Bootstrap | | | | Bootstrap | | |
| Effect | Coef. | S.E. | normal 95% CI | | Coef. | S.E. | normal 95% CI | | p-value |
| All instruments - OLS | | | | | | | | | |
| Randomisation | -4.862 | 4.199 | -13.13 | 3.334 | -5.322 | 3.458 | -12.10 | 1.45 | 0.125 |
| % homework | -0.034 | 0.074 | -0.176 | 0.113 | -0.055 | 0.056 | -0.16 | 0.06 | 0.330 |
| LASSO selection of instruments - 2SLS | | | | | | | | | |
| Randomisation | 3.441 | 7.693 | -38.68 | 49.36 | -0.013 | 6.337 | -12.43 | 12.41 | 0.998 |
| % homework | -0.218 | 0.153 | -1.151 | 0.609 | -0.201 | 0.149 | -0.49 | 0.09 | 0.182 |

When the proportion of sessions involving homework is dichotomised at 50% (Table 4.16) the conclusions drawn are the same. The effect estimate for homework use in more than half of sessions is not significant; effect estimate with LASSO selected instruments=-15, p=0.182. The direct effect of therapy when homework is not given is estimated to reduce symptoms but the effect is not significant.

**Table 4.16: The effect of randomisation and homework in more than half of sessions on symptom severity at 12 months outcome for complete case and imputed datasets**

| | Complete Case N=126 | | | | Imputed N=188 | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | Bootstrap | | | | Bootstrap | | |
| Effect | Coef. | S.E. | normal 95% CI | | Coef. | S.E. | normal 95% CI | | p-value |
| All instruments - OLS | | | | | | | | | |
| Randomisation | -3.44 | 3.88 | -11.04 | 4.17 | -9.89 | 2.87 | -15.51 | -4.26 | 0.001 |
| >50% homework | -5.24 | 5.14 | -15.20 | 4.97 | 3.59 | 3.65 | -3.56 | 10.75 | 0.325 |
| LASSO selection of instruments - 2SLS | | | | | | | | | |
| Randomisation | 7.30 | 8.30 | -3.11 | 29.45 | -1.11 | 5.69 | -12.26 | 10.04 | 0.845 |
| >50% homework | -23.06 | 13.40 | -59.73 | -7.21 | -15.18 | 11.30 | -37.32 | 6.96 | 0.182 |

Active change strategies

The direct effect of randomisation and the effect of active change strategies in therapy are presented in Table 4.17 and Table 4.18. There is no significant direct effect of randomisation when use of active change strategies is accounted for. Each increase in the

proportion of therapy sessions involving change strategies is estimated to decrease symptom severity by approximately 0.2 points on the CAARMS symptom scale (effect estimate for LASSO selected instruments -0.21, 95% CI -0.47 to 0.05, p=0.113). The average first-stage F-statistic of the instruments selected by the LASSO method is 5.2.

**Table 4.17: The effect of randomisation and proportion of sessions involving active change strategies on symptom severity at 12 months outcome for complete case and imputed datasets**

| | Complete Case N=126 | | | | Imputed N=188 | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | Bootstrap | | | | Bootstrap | | |
| Effect | Coef. | S.E. | normal 95% CI | | Coef. | S.E. | normal 95% CI | | p-value |
| All instruments – OLS | | | | | | | | | |
| Randomisation | -9.545 | 4.516 | -18.69 | -0.99 | -6.437 | 3.442 | -13.18 | 0.31 | 0.062 |
| % change strategies | 0.063 | 0.081 | -0.09 | 0.23 | -0.025 | 0.062 | -0.15 | 0.10 | 0.686 |
| LASSO selection of instruments - 2SLS | | | | | | | | | |
| Randomisation | 8.908 | 8.116 | 4.33 | 36.15 | 1.507 | 6.149 | -10.55 | 13.56 | 0.807 |
| % change strategies | -0.263 | 0.148 | -0.75 | -0.17 | -0.211 | 0.131 | -0.47 | 0.05 | 0.113 |

The binary measure loses power and the estimate of a decrease in symptom severity is not significant at the 5% level (Table 4.18). It is estimated that using active change strategies in more than half of sessions decreases symptom severity but the effect is not statistically significant (change strategies effect estimate, LASSO selected instruments -16, 95% CI -38 to 5.4, p=0.152).

**Table 4.18: The effect of randomisation and active change strategies in more than half of sessions on symptom severity at 12 months outcome for complete case and imputed datasets**

| | Complete Case N=126 | | | | Imputed N=188 | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | Bootstrap | | | | Bootstrap | | |
| Effect | Coef. | S.E. | normal 95% CI | | Coef. | S.E. | normal 95% CI | | p-value |
| All instruments - OLS | | | | | | | | | |
| Randomisation | -8.25 | 3.70 | -15.55 | -1.03 | -8.21 | 2.90 | -13.91 | -2.52 | 0.005 |
| >50% change strategies | 3.15 | 4.81 | -6.12 | 12.73 | 0.99 | 3.84 | -6.54 | 8.52 | 0.796 |
| LASSO selection of instruments - 2SLS | | | | | | | | | |
| Randomisation | 1.76 | 6.77 | -6.66 | 19.90 | 0.46 | 6.01 | -11.31 | 12.24 | 0.939 |
| >50% change strategies | -12.64 | 9.97 | -40.42 | -1.35 | -16.37 | 11.08 | -38.09 | 5.36 | 0.152 |

Composite measures of therapy content

Formulation has been shown to improve severity of symptoms in people at high risk of psychosis. There is some indication that both homework and active change strategies are important factors to reduce symptoms though the results are not statistically significant. The magnitudes of the effect estimates for each of the components are similar and there is high correlation between the components, this indicates that the individual estimates may be measuring the same effect. To gain further insight we consider the impact on symptom severity at 12 months of having all required components of therapy present versus some or none of the components of therapy; the results are shown in Table 4.19. The results again are similar to those seen for the separate components of therapy with approximately a twenty point decrease in symptom severity when all aspects of therapy are received compared to some or none of the components, the result is of borderline significance when instruments are selected by the LASSO (imputed estimate of receiving full therapy= -20, 95% CI -40 to 0.06, p=0.051). The direct effect of treatment in this situation is the effect when only some or none but not all of the specific components of therapy have been used in the intervention given. As in the other analyses the direct effect of randomisation is not statistically significant though the direction of the estimates is to reduce symptoms at 12 months.

**Table 4.19: The effect of randomisation and all components of therapy on symptom severity at 12 months outcome for complete case and imputed datasets**

| | Complete Case N=126 | | | | Imputed N=188 | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | Bootstrap | | | | Bootstrap | | |
| Effect | Coef. | S.E. | normal 95% CI | | Coef. | S.E. | normal 95% CI | | p-value |
| All instruments - OLS | | | | | | | | | |
| Randomisation | -4.92 | 5.73 | -16.46 | 6.02 | -7.21 | 2.63 | -12.36 | -2.06 | 0.006 |
| All components | -1.64 | 6.37 | -13.73 | 11.26 | -1.41 | 4.41 | -10.06 | 7.24 | 0.750 |
| LASSO selection of instruments - 2SLS | | | | | | | | | |
| Randomisation | 27.97 | 130.23 | -211.6 | 298.9 | -2.80 | 3.52 | -9.70 | 4.10 | 0.427 |
| All components | -38.42 | 149.50 | -349.4 | 236.7 | -20.21 | 10.34 | -40.49 | 0.06 | 0.051 |

Exploring the effects of receiving different levels of therapy further the three-category composite measure is analysed to compare outcomes in participants receiving some of the key components of therapy and all components to participants that receive none. It is

assumed that if participants do not receive any of the key components of therapy they will not experience an effect of treatment, this is the exclusion restriction applied to the effect of randomisation. The results of the analysis presented in Table 4.20 show that receiving all components of therapy reduce symptom severity at twelve months by approximately 23 points on the CAARMS severity scale (LASSO selected imputed coefficient=-23, standard error=8.7, p=0.008) whereas there is no significant effect on symptoms if only some but not all components are received.

**Table 4.20: The effect of receiving some and all components compared to no components of therapy on symptom severity at 12 months outcome for complete case and imputed datasets**

| | Complete Case N=126 | | | | Imputed N=188 | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | Bootstrap | | | | Bootstrap | | |
| Effect | Coef. | S.E. | normal 95% CI | | Coef. | S.E. | normal 95% CI | | p-value |
| All instruments - OLS | | | | | | | | | |
| Some components | -6.87 | 3.36 | -13.51 | -0.33 | 4.80 | 3.00 | -1.09 | 10.69 | 0.110 |
| All components | 1.66 | 5.79 | -9.18 | 13.52 | 5.43 | 2.57 | 0.39 | 10.46 | 0.036 |
| LASSO selection of instruments - 2SLS | | | | | | | | | |
| Some components | -0.33 | 4.48 | -6.25 | 11.31 | -3.70 | 5.15 | -13.80 | 6.40 | 0.473 |
| All components | -15.73 | 9.88 | -43.91 | -5.17 | -23.15 | 8.67 | -40.13 | -6.16 | 0.008 |

Each of these analyses has allowed for a direct effect of treatment meaning that participants could potentially experience an effect of treatment even if they did not receive the specific aspect of therapy in question. In all cases the direct effect of treatment does not have a significant effect on symptoms and the confidence interval around the estimate is very wide. In Table 4.21 the effect estimates for the post-randomisation process variables are presented where no direct effect of randomisation is allowed. These instrumental variables analyses are implemented using 2SLS estimation with randomisation group as the instrument. Restricting the direct effect of randomisation to equal zero generates more confidence in the estimates of the effects of the post-randomisation mediators and all are estimated to be strongly associated with an improvement in symptoms. Although the direct

effect was not statistically significant it would be a strong assumption to not allow it and would result in overly confident conclusions.

**Table 4.21: Estimates without the inclusion of a direct effect of randomisation (only the effect of process variable shown), imputed data**

| | Coefficient | Std. Error | 95% CI | | p-value |
|---|---|---|---|---|---|
| Problem agreement | -10.01 | 3.29 | -16.46 | -3.57 | 0.002 |
| Formulation | -11.49 | 3.90 | -19.12 | -3.85 | 0.003 |
| % sessions involving homework | -0.17 | 0.06 | -0.28 | -0.06 | 0.003 |
| >50% sessions involving homework | -14.15 | 4.76 | -23.47 | -4.83 | 0.003 |
| % sessions involving active change strategies | -0.15 | 0.05 | -0.24 | -0.05 | 0.003 |
| >50% sessions involving change strategies | -12.70 | 4.34 | -21.21 | -4.19 | 0.003 |

### 4.3.5 Stage 4: interaction of sessions and process variables

#### *4.3.5.1 Statistical methods*

The model depicted in the path diagram of Stage 4 involves two post-randomisation process variables, attendance at therapy and the interaction of attendance with content of therapy. Participants that attend more therapy sessions and experience therapy that adheres to the protocol are likely to be different in some underlying way to those that attend fewer sessions or whose therapy does not follow protocol and these underlying traits may also influence the participant's outcome; unmeasured confounding between both of these measures and the outcome is expected. To remove bias due to unmeasured confounding and allow a causal interpretation of the results instrumental variables analysis is used. Values of the post-randomisation process variables are predicted based on the instruments. These predicted values are used instead of the observed values in a regression analysis. For the model to be identified instruments must be found for both process variables. The LASSO method is used to select instruments for the attendance by content interaction and used as instruments for attendance as well.

All analyses are conducted on complete cases as well as imputed data and bootstrapped with 1000 replications.

#### *4.3.5.2 Results*

When the effect of attendance and the multiplicative effect of attendance with a process variable are modelled together no significant effect is seen on symptom severity at 12 month follow-up for either (Table 4.22). The models for the interaction of problem agreement by attendance estimate the effect of attendance when the component in question is not present to be positive on symptom severity (indicated by the negative sign on the coefficient) but the additional effect of attendance when problem agreement is present has a negative effect (positive coefficient) on symptoms. This opposes the previous results which indicated a positive (though non-significant) effect of problem agreement on symptom severity. However, as the estimate is non-significant, the effect could be in the opposite direction. This pattern is also seen for the effect of attendance and homework interaction though again the finding is not significant.

The interactions between attendance and formulation or active change strategies both provide an estimated treatment effect in the expected direction. The results estimate a reduction in symptoms with increasing attendance and an additional reduction if the therapy has contained the specific component of interest. None of these produce a significant result for either the direct effect of attendance or the multiplicative effect of attendance and therapy component.

To help explain these results the data are analysed using principal stratification methods estimating the effect of attendance at therapy within those that do not experience the process of interest as part of their therapy (never-takers as defined in the CACE analysis) and those that do (compliers as defined in the CACE analysis). The analyses indicate that the effect of attendance reduces symptoms within strata but the effect is similar whether the component of therapy of interest is present or not. The positive effect estimates for the interactions of problem agreement and homework with attendance do not necessarily indicate that attendance increases symptoms if these components are present, only that the reduction in symptoms due to attendance is less when these components are present. This result is in agreement with that of the IV regression and explains that although greater attendance may lead to a greater reduction in severity of symptoms the effect is not significantly different if the component of therapy is present or not.

**Table 4.22: Attendance and attendance by process interactions on severity of symptoms at 12 months. Complete case and imputed data results, LASSO selected instruments.**

| | Complete Case | | | | Imputed | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | Bootstrap Normal 95% | | | | Bootstrap Normal 95% | | |
| Estimated effect | Coef. | S.E. | CI | | Coef. | S.E. | CI | | p-value |
| Assessment of problems and goals | | | | | | | | | |
| Sessions | -1.409 | 1.718 | -5.023 | 1.710 | -1.819 | 1.572 | -4.900 | 1.262 | 0.248 |
| Sessions*Problem agreement | 1.046 | 1.836 | -2.308 | 4.889 | 1.084 | 1.752 | -2.350 | 4.518 | 0.537 |
| Formulation | | | | | | | | | |
| Sessions | -0.475 | 1.118 | -2.456 | 1.927 | -0.068 | 1.449 | -2.908 | 2.772 | 0.963 |
| Sessions*Formulation | 0.066 | 1.289 | -2.698 | 2.357 | -1.014 | 1.791 | -4.524 | 2.496 | 0.574 |
| Homework (more than half of sessions) | | | | | | | | | |
| Sessions | 0.096 | 0.594 | -0.962 | 1.367 | -1.029 | 0.687 | -2.375 | 0.316 | 0.135 |
| Sessions*Homework | -0.821 | 0.801 | -2.563 | 0.579 | 0.220 | 1.018 | -1.775 | 2.215 | 0.829 |
| Active change strategies (more than half of sessions) | | | | | | | | | |
| Sessions | 0.140 | 0.965 | -1.028 | 2.753 | -0.499 | 0.909 | -2.280 | 1.282 | 0.584 |
| Sessions*Change strategies | -0.764 | 1.134 | -3.910 | 0.537 | -0.560 | 1.196 | -2.903 | 1.784 | 0.640 |

As with the previous analyses if it is assumed that there is no effect of attending therapy unless the component of interest is present in the therapy i.e. an exclusion restriction on attendance, then attendance with the presence of each component of therapy significantly improves outcomes. The magnitude of the effects are similar, if slightly larger than that seen for the direct effect of attendance detailed at the beginning of this results section.

### 4.3.6 Stage 5: mediators of the treatment process

#### *4.3.6.1 Methods*

It is hypothesised that cognitive therapy will improve symptoms by changing certain beliefs that the patient has about themselves and their illness. Specifically it is thought that CBT will change clients meta-cognitive appraisal of their illness, their beliefs about their illness, beliefs about paranoia and core beliefs.

Details of the scales used to measure the belief mediators are given in chapter 4.1.4. The belief mediator scales used are the MetaCognition Questionnaire (MCQ - 4 subscales analysed), Brief Core Schema Scale (BCSS - 2 subscales analysed), Beliefs About

Paranoia Scale (BAPS - 2 subscales analysed), Personal Beliefs about Experiences Scale (PBEQ – 2 subscales). CBT therapy is expected to lead to an improvement in outcomes by: reducing scores on the four metacognition scales, the BCSS negative thoughts about self, the BAPS negative thoughts about paranoia and the PBEQ negative appraisal of experiences; increasing scores on the BCSS positive thoughts about self, the BAPS normal thoughts about paranoia and the PBEQ social acceptance of experiences scales. The belief measures were recorded at one and six months after randomisation. Change is expected by the six month follow-up and this time point is used in the analyses.

The first step to test if changes in beliefs are mediators on the pathway from therapy to symptoms is to test the pathway from therapy to beliefs. To do this the intention to treat (Stage 1), attendance at therapy (Stage 2) and process (Stage 3 and 4) analyses are repeated with beliefs as the outcome instead of symptoms. If it is shown that therapy does have an effect on beliefs the next step will incorporate beliefs as a mediator on outcome (Stage 5).

The intention to treat analysis is summarised by the mean summed scores at 6 months follow-up for each belief scale in the two treatment arms. Differences are tested using a t-test as well as a linear regression adjusting for treatment centre. The analyses are not adjusted for baseline measures of the outcome scores as they were not taken at baseline but at one month after randomisation.

The impact of attendance is investigated in an instrumental variables analysis with randomisation group as the instrument using both the continuous measure of attendance and the binary measure dichotomised at 4 sessions (CACE analysis). The analyses will be adjusted for centre and no direct effect of randomisation is modelled (exclusion restriction is applied).

Stage 3 analyses investigating the effect of process variables are repeated with the belief measures as outcomes. Instrumental variable analysis is applied with instruments chosen by the LASSO. A direct effect of randomisation on beliefs at 6 months is allowed.

Stage 4 analyses investigating the effect of attendance and attendance by process variable interactions are repeated with the belief measures as outcomes. As previously, instrumental variable analysis is applied with instruments chosen by the LASSO. No direct effect of randomisation on beliefs at 6 months is allowed.

All analyses are adjusted for participant site and carried out on participants with complete data only as well as an analysis of imputed data. There is no adjustment for baseline score of the mediator as these were not measured prior to randomisation. Results are bootstrapped with 1000 replications.

#### *4.3.6.2 Results*

##### *4.3.6.2.1 Intention to treat*

No significant effect of randomisation group was found on any of the belief mediators at six months follow-up with or without adjustment for centre. Mean scores by group, regression coefficients and test results are presented in Table 4.23. The implication of this finding is that the beliefs cannot be mediators of the treatment effects, but the results of the more complex analyses will be presented in any case, with the expectation that they will be consistent with this initial judgement.

##### *4.3.6.2.2 Attendance at therapy*

Attendance at therapy as either a continuous or categorical measure showed no significant effect on any belief measures when analysed in an instrumental variables regression. The results for the continuous outcome, are presented in the final columns of Table 4.23.

##### *4.3.6.2.3 Post-randomisation process variables*

The direct impact of attending therapy with the indirect effect of receiving each specific component of therapy on each of the hypothesised belief mediators is calculated and the results given in Table 4.24 for instruments selected by the LASSO procedure. There are a great number of tests carried out here and so caution is needed when interpreting the results. In many of the analyses the estimates for the group effect and post-randomisation process effect are in different directions. We would expect them to both be in the same direction whether that is positive or negative. However very few results are significant and the confidence intervals are wide indicating that they could act in either direction. When instruments are selected by the LASSO three effects are seen to be significant at the 5% level. The direct effect of randomisation that is not through problem agreement increases BCSS negative thoughts about self whereas inclusion of a problem agreement decreases the score on the negative thoughts scale; CBT is expected to reduce this measure. The

direct effect of randomisation that is not through homework increases negative beliefs about paranoia but an increase in the proportion of sessions that involve homework decreases negative beliefs about paranoia; this measure is expected to decrease with CBT. There is no significant impact of homework on negative beliefs about paranoia when the binary measure of homework in more than half of sessions is considered. Similarly a decrease in negative thoughts about paranoia is expected as more therapy sessions involve active change strategies but there is an increase if therapy does not involve active change strategies. No significant effect is found for having more than half of sessions involve an active change strategy. Additionally if the exclusion restriction is placed on randomisation so that no direct effect of randomisation on the belief mediators is allowed, randomisation is used as the instrument for the process variables and no effect is found for any of the process variables on beliefs at six months.

#### *4.3.6.2.4 Attendance and post-randomisation process interaction on beliefs*

No significant effect of attendance or attendance by post-randomisation process variable interaction was found on any of the belief measures at six months. The results have not been shown here since there were again a large number of tests carried out but are provided in the Appendix 2. The estimates suffered from large standard errors indicating a lack of confidence in the results.

### 4.3.7 Stage 6: interaction of sessions and process variable through mediators

Since there is no effect of therapy on the hypothesised belief mediators there will be no causal effect of these on outcomes and so the hypothesised model illustrated in Stage 5 and Stage 6 are not valid from these results. Reminding ourselves of these models the causal links between treatment and the belief mediators analysed can be removed as shown in Figure 4.8 below.

**Figure 4.8: Revised causal diagrams of Stage 5 and Stage 6**

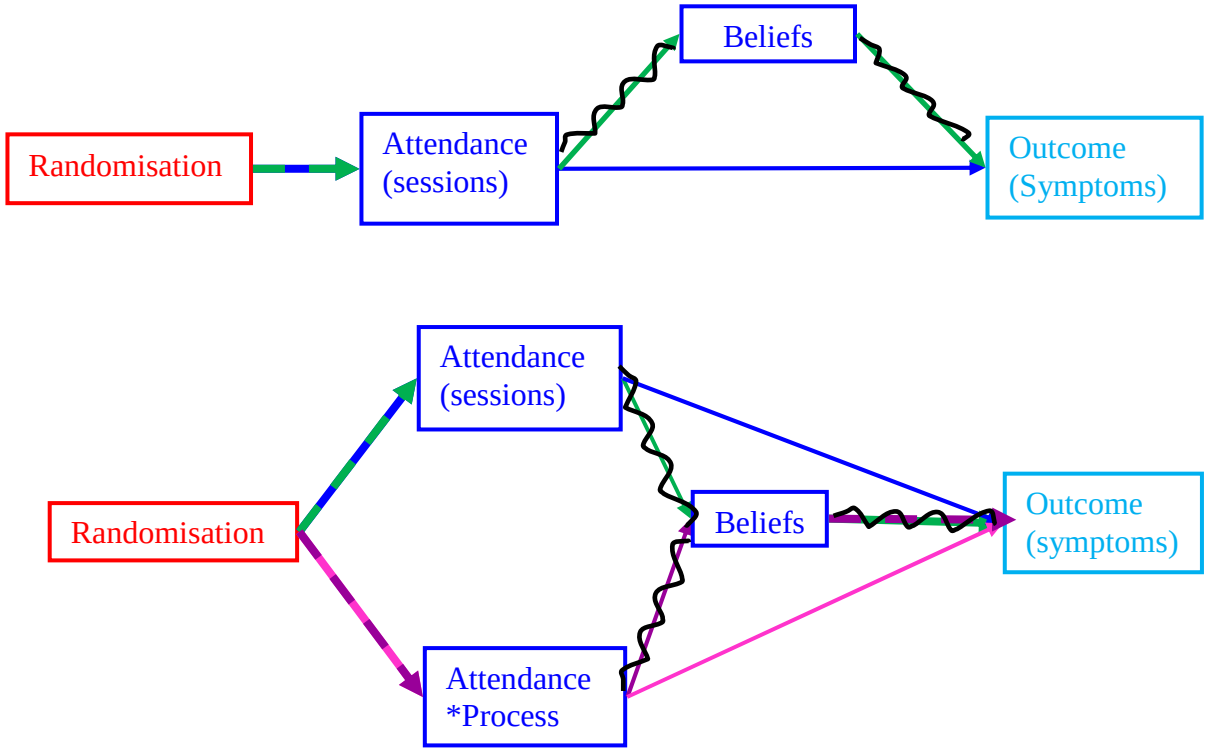


**Table 4.23: Treatment effect of CBT on belief mediators at six months follow-up, intention to treat and CACE analysis**

| | Belief measure at 6 months | | | Intention to treat analysis | | Instrumental variables analysis Effect of attendance at sessions | |
|---|---|---|---|---|---|---|---|
| | Total N, mean (sd) | Control N, mean (sd) | Treatment N, mean (sd) | Treatment effect (95% CI) | p-value* | Coefficient (95% CI) | p-value |
| MCQ Cog Con | 148, 11.91 (4.67) | 75, 11.85 (4.46) | 73, 11.97 (4.90) | 0.11 (-1.40 - 1.62) | 0.886 | 0.01 (-0.13 - 0.15) | 0.883 |
| MCQ Cog Self | 148, 15.09 (4.89) | 74, 15.09 (4.52) | 74, 15.09 (5.27) | 0.002 (-1.59 - 1.59) | 0.998 | 0.002 (-0.14 - 0.14) | 0.998 |
| MCQ Neg Thoughts | 144, 14.41 (5.31) | 73, 14.38 (5.19) | 71, 14.44 (5.47) | 0.01 (-1.68 - 1.70) | 0.990 | 0.001 (-0.15 - 0.15) | 0.990 |
| MCQ Neg Control | 147, 11.79 (4.22) | 74, 11.96 (4.15) | 73, 11.62 (4.32) | -0.32 (-1.68 - 1.05) | 0.645 | -0.03 (-0.15 - 0.09) | 0.638 |
| BCSS Neg Self | 152, 5.74 (5.80) | 75, 5.43 (5.60) | 77, 6.04 (6.01) | 0.73 (-1.14 - 2.60) | 0.442 | 0.07 (-0.10 - 0.24) | 0.432 |
| BCSS Pos Self | 150, 8.33 (6.25) | 73, 9.10 (6.57) | 77, 7.60 (5.88) | -1.62 (-3.65 - 0.41) | 0.118 | -0.15 (-0.33 - 0.03) | 0.107 |
| BAPS negative | 159, 13.73 (5.39) | 78, 13.10 (5.13) | 81, 14.33 (5.60) | 1.34 (-0.34 - 3.03) | 0.118 | 0.12 (-0.03 - 0.27) | 0.106 |
| BAPS normal | 162, 16.04 (5.01) | 80, 15.48 (5.05) | 82, 16.60 (4.93) | 1.09 (-0.48 - 2.66) | 0.171 | 0.10 (-0.04 - 0.24) | 0.163 |
| PBEQ NAE | 164, 20.78 (5.09) | 83, 21.12 (4.96) | 81, 20.43 (5.23) | -0.58 (-2.10 - 0.95) | 0.454 | -0.05 (-0.19 - 0.08) | 0.447 |
| PBEQ SAE | 165, 10.15 (2.05) | 83, 10.06 (2.04) | 82, 10.24 (2.06) | 0.16 (-0.47 - 0.79) | 0.609 | 0.02 (-0.04 - 0.07) | 0.602 |

*adjusted for site

**Table 4.24: Instrumental variables (2SLS) analysis of group and process variable effects on beliefs at 6 months; instruments selected via LASSO, imputed data, 1000 bootstraps**

| Belief measure | Effect | Agreement of problems and goals | | | | Formulation | | | | % homework | | | | % active change strategies | | | | Expected direction |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Coef. | 95% CI | | | Coef. | 95% CI | | | Coef. | 95% CI | | | Coef. | 95% CI | | | |
| MCQ Cog Con | Group | 1.43 | -2.85 | 5.71 | | 0.53 | -3.37 | 4.43 | | 0.18 | -3.83 | 4.19 | | -0.18 | -3.66 | 3.31 | | |
| | Process | -2.19 | -7.85 | 3.47 | | -1.11 | -7.60 | 5.38 | | -0.01 | -0.10 | 0.09 | | 0.00 | -0.07 | 0.07 | | Neg |
| MCQ Cog Self | Group | -2.83 | -7.11 | 1.45 | | -2.32 | -6.28 | 1.64 | | -0.82 | -5.01 | 3.36 | | -1.93 | -5.69 | 1.84 | | |
| | Process | 3.59 | -2.17 | 9.36 | | 3.74 | -2.93 | 10.42 | | 0.00 | -0.10 | 0.10 | | 0.03 | -0.05 | 0.11 | | Neg |
| MCQ Neg Thoughts | Group | 2.29 | -2.31 | 6.89 | | 2.20 | -2.16 | 6.55 | | 3.48 | -1.08 | 8.05 | | 2.72 | -1.38 | 6.82 | | |
| | Process | -3.97 | -10.14 | 2.20 | | -4.52 | -11.78 | 2.73 | | -0.10 | -0.21 | 0.00 | | -0.07 | -0.16 | 0.01 | | Neg |
| MCQ Neg Control | Group | -0.88 | -4.21 | 2.46 | | 0.06 | -2.93 | 3.05 | | 0.26 | -3.37 | 3.89 | | 0.19 | -2.97 | 3.35 | | |
| | Process | 0.48 | -4.26 | 5.23 | | -1.03 | -6.31 | 4.24 | | -0.03 | -0.11 | 0.06 | | -0.02 | -0.09 | 0.05 | | Neg |
| BCSS Neg Self | Group | 5.05 | 0.48 | 9.63 | * | 3.79 | -1.16 | 8.74 | | 3.76 | -0.93 | 8.45 | | 2.97 | -1.02 | 6.97 | | |
| | Process | -7.45 | -14.08 | -0.81 | * | -7.01 | -15.44 | 1.42 | | -0.10 | -0.22 | 0.01 | | -0.07 | -0.16 | 0.02 | | Neg |
| BCSS Pos Self | Group | -3.08 | -8.98 | 2.82 | | -1.90 | -7.19 | 3.40 | | -1.31 | -6.66 | 4.05 | | -0.94 | -6.19 | 4.31 | | |
| | Process | 2.96 | -5.01 | 10.93 | | 0.96 | -8.07 | 9.98 | | 0.00 | -0.13 | 0.13 | | -0.01 | -0.12 | 0.10 | | Pos |
| BAPS negative | Group | 2.84 | -1.21 | 6.89 | | 3.32 | -0.30 | 6.94 | | 5.17 | 0.65 | 9.68 | * | 5.12 | 1.23 | 9.01 | * | |
| | Process | -3.06 | -8.72 | 2.61 | | -4.67 | -11.38 | 2.04 | | -0.12 | -0.23 | -0.01 | * | -0.10 | -0.18 | -0.02 | * | Neg |
| BAPS normal | Group | -0.51 | -4.65 | 3.63 | | 1.08 | -2.76 | 4.92 | | 1.92 | -2.37 | 6.21 | | 2.54 | -1.57 | 6.65 | | |
| | Process | 2.22 | -3.70 | 8.14 | | 0.06 | -6.67 | 6.79 | | -0.02 | -0.13 | 0.09 | | -0.03 | -0.12 | 0.06 | | Pos |
| PBEQ NAE | Group | -0.35 | -3.14 | 2.43 | | -1.06 | -3.58 | 1.47 | | -0.85 | -3.82 | 2.13 | | -0.68 | -3.87 | 2.52 | | |
| | Process | 0.44 | -3.49 | 4.36 | | 1.85 | -1.57 | 5.28 | | 0.02 | -0.04 | 0.08 | | 0.01 | -0.04 | 0.07 | | Neg |
| PBEQ SAE | Group | 2.17 | -3.00 | 7.34 | | 1.39 | -2.72 | 5.50 | | -0.21 | -6.07 | 5.64 | | 0.94 | -4.50 | 6.37 | | |
| | Process | -3.30 | -10.15 | 3.56 | | -2.76 | -8.80 | 3.27 | | 0.00 | -0.13 | 0.12 | | -0.03 | -0.13 | 0.07 | | Pos |

*p<0.05, **p<0.01

## 4.4 Conclusions

The analysis of the mediation process of cognitive therapy on the reduction of symptom severity has been broken down into a series of stages but these must be considered together to understand the fuller picture.

The inclusion of formulation during therapy has been shown to have a positive impact on severity of symptoms at 12 months follow-up as well as an improvement when more sessions are attended. There is some evidence to suggest that the inclusion of homework and active change strategies improve symptom severity at twelve months though the effects are not found to be statistically significant. These aspects of therapy are expected by therapists to be important factors in the success of CBT according to the Delphi study by Morrison and Barratt[49]. It is logical to believe that attending more therapy sessions that contain the effective components of therapy would have an even greater effect. Unfortunately the results do not support this hypothesis and we find that when considering these together neither has an effect. The number of sessions attended is highly associated with whether or not specific components of therapy are included and it is not possible to untangle the effects of one from another. Additionally, participants who experience one aspect of therapy are also more likely to experience other aspects. The content of therapy variables seem to be explaining the same variation in symptoms and this post-hoc analysis cannot distinguish the affect attributed to one from another. Sample size may also cause a problem as large samples are needed to show interaction effects and less than 200 participants have information at follow-up.

Cognitive behavioural therapy for the prevention of psychosis is hypothesised to change the patient's beliefs about their thoughts and symptoms and therefore improve their symptoms. The results have shown that the impact of CBT on patient's beliefs is no greater than that of the control condition. Delving further it has been shown that the lack of an intention to treat effect on any of the hypothesised belief mediators is not due to participants not complying with therapy and not attending enough therapy sessions. The treatment effect in compliers, defined as those that have attended at least 4 sessions remains non-significant. An analysis of the inclusion of specific components of therapy indicates that some components of therapy may have an impact on some belief mediators. However, since so many tests have been completed and the findings are not consistent across either the belief outcomes or the post-randomisation processes these findings should be interpreted with caution and may be a result of multiple testing. When stricter

assumptions are applied excluding a direct effect of randomisation and forcing all of the effect through the process variable there is still no significant impact on beliefs for any component of therapy.

Though the therapy is effective in improving symptoms by 12 months it does not appear to change beliefs by the 6 month follow-up as expected. The analysis is limited to the follow-up times of the study and the psychometric measures recorded. It may be that CBT does have an effect on beliefs but this is seen at a later time after the 6 month follow-up period or the mechanism by which the therapy works is not as hypothesised but effects other beliefs or mediators not captured in the study.

The instrumental variables analysis of the post randomisation processes uses interactions of group by baseline covariates as the instrument for content of therapy received. The instrument is expected to be associated with the component of therapy but to have no direct association with symptom severity other than through the component of therapy received. In order to satisfy the first requirement, baseline covariates that explain variation in the process variable are selected using the LASSO procedure. In this way a balance is achieved between selecting variables to increase the amount of variation explained and reducing the number of instruments selected. We expect any baseline covariate by group interaction to satisfy this second requirement because of randomisation. Since the covariates are balanced between the treatment groups we expect that any relationship between a covariate and outcome will be the same between the two groups unless it is related to the treatment received which then alters the outcome. These interactions have been shown by Small to be valid (Chapter 1.2.8.2) if additionally there is no moderating effect of the covariates on the mediator effect or the direct effect of randomisation. There is no evidence in the literature to suggest that CBT for psychosis effects different subgroups of patients in different ways, however this is not say that there are not effect modifiers present that have not yet been investigated. We assume that the variables considered here are not effect modifiers.

# 5 COMMAND trial

## 5.1 Trial design

The design of the COMMAND trial is described in detail in the study protocol paper[65]. A short summary is provided here. The COMMAND trial is a two arm randomised controlled trial of Cognitive Therapy for Command Hallucinations (CTCH) versus treatment as usual. Participants were patients with a diagnosis of schizophrenia or schizoaffective disorders who had auditory hallucinations (hearing voices) that had previously been treated unsuccessfully. Patients were recruited from inpatient wards and community mental health teams in Birmingham & Leicester, London and Manchester.

The intervention is individual cognitive therapy based on the hypothesised cognitive model of command hallucinations. The general approach of the therapy is to target a patient's beliefs about the voices power, to weaken the voices power and thereby transfer power to the patient enabling them to disobey the voice. The therapy targets four factors of voice power: the power and control that the voice has; that the voice must be obeyed or appeased or the individual will be punished; the identity of the voice and the meaning of the experience i.e. a punishment for past behaviour. A maximum of 25 sessions of therapy is given over, at most, a 9 month period.

Primary outcome

Compliance with each command is the primary outcome of the trial and is assessed by interviews with the participant and their carer/care-coordinator. The participant is asked to describe the voices and their behaviours and reactions to them, this information along with that of the other informants is used together to describe the patients compliance using the Voice Compliance Scale (VCS). Compliance is summarised by a 5 point scale

1. Neither appeasement nor compliance
2. Symbolic appeasement i.e. compliant with harmless commands
3. Actual appeasement i.e. preparatory acts or gestures
4. Partial compliance with at least one severe command
5. Full compliance with at least one severe command

Outcomes are measured at 9 and 18 months after randomisation. Therapy will have been completed by the 9 month measurement point, maintaining the temporal order of events.

The primary analysis dichotomises the compliance measure as not full compliance (grades 1-4) versus full compliance (grade 5).

Mediator

It is hypothesised that CTCH will give the patient more power over the voice and the ability to resist it's commands and so make them less likely to comply. The power differential between the voice and the patient is measured by the Voice Power Differential Scale (VPD)[173,174]. The total scale consists of seven statements: I'm more powerful than My voice; I'm stronger than My voice; I'm more confident than My voice; I respect My voice more than it respects me; I'm more able to harm My voice than it is to harm me; I'm superior to My voice; I'm more knowledgeable than My voice. Each statement is rated on a 1-5 scale indicating whether influence in each construct is with the individual or the voice resulting in a total score of 7-35. The power subscale is a 1-5 rating where a score of one indicates power with the individual and 5 indicates that the power lies with the voice. An improvement would therefore be indicated by a reduction in the total VPD and power subscale. The VPD is measured at baseline and at 9 and 18 months follow-up.

Other measures

In addition to the outcome and mediator of interest other clinical measures were also taken at baseline, 9 and 18 months as well as demographic measures at baseline. Demographic measures of age at onset of psychosis, gender, employment, ethnicity, living situation and treatment centre were taken at baseline. Clinical indicators of symptoms were measured by the Positive and Negative Syndrome Scale (PANSS)[175] and Psychotic Symptoms Rating Scales (PSYRATS)[176], depression by the Calgary Depression Scale (CDSS)[177], hopelessness by the Beck Hopelessness Scale (BHS)[178], suicidal ideation by the Beck Suicidal Ideation scale (BSI)[179] and beliefs about voices by the Beliefs about Voices Questionnaire[180].

## 5.2 Statistical methods

### 5.2.1 Intention to treat analysis

The primary analysis of the trial was carried out by intention-to-treat methods. Treatment effects are estimated at 9 and 18 month follow-up and combined in a cross-sectional time series logistic regression of compliance with command hallucinations on randomisation group. The random effects model accounts for trial site, baseline severity of command

hallucinations and month of follow-up. The trial results which are not yet published but have been submitted[66] show a significant treatment effect on compliance with command hallucinations.

### 5.2.2 Mediation analysis

In order to determine mediation the temporal order of events must be established. Therapy is experienced in the nine months post randomisation, the mediator and outcome is measured at both 9 and 18 months follow-up. Mediation analyses that make use of the measures at both of these time points must account for complexities in the associations between measures over time. The mediator and outcome will be associated between and within each time point and there may not only be an effect of the mediator on the outcome but also of the outcome on the mediator. The causal analysis of these complex longitudinal models is not yet established but is a focus of future developments. The analysis presented here is therefore a simplified model to illustrate the procedures described in the rest of the thesis. The impact of voice power measured at 9 months is therefore determined on compliance measured at 18 months. The first stage of the mediation process is to establish if power of voice is affected by treatment received. The power subscale of the VPD scale at 9 months is regressed on randomisation group in an intention-to-treat analysis with adjustment for trial site and baseline power.

An instrumental variables analysis is conducted to estimate the indirect effect of power of voice on compliance with commands. It is assumed that there may be a direct effect of treatment that is not through the impact on power of voice. Interaction of baseline total VPD and the power of voice subscale by randomisation group are expected to be effective instruments since they are not expected to be associated with the outcome other than through the VPD and power of voice due to the properties of randomisation. Instruments will also be selected using the LASSO method so as to compare with results using instruments that are selected because they are associated with the mediator. It is expected that interactions of baseline covariates with randomisation arm will be valid instruments because, from the properties of randomisation, any association between a covariate and outcome would not be expected to be any different in one randomisation group to another other than through the treatment received.

The outcome compliance with the voice is a binary measure and will be analysed using an instrumental variables probit model. The probit uses maximum likelihood estimation, modelling the probability of the outcome transformed to the inverse normal distribution. It

is therefore not as intuitive in its interpretation. The coefficient is the estimated change in z-score of the outcome that a one unit change in the predictor brings about. Average marginal probabilities are provided to aid interpretation. The average marginal probability is the average change in probability of the outcome associated with a one point change in the variable of interest. The instrumental variable probit regression requires a continuous mediator and assumes that the model for the mediator follows the normality assumptions of a linear regression. The mediator of interest here is the power subscale of the VPD and is a 1-5 rating of the power of the voice. It is assumed that the underlying concept of the power of the voice is a continuous measure and that the interval between a rating of 1 and 2 is the same as the interval between 3 and 4. The measure can therefore be considered as a continuous measure and analysed as such.

The analysis of the power of voice mediator will use both continuous and binary models of the outcome using both instrumental variable regressions and the Baron and Kenny type OLS mediation model. The probit regression will be carried out in one-step using only those with information on both the mediator and outcome. Results will be shown for instruments selected via the LASSO and prior hypotheses. All analyses will be adjusted for baseline compliance, baseline power of voice and centre.

As a sensitivity analysis a linear prediction estimator will be used with parameters estimated using the 2SLS, LIML and Fullers adjustment in a one-step procedure with only participants that have both mediator and outcome values, the results will not be presented here but in the appendix.

### 5.3 Results

The sample description and primary analysis results have been submitted to the reviewing process[66]; a summary is given in Table 5.1 to inform the mediation analysis. 197 participants were recruited and underwent randomisation, of these 164 (83%) completed the 18 month follow-up assessment. No baseline characteristics or clinical measures were associated with missing follow-up at 18 months. Participants tended to be male (57%) and unemployed (74%), the average age of psychosis onset was around 22 years of age. The balance in demographic and clinical measures between the randomisation groups indicates that the randomisation procedure was successful. There are some missing data in the baseline measures but this is very small and so multiple imputation was not carried out. The largest amount of missing data was in the age of onset of psychosis of which 10% was

missing. The analyses were carried out with and without this variable and the results were comparable. Additionally, age of onset was not selected as a predictor of the mediator under any procedures. The mediation results presented here, therefore, do not include age of onset so as to make use of the largest sample possible.

### 5.3.1 Intention to treat analysis

At nine months follow-up 48% of the CTCH group fully-complied with the voices compared to 55% in the TAU group; this reduced in both arms to 28% and 46% by the 18 month follow-up respectively (Table 5.2). The odds ratio at 9 months was estimated to be 0.75 (95% confidence interval 0.40 to 1.39, p=0.353) and at 18 months was estimated to be 0.46 (95% confidence interval 0.23 to 0.89, p=0.021). The treatment by time interaction was not significant, odds ratio=0.54 (95% confidence interval 0.27 to 1.1, p=0.091) indicating that the treatment effect seen at 9 months is not significantly different to the effect seen at 18 months; it was therefore not included in the analysis. The average treatment effect estimate common to both time points, adjusted for baseline compliance and treatment centre, was 0.574, (95% confidence interval 0.33 to 0.98, p=0.042) indicating a significant reduction in the odds of full-compliance amongst those receiving CTCH compared to TAU.

**Table 5.1: Description of COMMAND trial sample at baseline by randomisation group**

| | CTCH+TAU | TAU | Total | Missing |
|---|---|---|---|---|
| Age at onset | 21.8 (10.3) | 22.5 (10.9) | 22.1 (10.6) | 19 (10%) |
| Gender (male): | 61 (62%) | 52 (53%) | 113 (57%) | 0 |
| Site: | | | | |
| Birmingham/Leicester | 43 (44%) | 44 (44%) | 87 (44%) | |
| Manchester | 23 (23%) | 25 (25%) | 48 (24%) | |
| London | 30 (30%) | 32 (33%) | 62 (31%) | 0 |
| Ethnicity: non-white | 31 (32%) | 37 (38%) | 68 (35%) | 1 (<1%) |
| Employment: | | | | |
| Employed | 10 (10%) | 22 (23%) | 16 (32%) | |
| Unemployed | 77 (78%) | 67 (69%) | 144 (74%) | |
| Other | 11 (11%) | 8 (8%) | 19 (10%) | 2 (1%) |
| Cohabiting: yes | 13 (13%) | 7 (7%) | 20 (10%) | 1 (<1%) |
| | | | | |
| PANSS total | 70.7 (17.1) | 72.7 (16.0) | 71.7 (16.6) | 2 (1%) |
| CDSS | 12.4 (6.3) | 11.7 (5.7) | 12.1 (6.0) | 0 |
| BHS | 11.1 (5.30 | 10.4 (5.5) | 10.7 (5.4) | 4 (2%) |
| BSI | 10.6 (9.6) | 9.2 (9.1) | 9.9 (9.4) | 1 (<1%) |
| VDP power | 3.9 (1.2) | 4.0 (1.0) | 3.9 (1.1) | 2 (1%) |
| VDP total | 26.5 (5.5) | 27.4 (4.6) | 27.0 (5.1) | 6 (3%) |

**Table 5.2: Summary of outcome and mediator measures over time by randomisation group**

| | | Baseline | 9 months | 18 months |
|---|---|---|---|---|
| Full compliance | CTCH | 84 (86%) | 41 (48%) | 22 (28%) |
| | TAU | 81 (81%) | 49 (55%) | 39 (46%) |
| VPD power: mean (sd) | CTCH | 3.9 (1.2) n=97 | 2.8 (1.2) n=87 | 2.8 (1.3) n=76 |
| | TAU | 4.0 (1.0) n=98 | 3.3 (1.4) n=86 | 3.2 (1.4) n=81 |
| VPD total: mean (sd) | CTCH | 26.5 (5.5) n=95 | 21.3 (5.9) n=87 | 22.4 (6.2) n=75 |
| | TAU | 27.4 (4.6) n=96 | 24.0 (6.4) n=85 | 23.4 (6.9) n=81 |

### 5.3.2 Mediation analysis

Average scores on the VDP total and power subscales reduce in both groups between baseline and 9 months flattening out at 18 months. An intention-to-treat analysis of treatment on VPD power at 9 months adjusting for centre and baseline power estimates that CTCH reduces power of the voice by 0.56 points (coefficient= -0.56, 95% CI -0.94 to -0.18, p=0.004). Repeating this analysis for total VDP the treatment effect at 9 months is estimated at -2.51 (95% CI -4.36 to -0.66, p=0.008). The analysis has shown that CTCH causes a significant improvement in the power of the voice and therefore may potentially mediate the treatment effect of CTCH on compliance.

To test for mediation the Baron and Kenny mediation model with a probit link function and an instrumental variables probit model are used. Interactions of randomisation group with baseline VPD power and total scores were considered as instruments as they would be expected to only affect outcome through VPD power at 9 months. The other requirement of an instrument is that it is associated with the mediator. First stage regression results indicate that the interaction of baseline VPD power with randomisation group is not significantly associated with VPD power at 9 months and so does not fulfil the requirements of an instrument. The interaction of group with VPD total score at baseline is significantly associated with VPD power at 9 months after accounting for treatment centre, baseline VPD power, baseline VPD total and treatment (coefficient=-0.11, 95% CI -0.19 to -0.04, p=0.004).

The coefficient estimates of the probit model do not provide an intuitive interpretation so marginal effects are given along with the coefficient estimates in Table 5.3. The estimates of the average marginal effect are negative for randomisation and positive for VPD power. This is as expected indicating that the risk of compliance is lower in the CTCH group and increases as the power of the voice increases. The effect estimates of the instrumental variables probit analyses with VPD total only are quite different in magnitude from the probit regression without instrumental variables (Baron and Kenny probit link) and the IV with LASSO selected variables. When VPD total is used as an instrument a greater effect of power of voice is estimated and a smaller direct effect of randomisation.

The first stage F-statistic attributable to the instruments is 11.7 and 3.2 for the models with VPD total only and LASSO selected instruments respectively. If the model with the greatest F-statistic is preferred then this would support the use of VPD total score as an instrument which estimates an average marginal 7% decrease in probability of compliance in the CTCH arm (95% CI -0.27 to 0.13) and an average marginal 14% increase in risk of compliance for every increment on the VPD power scale (95% CI -0.05 to 0.32), neither of which are statistically significant.

**Table 5.3: Instrumental variables analysis of the voice power mediator: compliance outcome modelled as a categorical measure, comparison of instruments used**

| Effect | Coef. | Std. Err. | 95% CI | | p-value | average marginal effect | 95% CI of marginal effect | |
|---|---|---|---|---|---|---|---|---|
| All variables – probit (Baron and Kenny) | | | | | | | | |
| Randomisation | -0.466 | 0.278 | -1.011 | 0.079 | 0.094 | -0.131 | -0.279 | 0.017 |
| Power of voice | 0.168 | 0.118 | -0.062 | 0.399 | 0.153 | 0.047 | -0.016 | 0.111 |
| VPD total by group interaction as instrument – ivprobit | | | | | | | | |
| Randomisation | -0.228 | 0.324 | -0.863 | 0.408 | 0.483 | -0.068 | -0.270 | 0.133 |
| Power of voice | 0.459 | 0.365 | -0.257 | 1.174 | 0.209 | 0.138 | -0.046 | 0.322 |
| LASSO variables by group interaction as instruments – ivprobit | | | | | | | | |
| Randomisation | -0.449 | 0.291 | -1.019 | 0.122 | 0.123 | -0.147 | -0.346 | 0.053 |
| Power of voice | 0.175 | 0.401 | -0.612 | 0.961 | 0.663 | 0.057 | -0.190 | 0.304 |

Applying alternative estimation methods does not change the overall conclusions of the analysis though there are small differences in the effect estimates. These are not shown here but results of the linear probability including the OLS and instrumental variables estimates are given in Appendix 3. When the instruments are selected by the LASSO the

effect of voice power is lower and the direct treatment effect is greater, however they remain non-significant.

In these analyses the exclusion restriction was relaxed on the effect of randomisation. The results do not suggest that there is a direct effect of randomisation. If the exclusion restriction is enforced so that there is no direct effect of randomisation on compliance then randomisation can be used as the instrument. This results in a significant mediating effect of power of voice, as is expected since all of the ITT effect is directed through this mediator. When compliance is modelled as a categorical outcome in a probit analysis the coefficient of voice power is 0.77 (95% CI 0.50 to 1.04, $p<0.001$) which indicates an average marginal effect of 0.21, a 21% increase in risk of compliance with each point increase in VPD power (95% CI 0.16 to 0.26).

## 5.4 Conclusions

The results support those in current literature that CBT reduces the power of the voice and that CBT also reduces compliance with the voice. No other studies have taken the next step to determine if power of the voice is a mediator of the treatment process.

The instrumental variables analysis of the mediation model is used so that the assumption of no unmeasured confounding can be relaxed. In doing so some additional assumptions are made, namely that the instrument is associated with the mediator and that it is not associated with the outcome other than through its effect on the mediator. The interaction of treatment group and total score on the Voice Power Differential scale has been shown to be significantly associated with the power subscale of the VPD at 9 months and the F-statistic attributable to the interaction is 11.7. The second assumption is that the interaction is only associated with the outcome through the mediator. This states that the association between VPD total score and compliance is no different between the treatment groups other than through the change in VPD power at 9 months. This assumption cannot be tested but must be assessed. VPD total and power scores are closely related so an association between baseline VPD and compliance is likely to be seen through a change in power score at 9 months. When all variables or those selected by the LASSO or stepwise are selected as instruments the validity of the second assumption may be questioned. Baseline covariate by randomisation arm interactions that are associated with power of the voice may also be associated with other mediators on the pathway and if so would not be appropriate to use as instruments. For this reason VPD total score may be preferred.

# 6 Discussion

## 6.1 Discussion regarding statistical methodology

Randomised controlled trials are designed to enable unbiased estimation of treatment effects and the interpretation of causal links. However, when investigating the causal pathway of the treatment through mediating factors and process variables these are no longer subject to randomisation but are influenced by external factors which may also influence the outcome. Popular statistical methods implemented in mediation analyses often depend upon unverifiable assumptions. This thesis has applied lesser used methods that have been developed to relax some of these assumptions so that causation can still be inferred. In this study, mediation methods have been used to establish the mechanisms and pathways through which a complex therapy works by looking at post-randomisation process and mediating variables on the causal pathway.

Initial comparisons of estimation methods for the effect of attendance at therapy have shown the different approaches give very similar results strengthening our confidence in the conclusions drawn. The G-estimation and 2SLS approaches are essentially the same if the same instruments are used which reflects work published by Maracy and Dunn[109]. There are small differences in the estimates which can be explained by the software package used and are generally attributed to the estimation methods implemented (maximum likelihood or least squares) or the treatment of missing data. The instrumental variable regression command in Stata, for example, will only use patient observations that have complete data, though the two-stage least squares estimator can be calculated in two stages maximising the sample at each stage, whereas in the software package Mplus all patients with some observed data will be used. Use of all observed data may well be preferred over the complete case analysis, though of course the assumptions of the missing data mechanisms must still be considered. When applied to the EDIE-II dataset the missing data in the baseline measurements have been imputed using multiple imputation by chained equations in order to improve efficiency.

The instrumental variable method was used in this study to conduct mediation analyses. Unfortunately, though these methods overcome problems of unmeasured confounding that cause bias in standard regression approaches they are not infallible. When a mediation analysis is not included in the study design and no prior beliefs exist regarding a valid instrument then instruments must be found statistically. The statistical component of this

thesis sought to investigate further the selection of instruments and specifically the use of weak instruments and multiple instruments in the instrumental variables context, the objectives being to:

- Compare methods for instrumental variable selection in a clinical trial design when there is no prior hypothesis and many potential candidates
- Compare estimation methods in the presence of many, potentially weak instruments
- Apply these methods to estimate mediation effects in the EDIE-II and COMMAND trial data

### 6.1.1 Comparison of selection methods

When applying the instrumental variables analysis to estimate the effect of receiving specific components of therapy on outcome it was decided that randomisation could no longer be used as an instrument as it had been in the analysis of attendance. To use randomisation as an instrument would mean assuming that there is no effect of randomisation on outcome other than through the effect on the component of therapy being analysed. If this assumption is applied and randomisation used as the instrument for the post randomisation variables the magnitudes of effect of the process variables are smaller as are the standard errors giving strongly significant results. In practice this would state that participants who attended any number of therapy sessions but did not receive a particular part of the therapy, for example homework, or saw no change in beliefs such as the power of the voice would incur no benefit from those sessions. It is not possible to test this assumption but the validity of it must be considered based on knowledge and experience. It was decided that this was a very strong assumption that could not be defended. The results would give standard errors reflecting over confidence in the estimates; as such randomisation was rejected as an instrument and other alternatives needed to be identified.

Work by Small[106] demonstrates that randomisation by covariate interactions are valid instruments if certain assumptions are made, one being that the interactions explain some part of the mediator or process variable. The interactions are generally weaker instruments than randomisation and in the real data the use of individual instruments resulted in large standard errors and the results were found to be affected by the choice of instrument. This highlighted issues with weak instruments; if an instrument is only weakly associated with the mediator or process variable that it is trying to estimate then the estimates will be

biased. The choice of instruments is a key part of instrumental variables analysis and a balance must be achieved between improving the fit of the first stage regression model and minimising the number of instruments used. Introducing more instruments can improve the amount of variation in the mediator/process variable explained and reduce standard errors but additional instruments can also be a source of bias. A rule of thumb that a first stage F-statistic greater than ten is indicative of a good instrument and instruments with a smaller F-statistic are weak has become popular from work by Staiger and Stock[128] and Stock and Yogo[131]. The F-statistic is a measure of the fit of the instruments to the mediator/process variable which is penalised for having more variables included. The rule of thumb is based on the bias of the IV estimate being at most 10% of the bias of the OLS estimate. However, the IV estimate may still be preferred over the OLS estimate when the difference is not as extreme. The two-stage least squares method is unbiased when only one instrument is used but this is not a safe fall back if the instrument is very weak and the standard errors may be so large as to provide no useful information. Several authors[132,133,136,181] warn against selecting instruments by the F-statistic, in fact they do not advise selecting instrument post-hoc at all and state that instruments should be determined at the design stage. Unfortunately this is not always possible, or it may turn out that the instrument stated due to prior literature or theory is also weak. Work in the econometrics field has suggested that variable selection techniques from prediction modelling can effectively select instruments to reduce bias in instrumental variables analysis. Work by Belloni and colleagues[142] considered these methods in the situation where they had a large set of potential instruments relative to sample size and compared shrinkage methods to including all variables within an observational study setting. In this thesis their findings have been extended by applying the selection methods to simulated data which replicates a randomised controlled trial with post-randomisation process variables and unmeasured confounding. The LASSO selection method is compared to results when including all potential variables or selecting variables by the elastic net and stepwise methods at varying strengths of the instruments available.

The simulations in this study showed that the two-stage least squares method outperformed the ordinary least squares mediation model in terms of minimising the mean-squared error regardless of the strength of the instruments. In line with the rule of thumb[131] the amount of bias in the IV estimates increased relative to the OLS bias as the instruments became weaker. Selecting multiple variables as instruments reduced the bias in the direct effect of

randomisation compared to models using only one of the known explanatory variables as an instrument. Bias was comparable for the mediator but the mean-squared error was much worse when one instrument was used compared to multiple instruments. In terms of the selection methods all methods gave similar results when the level of unexplained variation in the post-randomisation variable was small. The most parsimonious LASSO models gave the best results in terms of bias but at the cost of precision and struggled to find any variables when the level of unmeasured variation was high. When the explanatory variables explain less of the variation in the process variable the LASSO with cross-validation to minimise mean-squared error for the first-stage regression was the most effective in terms of bias and precision using the two-stage least square estimator. However, it was shown that when instruments are weak including all potential interactions as instruments (true and noise) remains better than using the Baron and Kenny mediation model, though the bias is greater than ten percent of the Baron and Kenny model. This finding may be because within the potential variables are variables that are truly associated with the process. As a comparison we can assess the fit of the models if only variables that are not truly associated with the process variable are used as instruments when interacted with randomisation. These results were not shown in the main body of the text but have been provided in Appendix 1. As expected these gave the poorest results in terms of MSE, bias and the first-stage F-statistic though the 2SLS analysis with non-explanatory variables was still preferable to the OLS method adjusting for the same variables. This finding is likely to be due to the design of the simulation as some of the noise variables that did not contribute to the process variable were correlated with those that did to differing extents, therefore providing some weak association with the process variable. The simulation was designed in this way to replicate a randomised trial where baseline measures are likely to be correlated. It would be of interest to assess the impact on results of the failure of the over-riding assumption that there are instruments available that are associated with the process variable/mediator. In practice, selection techniques would be used to improve the first-stage estimation under the assumption that the set of variables contain some that are truly associated with the process-variable.

In terms of the F-statistic it is clear that the mean-squared error of the estimate does increase as the F-statistic decreases. However, the highest first-stage F-statistic does not always indicate the model with the lowest mean-squared error. This may be because the F-statistic comparison favours a more parsimonious model. This result agrees with that of

Burgess and Thompson[132] who also found that using all relevant variables simulated to be associated with the mediator in their models was better than using just one. The Staiger and Stock rule of thumb defines instruments as weak if the bias of the IV estimate is more than 10% of the bias of the OLS estimate. This is an arbitrary cut-off and although the instruments may be labelled as 'weak' the IV estimates are still less biased even with these weak instruments than using an OLS mediation model. Stock and Yogo give a range of cut-offs and so it is possible to assess the level of bias relative to the OLS estimator dependent on the fit of the first stage model. In terms of deciding if the analysis is appropriate, it is up to the analyst to determine whether less improvement with regard to bias compared to the standard OLS estimate is acceptable or if the 10% marker is adhered to.

The present study has investigated methods for the selection of instruments where there are several true and several bogus possibilities. The LASSO method strives for a balance between increasing model fit and decreasing the number of variables used. The R program for the LASSO applied in this study allowed a direct comparison between models selected using the same methods but different levels of the shrinkage factor to provide a different number of selected variables. Interestingly the methods that select the most parsimonious model are not always preferred as we would expect from the literature. The stepwise method which has been criticised for over-fitting performs well for this purpose, possibly suggesting that the addition of variables that provide even a small amount of information is not detrimental. This is especially clear in the case where the instruments are weak; in this situation the larger shrinkage factors that produce the most parsimonious models have difficulty finding instruments. Although the IV methods are still less biased than the OLS mediation model even with weak instruments the amount of bias increases in the face of weak instruments and the validity of the results may be questioned. The results indicate that using many instruments or even all of the potential variables although they are not associated with the endogenous variable may not be as detrimental to the bias of the effect estimates as expected. The mean-squared error and mean bias are comparable to those when selection methods are applied. The simulation results also show that when there are multiple instruments using only one variable that is known to be associated with the process variable is not as effective as using multiple associated instruments at any strength of the instruments to minimise the MSE. This situation can be compared to having one pre-

specified instrument and suggests that there are benefits to exploring potential additional instruments.

The selection of instruments does not necessarily supersede the use of a pre-defined instrument which has been built in to the design of the study. If a pre-specified instrument is found to be strongly associated with the process/mediator then this should be used but it may also be useful as a sensitivity analysis to select instruments from a valid pool of variables, choosing those that are strongly associated with the mediator. The selection techniques provide an informed method of choosing instruments and can increase confidence in selecting the correct instruments to minimise bias and increase precision when there are several possible variables to choose from, but the results of the analysis must still be questioned if they can change drastically with the application of a different instrument. Thought should be given to the measurement of potential instruments at the outset of a trial, but these methods can be a useful tool in post-hoc mediation analyses or when the hypothesised instrument is only weakly associated with the mediator. In the COMMAND trial, for example, the baseline measure of the mediator was measured and the interaction with randomisation group was expected to be a good instrument. It was shown, however, that the interaction was not strongly associated with the mediator. In this case, the total scale score rather than the subscale score provided a good instrument; this is likely to be because the subscale and the total scale are measuring a similar overall construct but the total scale is more reliable than the subscale.

### 6.1.2 Comparison of estimation methods

Several estimation methods are available for instrumental variables analysis; two-stage least-squares is the standard method used as the default in most packages. The LIML estimator has been shown in the literature to outperform the 2SLS in terms of the median bias in the presence of weak instruments. Additionally, the Fuller adjustment to the LIML estimator and the GMM are available in several software packages. In this study, the methods were applied to simulations designed to replicate an RCT with a post-randomisation process variable available only in the treatment arm; the 2SLS, LIML and Fuller estimators were compared. For the continuous process variable the LIML and Fuller adjusted LIML gave the best results in terms of the bias but the 2SLS was the preferred estimator to reduce variance and the overall mean-squared error. An estimator with only a small amount of bias may still give a biased result in a given sample if it has large variation. An estimator that is a little more biased but has smaller variation may actually

reduce the chance of producing a biased estimate in a sample. The estimator that gives a balance of bias and precision by minimising the mean-squared error is therefore preferred. When a categorical process variable is analysed the 2SLS outperforms the LIML and Fuller estimators in terms of bias and precision at all strengths of instruments.

The 2SLS has benefits over the other estimation methods in addition to producing a lower mean-squared error. Other than the availability of software packages to conduct the analysis 2SLS allows for a simple OLS calculation of the estimate in two separate stages. In the analysis of real data sets this means that each stage can use as much information as is available. In terms of a mediation analysis in a trial where the mediator is measured at some time prior to the final outcome, participants drop out over time and the number available to analyse will change. Taking the analysis in two stages means that information can be used from all participants who have information regarding the mediator/process even if they have not completed the final outcome. The first stage estimates will, therefore, be better informed than if only complete cases are included. Analysis of data by the different estimation methods will help to convince us of the accuracy of our results if we obtain similar estimates and can draw the same conclusions under the different estimation procedures. This is a practice advocated by Angrist and Pishke[136] and would serve as a sensitivity analysis. The results are not reported in the main body of text for the different estimation methods but are given in Appendix 2 and 3. In summary the conclusions of the analyses do not change when the 2SLS, LIML or Fuller methods are used. When the LIML and Fuller estimators are applied to the analysis of the post-randomisation process variables the treatment effects tend to be larger than under 2SLS estimation and the standard errors of the estimates are also larger. There is greater uncertainty in the LIML and Fuller estimates giving more confidence in the 2SLS results.

Instrumental variables analysis can be a very effective tool but is not a solution to all problems. As with any analysis, it is important to consider the validity of the assumptions being made and to determine the sensitivity of results to changes in the analysis. The standard OLS mediation model makes assumptions that were deemed invalid in the context of the trials investigated here and are replaced with different assumptions necessary for the IV model. The validity of these new assumptions must be questioned and it may be that a lack of true or strong instruments means that the method is not of use.

## 6.2 Discussion from the substantive questions

The statistical methods investigated through this piece of work were applied to two RCTs of cognitive behavioural therapy for the prevention and treatment of psychosis. The specific substantive aims that the application of these methods hoped to address were to:

1. Estimate the effect of dose of therapy on symptom reduction in those at high risk of psychosis
2. Estimate the effectiveness of including the following aspects of therapy on reduction of symptoms in those at high risk of psychosis; agreement of problems and goals, formulation, homework, active change strategies
3. Determine if the effectiveness of CBT for high-risk individuals is mediated by changes in beliefs
4. Determine if there is a causal pathway from the amount and content of therapy received through changes in beliefs to a reduction in symptom severity in those at high risk
5. Determine if the effectiveness of CBT on compliance with command hallucinations is mediated by changes in appraisals of the power of voice

### 6.2.1 Overall findings

The aims of the applied aspect of the project are defined above as separate questions to determine from the data, however to achieve a full understanding of the results it is important to consider the questions of the same trial as one.

#### *6.2.1.1 EDIE-II trial findings*

The instrumental variable analysis was first applied to investigate the dose-response effect of attending therapy and found a significant improvement in symptom severity as more sessions were attended. This is expected as there was a significant intention-to-treat effect of randomisation on severity. The instrumental variable analysis when no direct effect of randomisation is allowed forces the effect seen in the intention-to-treat analysis through the number of sessions. When randomisation is used as the instrument the treatment effect per session becomes the overall treatment effect divided by the average number of sessions. Non-linear associations were investigated in order to determine if there was an optimal range of sessions that should be attended to achieve some benefit. It was found that attending at least four sessions improved symptoms at 12 months compared to attending less than four sessions but no plateauing effect was found.

The next step was to investigate the impact of adherence to therapy protocol by determining if the involvement of specific components of therapy improved outcome. The analysis indicated that formulation improved severity of symptoms and there was a suggestion that homework and active change strategies also led to an improvement but there was no significant improvement if an agreement regarding problems and goals had been achieved.

In the initial analyses no restriction was placed on the direct treatment effect as it was not believed to be valid to assume that a participant would experience no benefit of treatment without the specific component of therapy in question. The results however indicated that the direct effect of treatment may be zero and there was great uncertainty in the estimates characterised by wide confidence intervals. The effects of the individual process variables were therefore re-calculated with the assumption of no direct effect applied. This amounts to applying the exclusion restriction to randomisation status and means that this can now be used as the instrument. The estimates of the effect became more certain and the significance much stronger. A comparison of these estimates with and without the direct treatment effect is provided in Table 6.1, the estimates are just for comparison of the process variable effect and so the direct effect of randomisation is not provided. The magnitude of the estimates are larger when the direct effect of randomisation is included, this is probably because, although non-significant, randomisation without the inclusion of the components of therapy is estimated to have a negative effect by increasing symptoms. The standard errors of the effect estimates are much larger when the direct effect of randomisation is allowed. This is because randomisation is no longer an instrument and the covariate by randomisation arm interactions are used as instruments instead. The interactions are weaker instruments giving larger standard errors. Even though the effect estimates with the exclusion restriction applied are smaller in magnitude the overall conclusions drawn are similar with regard to all of the components and indicate that the inclusion of each improves symptoms at 12 months. The direct effect was not statistically significant in any of the analyses but it would be a strong assumption to not allow it and would result in overly confident conclusions.

**Table 6.1: Comparison of estimates with and without the inclusion of a direct effect of randomisation (only the effect of process variable shown), imputed data**

| | | Coefficient | Std. Error | 95% CI | | p-value |
|---|---|---|---|---|---|---|
| Problem Agreement | No direct effect | -10.02 | 3.29 | -16.47 | -3.57 | 0.002 |
| | Direct effect allowed | -12.16 | 8.80 | -29.40 | 5.08 | 0.167 |
| Formulation | No direct effect | -11.49 | 3.90 | -19.13 | -3.85 | 0.003 |
| | Direct effect allowed | -23.21 | 9.83 | -42.48 | -3.94 | 0.018 |
| % sessions with homework | No direct effect | -0.17 | 0.06 | -0.28 | -0.06 | 0.003 |
| | Direct effect allowed | -0.25 | 0.12 | -0.49 | -0.01 | 0.041 |
| >50% sessions with homework | No direct effect | -14.15 | 4.76 | -23.47 | -4.83 | 0.003 |
| | Direct effect allowed | -19.20 | 10.80 | -40.37 | 1.96 | 0.076 |
| % sessions with change strategies | No direct effect | -0.15 | 0.05 | -0.24 | -0.05 | 0.003 |
| | Direct effect allowed | -0.26 | 0.11 | -0.48 | -0.04 | 0.021 |
| >50% sessions with change strategies | No direct effect | -12.70 | 4.34 | -21.20 | -4.19 | 0.003 |
| | Direct effect allowed | -21.33 | 10.11 | -41.15 | -1.52 | 0.038 |

Since attending more therapy improves outcomes and receiving specific aspects of therapy improve outcomes we sought to incorporate both together. An interaction between these two aspects of therapy was hypothesised meaning that attending more therapy that includes the pre-specified components would be even better. When both attendance and attendance by component interaction were included neither were significantly associated with symptom severity. Analysis of attendance within those that would receive the component of therapy and those that would not indicates that although attendance may improve symptoms within these groups the effect of attendance on outcome is not significantly different between the groups.

This finding highlighted a problem with the individual components of therapy that they are all highly correlated with each other. Attending more therapy sessions increases the likelihood of receiving these particular aspects of therapy and participants that receive one aspect of therapy are likely to receive other aspects as well. This is exemplified by the similarities in certain aspects of the components considered, for example normalising which is considered as an active change strategy shares aspects of formulation which has been considered as a separate component in its own right. It is not possible to untangle the effects of attendance and the different components of therapy since they are so closely related.

Extending this analysis to a composite measure of the number of components included was attempted but unfortunately it proved to be problematic. The summary measure could not be considered as a continuous variable and as a categorical measure the instrumental variables analysis broke down. A categorical mediator must be considered as multiple binary variables and a separate instrument found for each one to identify the model. Similar characteristics are found to be associated with each of the process variables and so finding four distinct instruments that are strongly associated with the processes is difficult. The confidence intervals around the effect estimates of a summary measure are so wide as to not be informative. It is possible that in addition to multiple instruments a larger sample would allow this analysis to be conducted. Instead two and three category summary measures were analysed to quantify the impact of receiving all components of therapy to some or no components. The results showed that receiving all four components of therapy significantly improved symptoms at twelve months compared to not receiving any components. There was no significant improvement in symptoms when only some of the components of therapy were experienced compared to none. This provided an analysis of the composite measure indicating the impact of adherence to the therapeutic model and highlights the importance of treatment fidelity.

A summary measure indicating the degree to which the therapy received met the protocol does not answer the question regarding the importance of specific aspects of therapy and does not seek to identify the active ingredient. The strongest treatment effect was seen for formulation and the effects of homework and active change strategies, though suggesting an effect, were not significantly associated with an improvement. This would suggest that formulation is the driving force in the cognitive therapy provided in the EDIE-II trial. However, due to correlation between the components of therapy received it seems unlikely that we are really seeing the effect attributable to this one component. This is supported by the similarity in the magnitude of the effect estimates for each component as well as for the combined none/some versus all components and none versus some versus all component analyses. In order to determine the effect of particular parts of therapy or dose of therapy these would have to be incorporated into the design of the trial, for example, randomisation to receive therapy with or without homework or randomisation to different number of therapy sessions.

Repeating the analyses on the hypothesised mediators within the EDIE-II trial provided no significant results. We found no evidence that CBT reduces symptom severity by changing

patient's beliefs about themselves or their illness immediately after the therapy window. It is possible that CBT changes the patient's beliefs at a later time point though this was not analysed since it did not provide information for a mediation analysis. The later time points at which the mediators were measured coincided with the outcome time point removing the temporal order of events which is needed for causal inference.

The treatment effect estimates seem large, estimating reduction in severity of symptoms of between 10 and 20 points on a scale that could range from 0-144 and in this sample ranges from 0-90. It is important to consider the clinical significance of this improvement in order to really gauge the benefits of CBT therapy. It is also important to note that the confidence intervals surrounding the effect estimates are very wide even when the effect is found to be significantly different from zero. This is a problem with the instrumental variable technique in that by using predicted rather than observed values more uncertainty is introduced into the estimation process.

#### *6.2.1.2 COMMAND trial findings*

In the COMMAND trial the cognitive therapy was designed to change the voice power differential between the voice and the voice hearer and in so doing reduce compliance with the voice. The results found a significant direct effect of CBT on the mediator, power of voice, and on compliance as hypothesised by the therapeutic model. When power of voice was analysed as a mediator in an instrumental variables analysis with a direct effect of randomisation group the only significant mediating effect was found when all instruments were used in the analysis, but the first stage F-statistic in this model was very small at only 1.6. The use of VPD total score in an interaction with group as the instrument gives an F-statistic attributable to the instrument of 11.7 indicating a good instrument and higher than the F-statistic with other instruments selected. It is also more likely that the VPD total score only affects the outcome through the effect on the mediator which is more questionable for other baseline characteristics.

Applying the exclusion restriction to randomisation meant that it could be used as an instrument. This resulted in a significant mediating effect of power of voice. This is as expected since all of the ITT effect is directed through this mediator. In practical terms implementing the exclusion restriction on randomisation would state that randomisation to CTCH only effects compliance with commands through its influence on the voice power differential. It may be that this is true since that is the aim of the therapy but it is very

unlikely that the therapy has no effect on any other aspect of beliefs or clinical characteristics. The finding suggests that either the power of voice does not work in the expected way to reduce compliance or that it may work through another process which is nullifying the effect on compliance.

The analysis presented here uses only the intermediary measure of the hypothesised mediator taken at the end of the therapy window and the outcome measure at the end of the trial rather than the repeated measures of both which are available. A full analysis would incorporate both the hypothesised mediator and outcome at both the nine and eighteen month time points. This analysis is complicated in that, not only are the repeated measures of each variable not independent but that the measures are likely to be related to each other. Power of voice at nine months will be correlated with compliance at nine months and is hypothesised to effect compliance at eighteen months and power of voice at eighteen months. However, compliance at nine months may also influence power of voice at eighteen months which is also associated with compliance at eighteen months; an example of time-varying confounding.

A general assumption of both the IV and OLS methods is that the individuals are independent of each other. Although the treatment tested is individual rather than group therapy it is possible that therapist effects will be seen with similarities in characteristics of patients and treatment effects produced in patients treated by the same therapist. The specific effect of therapist was not investigated in the analysis of the EDIE-II or COMMAND trials but assumed to be accounted for by location. In the EDIE-II trial each study site had one main therapist that treated most of the patients in the study; in two of the five sites this was the only therapist. Where additional therapists were used the number of patients they saw ranged from 2-9 compared with a range of 14-35 clients for the main therapists. Since most patients are seen by a lead therapist in each site and the number of patients per additional therapist is low it would seem acceptable to adjust for site in the analysis as a proxy for therapist. Therapist information was not available for the COMMAND data analysis and so it is assumed that an adjustment for centre effects is sufficient.

#### *6.2.1.3 Comparison of findings with the literature*

The motivation of the EDIE-II analysis comes from the heterogeneity in the treatment received by participants. There has been very little experimental work investigating the

active ingredients of CBT therapy. Most has concentrated on the impact of homework though this has generally been within anxiety and depression rather than psychosis[84]. The findings generally support the importance of homework but do not provide conclusive evidence of causative links. Within the population with psychosis, investigations have been limited to looking at associations within those receiving therapy, the results of which are mixed; however, the samples are small and they have not accounted for confounders that are associated with both complying with homework and outcome[85,86]. The current studies of homework have considered amount of homework given during therapy and compliance with homework. In this study the amount of homework given has been quantified by the proportion of sessions in which homework was reviewed. This is not a measure of compliance but accounts not only for homework being given but that follow-up has taken place at the proceeding session. In this way the effect is of the application of homework as part of therapy rather than the effect of a patient carrying out the homework tasks.

Only one study has specifically considered the impact of formulation but the study was flawed in that it again had no control group and a very small sample size and no association was found. No studies have specifically considered problem agreement or active change strategies as mechanisms of the treatment process. It may be that problem agreement is such an inherent part of therapy that the role of this as an active ingredient has not been considered. It is surprising that a significant effect of problem agreement is not found in the results since it is the first stage of cognitive therapy. It may be that an agreement though necessary as a foundation of the following therapeutic process is not an active ingredient by itself.

It has been shown by Dunn et al[79] that receiving full therapy improves outcome but having only partial therapy may be detrimental. The EDIE-II results do not indicate a detrimental effect of receiving only partial therapy but show no significant effect of therapy when only some components of therapy are received (some therapy vs no therapy on severity of symptoms =-3.7, -13.8 to 6.4, p=0.473) compared to a large effect when full therapy is received (some therapy vs no therapy on severity of symptoms =-23.2, -40.1 to -6.2, p=0.008). The reasons for non-adherence to protocol are therefore of importance; is CBT therapy only suitable for certain types of people, are there treatment effect moderators to consider when deciding if patients should be prescribed CBT? The study of patient-therapist pairs considered patient suitability to therapy as a predictor of outcome[86]. The assessment of suitability incorporated perceptions of whether the patient had accepted their

own responsibility for change, their potential for alliance and how well they suited the CBT rationale. The authors did not find a significant relationship between suitability and outcome when alliance and homework compliance were accounted for, however these three predictors were shown to be correlated with each other and so may be explaining the same variation in outcome. Rector[83] summarises evidence of barriers to homework therapy as consisting of factors associated with negative symptoms, namely a lack of motivation, inability to make decisions, social withdrawal, lack of initiative and distractibility. The author suggests that since these barriers are also symptoms of the illness it is possible to target them in the course of therapy thereby improving homework compliance and general symptoms.

There has been little research into CBT for command hallucinations with only three RCTs reported including the COMMAND trial analysed in this thesis. The precursor to the COMMAND trial[55] found that as well as being effective in reducing compliance to commands CTCH also reduced the perceived power of the voice, the belief in the voices omnipotence and increased patients control over the voice. This supports the finding of the ITT effect of CTCH on power of the voice. The authors also found that adjustment for power of voice removed the direct effect of treatment but they did not report the effect of power of voice in this analysis which is the equivalent of the Baron and Kenny OLS mediation. The finding that CTCH also has an impact on beliefs about the voices omnipotence and the control that the patient has over the voice suggests that the therapy may work through other mediators and supports the hypothesis that there may be an effect of randomisation that is not through the power of the voice. The TORCH study[67] that compares an acceptance based CBT to befriending found no difference in the participants confidence to resist obeying harmful commands, confidence in coping with commands or with measures of symptoms, distress or quality of life. The authors found that the rate of compliance with harmful command hallucinations at baseline was too low in their sample to consider this as an outcome and did not measure the voice power differential. This study was small with only 44 randomised at the start of the trial which may have impacted on the lack of significant findings.

### 6.3 Strengths of the present study

This thesis sought to improve the instrumental variables estimation applied in real datasets by considering the methods used to choose instruments and the estimation methods applied in the analysis. To this end simulation studies were designed and implemented to imitate

the set-up of a randomised controlled trial with a treatment effect working through a post-randomisation process variable. Several methods of instrument selection were compared thoroughly for different numbers and strengths of true instrument. The simulations were designed to keep all factors the same other than the aspect being tested, so for example when the amount of variation in the process variable explained by the instruments (strength of the instrument) varied the total variation in the process variable and the level of unmeasured confounding was held constant. The simulations tested the selection methods for continuous and categorical process variables and at different sample sizes. The samples sizes chosen were 200 and 400 to be representative of a randomised trial. Comparison of instrumental variable estimation methods was conducted in several different contexts: when all true explanatory and noise variables were used; when only relevant variables were used and when those chosen by selection methods were used as instruments. The selection and estimation methods were both assessed by the bias, variance and mean-squared error of the final effect estimates so a judgement could be made based on both important criteria.

Previous studies have shown the effectiveness of cognitive therapy in the treatment of various mental illnesses including psychosis but there is little evidence for this form of treatment in the prevention of psychosis. Although there are theoretical processes to explain how therapy is effective very few studies attempt to test these theories empirically. Secondary analyses tend to be limited to investigating baseline predictors of outcome[68,72,73,75-78]. The EDIE-II trial of patients at-risk of psychosis allowed us to test the effectiveness of CBT and also the mechanism by which it works. Detailed data had been collected on the content of each therapy session for the individuals receiving CBT, this level of detail allowed for a mediation type analysis to investigate the mechanisms by which CBT is effective. The content of the therapy sessions was recorded by the therapist in the patient notes as standard practice at the end of each session. The notes were then reviewed by therapists involved in the trial who looked for evidence of particular parts of therapy being present. The data was therefore not subject to recall bias or social desirability bias. There may have been some under-reporting of therapy components taking place but this would only serve to give a conservative estimate of the effects. The way in which CBT works has also been assumed in theoretical models of the therapy but not tested experimentally. The COMMAND study and the EDIE-II study allowed for a full mediation analysis to determine if the treatment works through the expected pathways.

This builds on current literature in the treatment of command hallucinations that has so far been limited to investigating the power of the voice purely as a prognostic factor.

In the analysis of the treatment effect pathways we have acknowledged and assessed the assumptions made in the statistical models used. Applying an instrumental variables approach to determine the mechanisms of CBT has allowed a relaxation of the assumption of no unmeasured confounding between the process variable/mediator and outcome. There is disparity in the treatment estimates under OLS and IV. If it is accepted that the assumptions of the OLS model are invalid and those of the IV model are valid then the results would suggest that the IV methods have improved upon the OLS model providing less biased estimates of the treatment effect. The statistical component of the study has given confidence that instruments selected will provide estimates of the mediator/process variable effect that improve on those given using the Baron and Kenny mediation model.

### 6.4 Limitations of the present study

The instrumental variables methods have been used to relax the assumption of unmeasured confounding necessary for unbiased estimates in an OLS mediation analysis. However there are other assumptions implied in the IV analysis, the validity of which must be considered. Specifically, it is assumed that the instrument is only associated with the outcome through its effect on the mediator/post-randomisation process variable, which is not testable. When considering attendance at therapy sessions, randomisation is used as the instrument and it seems valid to assume that the only effect of randomisation on outcome is through the participants actually attending the therapy and therefore receiving the intervention. When post-randomisation process variables are included this assumption is questionable and so a direct effect of randomisation is permitted. In this instance interactions of baseline covariates with randomisation arm are used as instruments. The exclusion restriction now applies to the interaction and states that the interaction is only associated with the outcome through the influence on the mediator/process variable. Any difference in covariate effect on the outcome between the treatment groups is assumed to be entirely due to the difference in effects on the mediator/process variable. If the instrument (interaction) is associated with other mediators/processes then it may not be valid. This could be a problem in our real data scenario as the strongest predictors of the post-randomisation process variables tend to be associated with more than one process. If the interaction of group and anxiety is associated with both formulation and homework then when considering just one of these processes the instrument may have an influence on

outcome that is not through the single process being analysed. The process variables that are considered in this analysis are all highly correlated with each other and so it is unsurprising that covariates associated with the presence of one process are also associated with the presence of other processes in a subject's treatment. It may be however, that because the processes are all related the instruments are still valid in that they affect the outcome only through the process but that the process, rather than giving us specific information regarding a particular active ingredient is instead just a general measure of compliance or treatment fidelity. This hypothesis is supported by the similarity in effect estimates for each of the components.

In this situation where there are multiple parallel processes and the statistical methods are unable to decipher the effect due to each specific process the assumptions of the instrumental variables approach may be questionable. The composite measures of the components may give more validity to the assumptions required of the instruments since it is unlikely that the instruments will affect the outcome through other processes that have not been included. In this analysis a direct effect of treatment was not allowed in those that would not receive any of the components of therapy stated. This again may not be a valid assumption as there is potential that participants may attend therapy without receiving any components and still experience some effect, however it is a more convincing assumption when applied to the composite measure than individual components.

The validity of the exclusion restriction is equally as important in the analysis of mediators in both studies. The COMMAND trial benefitted from having baseline measures of the hypothesised mediator and the interaction of the baseline measure with randomisation group was used as an instrument. In this situation the assumption that any effect of the baseline measure interaction on outcome works solely through the effect on the same measure at an intermediary time point seems plausible. Unfortunately in the EDIE-II trial the hypothesised mediators were not recorded at baseline and so other instruments had to be found. These instruments were interactions of baseline covariates with randomisation and could potentially suffer from the problems described above relating to instruments for the process variables. In this case the hypothesised belief mediators were shown not be effected by the treatment received and so a mediation analysis was not carried out.

An assumption of the IV analysis states that the mediator/process effect is consistent and does not increase/decrease with the value of the mediator/process. In terms of the post-

randomisation process variables, relaxation of the interaction of process and process effect would indicate for example that the improvement in outcome with the involvement of homework in therapy sessions does not increase linearly as more sessions involve homework but that there is a multiplicative effect. This does not seem to be warranted and the assumption appears to be fair. The sensitivity to violations of these assumptions has not been assessed here but may be something for future investigations.

Treatment effect moderators are an area of current research interest as they inform the use of stratified medicine. It is expected that in future treatment will be personalised and targeted to specific groups in whom it will be most effective. So far research has focussed on overall prognostic factors rather than specifically testing if treatment effects differ by subject characteristics. It has been suggested that gender, insight, severity of illness, duration of illness, and length of untreated illness are all associated with outcome in patients receiving CBT for psychosis[68,72,73,75-78], although so far these have only been identified as general prognostic factors it may be that they act as treatment effect modifiers though this is yet to be examined.

The simulations in this study have attempted to imitate conditions similar to a real trial scenario however the continuous variables simulated are all standard normal variables and the associations between covariates, mediators and outcomes are all linear. The study does not involve non-normal variables, non-linear associations or missing data which are likely to lead to more complicated findings. The GMM estimator is expected to perform better than the 2SLS estimator in the presence of heteroskedasticity[115] but this was not applied in these simulations. If this had been modelled in simulations and the GMM applied it may have been preferred for estimation in our examples, however when GMM was used in the real data the results were very similar to the 2SLS suggesting that heteroskedasticity is not a problem in this data.

In the simulation studies the mediation model with unmeasured confounding was extended to include a process variable and a mediator effect. Unfortunately in the real data studies this analysis was not warranted as no significant treatment effects were found on the mediators. It is likely that, though the methods have been shown to be effective in simulated example, when applied to a real data set problems may arise.

Instrumental variable estimation has low power to detect an effect especially in small samples. Weak instruments increase uncertainty in the estimates even further and increase

the potential for bias. Larger sample sizes are needed for instrumental variables analysis and the problems associated with weak instruments reduce as the sample size increases, as illustrated in the simulations. The COMMAND trial recruited 197 participants and the EDIE-II recruited 288, by follow-up these numbers had fallen to 164 and 188. These are relatively small sample sizes for an instrumental variables analysis and may contribute to the large standard errors and non-significant results.

In addition to the problems of sample size the analysis of these studies suffers from limitations associated with many RCTs and observational studies, namely missing covariate data and drop-out. In both studies the planned analysis to deal with bias due to drop out was to include any baseline variables associated with missingness and in both cases there were none so no adjustment was made. The analyses assume that missing outcome data are missing at random, although none of the measured variables were associated with missing drop-out it is possible that some unmeasured mechanism is associated with missing follow-up which could cause bias. The validity of the assumption cannot be assessed statistically but must be considered in terms of the study design. If for example the individuals that did not attend the follow-up interview are the ones that are more ill and could not attend, or alternatively are the ones that are better and so did not feel the need to return then the results will be biased[154]. We do not know if either of these situations are the case but since none of the baseline clinical measures were associated with missing outcome it suggests that missing at random may be a plausible assumption. Missing baseline covariates was more of an issue in the EDIE data rather than the COMMAND data and so multiple imputation was carried out to fill in the gaps and make the most of the information available. Neither study specifically measured hypothesised instruments for a mediation analysis. The COMMAND study did measure the mediator at baseline which is useful for an instrument but in the EDIE-II trial the mediators were first measured at one month rather than baseline so could not be used as instruments.

Measurement error is another problem common to many studies. This has not been tackled explicitly here but the IV methods, in addition to accounting for bias due to unmeasured confounding also allows for bias due to random measurement error. IV does not however account for systematic measurement error. Psychometric scales are not objective measurements though the scales used have been shown to have good properties of internal validity and consistency. The treatment fidelity measures were recorded from therapist notes after the trial and the therapists were not aware that this was to be carried out prior to

the trial. The notes therefore may not have been as comprehensive as they could have been and some evidence may be lost but alternatively the treatment given and notes made would not have been influenced by the knowledge that they would be checked. It is assumed in these analyses that any measurement error in the mediators and process variables are randomly occurring. Vansteelandt et al.[182] investigate the use of instrumental variables and prior information to correct for systematic measurement error. They find that using an instrument for measurement error is not very efficient and there is only a real benefit if the measurement error is large and there is a strong instrument associated with the error. It may be useful in studies where there is some knowledge of the error; the authors give the example of a drug trial where a placebo is given for a run in period and adherence is measured which can be applied in the full trial. The use of prior knowledge in a Bayesian analysis may be of more use and the authors conclude that applying some prior information to nuisance parameters such as measurement error may improve efficiency of estimation for the parameter of interest.

### 6.5 Future work

Instrumental variables analysis is subject to many assumptions and its effectiveness is dependent on the strength of the instruments. The sensitivity to the assumptions of the exclusion restriction may be of interest as this is an assumption that cannot be tested in the data. In the analyses of mediation and process variables conducted in this study a direct effect of randomisation was allowed though statistically it was not shown to be significant. Allowing the direct effect to take any value introduces a large amount of uncertainty as has been shown by comparing the standard errors of the models with and without it. If there is some idea of the magnitude it may be warranted to apply Bayesian methods to limit the effect. A prior distribution could be applied to the direct effect; the method would effectively be providing an average of the results with no direct effect and a freely calculated direct effect. However, the decision as to what this is and the distribution that it takes will alter the conclusions of the analysis and so should be supported with strong evidence. The application of prior knowledge mentioned earlier in relation to measurement error can be considered in this situation regarding the direct effect of treatment, Vansteelandt et al[182] propose that any prior information can improve efficiency.

In this thesis instrument selection techniques were applied, but other authors advocate model averaging when there are many potential instruments, averaging the parameter estimates over all models with different instruments used. The models are weighted with

more weight given to the better models. Various methods are suggested to determine the weights, for example, the AIC or BIC which are measures of model fit, model averaging would weight the models to minimise these measures. Hansen[183] suggests the application of Mallow's criterion for weighting which again is a measure of model fit. When considering this as a method instead of model selection it implies that all of the variables included are valid instruments rather than selecting those that are valid, though Hansen suggests that the technique is similar to shrinkage methods as some weights will approach zero. Rather than averaging a host of IV models the method can be applied to average the IV model with the OLS model. However, if the OLS model is believed to be flawed then incorporating these results in to the estimate is questionable.

The work carried out in this thesis is a simplified scenario with one measure of the covariates, mediator/process and outcome however, many studies collect information on all of these parameters at multiple time points. The EDIE-II trial allows for analysis of changes in the treatment received and symptom outcome over time. The COMMAND data has measures of both the proposed mediator and outcome at two follow-up points. Extending the current analysis to including a longitudinal aspect of exposure or mediator changing over time with outcome would be an interesting development and may add power to the findings, but is not a trivial analysis. This type of longitudinal analysis where both the exposure and outcome change over time is complicated due to the multiple inter-related measures; the problems were highlighted earlier. Simple regression of the outcome on mediator at time 1 and 2 with adjustment for covariates will give biased effect estimates. These scenarios can be considered as examples of time-varying confounding. Several methods have been suggested for unbiased analysis of treatment effects in the presence of time-dependent confounding[153,184]. These methods assume that there is no unmeasured confounding and do not support causal inferences if this assumption is not valid. Causal inference for longitudinal analysis is the focus of future research projects.

As shown in the study results and described here, the application of statistical methods to investigate complex models in a post-hoc analysis has its limitations. To understand more fully the mechanism of action of a complex intervention it must be considered in the study design. Ideally this would involve randomisation to specific hypothesised mechanisms of the treatment, potentially to involve multiple randomised arms. In the example here this could involve randomising individuals to receive exactly the same therapeutic intervention with the only difference being that homework is involved or not, or formulation applied. If

this is not possible due to practical or ethical reasons and overall treatment trials must be used then larger studies and the measurement of hypothesised instruments should be incorporated in to the study design. Treatment effect moderators that have been indicated in the literature should be considered as instruments or some form of run-in period similar to placebo adherence in a drug trial which may act as a proxy for true adherence should be incorporated. However as we have seen even with a known pre-specified instrument if this is weak there may be benefits to exploring the possibility of additional instruments for estimation.

# 7 References


1. APA. *Diagnostic and Statistical Manual of Mental Disorders*. 4 ed: Arlington VA, USA: American Psychiatric Association, 2000.
2. WHO. The ICD-10 classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines. Geneva: World Health Organisation, 1992.
3. Jablensky A, Sartorius N, Ernberg G, et al. Schizophrenia: manifestations, incidence and course in different cultures A World Health Organization Ten-Country Study. *Psychological Medicine Monograph Supplement* 1992;Supplement:1-97.
4. McGrath J, Saha S, Welham J, El Saadi O, MacCauley C, Chant D. A systematic review of the incidence of schizophrenia: the distribution of rates and the influence of sex, urbanicity, migrant status and methodology. *BMC Med* 2004;2:13.
5. APA. *Diagnostic criteria from DSM-IV*. Washington, D.C.: American Psychiatric Association, 1994.
6. Saha S, Chant D, Welham J, McGrath J. A systematic review of the prevalence of schizophrenia. *PLoS Med* 2005;2(5):e141.
7. Statistics OfN, Health Statistics Quarterly: Comparison of area-based inequality measures and disease morbidity in England, 1994-1998, Summer, 2003,
8. Olin S-cS, Mednick SA. Risk Factors of Psychosis: Identifying Vulnerable Populations Premorbidly. *Schizophrenia Bulletin* 1996;22(2):223-240.
9. Semple DM, McIntosh AM, Lawrie SM. Cannabis as a risk factor for psychosis: systematic review. *Journal of Psychopharmacology* 2005;19(2):187-194.
10. Jones P, Rodgers B, Murray R, Marmot M. Child development risk factors for adult schizophrenia in the British 1946 birth cohort. *Lancet* 1994;344(8934):1398-402.
11. Isohanni I, Jarvelin MR, Nieminen P, et al. School performance as a predictor of psychiatric hospitalization in adult life. A 28-year follow-up in the Northern Finland 1966 Birth Cohort. *Psychol Med* 1998;28(4):967-74.
12. Lieberman JA, Fenton WS. Delayed Detection of Psychosis: Causes, Consequences, and Effect on Public Health. *Am J Psychiatry* 2000;157(11):1727-1730.
13. Yung AR, McGorry PD. The initial prodrome in psychosis: descriptive and qualitative aspects. *Aust N Z J Psychiatry* 1996;30(5):587-99.
14. Yung AR, McGorry PD. The prodromal phase of first-episode psychosis: past and current conceptualizations. *Schizophr Bull* 1996;22(2):353-70.
15. McGlashan TH. Duration of untreated psychosis in first-episode schizophrenia: marker or determinant of course? *Biol Psychiatry* 1999;46(7):899-907.
16. Luoma S, Hakko H, Ollinen T, Järvelin M-R, Lindeman S. Association between age at onset and clinical features of schizophrenia: The Northern Finland 1966 birth cohort study. *European Psychiatry* 2008;23(5):331-335.
17. McGorry PD, Yung AR, Phillips LJ. The "close-in" or ultra high-risk model: A safe and effective strategy for research and clinical intervention in prepsychotic mental disorder. *Schizophrenia Bulletin* 2003;29(4):771-790.
18. Yung AR, McGorry PD, McFarlane CA, Jackson HJ, Patton GC, Rakkar A. Monitoring and Care of Young People at Incipient Risk of Psychosis. *Schizophrenia Bulletin* 1996;22(2):283-303.
19. Yung AR, Yuen HP, McGorry PD, et al. Mapping the onset of psychosis: the Comprehensive Assessment of At-Risk Mental States. *Australian and New Zealand Journal of Psychiatry* 2005;39(11-12):964-971.
20. Miller TJ, McGlashan TH, Rosen JL, et al. Prospective diagnosis of the initial prodrome for schizophrenia based on the Structured Interview for Prodromal Syndromes: preliminary evidence of interrater reliability and predictive validity. *American Journal of Psychiatry* 2002;159(5):863-865.
21. Miller T, McGlashan T, Woods S, et al. Symptom Assessment in Schizophrenic Prodromal States. *Psychiatric Quarterly* 1999;70(4):273-287.

22. Ruhrmann S, Schultze-Lutter F, Klosterkötter J. Probably at-risk, but certainly ill -- Advocating the introduction of a psychosis spectrum disorder in DSM-V. *Schizophr Res* 2010;120(1-3):23-37.
23. Klosterkotter J, Hellmich M, Steinmeyer EM, Schultze-Lutter F. Diagnosing schizophrenia in the initial prodromal phase. *Arch Gen Psychiatry* 2001;58(2):158-64.
24. Haroun N, Dunn L, Haroun A, Cadenhead KS. Risk and protection in prodromal schizophrenia: ethical implications for clinical practice and future research. *Schizophr Bull* 2006;32(1):166-78.
25. Ruhrmann S, Schultze-Lutter F, Salokangas RK, et al. Prediction of psychosis in adolescents and young adults at high risk: results from the prospective European prediction of psychosis study. *Arch Gen Psychiatry* 2010;67(3):241-51.
26. Cannon TD, Cadenhead K, Cornblatt B, et al. Prediction of psychosis in youth at high clinical risk: a multisite longitudinal study in North America. *Arch Gen Psychiatry* 2008;65(1):28-37.
27. Morrison AP, Bentall RP, French P, et al. Randomised controlled trial of early detection and cognitive therapy for preventing transition to psychosis in high-risk individuals. Study design and interim analysis of transition rate and psychological risk factors. *Br J Psychiatry Suppl* 2002;43:s78-84.
28. Yung AR, Yuen HP, Berger G, et al. Declining Transition Rate in Ultra High Risk (Prodromal) Services: Dilution or Reduction of Risk? *Schizophrenia Bulletin* 2007;33(3):673-681.
29. Birchwood M, Trower P. The future of cognitive-behavioural therapy for psychosis: not a quasi-neuroleptic. *The British Journal of Psychiatry* 2006;188(2):107-108.
30. Pitt L, Kilbride M, Nothard S, Welford M, Morrison AP. Researching recovery from psychosis: a user-led project. *Psychiatric Bulletin* 2007;31(2):55-60.
31. Wood L, Price J, Morrison A, Haddock G. Conceptualisation of recovery from psychosis: a service-user perspective. *The Psychiatrist* 2010;34(11):465-470.
32. Yung AR, Phillips LJ, Yuen HP, McGorry PD. Risk factors for psychosis in an ultra high-risk group: psychopathology and clinical features. *Schizophr Res* 2004;67(2-3):131-42.
33. Morrison AP, Renton JC. Cognitive therapy for auditory hallucinations: A theory-based approach. *Cognitive and Behavioral Practice* 2001;8(2):147-160.
34. Shawyer F, Mackinnon A, Farhall J, Trauer T, Copolov D. Command Hallucinations and Violence: Implications for Detention and Treatment. *Psychiatry, Psychology and Law* 2003;10(1):97-107.
35. Braham LG, Trower P, Birchwood M. Acting on command hallucinations and dangerous behavior: A critique of the major findings in the last decade. *Clinical Psychology Review* 2004;24(5):513-528.
36. Chadwick P, Birchwood M. The omnipotence of voices. A cognitive approach to auditory hallucinations. *The British Journal of Psychiatry* 1994;164(2):190-201.
37. Junginger J. Command hallucinations and the prediction of dangerousness. *Pyschiatric Services* 1995;46(9):911-914.
38. Barrowcliff AL, Haddock G. The relationship between command hallucinations and factors of compliance: A critical review of the literature. *Journal of Forensic Psychiatry & Psychology* 2006;17(2):266-298.
39. Gilbert P. *Depression : the evolution of powerlessness / Paul Gilbert*. Hove: Hove : Lawrence Erlbaum, 1992.
40. Singer AR, Addington DE. The Application of Cognitive Therapy for Command Hallucinations. *Cognitive and Behavioral Practice* 2009;16(1):73-83.
41. Bentall RP, Morrison AP. More harm than good: the case against using antipsychotic drugs to prevent severe mental illness. *Journal of Mental Health* 2002;11(4):351-356.
42. Helmchen H, Sartorius N, Klosterkotter J, Schultze-Lutter F. Prevention and Early Treatment. *Ethics in Psychiatry*: Springer Netherlands:235-262.

43. Corcoran C, Malaspina D, Hercher L. Prodromal interventions for schizophrenia vulnerability: the risks of being "at risk". *Schizophrenia Research* 2005;73(2-3):173-184.
44. Beck AT. The Current State of Cognitive Therapy: A 40-Year Retrospective. *Arch Gen Psychiatry* 2005;62(9):953-959.
45. Garety PA, Kuipers E, Fowler D, Freeman D, Bebbington PE. A cognitive model of the positive symptoms of psychosis. *Psychol Med* 2001;31(2):189-95.
46. Morrison AP. THE INTERPRETATION OF INTRUSIONS IN PSYCHOSIS: AN INTEGRATIVE COGNITIVE APPROACH TO HALLUCINATIONS AND DELUSIONS. *Behavioural and Cognitive Psychotherapy* 2001;29(03):257-276.
47. Cognitive behavioural therapy for the management of common mental health problems. Commissioning Guide, 2008, National Institute of Health and Care Excellence
48. National Institute of Health and Care Excellence N, Schizophrenia. Core interventions in the treatment and management of schizophrenia in adults in primary and secondary care NICE clinical guideline 82, 2009, National Institute of Health and Care Excellence, NICE
49. Morrison AP, Barratt S. What Are the Components of CBT for Psychosis? A Delphi Study. *Schizophrenia Bulletin* 2010;36(1):136-142.
50. Morrison AP, French P, Walford L, et al. Cognitive therapy for the prevention of psychosis in people at ultra-high risk: randomised controlled trial. *Br J Psychiatry* 2004;185:291-7.
51. Addington J, Epstein I, Liu L, French P, Boydell KM, Zipursky RB. A randomized controlled trial of cognitive behavioral therapy for individuals at clinical high risk of psychosis. *Schizophrenia Research* 2011;125(1):54-61.
52. van der Gaag M, Nieman DH, Rietdijk J, et al. Cognitive Behavioral Therapy for Subjects at Ultrahigh Risk for Developing Psychosis: A Randomized Controlled Clinical Trial. *Schizophrenia Bulletin* 2012;38(6):1180-1188.
53. Morrison AP, Stewart SLK, French P, et al. Early detection and intervention evaluation for people at high-risk of psychosis-2 (EDIE-2): trial rationale, design and baseline characteristics. *Early Intervention in Psychiatry* 2011;5(1):24-32.
54. French P, Morrison AP. *Early Detection and Cognitive Therapy for People at High Risk of Developing Psychosis: A Treatment Approach*: Wiley, 2004.
55. Trower P, Birchwood M, Meaden A, Byrne S, Nelson A, Ross K. Cognitive therapy for command hallucinations: randomised controlled trial. *The British Journal of Psychiatry* 2004;184(4):312-320.
56. Byrne S. *A casebook of cognitive behaviour therapy for command hallucinations : a social rank theory approach / Sarah Byrne ... [et al.]*. London: London : Routledge, 2006.
57. Rector NA, Beck AT. Cognitive behavioral therapy for schizophrenia: an empirical review. *J Nerv Ment Dis* 2001;189(5):278-87.
58. Cohen J. *Statistical Power Analysis for the Behavioural Sciences*. 2nd Edition ed: Psychology Press, 1988.
59. Wykes T, Steel C, Everitt B, Tarrier N. Cognitive behavior therapy for schizophrenia: effect sizes, clinical models, and methodological rigor. *Schizophr Bull* 2008;34(3):523-37.
60. Morrison A. Cognitive behaviour therapy for first episode psychosis: Good for nothing or fit for purpose? *Psychosis* 2009;1:103-112.
61. Stafford MR, Jackson H, Mayo-Wilson E, Morrison AP, Kendall T. Early interventions to prevent psychosis: systematic review and meta-analysis. *BMJ: British Medical Journal* 2013;346:f185.
62. Hutton P, Taylor PJ. Cognitive behavioural therapy for psychosis prevention: a systematic review and meta-analysis. *Psychological Medicine* 2013;FirstView:1-20.
63. Bechdolf A, Wagner M, Ruhrmann S, et al. Preventing progression to first-episode psychosis in early initial prodromal states. *The British Journal of Psychiatry* 2012;200(1):22-29.
64. Morrison AP, French P, Stewart SLK, et al. Early detection and intervention evaluation for people at risk of psychosis: multisite randomised controlled trial. *BMJ* 2012;344.

65. Birchwood M, Peters E, Tarrier N, et al. A multi-centre, randomised controlled trial of cognitive therapy to prevent harmful compliance with command hallucinations. *BMC Psychiatry* 2011;11(1):155.
66. Birchwood M, Michail M, Meaden A, et al. The MRC COMMAND trial: results of a multi-centre, randomised controlled trial of a cognitive therapy to prevent harmful compliance with command hallucinations. Submitted.
67. Shawyer F, Farhall J, Mackinnon A, et al. A randomised controlled trial of acceptance-based cognitive behavioural therapy for command hallucinations in psychotic disorders. *Behaviour Research and Therapy* 2012;50(2):110-121.
68. Garety P, Fowler D, Kuipers E, et al. London-East Anglia randomised controlled trial of cognitive-behavioural therapy for psychosis. II: Predictors of outcome. *Br J Psychiatry* 1997;171:420-6.
69. Lewis S, Tarrier N, Haddock G, et al. Randomised controlled trial of cognitive--behavioural therapy in early schizophrenia: acute-phase outcomes. *The British Journal of Psychiatry* 2002;181(43):s91-97.
70. Tarrier N, Lewis S, Haddock G, et al. Cognitive-behavioural therapy in first-episode and early schizophrenia. 18-month follow-up of a randomised controlled trial. *Br J Psychiatry* 2004;184:231-9.
71. Haddock G, Lewis SN, Bentall R, Dunn G, Drake R, Tarrier N. Influence of age on outcome of psychological treatments in first-episode psychosis. *The British Journal of Psychiatry* 2006;188(3):250-254.
72. Naeem F, Kingdon D, Turkington D. Predictors of Response to Cognitive Behaviour Therapy in the Treatment of Schizophrenia: A Comparison of Brief and Standard Interventions. *Cognitive Therapy and Research* 2008;32(5):651-656.
73. Turkington D, Kingdon D, Turner T. Effectiveness of a brief cognitive--behavioural therapy intervention in the treatment of schizophrenia. *The British Journal of Psychiatry* 2002;180(6):523-527.
74. Drake RJ, Dunn G, Tarrier N, Bentall RP, Haddock G, Lewis SW. Insight as a predictor of the outcome of first-episode nonaffective psychosis in a prospective cohort study in England. *J Clin Psychiatry* 2007;68(1):81-6.
75. Morrison AP, Renton JC, Williams S, et al. Delivering cognitive therapy to people with psychosis in a community mental health setting: an effectiveness study. *Acta Psychiatrica Scandinavica* 2004;110(1):36-44.
76. Drury V, Birchwood M, Cochrane R, Macmillan F. Cognitive therapy and recovery from acute psychosis: a controlled trial. II. Impact on recovery time. *The British Journal of Psychiatry* 1996;169(5):602-607.
77. Tarrier N, Yusupoff L, Kinney C, et al. Randomised controlled trial of intensive cognitive behaviour therapy for patients with chronic schizophrenia. *BMJ* 1998;317(7154):303-307.
78. Brabban A, Tai S, Turkington D. Predictors of Outcome in Brief Cognitive Behavior Therapy for Schizophrenia. *Schizophrenia Bulletin* 2009;35(5):859-864.
79. Dunn G, Fowler D, Rollinson R, et al. Effective elements of cognitive behaviour therapy for psychosis: results of a novel type of subgroup analysis based on principal stratification. *Psychological Medicine* 2012;42(05):1057-1068.
80. Freeman D, Dunn G, Garety P, et al. Patients' beliefs about the causes, persistence and control of psychotic experiences predict take-up of effective cognitive behaviour therapy for psychosis. *Psychological Medicine*;43(02):269-277.
81. Chadwick P, Williams C, Mackenzie J. Impact of case formulation in cognitive behaviour therapy for psychosis. *Behaviour Research and Therapy* 2003;41(6):671-680.
82. Dunn H, Morrison A. Psychosis. In: Kazantzis N, LL'Abate L, editors. *Handbook of Homework Assignments in Psychotherapy: Research, Practice and Prevention*: Springer US, 2007:335-349.

83. Rector NA. Homework Use in Cognitive Therapy for Psychosis: A Case Formulation Approach. *Cognitive and Behavioral Practice* 2007;14(3):303-316.
84. Mausbach BT, Moore R, Roesch S, Cardenas V, Patterson TL. The Relationship Between Homework Compliance and Therapy Outcomes: An Updated Meta-Analysis. *Cognit Ther Res.* 2010;34(5):429-438.
85. Granholm E, Auslander L, Gottlieb J, McQuaid J, McClure F. Therapeutic Factors Contributing to Change in Cognitive-Behavioral Group Therapy for Older Persons with Schizophrenia. *Journal of Contemporary Psychotherapy* 2006;36(1):31-41.
86. Dunn H, Morrison AP, Bentall RP. The relationship between patient suitability, therapeutic alliance, homework compliance and outcome in cognitive therapy for psychosis. *Clinical Psychology & Psychotherapy* 2006;13(3):145-152.
87. Kelly PJ, Deane FP. Relationship between therapeutic homework and clinical outcomes for individuals with severe mental illness. *Australian and New Zealand Journal of Psychiatry* 2009;43(10):968-975.
88. Svensson B, Hansson L. Therapeutic alliance in cognitive therapy for schizophrenic and other long-term mentally ill patients: development and relationship to outcome in an in-patient treatment programme. *Acta Psychiatrica Scandinavica* 1999;99(4):281-287.
89. Dunn G, Bentall R. Modelling treatment-effect heterogeneity in randomized controlled trials of complex interventions (psychological treatments). *Statistics in Medicine* 2007;26(26):4719-4745.
90. Steel C. Cognitive Behaviour Therapy for Psychosis: Current Evidence and Future Directions. *Behavioural and Cognitive Psychotherapy* 2008;36(Special Issue 06):705-712.
91. Pearl J. *Causality: models, reasoning and inference*: Cambridge: Cambridge University Press, 2000.
92. Greenland S, Robins JM, Pearl J. Confounding and Collapsibility in Causal Inference. *Statistical Science* 1999;14(1):29-46.
93. Shrier I, Platt R. Reducing bias through directed acyclic graphs. *BMC Medical Research Methodology* 2008;8(1):70.
94. Greenland S, Pearl J, Robins JM. Causal Diagrams for Epidemiologic Research. *Epidemiology* 1999;10(1):37-48.
95. Shipley B. *Cause and correlation in biology: a user's guide to path analysis, structural equations and causal inference*: Cambridge University Press, 2002.
96. Wright S. Correlation and causation. *Journal of agricultural research* 1921;20(7):557-585.
97. Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* 1974;66(5):688-701.
98. Rubin DB. Causal Inference Using Potential Outcomes. *Journal of the American Statistical Association* 2005;100(469):322-331.
99. Holland PW. Statistics and Causal Inference. *Journal of the American Statistical Association* 1986;81(396):945-960.
100. Sobel ME. Identification of Causal Parameters in Randomized Studies With Mediating Variables. *Journal of Educational and Behavioral Statistics* 2008;33(2):230-251.
101. Emsley R, Dunn G, White IR. Mediation and moderation of treatment effects in randomised controlled trials of complex interventions. *Statistical Methods in Medical Research* 2010;19(3):237-270.
102. Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J Pers Soc Psychol* 1986;51(6):1173-82.
103. Judd CM, Kenny DA. Process Analysis. *Evaluation Review* 1981;5(5):602-619.
104. Kraemer HC, Kiernan M, Essex M, Kupfer DJ. How and why criteria defining moderators and mediators differ between the Baron & Kenny and MacArthur approaches. *Health Psychol* 2008;27(2 Suppl):S101-8.

105. Angrist JD, Krueger AB. Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments. *The Journal of Economic Perspectives* 2001;15(4):69-85.
106. Small DS. Mediation analysis without sequential ignorability: using baseline covariates interacted with random assignment as instrumental variables. *Journal of Statistical Research* 2012;46(2):91-103.
107. Conditional mathematical expectation. Encyclopedia of Mathematics. , N.G. Ushakov o, 7 February 2011, http://www.encyclopediaofmath.org/index.php?title=Conditional_mathematical_expectation&oldid=15801
108. Dunn G, Maracy M, Tomenson B. Estimating treatment effects from randomized clinical trials with noncompliance and loss to follow-up: the role of instrumental variable methods. *Statistical Methods in Medical Research* 2005;14(4):369-395.
109. Maracy M, Dunn G. Estimating dose-response effects in psychological treatment trials: the role of instrumental variables. *Statistical Methods in Medical Research* 2008.
110. StataCorp. *Stata 12 Base Reference Manual*. College Station TX: Stata Press, 2011.
111. StataCorp. Stata Statistical Software: Release 12. [program]. College Station, TX: StataCorp LP. , 2011.
112. Robins JM. Correcting for non-compliance in randomized trials using structural nested mean models. *Communications in Statistics - Theory and Methods* 1994;23(8):2379 - 2412.
113. Fischer-Lapp K, Goetghebeur E. Practical Properties of Some Structural Mean Analyses of the Effect of Compliance in Randomized Trials. *Controlled Clinical Trials* 1999;20(6):531-546.
114. Moment. Encyclopedia of Mathematics, 4 May 2012, http://www.encyclopediaofmath.org/index.php?title=Moment&oldid=25957
115. Baum CF, Schaffer ME, Stillman S. Instrumental variables and GMM: Estimation and testing. *Stata Journal* 2003;3(1):1-31.
116. Frangakis CE, Rubin DB. Principal Stratification in Causal Inference. *Biometrics* 2002;58(1):21-29.
117. Frangakis CE, Rubin DB. Addressing complications of intention-to-treat analysis in the combined presence of all-or-none treatment-noncompliance and subsequent missing outcomes. *Biometrika* 1999;86(2):365-379.
118. Rubin DB. Direct and Indirect Causal Effects via Potential Outcomes. *Scandinavian Journal of Statistics* 2004;31(2):161-170.
119. Jin H, Rubin DB. Principal Stratification for Causal Inference With Extended Partial Compliance. *Journal of the American Statistical Association* 2008;103(481):101-111.
120. Angrist JD, Imbens GW. Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity. *Journal of the American Statistical Association* 1995;90(430):431-442.
121. Angrist JD, Imbens GW, Rubin DB. Identification of Causal Effects Using Instrumental Variables. *Journal of the American Statistical Association* 1996;91(434):444-455.
122. Freeman G, Cowling BJ, Schooling CM. Power and sample size calculations for Mendelian randomization studies using one genetic instrument. *International Journal of Epidemiology*;42(4):1157-1163.
123. Angrist JD, Keueger AB. Does Compulsory School Attendance Affect Schooling and Earnings? *The Quarterly Journal of Economics* 1991;106(4):979-1014.
124. Angrist JD. Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records. *The American Economic Review* 1990;80(3):313-336.
125. Bound J, Jaeger DA, Baker RM. Problems with Instrumental Variables Estimation when the Correlation between the Instruments and the Endogenous Explanatory Variable is Weak. *Journal of the American Statistical Association* 1995;90(430):443-450.

126. Murray MP. Avoiding Invalid Instruments and Coping with Weak Instruments. *Journal of Economic Perspectives* 2006;20(4):111-132.
127. Hahn J, Hausman J. Estimation with valid and invalid instruments. *Annales d'Economie et de Statistique* 2005:25-57.
128. Staiger D, Stock JH. Instrumental Variables Regression with Weak Instruments. *Econometrica* 1997;65(3):557-586.
129. Burgess S, Thompson SG. Improving bias and coverage in instrumental variable analysis with weak instruments for continuous and binary outcomes. *Statistics in Medicine* 2012;31(15):1582-1600.
130. Stock JH, Wright JH, Yogo M. A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments. *Journal of Business and Economic Statistics* 2002;20(4):518-529.
131. Stock JH, Yogo M. *Testing for Weak Instruments in Linear IV Regression Identification and Inference for Econometric Models*: Cambridge University Press, 2005.
132. Burgess S, Thompson SG. Bias in causal estimates from Mendelian randomization studies with weak instruments. *Stat Med* 2011;30(11):1312-23.
133. Burgess S, Thompson SG. Avoiding bias from weak instruments in Mendelian randomization studies. *Int J Epidemiol* 2011;40(3):755-64.
134. Davidson R, MacKinnon JG, Russell D. *Estimation and inference in econometrics / Russell Davidson, James G. MacKinnon*. New York: Oxford University Press, 1993.
135. Greene W. *Econometric Analysis*. 7th ed: Pearson Education Ltd, 2012.
136. Angrist JD, Pischke J-S. *Mostly harmless econometrics: an empiricist's companion*. US: Princeton University Press, 2009.
137. Hahn J, Hausman J, Kuersteiner G. Estimation with weak instruments: Accuracy of higher-order bias and MSE approximations. *Econometrics Journal* 2004;7(1):272-306.
138. Derksen S, Keselman HJ. Backward, forward and stepwise automated subset selection algorithms: Frequency of obtaining authentic and noise variables. *British Journal of Mathematical and Statistical Psychology* 1992;45(2):265-282.
139. Donald SG, Newey WK. Choosing the Number of Instruments. *Econometrica* 2001;69(5):1161-1191.
140. Ng S, Bai J. Selecting Instrumental Variables in a Data Rich Environment. *Journal of Time Series Econometrics*, 2009.
141. Odondi L, McNamee R. Applying optimal model selection in principal stratification for causal inference. *Stat Med* 2013;32(11):1815-28.
142. Belloni A, Chen D, Chernozhukov V, Hansen C. Sparse Models and Methods for Optimal Instruments With an Application to Eminent Domain. *Econometrica* 2012;80(6):2369-2429.
143. Hastie T, Tibshirani RJ, Friedman JH. *The Elements of Statistical Learning: Data mining, inference and prediction*. 2nd Edition ed: Springer, 2009.
144. Zou H, Hastie T. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2005;67(2):301-320.
145. Friedman JH, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software* 2010;33(1):1-22.
146. Team RC. R: A language and environment for statistical computing [program]. Vienna, Austria: R Foundation for Statistical Computing, 2013. http://www.R-project.org/
147. Burton A, Altman DG, Royston P, Holder RL. The design of simulation studies in medical statistics. *Stat Med* 2006;25(24):4279-92.
148. Pitman EJG. A note on normal correlation. *Biometrika* 1939;31(1-2):9-12.
149. Morgan WA. TEST FOR THE SIGNIFICANCE OF THE DIFFERENCE BETWEEN THE TWO VARIANCES IN A SAMPLE FROM A NORMAL BIVARIATE POPULATION. *Biometrika* 1939;31(1-2):13-19.

150. Dunn G. Review papers : Design and analysis of reliability studies. *Statistical Methods in Medical Research* 1992;1(2):123-157.
151. Pickles A, Croudace T. Latent mixture models for multivariate and longitudinal outcomes. *Statistical Methods in Medical Research* 2010;19(3):271-289.
152. Muthen L.K. MBO. *Mplus User's Guide.* Sixth Edition ed. Los Angeles, CA: Muthen & Muthen, 1998-2010.
153. Robins JM, Hernan MA. Estimation of the causal effects of time-varying exposures. In: Fitzmaurice G, Davidian M, G V, Molenberghs G, editors. *Longitudinal data analysis*. New York: Chapman and Hall/ CRC press, 2009:553-599.
154. Sterne J, White I, Carlin J, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ* 2009;338.
155. Royston P. Multiple imputation of missing values: update. *Stata Journal* 2005;5(2):188-201.
156. Rubin DB. *Multiple imputation for nonresponse in surveys / Donald B. Rubin*. New York
Chichester: New York
Chichester : Wiley, 1987.
157. Little RJA. Regression with missing Xs - A review. *Journal of the American Statistical Association* 1992;87(420):1227-1237.
158. Von Hippel PT. Regression with missing Ys: An improved strategy for analyzing multiply imputed data. *Sociological Methodology* 2007;37(1):83-117.
159. Schafer JL. Multiple imputation: a primer. *Statistical Methods in Medical Research* 1999;8(1):3-15.
160. Tabachnick BG, Fidell LS. *Using Multivariate Statistics*. 6th edition ed: Pearson, 2012.
161. Efron B, Tibshirani RJ. *An introduction to the bootstrap*
New York: Chapman Hall, 1993.
162. Beck AT, Guth D, Steer RA, Ball R. Screening for major depression disorders in medical inpatients with the Beck Depression Inventory for Primary Care. *Behaviour Research and Therapy* 1997;35(8):785-791.
163. Mattick RP, Clarke JC. Development and validation of measures of social phobia scrutiny fear and social interaction anxiety. *Behaviour Research and Therapy* 1998;36(4):455-470.
164. Kind P, Hardman G, Macran S. UK population norms for EQ-5D. *Working Papers, Centre for Health Economics*. York: University of York, 1999.
165. EQ-5D, EuroQol, http://www.euroqol.org/home.html
166. Priebe S, Huxley P, Knight S, Evans S. Application and Results of the Manchester Short Assessment of Quality of Life (Mansa). *International Journal of Social Psychiatry* 1999;45(1):7-12.
167. Cartwright-Hatton S, Wells A. Beliefs about Worry and Intrusions: The Meta-Cognitions Questionnaire and its Correlates. *Journal of Anxiety Disorders* 1997;11(3):279-296.
168. Fowler D, Freeman D, Smith BEN, et al. The Brief Core Schema Scales (BCSS): psychometric properties and associations with paranoia and grandiosity in non-clinical and psychosis samples. *Psychological Medicine* 2006;36(06):749-759.
169. Gumley AI, Gillan K, Morrison AP, Schwannauer M. The Development and Validation of the Beliefs about Paranoia Scale (Short Form). *Behavioural and Cognitive Psychotherapy* 2011;39(01):35-53.
170. Morrison AP, Gumley AI, Schwannauer M, et al. The Beliefs about Paranoia Scale: Preliminary Validation of a Metacognitive Approach to Conceptualizing Paranoia. *Behavioural and Cognitive Psychotherapy* 2005;33(02):153-164.
171. Birchwood M, Jackson C, Brunet K, Holden J, Barton K. Personal beliefs about illness questionnaire-revised (PBIQ-R): reliability and validation in a first episode sample. *Br J Clin Psychol* 2012;51(4):448-58.
172. Excel file for combining results from multiply imputed datasets. [program]. version 1.1 version, 2011. www.tufis.ro/paula/QR/2010_files/calcul_coef_MI_v1_1.xls

173. Birchwood M, Meaden A, Trower P, Gilbert P, Plaistow J. The power and omnipotence of voices: subordination and entrapment by voices and significant others. *Psychological Medicine* 2000;30(02):337-344.
174. Birchwood M, Gilbert P, Gilbert J, et al. Interpersonal and role-related schema influence the relationship with the dominant 'voice' in schizophrenia: a comparison of three models. *Psychological Medicine* 2004;34(08):1571-1580.
175. Kay SR, Fiszbein A, Opler LA. The Positive and Negative Syndrome Scale (PANSS) for Schizophrenia. *Schizophrenia Bulletin* 1987;13(2):261-276.
176. Haddock G, McCarron J, Tarrier N, Faragher EB. Scales to measure dimensions of hallucinations and delusions: the psychotic symptom rating scales (PSYRATS). *Psychological Medicine* 1999;29(04):879-889.
177. Addington D, Addington J, Maticka-Tyndale E. Assessing depression in schizophrenia: the Calgary Depression Scale. *Br J Psychiatry Suppl* 1993;22:39-44
178. Beck AT, Steer RA. *BHS, Beck Hopelessness Scale: Manual*: Psychological Corporation, 1988.
179. Beck AT, Steer RA. *BSI, Beck Scale for Suicide Ideation: Manual*: Psychological Corporation, 1991.
180. Chadwick P, Birchwood M. The omnipotence of voices. II: The Beliefs About Voices Questionnaire (BAVQ). *The British Journal of Psychiatry* 1995;166(6):773-6.
181. Hall AR, Rudebusch GD, Wilcox DW. Judging Instrument Relevance in Instrumental Variables Estimation. *International Economic Review* 1996;37(2):283-298.
182. Vansteelandt SB, Manoochehr; and Goetghebeur, Els;. Correcting Instrumental Variables Estimators for Systematic Measurement Error. *Harvard University Biostatistics Working Paper Series*: Harvard University 2007.
183. Hansen BE. Least squares model averaging. *Econometrica* 2007;75(4):1175-1189.
184. Daniel RM, Cousens SN, De Stavola BL, Kenward MG, Sterne JA. Methods for dealing with time-dependent confounding. *Stat Med* 2013;32(9):1584-618.

# Appendix

**Appendix 1: Full results of simulation study**

**1.1 Results of simulation study 1a with sample size of 200**

**A 1: Full results simulation study 1a, 5 explanatory variables of the continuous process variable, average F-statistic of relevant instruments=341, average F-statistic of individual instrument=24, N=200, simulations=1000**

| | Group effect | | | | | | Post randomisation mediator effect | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean Coef. | Mean Bias | bias 2SLS/ bias OLS | SD coef. | Absolute Difference | MSE | Mean Coef. | Mean Bias | bias 2SLS/ bias OLS | SD coef. | Absolute Difference | MSE | F-stat |
| OLS All | -10.199 | 0.199 | | 1.441 | 1.158 | 2.115 | 50.507 | -0.507 | | 0.241 | 0.511 | 0.316 | |
| 2SLS All | -9.928 | **-0.072** | -0.361 | 1.453 | 1.158 | 2.114 | 50.049 | -0.049 | 0.096 | 0.253 | 0.206 | 0.066 | 77.8 |
| OLS relevant | -10.224 | 0.224 | | 1.448 | 1.182 | 2.146 | 50.510 | -0.510 | | 0.234 | 0.513 | 0.315 | |
| 2SLS relevant | -9.925 | -0.075 | -0.335 | 1.463 | 1.179 | 2.144 | 50.002 | -0.002 | 0.003 | 0.250 | 0.201 | 0.063 | 341.4 |
| OLS LASSO | -10.210 | 0.210 | | 1.459 | 1.184 | 2.172 | 50.494 | -0.494 | | 0.237 | 0.498 | 0.300 | |
| 2SLS LASSO | -9.927 | **-0.073** | -0.346 | 1.474 | 1.187 | 2.176 | 50.011 | -0.011 | 0.023 | 0.252 | 0.202 | **0.064** | **252.0** |
| OLS LASSO-m | -10.182 | 0.182 | | 1.435 | 1.153 | **2.092** | 50.490 | -0.490 | | 0.240 | 0.494 | 0.297 | |
| 2SLS LASSO - m | -9.912 | -0.088 | -0.486 | 1.447 | 1.160 | 2.101 | 50.031 | -0.031 | 0.063 | 0.253 | 0.204 | 0.065 | 160.8 |
| OLS Elastic Net | -10.197 | 0.197 | | 1.459 | 1.182 | 2.165 | 50.494 | -0.494 | | 0.239 | 0.498 | 0.301 | |
| 2SLS Elastic Net | -9.918 | -0.082 | -0.417 | 1.473 | 1.184 | 2.174 | 50.019 | -0.019 | 0.038 | 0.253 | 0.203 | 0.064 | 186.0 |
| OLS Elastic Net - m | -10.194 | 0.194 | | 1.444 | 1.160 | 2.119 | 50.493 | -0.493 | | 0.241 | 0.497 | 0.301 | |
| 2SLS Elastic Net - m | -9.924 | -0.076 | -0.391 | 1.455 | 1.164 | 2.120 | 50.035 | -0.035 | 0.071 | 0.254 | 0.205 | 0.066 | 129.5 |
| OLS Stepwise | -10.213 | 0.213 | | 1.456 | 1.183 | 2.164 | 50.505 | -0.505 | | 0.238 | 0.509 | 0.312 | |
| 2SLS Stepwise | -9.921 | -0.079 | -0.370 | 1.470 | 1.183 | 2.165 | 49.992 | **0.008** | -0.016 | 0.253 | 0.202 | 0.064 | 223.0 |
| OLS no explanatory | -10.104 | 0.104 | | 1.480 | 1.196 | 2.200 | 50.348 | -0.348 | | 0.199 | 0.356 | 0.161 | |
| 2SLS no explanatory | -9.928 | -0.072 | -0.693 | 1.504 | 1.211 | 2.265 | 50.048 | -0.048 | 0.139 | 0.345 | 0.277 | 0.121 | 7.01 |

**A 2: Simulation study 1a, 5 explanatory variables of the continuous process variable, average F-statistic of relevant instruments=341, average F-statistic of individual instrument=24, N=200, simulations=1000, individual explanatory variables as instrument**

| | Group effect | | | | | Post-randomisation mediator effect | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean coefficient | Mean bias | SD coefficent | Absolute Difference | MSE | Mean coefficient | Mean bias | SD coefficent | Absolute Difference | MSE | F-statistic of instrument |
| 2SLS X1 only | -9.903 | -0.097 | 1.530 | 1.217 | 2.347 | 49.962 | 0.038 | 0.635 | 0.487 | 0.404 | 23.926 |
| 2SLS X2 only | -9.894 | -0.106 | 1.540 | 1.235 | 2.381 | 49.965 | 0.035 | 0.613 | 0.468 | 0.377 | 24.496 |
| 2SLS X3 only | -9.899 | -0.101 | 1.525 | 1.216 | 2.335 | 49.976 | 0.024 | 0.605 | 0.467 | 0.366 | 24.470 |
| 2SLS X4 only | -9.903 | -0.097 | 1.547 | 1.224 | 2.399 | 49.982 | 0.018 | 0.693 | 0.512 | 0.480 | 24.335 |
| 2SLS X5 only | -9.908 | -0.092 | 1.533 | 1.231 | 2.356 | 49.980 | 0.020 | 0.605 | 0.469 | 0.366 | 24.484 |

**A 3: Full results simulation study 1a, 5 explanatory variables of the continuous process variable, average F-statistic of relevant instruments=71, average F-statistic of individual instrument=20, N=200, simulations=1000**

| | Group effect | | | | | | Post randomisation mediator effect | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean Coef. | Mean Bias | bias 2SLS/ bias OLS | SD coef. | Absolute Difference | MSE | Mean Coef. | Mean Bias | bias 2SLS/ bias OLS | SD coef. | Absolute Difference | MSE | F-stat |
| OLS All | -10.426 | 0.426 | | 1.405 | 1.162 | 2.155 | 50.886 | -0.886 | | 0.214 | 0.886 | 0.831 | |
| 2SLS All | -9.976 | **-0.024** | -0.056 | 1.443 | 1.148 | 2.082 | 50.126 | -0.126 | 0.142 | 0.268 | 0.238 | 0.087 | 16.9 |
| OLS relevant | -10.450 | 0.450 | | 1.404 | 1.184 | 2.173 | 50.889 | -0.889 | | 0.207 | 0.889 | 0.833 | |
| 2SLS relevant | -9.935 | -0.065 | -0.145 | 1.456 | 1.172 | 2.123 | 50.014 | -0.014 | 0.015 | 0.273 | 0.220 | 0.075 | 71.3 |
| OLS LASSO | -10.437 | 0.437 | | 1.416 | 1.185 | 2.194 | 50.875 | -0.875 | | 0.210 | 0.875 | 0.809 | |
| 2SLS LASSO | -9.952 | -0.048 | -0.110 | 1.465 | 1.178 | 2.147 | 50.049 | -0.049 | 0.056 | 0.272 | 0.222 | 0.076 | **53.1** |
| OLS LASSO-m | -10.413 | 0.413 | | 1.398 | 1.159 | 2.124 | 50.869 | -0.869 | | 0.213 | 0.869 | 0.800 | |
| 2SLS LASSO - m | -9.954 | -0.046 | -0.111 | 1.439 | 1.150 | 2.071 | 50.093 | -0.093 | 0.107 | 0.271 | 0.231 | 0.082 | 34.3 |
| OLS Elastic Net | -10.428 | 0.428 | | 1.418 | 1.186 | 2.193 | 50.874 | -0.874 | | 0.211 | 0.874 | 0.809 | |
| 2SLS Elastic Net | -9.949 | -0.051 | -0.119 | 1.465 | 1.181 | 2.146 | 50.062 | -0.062 | 0.071 | 0.271 | 0.223 | 0.077 | 42.0 |
| OLS Elastic Net - m | -10.415 | 0.415 | | 1.398 | 1.156 | 2.126 | 50.871 | -0.871 | | 0.214 | 0.871 | 0.804 | |
| 2SLS Elastic Net - m | -9.960 | -0.040 | -0.096 | 1.439 | 1.148 | **2.070** | 50.100 | -0.100 | 0.115 | 0.271 | 0.232 | 0.083 | 28.9 |
| OLS Stepwise | -10.442 | 0.442 | | 1.407 | 1.177 | 2.173 | 50.885 | -0.885 | | 0.211 | 0.885 | 0.827 | |
| 2SLS Stepwise | -9.952 | -0.048 | -0.108 | 1.452 | 1.166 | 2.109 | 50.026 | **-0.026** | 0.030 | 0.274 | 0.221 | **0.076** | 47.3 |
| OLS no explanatory | -10.302 | 0.302 | | 1.447 | 1.186 | 2.182 | 50.671 | -0.671 | | 0.186 | 0.671 | 0.484 | |
| 2SLS no explanatory | -9.986 | -0.014 | -0.048 | 1.486 | 1.193 | 2.205 | 50.135 | -0.135 | 0.201 | 0.364 | 0.312 | 0.151 | 5.3 |

**A 4: Simulation study 1a, 5 explanatory variables of the continuous process variable, average F-statistic of relevant instruments=71, average F-statistic of individual instrument=20, N=200, simulations=1000, individual explanatory variables as instrument**

| | Group effect | | | | | Post-randomisation mediator effect | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean coefficient | Mean bias | SD coefficent | Absolute Difference | MSE | Mean coefficient | Mean bias | SD coefficent | Absolute Difference | MSE | F-statistic of instrument |
| 2SLS X1 only | -9.880 | -0.120 | 1.559 | 1.231 | 2.444 | 49.929 | 0.071 | 0.744 | 0.549 | 0.558 | 19.139 |
| 2SLS X2 only | -9.882 | -0.118 | 1.563 | 1.244 | 2.453 | 49.939 | 0.061 | 0.690 | 0.519 | 0.479 | 19.655 |
| 2SLS X3 only | -9.886 | -0.114 | 1.559 | 1.232 | 2.442 | 49.946 | 0.054 | 0.706 | 0.525 | 0.500 | 19.784 |
| 2SLS X4 only | -9.951 | -0.049 | 2.208 | 1.283 | 4.873 | 50.001 | -0.001 | 1.492 | 0.609 | 2.224 | 19.559 |
| 2SLS X5 only | -9.891 | -0.109 | 1.553 | 1.241 | 2.420 | 49.954 | 0.046 | 0.683 | 0.523 | 0.469 | 19.750 |

**A 5: Full results simulation study 1a, 5 explanatory variables of the continuous process variable, average F-statistic of relevant instruments=21, average F-statistic of individual instrument=12, N=200, simulations=1000**

| | Group effect | | | | | | Post randomisation mediator effect | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean Coef. | Mean Bias | bias 2SLS/ bias OLS | SD coef. | Absolute Difference | MSE | Mean Coef. | Mean Bias | bias 2SLS/ bias OLS | SD coef. | Absolute Difference | MSE | F-stat |
| OLS All | -10.552 | 0.552 | | 1.357 | 1.154 | 2.143 | 51.090 | -1.090 | | 0.176 | 1.090 | 1.218 | |
| 2SLS All | -10.063 | 0.063 | 0.115 | 1.419 | 1.127 | 2.015 | 50.267 | -0.267 | 0.245 | 0.302 | 0.333 | 0.162 | 5.7 |
| OLS relevant | -10.576 | 0.576 | | 1.340 | 1.166 | 2.126 | 51.091 | -1.091 | | 0.169 | 1.091 | 1.219 | |
| 2SLS relevant | -9.954 | -0.046 | -0.080 | 1.445 | 1.159 | 2.088 | 50.040 | -0.040 | 0.037 | 0.331 | 0.269 | 0.111 | 21.3 |
| OLS LASSO | -10.563 | 0.563 | | 1.354 | 1.170 | 2.149 | 51.080 | -1.080 | | 0.171 | 1.080 | 1.196 | |
| 2SLS LASSO | -9.892 | -0.108 | -0.192 | 3.849 | 1.266 | 14.810 | 50.134 | **-0.134** | 0.124 | 0.318 | 0.279 | **0.119** | **16.8** |
| OLS LASSO-m | -10.543 | 0.543 | | 1.341 | 1.149 | 2.092 | 51.079 | -1.079 | | 0.175 | 1.079 | 1.196 | |
| 2SLS LASSO - m | -10.028 | 0.028 | 0.051 | 1.414 | 1.130 | **1.997** | 50.211 | -0.211 | 0.195 | 0.311 | 0.309 | 0.141 | 11.1 |
| OLS Elastic Net | -10.562 | 0.562 | | 1.357 | 1.173 | 2.154 | 51.081 | -1.081 | | 0.172 | 1.081 | 1.198 | |
| 2SLS Elastic Net | -10.009 | **0.009** | 0.015 | 1.437 | 1.153 | 2.062 | 50.146 | -0.146 | 0.135 | 0.320 | 0.286 | 0.124 | 14.7 |
| OLS Elastic Net - m | -10.548 | 0.548 | | 1.347 | 1.155 | 2.113 | 51.080 | -1.080 | | 0.176 | 1.080 | 1.198 | |
| 2SLS Elastic Net - m | -10.036 | 0.036 | 0.065 | 1.421 | 1.131 | 2.018 | 50.216 | -0.216 | 0.200 | 0.310 | 0.310 | 0.143 | 10.0 |
| OLS Stepwise | -10.558 | 0.558 | | 1.345 | 1.161 | 2.119 | 51.084 | -1.084 | | 0.175 | 1.084 | 1.206 | |
| 2SLS Stepwise | -10.018 | 0.018 | 0.032 | 1.417 | 1.135 | 2.008 | 50.140 | -0.140 | 0.129 | 0.320 | 0.281 | 0.122 | 15.3 |
| OLS no explanatory | -10.466 | 0.466 | | 1.387 | 1.166 | 2.138 | 50.932 | -0.932 | | 0.164 | 0.932 | 0.895 | |
| 2SLS no explanatory | -10.087 | 0.087 | 0.188 | 1.448 | 1.159 | 2.101 | 50.290 | -0.290 | 0.311 | 0.407 | 0.408 | 0.250 | 3.3 |

**A 6: Simulation study 1a, 5 explanatory variables of the continuous process variable, average F-statistic of relevant instruments=21, average F-statistic of individual instrument=12, N=200, simulations=1000, individual explanatory variables as instrument**

| | Group effect | | | | | Post-randomisation mediator effect | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean coefficient | Mean bias | SD coefficent | Absolute Difference | MSE | Mean coefficient | Mean bias | SD coefficent | Absolute Difference | MSE | F-statistic of instrument |
| 2SLS X1 only | -9.807 | -0.193 | 1.823 | 1.303 | 3.356 | 49.832 | 0.168 | 1.271 | 0.748 | 1.643 | 12.038 |
| 2SLS X2 only | -9.838 | -0.162 | 1.732 | 1.302 | 3.024 | 49.861 | 0.139 | 1.009 | 0.680 | 1.037 | 12.382 |
| 2SLS X3 only | -9.809 | -0.191 | 2.342 | 1.371 | 5.517 | 49.807 | 0.193 | 3.277 | 0.827 | 10.767 | 12.607 |
| 2SLS X4 only | -9.861 | -0.139 | 2.039 | 1.336 | 4.173 | 49.892 | 0.108 | 1.256 | 0.775 | 1.589 | 12.387 |
| 2SLS X5 only | -9.839 | -0.161 | 1.690 | 1.295 | 2.880 | 49.875 | 0.125 | 0.994 | 0.691 | 1.002 | 12.537 |

**A 7: Full results simulation study 1a, 5 explanatory variables of the continuous process variable, average F-statistic of relevant instruments=11, average F-statistic of individual instrument=8, N=200, simulations=1000**

| | Group effect | | | | | | Post randomisation mediator effect | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean Coef. | Mean Bias | bias 2SLS/ bias OLS | SD coef. | Absolute Difference | MSE | Mean Coef. | Mean Bias | bias 2SLS/ bias OLS | SD coef. | Absolute Difference | MSE | F-stat |
| OLS All | -10.584 | 0.584 | | 1.331 | 1.147 | 2.110 | 51.138 | -1.138 | | 0.158 | 1.138 | 1.319 | |
| 2SLS All | -10.153 | 0.153 | 0.261 | 1.394 | 1.111 | 1.965 | 50.410 | -0.410 | 0.360 | 0.335 | 0.450 | 0.280 | 3.3 |
| OLS relevant | -10.608 | 0.608 | | 1.304 | 1.150 | 2.068 | 51.138 | -1.138 | | 0.151 | 1.138 | 1.318 | |
| 2SLS relevant | -9.978 | -0.022 | -0.036 | 1.440 | 1.152 | 2.071 | 50.075 | -0.075 | 0.066 | 0.399 | 0.327 | 0.165 | 10.8 |
| OLS LASSO | -10.575 | 0.575 | | 1.309 | 1.135 | 2.042 | 51.094 | -1.094 | | 0.154 | 1.094 | 1.221 | |
| 2SLS LASSO | -7.032 | -2.968 | -5.161 | 12.835 | 4.383 | 173.379 | 50.288 | -0.288 | 0.263 | 0.375 | 0.385 | 0.223 | **11.3** |
| OLS LASSO-m | -10.582 | 0.582 | | 1.313 | 1.143 | 2.060 | 51.130 | -1.130 | | 0.156 | 1.130 | 1.302 | |
| 2SLS LASSO - m | -10.009 | **0.009** | 0.015 | 3.504 | 1.214 | 12.264 | 50.332 | -0.332 | 0.294 | 0.354 | 0.404 | 0.235 | 6.5 |
| OLS Elastic Net | -10.583 | 0.583 | | 1.312 | 1.145 | 2.061 | 51.101 | -1.101 | | 0.154 | 1.101 | 1.235 | |
| 2SLS Elastic Net | -6.922 | -3.078 | -5.277 | 13.028 | 4.387 | 179.027 | 50.278 | **-0.278** | 0.253 | 0.374 | 0.382 | **0.217** | 10.1 |
| OLS Elastic Net - m | -10.585 | 0.585 | | 1.314 | 1.145 | 2.068 | 51.131 | -1.131 | | 0.156 | 1.131 | 1.304 | |
| 2SLS Elastic Net - m | -10.098 | 0.098 | 0.167 | 1.482 | 1.125 | 2.204 | 50.338 | -0.338 | 0.299 | 0.352 | 0.406 | 0.238 | 6.0 |
| OLS Stepwise | -10.579 | 0.579 | | 1.315 | 1.143 | 2.063 | 51.128 | -1.128 | | 0.157 | 1.128 | 1.297 | |
| 2SLS Stepwise | -10.097 | 0.097 | 0.167 | 1.393 | 1.113 | **1.949** | 50.281 | -0.281 | 0.249 | 0.381 | 0.388 | 0.224 | 8.7 |
| OLS no explanatory | -10.532 | 0.532 | | 1.346 | 1.153 | 2.094 | 51.035 | -1.035 | | 0.151 | 1.035 | 1.094 | |
| 2SLS no explanatory | -10.186 | 0.186 | 0.349 | 1.418 | 1.138 | 2.043 | 50.443 | -0.443 | 0.428 | 0.450 | 0.527 | 0.399 | 2.4 |

**A 8: Simulation study 1a, 5 explanatory variables of the continuous process variable, average F-statistic of relevant instruments=11, average F-statistic of individual instrument=8, N=200, simulations=1000, individual explanatory variables as instrument**

| | Group effect | | | | | Post-randomisation mediator effect | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean coefficient | Mean bias | SD coefficent | Absolute Difference | MSE | Mean coefficient | Mean bias | SD coefficent | Absolute Difference | MSE | F-statistic of instrument |
| 2SLS X1 only | -9.978 | -0.022 | 9.902 | 1.975 | 97.951 | 50.556 | -0.556 | 25.456 | 2.171 | 647.655 | 7.903 |
| 2SLS X2 only | -9.199 | -0.801 | 21.207 | 2.097 | 449.915 | 49.157 | 0.843 | 20.076 | 1.644 | 403.351 | 8.130 |
| 2SLS X3 only | -9.900 | -0.100 | 3.009 | 1.533 | 9.053 | 49.880 | 0.120 | 2.930 | 1.092 | 8.592 | 8.360 |
| 2SLS X4 only | -9.877 | -0.123 | 3.433 | 1.608 | 11.792 | 49.685 | 0.315 | 5.314 | 1.361 | 28.308 | 8.180 |
| 2SLS X5 only | -10.413 | 0.413 | 25.105 | 2.785 | 629.821 | 50.733 | -0.733 | 30.358 | 2.444 | 921.222 | 8.284 |

**A 9: Full results simulation study 1a, 5 explanatory variables of the continuous process variable, average F-statistic of relevant instruments=3.9, average F-statistic of individual instrument=3.7, N=200, simulations=1000**

| | Group effect | | | | | | Post randomisation mediator effect | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean Coef. | Mean Bias | bias 2SLS/ bias OLS | SD coef. | Absolute Difference | MSE | Mean Coef. | Mean Bias | bias 2SLS/ bias OLS | SD coef. | Absolute Difference | MSE | F-stat |
| OLS All | -10.601 | 0.601 | | 1.305 | 1.135 | 2.063 | 51.159 | -1.159 | | 0.142 | 1.159 | 1.364 | |
| 2SLS All | -10.346 | 0.346 | 0.575 | 1.356 | 1.107 | 1.957 | 50.718 | -0.718 | 0.620 | 0.399 | 0.735 | 0.675 | 1.75 |
| OLS relevant | -10.627 | 0.627 | | 1.265 | 1.129 | 1.991 | 51.159 | -1.159 | | 0.135 | 1.159 | 1.361 | |
| 2SLS relevant | -10.065 | 0.065 | 0.103 | 1.448 | 1.153 | 2.099 | 50.215 | -0.215 | 0.185 | 0.594 | 0.504 | 0.398 | 3.94 |
| OLS LASSO | -10.583 | 0.583 | | 1.263 | 1.105 | 1.933 | 51.091 | -1.091 | | 0.131 | 1.091 | 1.208 | |
| 2SLS LASSO | 14.763 | -24.763 | -42.487 | 29.676 | 29.203 | 1492.959 | 50.621 | **-0.621** | 0.569 | 0.504 | 0.676 | **0.638** | **10.90** |
| OLS LASSO-m | -10.597 | 0.597 | | 1.271 | 1.116 | 1.971 | 51.138 | -1.138 | | 0.140 | 1.138 | 1.316 | |
| 2SLS LASSO - m | -5.905 | -4.095 | -6.861 | 15.305 | 6.145 | 250.772 | 50.655 | -0.655 | 0.575 | 0.509 | 0.714 | 0.687 | 4.98 |
| OLS Elastic Net | -10.580 | 0.580 | | 1.265 | 1.106 | 1.935 | 51.091 | -1.091 | | 0.133 | 1.091 | 1.209 | |
| 2SLS Elastic Net | 14.977 | -24.977 | -43.077 | 29.508 | 29.303 | 1493.701 | 50.659 | -0.659 | 0.604 | 0.482 | 0.702 | 0.665 | 9.81 |
| OLS Elastic Net - m | -10.600 | 0.600 | | 1.275 | 1.120 | 1.984 | 51.141 | -1.141 | | 0.141 | 1.141 | 1.322 | |
| 2SLS Elastic Net - m | -6.072 | -3.928 | -6.548 | 15.441 | 6.139 | 253.604 | 50.640 | -0.640 | 0.561 | 0.485 | 0.695 | 0.644 | 4.66 |
| OLS Stepwise | -10.603 | 0.603 | | 1.270 | 1.115 | 1.975 | 51.146 | -1.146 | | 0.140 | 1.146 | 1.333 | |
| 2SLS Stepwise | -10.306 | **0.306** | 0.508 | 1.342 | 1.094 | **1.892** | 50.628 | -0.628 | 0.548 | 0.515 | 0.683 | 0.659 | 4.80 |
| OLS no explanatory | -10.589 | 0.589 | | 1.297 | 1.133 | 2.028 | 51.120 | -1.120 | | 0.137 | 1.120 | 1.274 | |
| 2SLS no explanatory | -10.381 | 0.381 | 0.647 | 1.382 | 1.140 | 2.054 | 50.753 | -0.753 | 0.672 | 0.528 | 0.794 | 0.845 | 1.55 |

**A 10: Simulation study 1a, 5 explanatory variables of the continuous process variable, average F-statistic of relevant instruments=3.9, average F-statistic of individual instrument=3.7, N=200, simulations=1000, individual explanatory variables as instrument**

| | Group effect | | | | | Post-randomisation mediator effect | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean coefficient | Mean bias | SD coefficent | Absolute Difference | MSE | Mean coefficient | Mean bias | SD coefficent | Absolute Difference | MSE | F-statistic of instrument |
| 2SLS X1 only | -5.833 | -4.167 | 79.994 | 7.158 | 6409.943 | 49.215 | 0.785 | 69.879 | 7.077 | 4878.785 | 3.556 |
| 2SLS X2 only | -9.309 | -0.691 | 15.038 | 3.153 | 226.395 | 49.868 | 0.132 | 25.520 | 4.165 | 650.643 | 3.647 |
| 2SLS X3 only | -12.421 | 2.421 | 77.810 | 6.233 | 6054.237 | 52.483 | -2.483 | 81.806 | 6.663 | 6691.729 | 3.827 |
| 2SLS X4 only | -4.708 | -5.292 | 148.708 | 8.663 | 22119.910 | 55.466 | -5.466 | 189.366 | 10.200 | 35853.340 | 3.725 |
| 2SLS X5 only | -10.222 | 0.222 | 14.379 | 3.131 | 206.606 | 50.778 | -0.778 | 23.681 | 3.956 | 560.842 | 3.763 |

**A 11: Full results simulation study 1a, 10 explanatory variables of the continuous process variable, average F-statistic of relevant instruments=191, average F-statistic of individual instruments=14, N=200, simulations=1000**

| | Group effect | | | | | | Post randomisation mediator effect | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean Coef. | Mean Bias | bias 2SLS/ bias OLS | SD coef. | Absolute Difference | MSE | Mean Coef. | Mean Bias | bias 2SLS/ bias OLS | SD coef. | Absolute Difference | MSE | F-stat |
| OLS All | -10.192 | 0.192 | | 1.585 | 1.266 | 2.547 | 50.448 | -0.448 | | 0.237 | 0.455 | 0.257 | |
| 2SLS All | -9.950 | -0.050 | -0.259 | 1.601 | 1.265 | 2.562 | 50.045 | -0.045 | 0.101 | 0.247 | 0.201 | 0.063 | 83.8 |
| OLS relevant | -10.214 | 0.214 | | 1.558 | 1.249 | 2.472 | 50.446 | -0.446 | | 0.232 | 0.452 | 0.253 | |
| 2SLS relevant | -9.953 | -0.047 | -0.220 | 1.576 | 1.254 | 2.485 | 50.010 | -0.010 | 0.022 | 0.246 | 0.198 | 0.061 | 191.4 |
| OLS LASSO | -10.192 | 0.192 | | 1.561 | 1.253 | 2.472 | 50.435 | -0.435 | | 0.232 | 0.442 | 0.243 | |
| 2SLS LASSO | -9.945 | -0.055 | -0.285 | 1.578 | 1.258 | 2.492 | 50.023 | -0.023 | 0.052 | 0.245 | 0.197 | **0.060** | **150.2** |
| OLS LASSO-m | -10.189 | 0.189 | | 1.564 | 1.258 | 2.479 | 50.438 | -0.438 | | 0.236 | 0.445 | 0.248 | |
| 2SLS LASSO - m | -9.950 | -0.050 | -0.264 | 1.580 | 1.258 | 2.495 | 50.037 | -0.037 | 0.085 | 0.247 | 0.200 | 0.062 | 118.3 |
| OLS Elastic Net | -10.183 | 0.183 | | 1.561 | 1.254 | **2.469** | 50.436 | -0.436 | | 0.235 | 0.443 | 0.245 | |
| 2SLS Elastic Net | -9.938 | -0.062 | -0.340 | 1.577 | 1.258 | 2.490 | 50.026 | -0.026 | 0.059 | 0.246 | 0.199 | 0.061 | 134.7 |
| OLS Elastic Net - m | -10.185 | 0.185 | | 1.573 | 1.265 | 2.506 | 50.440 | -0.440 | | 0.237 | 0.447 | 0.249 | |
| 2SLS Elastic Net - m | -9.945 | -0.055 | -0.297 | 1.588 | 1.266 | 2.523 | 50.039 | -0.039 | 0.089 | 0.247 | 0.201 | 0.063 | 110.0 |
| OLS Stepwise | -10.207 | 0.207 | | 1.563 | 1.251 | 2.483 | 50.445 | -0.445 | | 0.236 | 0.452 | 0.253 | |
| 2SLS Stepwise | -9.956 | **-0.044** | -0.212 | 1.578 | 1.252 | 2.490 | 50.012 | **-0.012** | 0.028 | 0.253 | 0.203 | 0.064 | 138.4 |
| OLS no explanatory | -10.075 | 0.075 | | 1.564 | 1.253 | 2.448 | 50.270 | -0.270 | | 0.178 | 0.281 | 0.104 | |
| 2SLS no explanatory | -9.953 | -0.047 | -0.627 | 1.604 | 1.268 | 2.571 | 50.060 | -0.060 | 0.221 | 0.412 | 0.327 | 0.173 | 4.2 |

**A 12: Simulation study 1a, 10 explanatory variables of the continuous process variable, average F-statistic of relevant instruments=191, average F-statistic of individual instruments=14, N=200, simulations=1000, individual explanatory variables as instrument**

| | Group Effect | | | | | Post randomisation mediator effect | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean coefficient | Mean bias | SD coefficent | Absolute Difference | MSE | Mean coefficient | Mean bias | SD coefficent | Absolute Difference | MSE | F-statistic |
| 2SLS X1 only | -9.764 | -0.236 | 3.461 | 1.457 | 12.023 | 50.019 | -0.019 | 2.872 | 0.775 | 8.240 | 11.177 |
| 2SLS X2 only | -9.783 | -0.217 | 5.018 | 1.535 | 25.206 | 49.780 | 0.220 | 5.858 | 0.891 | 34.332 | 11.382 |
| 2SLS X3 only | -9.931 | -0.069 | 1.735 | 1.340 | 3.013 | 49.964 | 0.036 | 0.946 | 0.672 | 0.895 | 11.947 |
| 2SLS X4 only | -9.835 | -0.165 | 2.169 | 1.425 | 4.729 | 49.905 | 0.095 | 1.206 | 0.725 | 1.462 | 11.464 |
| 2SLS X5 only | -9.874 | -0.126 | 1.849 | 1.386 | 3.433 | 49.922 | 0.078 | 1.164 | 0.697 | 1.358 | 11.889 |
| 2SLS X6 only | -9.920 | -0.080 | 1.770 | 1.379 | 3.137 | 49.972 | 0.028 | 1.050 | 0.704 | 1.103 | 11.606 |
| 2SLS X7 only | -9.897 | -0.103 | 1.832 | 1.395 | 3.362 | 49.999 | 0.001 | 1.382 | 0.723 | 1.908 | 11.495 |
| 2SLS X8 only | -9.920 | -0.080 | 1.889 | 1.395 | 3.572 | 49.944 | 0.056 | 1.460 | 0.746 | 2.134 | 11.708 |
| 2SLS X9 only | -10.003 | 0.003 | 3.141 | 1.463 | 9.853 | 50.101 | -0.101 | 4.578 | 0.837 | 20.952 | 11.196 |
| 2SLS X10 only | -9.886 | -0.114 | 1.783 | 1.390 | 3.188 | 49.932 | 0.068 | 1.423 | 0.751 | 2.027 | 11.212 |

**A 13: Full results simulation study 1a, 10 explanatory variables of the continuous process variable, average F-statistic of relevant instruments=42, average F-statistic of individual instrument=9.8, N=200, simulations=1000**

| | Group effect | | | | | | Post randomisation mediator effect | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean Coef. | Mean Bias | bias 2SLS/ bias OLS | SD coef. | Absolute Difference | MSE | Mean Coef. | Mean Bias | bias 2SLS/ bias OLS | SD coef. | Absolute Difference | MSE | F-stat |
| OLS All | -10.405 | 0.405 | | 1.537 | 1.262 | 2.524 | 50.795 | -0.795 | | 0.213 | 0.795 | 0.678 | |
| 2SLS All | -9.994 | -0.006 | -0.016 | 1.588 | 1.255 | 2.521 | 50.114 | -0.114 | 0.143 | 0.259 | 0.229 | 0.080 | 18.6 |
| OLS relevant | -10.427 | 0.427 | | 1.509 | 1.245 | 2.458 | 50.794 | -0.794 | | 0.208 | 0.794 | 0.673 | |
| 2SLS relevant | -9.971 | -0.029 | -0.068 | 1.569 | 1.247 | 2.461 | 50.038 | -0.038 | 0.048 | 0.265 | 0.215 | 0.072 | 41.3 |
| OLS LASSO | -10.405 | 0.405 | | 1.511 | 1.245 | 2.446 | 50.784 | -0.784 | | 0.208 | 0.784 | 0.658 | |
| 2SLS LASSO | -9.979 | -0.021 | -0.053 | 1.567 | 1.249 | **2.452** | 50.074 | -0.074 | 0.095 | 0.260 | 0.218 | **0.073** | **32.8** |
| OLS LASSO-m | -10.400 | 0.400 | | 1.520 | 1.253 | 2.466 | 50.785 | -0.785 | | 0.211 | 0.785 | 0.660 | |
| 2SLS LASSO - m | -9.989 | -0.011 | -0.028 | 1.572 | 1.253 | 2.469 | 50.101 | -0.101 | 0.129 | 0.259 | 0.225 | 0.078 | 26.0 |
| OLS Elastic Net | -10.399 | 0.399 | | 1.515 | 1.247 | 2.452 | 50.785 | -0.785 | | 0.210 | 0.785 | 0.660 | |
| 2SLS Elastic Net | -9.974 | -0.026 | -0.064 | 1.568 | 1.250 | 2.457 | 50.079 | -0.079 | 0.101 | 0.260 | 0.220 | 0.074 | 30.2 |
| OLS Elastic Net - m | -10.398 | 0.398 | | 1.523 | 1.258 | 2.477 | 50.786 | -0.786 | | 0.212 | 0.786 | 0.663 | |
| 2SLS Elastic Net - m | -9.987 | -0.013 | -0.032 | 1.576 | 1.258 | 2.482 | 50.103 | -0.103 | 0.132 | 0.260 | 0.226 | 0.078 | 24.6 |
| OLS Stepwise | -10.419 | 0.419 | | 1.513 | 1.246 | 2.463 | 50.789 | -0.789 | | 0.210 | 0.789 | 0.666 | |
| 2SLS Stepwise | -9.994 | **-0.006** | -0.014 | 1.568 | 1.243 | 2.455 | 50.062 | **-0.062** | 0.078 | 0.273 | 0.225 | 0.078 | 32.2 |
| OLS no explanatory | -10.234 | 0.234 | | 1.524 | 1.238 | 2.376 | 50.532 | -0.532 | | 0.169 | 0.532 | 0.311 | |
| 2SLS no explanatory | -10.005 | 0.005 | 0.021 | 1.588 | 1.253 | 2.518 | 50.144 | -0.144 | 0.271 | 0.438 | 0.364 | 0.213 | 3.6 |

**A 14: Simulation study 1a, 10 explanatory variables of the continuous process variable, average F-statistic of relevant instruments=42, average F-statistic of individual instruments=9.8, N=200, simulations=1000, individual explanatory variables as instrument**

| | Group Effect | | | | | Post randomisation mediator effect | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean coefficient | Mean bias | SD coefficent | Absolute Difference | MSE | Mean coefficient | Mean bias | SD coefficent | Absolute Difference | MSE | F-statistic |
| 2SLS X1 only | -9.797 | -0.203 | 3.777 | 1.530 | 14.292 | 49.903 | 0.097 | 2.673 | 0.873 | 7.148 | 9.580 |
| 2SLS X2 only | -13.801 | 3.801 | 107.033 | 5.358 | 11459.150 | 39.596 | 10.404 | 346.049 | 14.723 | 119738.500 | 9.640 |
| 2SLS X3 only | -9.902 | -0.098 | 1.810 | 1.374 | 3.283 | 49.909 | 0.091 | 1.102 | 0.750 | 1.222 | 10.194 |
| 2SLS X4 only | -9.791 | -0.209 | 2.961 | 1.523 | 8.805 | 49.872 | 0.128 | 1.894 | 0.871 | 3.599 | 9.721 |
| 2SLS X5 only | -9.847 | -0.153 | 1.878 | 1.418 | 3.548 | 49.868 | 0.132 | 1.221 | 0.779 | 1.508 | 10.114 |
| 2SLS X6 only | -9.857 | -0.143 | 1.912 | 1.432 | 3.672 | 49.869 | 0.131 | 1.489 | 0.826 | 2.231 | 9.817 |
| 2SLS X7 only | -9.843 | -0.157 | 1.937 | 1.436 | 3.772 | 49.936 | 0.064 | 1.746 | 0.823 | 3.050 | 9.823 |
| 2SLS X8 only | -9.723 | -0.277 | 4.682 | 1.581 | 21.977 | 49.946 | 0.054 | 5.689 | 1.025 | 32.340 | 9.852 |
| 2SLS X9 only | -10.523 | 0.523 | 20.141 | 2.048 | 405.514 | 50.315 | -0.315 | 11.967 | 1.219 | 143.154 | 9.494 |
| 2SLS X10 only | -10.071 | 0.071 | 14.462 | 2.040 | 208.937 | 49.192 | 0.808 | 16.528 | 1.513 | 273.541 | 9.539 |

**A 15: Full results simulation study 1a, 10 explanatory variables of the continuous process variable, average F-statistic of relevant instruments=14, average F-statistic of individual instrument=7.0, N=200, simulations=1000**

| | Group effect | | | | | | Post randomisation mediator effect | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean Coef. | Mean Bias | bias 2SLS/ bias OLS | SD coef. | Absolute Difference | MSE | Mean Coef. | Mean Bias | bias 2SLS/ bias OLS | SD coef. | Absolute Difference | MSE | F-stat |
| OLS All | -10.532 | 0.532 | | 1.477 | 1.251 | 2.461 | 50.995 | -0.995 | | 0.176 | 0.995 | 1.022 | |
| 2SLS All | -10.067 | 0.067 | 0.126 | 1.561 | 1.233 | 2.437 | 50.229 | -0.229 | 0.230 | 0.284 | 0.302 | 0.133 | 6.5 |
| OLS relevant | -10.553 | 0.553 | | 1.445 | 1.231 | 2.392 | 50.995 | -0.995 | | 0.172 | 0.995 | 1.019 | |
| 2SLS relevant | -10.003 | 0.003 | 0.006 | 1.555 | 1.234 | 2.415 | 50.090 | -0.090 | 0.090 | 0.307 | 0.259 | 0.102 | 13.6 |
| OLS LASSO | -10.538 | 0.538 | | 1.447 | 1.234 | 2.381 | 50.982 | -0.982 | | 0.171 | 0.982 | 0.994 | |
| 2SLS LASSO | -9.968 | **-0.032** | -0.059 | 3.166 | 1.306 | 10.016 | 50.180 | **-0.180** | 0.183 | 0.295 | 0.283 | 0.119 | 11.7 |
| OLS LASSO-m | -10.533 | 0.533 | | 1.456 | 1.242 | 2.402 | 50.989 | -0.989 | | 0.175 | 0.989 | 1.008 | |
| 2SLS LASSO - m | -10.062 | 0.062 | 0.117 | 1.542 | 1.229 | 2.381 | 50.210 | -0.210 | 0.212 | 0.287 | 0.293 | 0.126 | 9.1 |
| OLS Elastic Net | -10.531 | 0.531 | | 1.451 | 1.236 | 2.385 | 50.984 | -0.984 | | 0.171 | 0.984 | 0.998 | |
| 2SLS Elastic Net | -9.961 | -0.039 | -0.074 | 3.168 | 1.311 | 10.028 | 50.180 | -0.180 | 0.183 | 0.294 | 0.283 | **0.119** | 11.0 |
| OLS Elastic Net - m | -10.532 | 0.532 | | 1.461 | 1.246 | 2.416 | 50.990 | -0.990 | | 0.175 | 0.990 | 1.010 | |
| 2SLS Elastic Net - m | -10.061 | 0.061 | 0.115 | 1.548 | 1.233 | 2.396 | 50.212 | -0.212 | 0.215 | 0.287 | 0.295 | 0.127 | 8.8 |
| OLS Stepwise | -10.523 | 0.523 | | 1.451 | 1.226 | 2.376 | 50.981 | -0.981 | | 0.172 | 0.981 | 0.992 | |
| 2SLS Stepwise | -10.055 | 0.055 | 0.105 | 1.540 | 1.219 | **2.372** | 50.180 | -0.180 | 0.184 | 0.313 | 0.295 | 0.130 | **11.8** |
| OLS no explanatory | -10.384 | 0.384 | | 1.466 | 1.221 | 2.293 | 50.773 | -0.773 | | 0.153 | 0.773 | 0.620 | |
| 2SLS no explanatory | -10.092 | 0.092 | 0.239 | 1.558 | 1.232 | 2.435 | 50.279 | -0.279 | 0.361 | 0.486 | 0.449 | 0.314 | 2.7 |

**A 16: Simulation study 1a, 10 explanatory variables of the continuous process variable, average F-statistic of relevant instruments=14, average F-statistic of individual instruments=7.0, N=200, simulations=1000, individual explanatory variables as instrument**

| | Group Effect | | | | | Post randomisation mediator effect | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean coefficient | Mean bias | SD coefficent | Absolute Difference | MSE | Mean coefficient | Mean bias | SD coefficent | Absolute Difference | MSE | F-statistic of instrument |
| 2SLS X1 only | -9.857 | -0.143 | 2.888 | 1.628 | 8.351 | 49.870 | 0.130 | 2.661 | 1.173 | 7.092 | 6.943 |
| 2SLS X2 only | -10.060 | 0.060 | 8.961 | 2.037 | 80.221 | 50.632 | -0.632 | 19.036 | 1.963 | 362.407 | 6.891 |
| 2SLS X3 only | -9.941 | -0.059 | 2.726 | 1.547 | 7.426 | 49.932 | 0.068 | 2.574 | 1.043 | 6.623 | 7.347 |
| 2SLS X4 only | -9.036 | -0.964 | 20.993 | 2.570 | 441.214 | 49.393 | 0.607 | 12.759 | 1.789 | 163.002 | 6.954 |
| 2SLS X5 only | -10.222 | 0.222 | 10.280 | 2.111 | 105.632 | 50.629 | -0.629 | 20.077 | 1.996 | 403.066 | 7.257 |
| 2SLS X6 only | -8.893 | -1.107 | 38.398 | 3.395 | 1474.192 | 49.713 | 0.287 | 27.099 | 2.583 | 733.706 | 7.020 |
| 2SLS X7 only | -8.345 | -1.655 | 27.797 | 2.966 | 774.643 | 50.205 | -0.205 | 30.468 | 2.509 | 927.409 | 7.097 |
| 2SLS X8 only | -9.732 | -0.268 | 3.319 | 1.720 | 11.078 | 49.688 | 0.312 | 3.489 | 1.253 | 12.255 | 6.977 |
| 2SLS X9 only | -9.572 | -0.428 | 7.344 | 1.834 | 54.068 | 49.872 | 0.128 | 6.672 | 1.383 | 44.492 | 6.809 |
| 2SLS X10 only | -9.769 | -0.231 | 3.710 | 1.744 | 13.805 | 50.127 | -0.127 | 8.166 | 1.476 | 66.635 | 6.862 |

**A 17: Full results simulation study 1a, 10 explanatory variables of the continuous process variable, average F-statistic of relevant instruments=7.9, average F-statistic of individual instrument=5.3, N=200, simulations=1000**

| | Group effect | | | | | | Post randomisation mediator effect | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean Coef. | Mean Bias | bias 2SLS/ bias OLS | SD coef. | Absolute Difference | MSE | Mean Coef. | Mean Bias | bias 2SLS/ bias OLS | SD coef. | Absolute Difference | MSE | F-stat |
| OLS All | -10.567 | 0.567 | | 1.446 | 1.241 | 2.410 | 51.048 | -1.048 | | 0.157 | 1.048 | 1.123 | |
| 2SLS All | -10.133 | 0.133 | 0.235 | 1.535 | 1.216 | 2.371 | 50.334 | -0.334 | 0.318 | 0.305 | 0.383 | 0.204 | 4.0 |
| OLS relevant | -10.587 | 0.587 | | 1.411 | 1.219 | 2.333 | 51.048 | -1.048 | | 0.154 | 1.048 | 1.121 | |
| 2SLS relevant | -10.037 | 0.037 | 0.062 | 1.542 | 1.222 | 2.376 | 50.144 | -0.144 | 0.138 | 0.347 | 0.305 | 0.141 | 7.7 |
| OLS LASSO | -10.552 | 0.552 | | 1.410 | 1.214 | 2.291 | 51.007 | -1.007 | | 0.149 | 1.007 | 1.035 | |
| 2SLS LASSO | -8.807 | -1.193 | -2.159 | 9.830 | 2.752 | 97.947 | 50.295 | -0.295 | 0.293 | 0.340 | 0.374 | 0.202 | **8.6** |
| OLS LASSO-m | -10.570 | 0.570 | | 1.418 | 1.230 | 2.333 | 51.041 | -1.041 | | 0.155 | 1.041 | 1.108 | |
| 2SLS LASSO - m | -10.122 | 0.122 | 0.214 | 1.512 | 1.209 | 2.299 | 50.297 | -0.297 | 0.286 | 0.320 | 0.366 | **0.191** | 5.9 |
| OLS Elastic Net | -10.557 | 0.557 | | 1.406 | 1.214 | 2.285 | 51.009 | -1.009 | | 0.152 | 1.009 | 1.042 | |
| 2SLS Elastic Net | -8.218 | -1.782 | -3.199 | 11.607 | 3.418 | 137.773 | 50.301 | -0.301 | 0.299 | 0.330 | 0.375 | 0.200 | 8.0 |
| OLS Elastic Net - m | -10.569 | 0.569 | | 1.417 | 1.229 | 2.330 | 51.042 | -1.042 | | 0.156 | 1.042 | 1.111 | |
| 2SLS Elastic Net - m | -10.044 | **0.044** | 0.078 | 3.385 | 1.314 | 11.448 | 50.305 | -0.305 | 0.293 | 0.314 | 0.366 | 0.191 | 5.6 |
| OLS Stepwise | -10.546 | 0.546 | | 1.412 | 1.211 | 2.291 | 51.027 | -1.027 | | 0.154 | 1.027 | 1.079 | |
| 2SLS Stepwise | -10.121 | 0.121 | 0.221 | 1.504 | 1.198 | **2.274** | 50.287 | **-0.287** | 0.279 | 0.342 | 0.371 | 0.199 | 7.7 |
| OLS no explanatory | -10.455 | 0.455 | | 1.429 | 1.211 | 2.246 | 50.882 | -0.882 | | 0.143 | 0.882 | 0.798 | |
| 2SLS no explanatory | -10.167 | 0.167 | 0.367 | 1.534 | 1.221 | 2.379 | 50.396 | -0.396 | 0.449 | 0.520 | 0.530 | 0.427 | 2.2 |

**A 18: Simulation study 1a, 10 explanatory variables of the continuous process variable, average F-statistic of relevant instruments=7.9, average F-statistic of individual instruments=5.3, N=200, simulations=1000, individual explanatory variables as instrument**

| | Group Effect | | | | | Post randomisation mediator effect | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean coefficient | Mean bias | SD coefficent | Absolute Difference | MSE | Mean coefficient | Mean bias | SD coefficent | Absolute Difference | MSE | F-statistic |
| 2SLS X1 only | -10.213 | 0.213 | 11.788 | 2.313 | 138.853 | 48.966 | 1.034 | 22.554 | 2.405 | 509.263 | 5.279 |
| 2SLS X2 only | -9.699 | -0.301 | 7.214 | 2.241 | 52.075 | 50.147 | -0.147 | 9.402 | 2.137 | 88.328 | 5.190 |
| 2SLS X3 only | -9.355 | -0.645 | 11.668 | 2.110 | 136.430 | 49.311 | 0.689 | 8.086 | 1.660 | 65.787 | 5.562 |
| 2SLS X4 only | -9.670 | -0.330 | 11.834 | 2.582 | 140.010 | 49.066 | 0.934 | 28.839 | 2.936 | 831.739 | 5.241 |
| 2SLS X5 only | -9.882 | -0.118 | 4.245 | 1.963 | 18.019 | 49.791 | 0.209 | 5.074 | 1.664 | 25.762 | 5.475 |
| 2SLS X6 only | -10.250 | 0.250 | 28.831 | 3.631 | 830.457 | 50.843 | -0.843 | 42.978 | 3.702 | 1845.961 | 5.300 |
| 2SLS X7 only | -9.365 | -0.635 | 18.884 | 2.948 | 356.634 | 49.309 | 0.691 | 19.022 | 2.578 | 361.952 | 5.389 |
| 2SLS X8 only | -9.867 | -0.133 | 4.576 | 2.022 | 20.936 | 49.813 | 0.187 | 10.115 | 2.168 | 102.237 | 5.222 |
| 2SLS X9 only | -9.947 | -0.053 | 3.688 | 1.843 | 13.591 | 49.776 | 0.224 | 3.741 | 1.548 | 14.035 | 5.151 |
| 2SLS X10 only | -10.105 | 0.105 | 12.329 | 2.396 | 151.871 | 47.117 | 2.883 | 61.630 | 4.139 | 3802.783 | 5.196 |

**A 19: Full results simulation study 1a, 10 explanatory variables of the continuous process variable, average F-statistic of relevant instruments=4.0, average F-statistic of individual instrument=3.4, N=200, simulations=1000**

| | Group effect | | | | | | Post randomisation mediator effect | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean Coef. | Mean Bias | bias 2SLS/ bias OLS | SD coef. | Absolute Difference | MSE | Mean Coef. | Mean Bias | bias 2SLS/ bias OLS | SD coef. | Absolute Difference | MSE | F-stat |
| OLS All | -10.588 | 0.588 | | 1.416 | 1.228 | 2.347 | 51.077 | -1.077 | | 0.140 | 1.077 | 1.179 | |
| 2SLS All | -10.255 | 0.255 | 0.434 | 1.494 | 1.204 | 2.295 | 50.524 | -0.524 | 0.487 | 0.339 | 0.546 | 0.389 | 2.50 |
| OLS relevant | -10.607 | 0.607 | | 1.376 | 1.204 | 2.260 | 51.076 | -1.076 | | 0.137 | 1.076 | 1.177 | |
| 2SLS relevant | -10.112 | 0.112 | 0.184 | 1.521 | 1.206 | 2.323 | 50.270 | -0.270 | 0.251 | 0.421 | 0.413 | 0.250 | 4.06 |
| OLS LASSO | -10.548 | 0.548 | | 1.361 | 1.183 | 2.150 | 50.993 | -0.993 | | 0.138 | 0.993 | 1.005 | |
| 2SLS LASSO | 5.803 | -15.803 | -28.822 | 30.244 | 20.555 | 1163.489 | 50.557 | -0.557 | 0.561 | 0.394 | 0.585 | 0.465 | 5.01 |
| OLS LASSO-m | -10.586 | 0.586 | | 1.364 | 1.198 | 2.201 | 51.058 | -1.058 | | 0.136 | 1.058 | 1.137 | |
| 2SLS LASSO - m | -10.200 | **0.200** | 0.341 | 1.472 | 1.192 | **2.206** | 50.425 | **-0.425** | 0.402 | 0.395 | 0.495 | **0.337** | 4.59 |
| OLS Elastic Net | -10.550 | 0.550 | | 1.368 | 1.186 | 2.171 | 50.997 | -0.997 | | 0.137 | 0.997 | 1.012 | |
| 2SLS Elastic Net | 5.468 | -15.468 | -28.114 | 30.298 | 20.213 | 1156.331 | 50.536 | -0.536 | 0.538 | 0.391 | 0.568 | 0.440 | 4.92 |
| OLS Elastic Net - m | -10.594 | 0.594 | | 1.376 | 1.208 | 2.244 | 51.061 | -1.061 | | 0.136 | 1.061 | 1.144 | |
| 2SLS Elastic Net - m | -7.998 | -2.002 | -3.373 | 12.532 | 3.626 | 160.892 | 50.486 | -0.486 | 0.458 | 0.379 | 0.528 | 0.380 | 4.32 |
| OLS Stepwise | -10.555 | 0.555 | | 1.378 | 1.193 | 2.204 | 51.054 | -1.054 | | 0.137 | 1.054 | 1.130 | |
| 2SLS Stepwise | -10.219 | 0.219 | 0.396 | 1.470 | 1.187 | 2.207 | 50.473 | -0.473 | 0.448 | 0.405 | 0.526 | 0.387 | **5.62** |
| OLS no explanatory | -10.521 | 0.521 | | 1.387 | 1.201 | 2.194 | 50.982 | -0.982 | | 0.133 | 0.982 | 0.983 | |
| 2SLS no explanatory | -10.295 | 0.295 | 0.566 | 1.494 | 1.212 | 2.317 | 50.597 | -0.597 | 0.608 | 0.567 | 0.684 | 0.677 | 1.73 |

**A 20: Simulation study 1a, 10 explanatory variables of the continuous process variable, average F-statistic of relevant instruments=4.0, average F-statistic of individual instruments=3.4, N=200, simulations=1000, individual explanatory variables as instrument**

| | Group Effect | | | | | Post randomisation mediator effect | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean coefficient | Mean bias | SD coefficent | Absolute Difference | MSE | Mean coefficient | Mean bias | SD coefficent | Absolute Difference | MSE | F-statistic |
| 2SLS X1 only | -10.948 | 0.948 | 21.709 | 3.745 | 471.714 | 50.617 | -0.617 | 19.382 | 3.693 | 375.662 | 3.410 |
| 2SLS X2 only | -10.168 | 0.168 | 57.247 | 6.236 | 3273.937 | 51.562 | -1.562 | 49.138 | 5.998 | 2414.585 | 3.308 |
| 2SLS X3 only | -12.987 | 2.987 | 82.547 | 5.600 | 6816.063 | 47.362 | 2.638 | 93.835 | 6.108 | 8803.168 | 3.565 |
| 2SLS X4 only | -9.404 | -0.596 | 19.670 | 3.639 | 386.883 | 49.293 | 0.707 | 26.502 | 4.191 | 702.152 | 3.343 |
| 2SLS X5 only | -12.231 | 2.231 | 51.910 | 4.743 | 2696.895 | 52.145 | -2.145 | 36.193 | 4.539 | 1313.189 | 3.494 |
| 2SLS X6 only | -11.173 | 1.173 | 38.524 | 3.844 | 1484.022 | 51.509 | -1.509 | 32.397 | 3.663 | 1050.790 | 3.404 |
| 2SLS X7 only | -10.530 | 0.530 | 19.131 | 3.588 | 365.900 | 35.491 | 14.509 | 419.972 | 17.482 | 176410.800 | 3.483 |
| 2SLS X8 only | -9.652 | -0.348 | 10.318 | 3.008 | 106.486 | 49.922 | 0.078 | 13.227 | 3.291 | 174.797 | 3.303 |
| 2SLS X9 only | -9.379 | -0.621 | 39.302 | 4.505 | 1543.524 | 48.992 | 1.008 | 26.111 | 4.217 | 682.114 | 3.320 |
| 2SLS X10 only | -9.129 | -0.871 | 28.471 | 4.608 | 810.529 | 48.594 | 1.406 | 31.964 | 5.070 | 1022.631 | 3.349 |

**1.2 Results of simulation study 1a with sample size of 400**

**A 21: Full results simulation study 1a, 5 explanatory variables of the continuous process variable, average F-statistic of relevant instruments=696, average F-statistic of individual instrument=48, N=400, simulations=1000**

| | Group effect | | | | | | Post randomisation mediator effect | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean Coef. | Mean Bias | bias 2SLS/ bias OLS | SD coef. | Absolute Difference | MSE | Mean Coef. | Mean Bias | bias 2SLS/ bias OLS | SD coef. | Absolute Difference | MSE | F-stat |
| OLS All | -10.275 | 0.275 | | 1.039 | 0.858 | 1.155 | 50.514 | -0.514 | | 0.163 | 0.514 | 0.291 | |
| 2SLS All | -9.984 | **-0.016** | -0.057 | 1.044 | 0.840 | 1.089 | 50.028 | -0.028 | 0.055 | 0.176 | 0.142 | 0.032 | 162.5 |
| OLS relevant | -10.278 | 0.278 | | 1.032 | 0.853 | 1.141 | 50.511 | -0.511 | | 0.166 | 0.511 | 0.288 | |
| 2SLS relevant | -9.973 | -0.027 | -0.098 | 1.038 | 0.832 | 1.077 | 50.002 | -0.002 | 0.003 | 0.181 | 0.145 | 0.033 | 695.5 |
| OLS LASSO | -10.271 | 0.271 | | 1.034 | 0.856 | 1.142 | 50.506 | -0.506 | | 0.167 | 0.506 | 0.284 | |
| 2SLS LASSO | -9.973 | -0.027 | -0.101 | 1.040 | 0.834 | 1.082 | 50.007 | -0.007 | 0.013 | 0.181 | 0.145 | 0.033 | **553.0** |
| OLS LASSO-m | -10.268 | 0.268 | | 1.029 | 0.850 | 1.131 | 50.506 | -0.506 | | 0.165 | 0.506 | 0.283 | |
| 2SLS LASSO - m | -9.979 | -0.021 | -0.080 | 1.035 | 0.830 | 1.070 | 50.020 | -0.020 | 0.040 | 0.178 | 0.144 | 0.032 | 323.2 |
| OLS Elastic Net | -10.271 | 0.271 | | 1.034 | 0.854 | 1.141 | 50.506 | -0.506 | | 0.168 | 0.506 | 0.284 | |
| 2SLS Elastic Net | -9.975 | -0.025 | -0.092 | 1.040 | 0.834 | 1.081 | 50.012 | -0.012 | 0.023 | 0.182 | 0.146 | 0.033 | 395.1 |
| OLS Elastic Net - m | -10.272 | 0.272 | | 1.025 | 0.845 | 1.122 | 50.509 | -0.509 | | 0.164 | 0.509 | 0.286 | |
| 2SLS Elastic Net - m | -9.983 | -0.017 | -0.064 | 1.029 | 0.828 | **1.059** | 50.024 | -0.024 | 0.048 | 0.178 | 0.144 | **0.032** | 259.0 |
| OLS Stepwise | -10.282 | 0.282 | | 1.030 | 0.854 | 1.139 | 50.514 | -0.514 | | 0.165 | 0.514 | 0.291 | |
| 2SLS Stepwise | -9.980 | -0.020 | -0.071 | 1.036 | 0.830 | 1.072 | 50.002 | **-0.002** | 0.003 | 0.180 | 0.144 | 0.032 | 474.8 |
| OLS no explanatory | -10.176 | 0.176 | | 1.055 | 0.858 | 1.143 | 50.357 | -0.357 | | 0.141 | 0.357 | 0.147 | |
| 2SLS no explanatory | -9.978 | -0.022 | -0.124 | 1.068 | 0.859 | 1.140 | 50.023 | -0.023 | 0.065 | 0.243 | 0.196 | 0.060 | 13.2 |

**A 22: Full results simulation study 1a, 5 explanatory variables of the continuous process variable, average F-statistic of relevant instruments=696, average F-statistic of individual instrument=48, N=400, simulations=1000, individual explanatory variables only**

| | Group Effect | | | | | Post randomisation mediator effect | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean coefficient | Mean bias | SD coefficent | Absolute Difference | MSE | Mean coefficient | Mean bias | SD coefficent | Absolute Difference | MSE | F-statistic of instrument |
| 2SLS X1 only | -9.946 | -0.054 | 1.071 | 0.858 | 1.148 | 49.959 | 0.041 | 0.412 | 0.326 | 0.171 | 47.217 |
| 2SLS X2 only | -9.978 | -0.022 | 1.075 | 0.857 | 1.155 | 50.008 | -0.008 | 0.420 | 0.329 | 0.176 | 47.799 |
| 2SLS X3 only | -9.967 | -0.033 | 1.084 | 0.863 | 1.174 | 49.998 | 0.002 | 0.416 | 0.331 | 0.173 | 46.948 |
| 2SLS X4 only | -9.960 | -0.040 | 1.065 | 0.842 | 1.135 | 49.982 | 0.018 | 0.410 | 0.322 | 0.168 | 47.885 |
| 2SLS X5 only | -9.970 | -0.030 | 1.066 | 0.848 | 1.137 | 49.997 | 0.003 | 0.422 | 0.335 | 0.178 | 48.189 |

**A 23: Full results simulation study 1a, 5 explanatory variables of the continuous process variable, average F-statistic of relevant instruments=145, average F-statistic of individual instrument=38, N=400, simulations=1000**

| | Group effect | | | | | | Post randomisation mediator effect | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean Coef. | Mean Bias | bias 2SLS/ bias OLS | SD coef. | Absolute Difference | MSE | Mean Coef. | Mean Bias | bias 2SLS/ bias OLS | SD coef. | Absolute Difference | MSE | F-stat |
| OLS All | -10.502 | 0.502 | | 1.010 | 0.895 | 1.272 | 50.891 | -0.891 | | 0.144 | 0.891 | 0.814 | |
| 2SLS All | -10.007 | 0.007 | 0.013 | 1.042 | 0.838 | 1.084 | 50.066 | -0.066 | 0.074 | 0.190 | 0.162 | 0.041 | 34.5 |
| OLS relevant | -10.505 | 0.505 | | 0.999 | 0.894 | 1.252 | 50.887 | -0.887 | | 0.146 | 0.887 | 0.809 | |
| 2SLS relevant | -9.976 | -0.024 | -0.048 | 1.035 | 0.829 | 1.071 | 50.008 | -0.008 | 0.009 | 0.198 | 0.159 | 0.039 | 144.8 |
| OLS LASSO | -10.499 | 0.499 | | 1.003 | 0.895 | 1.253 | 50.884 | -0.884 | | 0.147 | 0.884 | 0.802 | |
| 2SLS LASSO | -9.981 | -0.019 | -0.037 | 1.037 | 0.832 | 1.075 | 50.023 | -0.023 | 0.026 | 0.197 | 0.159 | 0.039 | **115.5** |
| OLS LASSO-m | -10.494 | 0.494 | | 0.997 | 0.885 | 1.238 | 50.882 | -0.882 | | 0.145 | 0.882 | 0.799 | |
| 2SLS LASSO - m | -9.996 | -0.004 | -0.008 | 1.030 | 0.827 | 1.060 | 50.051 | -0.051 | 0.058 | 0.194 | 0.162 | 0.040 | 68.1 |
| OLS Elastic Net | -10.496 | 0.496 | | 1.003 | 0.896 | 1.251 | 50.883 | -0.883 | | 0.146 | 0.883 | 0.801 | |
| 2SLS Elastic Net | -9.985 | -0.015 | -0.031 | 1.036 | 0.831 | 1.073 | 50.031 | -0.031 | 0.035 | 0.197 | 0.160 | 0.040 | 88.4 |
| OLS Elastic Net - m | -10.498 | 0.498 | | 0.996 | 0.883 | 1.240 | 50.884 | -0.884 | | 0.144 | 0.884 | 0.802 | |
| 2SLS Elastic Net - m | -10.003 | **0.003** | 0.005 | 1.028 | 0.827 | **1.056** | 50.057 | -0.057 | 0.064 | 0.193 | 0.162 | 0.040 | 56.3 |
| OLS Stepwise | -10.507 | 0.507 | | 0.999 | 0.896 | 1.254 | 50.889 | -0.889 | | 0.145 | 0.889 | 0.812 | |
| 2SLS Stepwise | -9.987 | -0.013 | -0.025 | 1.033 | 0.827 | 1.067 | 50.012 | **-0.012** | 0.014 | 0.196 | 0.158 | **0.039** | 99.6 |
| OLS no explanatory | -10.368 | 0.368 | | 1.029 | 0.874 | 1.193 | 50.676 | -0.676 | | 0.129 | 0.676 | 0.474 | |
| 2SLS no explanatory | -10.005 | 0.005 | 0.015 | 1.063 | 0.855 | 1.130 | 50.068 | -0.068 | 0.101 | 0.264 | 0.219 | 0.074 | 9.8 |

**A 24: Full results simulation study 1a, 5 explanatory variables of the continuous process variable, average F-statistic of relevant instruments=145, average F-statistic of individual instrument=38, N=400, simulations=1000, individual explanatory variables only**

| | Group Effect | | | | | Post randomisation mediator effect | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean coefficient | Mean bias | SD coefficent | Absolute Difference | MSE | Mean coefficient | Mean bias | SD coefficent | Absolute Difference | MSE | F-statistic of instrument |
| 2SLS X1 only | -9.936 | -0.064 | 1.076 | 0.862 | 1.162 | 49.944 | 0.056 | 0.461 | 0.362 | 0.216 | 37.607 |
| 2SLS X2 only | -9.972 | -0.028 | 1.081 | 0.861 | 1.169 | 49.997 | 0.003 | 0.466 | 0.362 | 0.217 | 38.459 |
| 2SLS X3 only | -9.959 | -0.041 | 1.090 | 0.868 | 1.188 | 49.988 | 0.012 | 0.459 | 0.364 | 0.210 | 37.581 |
| 2SLS X4 only | -9.950 | -0.050 | 1.073 | 0.849 | 1.154 | 49.968 | 0.032 | 0.455 | 0.355 | 0.208 | 38.360 |
| 2SLS X5 only | -9.962 | -0.038 | 1.078 | 0.857 | 1.163 | 49.986 | 0.014 | 0.466 | 0.368 | 0.217 | 38.761 |

**A 25: Full results simulation study 1a, 5 explanatory variables of the continuous process variable, average F-statistic of relevant instruments=43, average F-statistic of individual instrument=24, N=400, simulations=1000**

| | Group effect | | | | | | Post randomisation mediator effect | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean Coef. | Mean Bias | bias 2SLS/ bias OLS | SD coef. | Absolute Difference | MSE | Mean Coef. | Mean Bias | bias 2SLS/ bias OLS | SD coef. | Absolute Difference | MSE | F-stat |
| OLS All | -10.627 | 0.627 | | 0.966 | 0.919 | 1.324 | 51.093 | -1.093 | | 0.119 | 1.093 | 1.209 | |
| 2SLS All | -10.053 | 0.053 | 0.084 | 1.034 | 0.831 | 1.071 | 50.142 | -0.142 | 0.130 | 0.225 | 0.218 | 0.071 | 11.5 |
| OLS relevant | -10.629 | 0.629 | | 0.949 | 0.914 | 1.296 | 51.091 | -1.091 | | 0.119 | 1.091 | 1.204 | |
| 2SLS relevant | -9.983 | -0.017 | -0.027 | 1.032 | 0.827 | 1.065 | 50.022 | -0.022 | 0.020 | 0.242 | 0.195 | 0.059 | 40.1 |
| OLS LASSO | -10.624 | 0.624 | | 0.952 | 0.913 | 1.294 | 51.089 | -1.089 | | 0.120 | 1.089 | 1.201 | |
| 2SLS LASSO | -10.001 | **0.001** | 0.001 | 1.031 | 0.826 | 1.061 | 50.057 | **-0.057** | 0.053 | 0.239 | 0.197 | 0.060 | **33.3** |
| OLS LASSO-m | -10.620 | 0.620 | | 0.951 | 0.906 | 1.288 | 51.088 | -1.088 | | 0.119 | 1.088 | 1.197 | |
| 2SLS LASSO - m | -10.034 | 0.034 | 0.055 | 1.022 | 0.822 | 1.044 | 50.114 | -0.114 | 0.105 | 0.232 | 0.209 | 0.067 | 20.8 |
| OLS Elastic Net | -10.622 | 0.622 | | 0.954 | 0.915 | 1.295 | 51.089 | -1.089 | | 0.120 | 1.089 | 1.201 | |
| 2SLS Elastic Net | -10.005 | 0.005 | 0.009 | 1.029 | 0.825 | 1.058 | 50.069 | -0.069 | 0.063 | 0.238 | 0.199 | 0.061 | 28.5 |
| OLS Elastic Net - m | -10.621 | 0.621 | | 0.949 | 0.902 | 1.286 | 51.089 | -1.089 | | 0.119 | 1.089 | 1.199 | |
| 2SLS Elastic Net - m | -10.039 | 0.039 | 0.063 | 1.018 | 0.818 | **1.036** | 50.121 | -0.121 | 0.111 | 0.230 | 0.210 | 0.068 | 18.4 |
| OLS Stepwise | -10.630 | 0.630 | | 0.949 | 0.913 | 1.297 | 51.092 | -1.092 | | 0.120 | 1.092 | 1.206 | |
| 2SLS Stepwise | -10.017 | 0.017 | 0.027 | 1.024 | 0.820 | 1.048 | 50.061 | -0.061 | 0.056 | 0.237 | 0.197 | **0.060** | 30.0 |
| OLS no explanatory | -10.526 | 0.526 | | 0.981 | 0.890 | 1.239 | 50.935 | -0.935 | | 0.113 | 0.935 | 0.886 | |
| 2SLS no explanatory | -10.060 | 0.060 | 0.114 | 1.051 | 0.846 | 1.108 | 50.157 | -0.157 | 0.168 | 0.312 | 0.284 | 0.122 | 6.0 |

**A 26: Full results simulation study 1a, 5 explanatory variables of the continuous process variable, average F-statistic of relevant instruments=43, average F-statistic of individual instrument=24, N=400, simulations=1000, individual explanators only**

| | Group Effect | | | | | Post randomisation mediator effect | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean coefficient | Mean bias | SD coefficent | Absolute Difference | MSE | Mean coefficient | Mean bias | SD coefficent | Absolute Difference | MSE | F-statistic of instrument |
| 2SLS X1 only | -9.912 | -0.088 | 1.108 | 0.884 | 1.235 | 49.903 | 0.097 | 0.612 | 0.462 | 0.384 | 23.321 |
| 2SLS X2 only | -9.961 | -0.039 | 1.108 | 0.878 | 1.228 | 49.970 | 0.030 | 0.618 | 0.456 | 0.382 | 24.212 |
| 2SLS X3 only | -9.942 | -0.058 | 1.121 | 0.891 | 1.258 | 49.965 | 0.035 | 0.582 | 0.456 | 0.339 | 23.499 |
| 2SLS X4 only | -9.929 | -0.071 | 1.108 | 0.873 | 1.231 | 49.935 | 0.065 | 0.594 | 0.448 | 0.356 | 24.009 |
| 2SLS X5 only | -9.942 | -0.058 | 1.124 | 0.885 | 1.266 | 49.957 | 0.043 | 0.606 | 0.463 | 0.369 | 24.376 |

**A 27: Full results simulation study 1a, 5 explanatory variables of the continuous process variable, average F-statistic of relevant instruments=21, average F-statistic of individual instrument=15, N=400, simulations=1000**

| | Group effect | | | | | | Post randomisation mediator effect | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean Coef. | Mean Bias | bias 2SLS/ bias OLS | SD coef. | Absolute Difference | MSE | Mean Coef. | Mean Bias | bias 2SLS/ bias OLS | SD coef. | Absolute Difference | MSE | F-stat |
| OLS All | -10.657 | 0.657 | | 0.940 | 0.919 | 1.315 | 51.140 | -1.140 | | 0.108 | 1.140 | 1.312 | |
| 2SLS All | -10.108 | 0.108 | 0.164 | 1.024 | 0.825 | 1.059 | 50.233 | -0.233 | 0.204 | 0.262 | 0.292 | 0.123 | 6.1 |
| OLS relevant | -10.659 | 0.659 | | 0.920 | 0.911 | 1.281 | 51.139 | -1.139 | | 0.106 | 1.139 | 1.087 | |
| 2SLS relevant | -9.994 | -0.006 | -0.010 | 1.033 | 0.827 | 1.066 | 50.042 | -0.042 | 0.037 | 0.296 | 0.239 | 0.197 | 20.0 |
| OLS LASSO | -10.649 | 0.649 | | 0.923 | 0.909 | 1.273 | 51.134 | -1.134 | | 0.106 | 1.134 | 1.308 | |
| 2SLS LASSO | -10.001 | **0.001** | 0.001 | 1.426 | 0.851 | 2.031 | 50.116 | **-0.116** | 0.102 | 0.286 | 0.248 | **0.089** | **17.3** |
| OLS LASSO-m | -10.652 | 0.652 | | 0.923 | 0.905 | 1.275 | 51.137 | -1.137 | | 0.107 | 1.137 | 1.298 | |
| 2SLS LASSO - m | -10.080 | 0.080 | 0.123 | 1.011 | 0.814 | 1.028 | 50.192 | -0.192 | 0.168 | 0.275 | 0.274 | 0.095 | 10.9 |
| OLS Elastic Net | -10.651 | 0.651 | | 0.925 | 0.911 | 1.278 | 51.136 | -1.136 | | 0.106 | 1.136 | 1.305 | |
| 2SLS Elastic Net | -10.004 | 0.004 | 0.007 | 1.426 | 0.852 | 2.032 | 50.119 | -0.119 | 0.105 | 0.289 | 0.252 | 0.112 | 15.7 |
| OLS Elastic Net - m | -10.652 | 0.652 | | 0.923 | 0.903 | 1.277 | 51.138 | -1.138 | | 0.107 | 1.138 | 1.302 | |
| 2SLS Elastic Net - m | -10.085 | 0.085 | 0.131 | 1.010 | 0.813 | **1.027** | 50.197 | -0.197 | 0.173 | 0.273 | 0.277 | 0.098 | 10.0 |
| OLS Stepwise | -10.652 | 0.652 | | 0.920 | 0.906 | 1.272 | 51.138 | -1.138 | | 0.107 | 1.138 | 1.306 | |
| 2SLS Stepwise | -10.058 | 0.058 | 0.089 | 1.013 | 0.811 | 1.029 | 50.137 | -0.137 | 0.121 | 0.279 | 0.252 | 0.113 | 15.7 |
| OLS no explanatory | -10.591 | 0.591 | | 0.950 | 0.895 | 1.250 | 51.037 | -1.037 | | 0.105 | 1.037 | 1.306 | |
| 2SLS no explanatory | -10.124 | 0.124 | 0.209 | 1.039 | 0.839 | 1.094 | 50.259 | -0.259 | 0.250 | 0.361 | 0.365 | 0.097 | 4.0 |

**A 28: Full results simulation study 1a, 5 explanatory variables of the continuous process variable, average F-statistic of relevant instruments=21, average F-statistic of individual instrument=15, N=400, simulations=1000, individual explanatory variables only**

| | Group Effect | | | | | Post randomisation mediator effect | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean coefficient | Mean bias | SD coefficent | Absolute Difference | MSE | Mean coefficient | Mean bias | SD coefficent | Absolute Difference | MSE | F-statistic of instrument |
| 2SLS X1 only | -9.838 | -0.162 | 1.518 | 0.968 | 2.328 | 49.760 | 0.240 | 2.235 | 0.680 | 5.048 | 14.991 |
| 2SLS X2 only | -9.937 | -0.063 | 1.180 | 0.917 | 1.394 | 49.998 | 0.002 | 2.706 | 0.672 | 7.316 | 15.785 |
| 2SLS X3 only | -9.916 | -0.084 | 1.186 | 0.933 | 1.413 | 49.927 | 0.073 | 0.762 | 0.581 | 0.586 | 15.230 |
| 2SLS X4 only | -9.895 | -0.105 | 1.180 | 0.916 | 1.401 | 49.875 | 0.125 | 0.876 | 0.584 | 0.782 | 15.569 |
| 2SLS X5 only | -9.929 | -0.071 | 1.268 | 0.941 | 1.611 | 49.931 | 0.069 | 0.985 | 0.606 | 0.974 | 15.872 |

**A 29: Full results simulation study 1a, 5 explanatory variables of the continuous process variable, average F-statistic of relevant instruments=7.2, average F-statistic of individual instrument=6.6, N=400, simulations=1000**

| | Group effect | | | | | | Post randomisation mediator effect | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean Coef. | Mean Bias | bias 2SLS/ bias OLS | SD coef. | Absolute Difference | MSE | Mean Coef. | Mean Bias | bias 2SLS/ bias OLS | SD coef. | Absolute Difference | MSE | F-stat |
| OLS All | -10.672 | 0.672 | | 0.913 | 0.912 | 1.284 | 51.162 | -1.162 | | 0.097 | 1.162 | 1.360 | |
| 2SLS All | -10.266 | 0.266 | 0.396 | 0.998 | 0.822 | 1.066 | 50.494 | -0.494 | 0.425 | 0.345 | 0.525 | 0.363 | 2.56 |
| OLS relevant | -10.674 | 0.674 | | 0.889 | 0.899 | 1.245 | 51.161 | -1.161 | | 0.094 | 1.161 | 1.356 | |
| 2SLS relevant | -10.039 | 0.039 | 0.058 | 1.048 | 0.837 | 1.098 | 50.121 | -0.121 | 0.104 | 0.458 | 0.379 | 0.224 | 7.21 |
| OLS LASSO | -10.646 | 0.646 | | 0.894 | 0.886 | 1.215 | 51.113 | -1.113 | | 0.093 | 1.113 | 1.247 | |
| 2SLS LASSO | 6.003 | -16.003 | -24.766 | 22.337 | 17.410 | 754.533 | 50.444 | -0.444 | 0.399 | 0.396 | 0.494 | 0.354 | **12.30** |
| OLS LASSO-m | -10.667 | 0.667 | | 0.895 | 0.896 | 1.244 | 51.157 | -1.157 | | 0.096 | 1.157 | 1.349 | |
| 2SLS LASSO - m | -9.786 | -0.214 | -0.321 | 4.600 | 1.216 | 21.183 | 50.416 | -0.416 | 0.360 | 0.396 | 0.487 | 0.330 | 5.26 |
| OLS Elastic Net | -10.649 | 0.649 | | 0.893 | 0.887 | 1.219 | 51.114 | -1.114 | | 0.093 | 1.114 | 1.249 | |
| 2SLS Elastic Net | 5.887 | -15.887 | -24.467 | 22.437 | 17.272 | 755.305 | 50.455 | -0.455 | 0.408 | 0.381 | 0.501 | 0.352 | 11.61 |
| OLS Elastic Net - m | -10.668 | 0.668 | | 0.896 | 0.897 | 1.249 | 51.158 | -1.158 | | 0.096 | 1.158 | 1.350 | |
| 2SLS Elastic Net - m | -9.724 | -0.276 | -0.413 | 4.885 | 1.293 | 23.916 | 50.420 | -0.420 | 0.363 | 0.384 | 0.483 | 0.324 | 4.91 |
| OLS Stepwise | -10.663 | 0.663 | | 0.894 | 0.895 | 1.239 | 51.156 | -1.156 | | 0.096 | 1.156 | 1.346 | |
| 2SLS Stepwise | -10.212 | **0.212** | 0.320 | 0.997 | 0.808 | **1.038** | 50.402 | **-0.402** | 0.348 | 0.400 | 0.474 | **0.322** | 6.76 |
| OLS no explanatory | -10.646 | 0.646 | | 0.913 | -10.635 | 1.249 | 51.122 | -1.122 | | 0.096 | 1.122 | 1.269 | |
| 2SLS no explanatory | -10.296 | 0.296 | 0.459 | 1.018 | -10.289 | 1.123 | 50.538 | -0.538 | 0.479 | 0.470 | 0.604 | 0.510 | 2.11 |

**A 30: Full results simulation study 1a, 5 explanatory variables of the continuous process variable, average F-statistic of relevant instruments=7.2, average F-statistic of individual instrument=6.6, N=400, simulations=1000, individual explanatory variables only**

| | Group Effect | | | | | Post randomisation mediator effect | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean coefficient | Mean bias | SD coefficent | Absolute Difference | MSE | Mean coefficient | Mean bias | SD coefficent | Absolute Difference | MSE | F-statistic of instrument |
| 2SLS X1 only | -9.981 | -0.019 | 5.751 | 1.524 | 33.043 | 49.778 | 0.222 | 9.475 | 1.806 | 89.742 | 6.208 |
| 2SLS X2 only | -9.856 | -0.144 | 1.948 | 1.175 | 3.813 | 49.764 | 0.236 | 2.752 | 1.230 | 7.623 | 6.773 |
| 2SLS X3 only | -9.831 | -0.169 | 5.276 | 1.437 | 27.836 | 49.760 | 0.240 | 5.121 | 1.439 | 26.256 | 6.448 |
| 2SLS X4 only | -9.883 | -0.117 | 6.337 | 1.430 | 40.126 | 49.832 | 0.168 | 7.424 | 1.475 | 55.094 | 6.597 |
| 2SLS X5 only | -10.375 | 0.375 | 10.814 | 1.827 | 116.969 | 50.401 | -0.401 | 12.253 | 1.856 | 150.139 | 6.784 |

**A 31: Full results simulation study 1a, 10 explanatory variables of the continuous process variable, average F-statistic of relevant instruments=390, average F-statistic of individual instrument=22, N=400, simulations=1000**

| | Group effect | | | | | | Post randomisation mediator effect | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean Coef. | Mean Bias | bias 2SLS/ bias OLS | SD coef. | Absolute Difference | MSE | Mean Coef. | Mean Bias | bias 2SLS/ bias OLS | SD coef. | Absolute Difference | MSE | F-stat |
| OLS All | -10.245 | 0.245 | | 1.033 | 0.847 | 1.127 | 50.454 | -0.454 | | 0.155 | 0.454 | 0.230 | |
| 2SLS All | -9.992 | -0.008 | -0.031 | 1.046 | 0.838 | 1.094 | 50.022 | -0.022 | 0.049 | 0.163 | 0.133 | 0.027 | 175.3 |
| OLS relevant | -10.247 | 0.247 | | 1.027 | 0.847 | 1.114 | 50.454 | -0.454 | | 0.156 | 0.454 | 0.230 | |
| 2SLS relevant | -9.986 | -0.014 | -0.059 | 1.041 | 0.840 | 1.082 | 50.007 | -0.007 | 0.015 | 0.164 | 0.132 | 0.027 | 390.5 |
| OLS LASSO | -10.245 | 0.245 | | 1.030 | 0.851 | 1.119 | 50.449 | -0.449 | | 0.156 | 0.449 | 0.226 | |
| 2SLS LASSO | -9.990 | -0.010 | -0.042 | 1.043 | 0.840 | 1.086 | 50.013 | -0.013 | 0.028 | 0.164 | 0.133 | 0.027 | **321.2** |
| OLS LASSO-m | -10.244 | 0.244 | | 1.029 | 0.846 | 1.117 | 50.449 | -0.449 | | 0.155 | 0.449 | 0.226 | |
| 2SLS LASSO - m | -9.993 | **-0.007** | -0.030 | 1.042 | 0.837 | 1.085 | 50.019 | -0.019 | 0.043 | 0.163 | 0.132 | 0.027 | 241.2 |
| OLS Elastic Net | -10.245 | 0.245 | | 1.033 | 0.852 | 1.127 | 50.450 | -0.450 | | 0.156 | 0.450 | 0.227 | |
| 2SLS Elastic Net | -9.991 | -0.009 | -0.036 | 1.046 | 0.842 | 1.092 | 50.015 | -0.015 | 0.033 | 0.164 | 0.132 | 0.027 | 287.0 |
| OLS Elastic Net - m | -10.244 | 0.244 | | 1.031 | 0.848 | 1.121 | 50.450 | -0.450 | | 0.155 | 0.450 | 0.227 | |
| 2SLS Elastic Net - m | -9.992 | -0.008 | -0.032 | 1.044 | 0.840 | 1.088 | 50.020 | -0.020 | 0.045 | 0.164 | 0.133 | 0.027 | 224.6 |
| OLS Stepwise | -10.250 | 0.250 | | 1.022 | 0.843 | 1.105 | 50.454 | -0.454 | | 0.155 | 0.454 | 0.230 | |
| 2SLS Stepwise | -9.991 | -0.009 | -0.037 | 1.035 | 0.834 | **1.071** | 50.005 | **-0.005** | 0.010 | 0.164 | 0.132 | **0.027** | 328.0 |
| OLS no explanatory | -10.149 | 0.149 | | 1.042 | 0.849 | 1.108 | 50.276 | -0.276 | | 0.119 | 0.277 | 0.090 | |
| 2SLS no explanatory | -10.004 | 0.004 | 0.026 | 1.063 | 0.854 | 1.129 | 50.030 | -0.030 | 0.109 | 0.302 | 0.241 | 0.092 | 7.4 |

**A 32: Simulation study 1a, 10 explanatory variables of the continuous process variable, average F-statistic of relevant instruments=390, average F-statistic of individual instrument=22, N=400, simulations=1000, individual explanatory variables only**

| | Group Effect | | | | | Post randomisation mediator effect | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean coefficient | Mean bias | SD coefficent | Absolute Difference | MSE | Mean coefficient | Mean bias | SD coefficent | Absolute Difference | MSE | F-statistic of instrument |
| 2SLS X1 only | -9.998 | -0.002 | 1.084 | 0.867 | 1.174 | 49.994 | 0.006 | 0.555 | 0.424 | 0.308 | 22.420 |
| 2SLS X2 only | -9.979 | -0.021 | 1.109 | 0.879 | 1.229 | 49.990 | 0.010 | 0.579 | 0.429 | 0.335 | 21.860 |
| 2SLS X3 only | -9.965 | -0.035 | 1.116 | 0.883 | 1.246 | 49.960 | 0.040 | 0.547 | 0.424 | 0.301 | 21.845 |
| 2SLS X4 only | -10.001 | 0.001 | 1.098 | 0.873 | 1.204 | 49.996 | 0.004 | 0.561 | 0.436 | 0.314 | 22.192 |
| 2SLS X5 only | -9.976 | -0.024 | 1.076 | 0.862 | 1.158 | 49.967 | 0.033 | 0.573 | 0.443 | 0.329 | 22.372 |
| 2SLS X6 only | -9.970 | -0.030 | 1.122 | 0.893 | 1.258 | 49.966 | 0.034 | 0.566 | 0.445 | 0.321 | 21.709 |
| 2SLS X7 only | -9.974 | -0.026 | 1.099 | 0.877 | 1.207 | 49.992 | 0.008 | 0.542 | 0.415 | 0.293 | 22.247 |
| 2SLS X8 only | -9.988 | -0.012 | 1.115 | 0.884 | 1.241 | 49.988 | 0.012 | 0.558 | 0.437 | 0.312 | 22.447 |
| 2SLS X9 only | -9.981 | -0.019 | 1.111 | 0.889 | 1.233 | 49.998 | 0.002 | 0.574 | 0.440 | 0.329 | 21.970 |
| 2SLS X10 only | -9.977 | -0.023 | 1.119 | 0.883 | 1.251 | 49.979 | 0.021 | 0.571 | 0.446 | 0.326 | 21.886 |

**A 33: Full results simulation study 1a, 10 explanatory variables of the continuous process variable, average F-statistic of relevant instruments=84, average F-statistic of individual instrument=18, N=400, simulations=1000**

| | Group effect | | | | | | Post randomisation mediator effect | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean Coef. | Mean Bias | bias 2SLS/ bias OLS | SD coef. | Absolute Difference | MSE | Mean Coef. | Mean Bias | bias 2SLS/ bias OLS | SD coef. | Absolute Difference | MSE | F-stat |
| OLS All | -10.450 | 0.450 | | 1.004 | 0.883 | 1.210 | 50.800 | -0.800 | | 0.142 | 0.800 | 0.659 | |
| 2SLS All | -10.012 | 0.012 | 0.026 | 1.043 | 0.836 | 1.088 | 50.055 | -0.055 | 0.069 | 0.175 | 0.148 | 0.034 | 38.1 |
| OLS relevant | -10.451 | 0.451 | | 0.996 | 0.878 | 1.195 | 50.800 | -0.800 | | 0.144 | 0.800 | 0.660 | |
| 2SLS relevant | -9.993 | -0.007 | -0.015 | 1.038 | 0.838 | 1.075 | 50.021 | -0.021 | 0.026 | 0.176 | 0.143 | 0.031 | 83.7 |
| OLS LASSO | -10.449 | 0.449 | | 1.000 | 0.882 | 1.200 | 50.795 | -0.795 | | 0.143 | 0.795 | 0.652 | |
| 2SLS LASSO | -10.004 | 0.004 | 0.008 | 1.039 | 0.837 | 1.079 | 50.036 | -0.036 | 0.045 | 0.177 | 0.146 | 0.032 | 69.2 |
| OLS LASSO-m | -10.449 | 0.449 | | 0.999 | 0.876 | 1.198 | 50.795 | -0.795 | | 0.142 | 0.795 | 0.652 | |
| 2SLS LASSO - m | -10.011 | 0.011 | 0.025 | 1.037 | 0.833 | 1.075 | 50.051 | -0.051 | 0.064 | 0.175 | 0.147 | 0.033 | 52.2 |
| OLS Elastic Net | -10.451 | 0.451 | | 1.005 | 0.885 | 1.211 | 50.796 | -0.796 | | 0.143 | 0.796 | 0.653 | |
| 2SLS Elastic Net | -10.008 | 0.008 | 0.018 | 1.043 | 0.839 | 1.086 | 50.040 | -0.040 | 0.050 | 0.176 | 0.146 | 0.033 | 63.1 |
| OLS Elastic Net - m | -10.448 | 0.448 | | 1.001 | 0.879 | 1.202 | 50.795 | -0.795 | | 0.142 | 0.795 | 0.653 | |
| 2SLS Elastic Net - m | -10.010 | 0.010 | 0.023 | 1.039 | 0.835 | 1.079 | 50.052 | -0.052 | 0.065 | 0.176 | 0.148 | 0.033 | 49.2 |
| OLS Stepwise | -10.454 | 0.454 | | 0.991 | 0.875 | 1.188 | 50.799 | -0.799 | | 0.142 | 0.799 | 0.659 | |
| 2SLS Stepwise | -10.004 | **0.004** | 0.010 | 1.031 | 0.830 | **1.063** | 50.024 | **-0.024** | 0.030 | 0.178 | 0.144 | **0.032** | **70.4** |
| OLS no explanatory | -10.303 | 0.303 | | 1.020 | 0.854 | 1.132 | 50.537 | -0.537 | | 0.113 | 0.537 | 0.301 | |
| 2SLS no explanatory | -10.031 | 0.031 | 0.103 | 1.061 | 0.853 | 1.125 | 50.077 | -0.077 | 0.144 | 0.319 | 0.263 | 0.108 | 6.2 |

**A 34: Simulation study 1a, 10 explanatory variables of the continuous process variable, average F-statistic of relevant instruments=84, average F-statistic of individual instrument=18, N=400, simulations=1000, individual explanatory variables only**

| | Group Effect | | | | | Post randomisation mediator effect | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean coefficient | Mean bias | SD coefficent | Absolute Difference | MSE | Mean coefficient | Mean bias | SD coefficent | Absolute Difference | MSE | F-statistic of instrument |
| 2SLS X1 only | -9.989 | -0.011 | 1.100 | 0.876 | 1.209 | 49.978 | 0.022 | 0.618 | 0.464 | 0.382 | 18.997 |
| 2SLS X2 only | -10.025 | 0.025 | 1.873 | 0.932 | 3.505 | 50.052 | -0.052 | 2.400 | 0.539 | 5.758 | 18.400 |
| 2SLS X3 only | -9.953 | -0.047 | 1.134 | 0.893 | 1.286 | 49.939 | 0.061 | 0.606 | 0.466 | 0.371 | 18.476 |
| 2SLS X4 only | -9.990 | -0.010 | 1.111 | 0.880 | 1.234 | 49.975 | 0.025 | 0.628 | 0.479 | 0.394 | 18.802 |
| 2SLS X5 only | -9.967 | -0.033 | 1.093 | 0.873 | 1.194 | 49.943 | 0.057 | 0.649 | 0.490 | 0.424 | 18.741 |
| 2SLS X6 only | -9.955 | -0.045 | 1.142 | 0.904 | 1.305 | 49.945 | 0.055 | 0.636 | 0.493 | 0.407 | 18.193 |
| 2SLS X7 only | -9.960 | -0.040 | 1.115 | 0.887 | 1.244 | 49.970 | 0.030 | 0.607 | 0.456 | 0.369 | 18.780 |
| 2SLS X8 only | -9.973 | -0.027 | 1.133 | 0.894 | 1.283 | 49.968 | 0.032 | 0.618 | 0.477 | 0.383 | 18.985 |
| 2SLS X9 only | -9.967 | -0.033 | 1.132 | 0.900 | 1.282 | 49.978 | 0.022 | 0.633 | 0.481 | 0.401 | 18.665 |
| 2SLS X10 only | -9.966 | -0.034 | 1.140 | 0.894 | 1.300 | 49.956 | 0.044 | 0.634 | 0.490 | 0.404 | 18.364 |

**A 35: Full results simulation study 1a, 10 explanatory variables of the continuous process variable, f-statistic of relevant instruments=27, average F-statistic of individual instrument=13, N=400, simulations=1000**

| | Group effect | | | | | | Post randomisation mediator effect | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean Coef. | Mean Bias | bias 2SLS/ bias OLS | SD coef. | Absolute Difference | MSE | Mean Coef. | Mean Bias | bias 2SLS/ bias OLS | SD coef. | Absolute Difference | MSE | F-stat |
| OLS All | -10.571 | 0.571 | | 0.969 | 0.908 | 1.264 | 50.999 | -0.999 | | 0.120 | 0.999 | 1.012 | |
| 2SLS All | -10.047 | 0.047 | 0.082 | 1.035 | 0.830 | 1.073 | 50.115 | -0.115 | 0.115 | 0.202 | 0.189 | 0.054 | 12.7 |
| OLS relevant | -10.572 | 0.572 | | 0.958 | 0.900 | 1.244 | 50.999 | -0.999 | | 0.121 | 0.999 | 1.012 | |
| 2SLS relevant | -10.008 | 0.008 | 0.015 | 1.032 | 0.832 | 1.064 | 50.046 | -0.046 | 0.046 | 0.206 | 0.170 | 0.044 | 26.9 |
| OLS LASSO | -10.571 | 0.571 | | 0.961 | 0.902 | 1.249 | 50.996 | -0.996 | | 0.121 | 0.996 | 1.006 | |
| 2SLS LASSO | -10.030 | **0.030** | 0.052 | 1.030 | 0.830 | 1.062 | 50.080 | -0.080 | 0.080 | 0.205 | 0.178 | **0.048** | 22.5 |
| OLS LASSO-m | -10.571 | 0.571 | | 0.961 | 0.900 | 1.250 | 50.996 | -0.996 | | 0.120 | 0.996 | 1.006 | |
| 2SLS LASSO - m | -10.045 | 0.045 | 0.078 | 1.028 | 0.826 | 1.058 | 50.107 | -0.107 | 0.107 | 0.202 | 0.186 | 0.052 | 17.2 |
| OLS Elastic Net | -10.573 | 0.573 | | 0.965 | 0.906 | 1.258 | 50.996 | -0.996 | | 0.121 | 0.996 | 1.007 | |
| 2SLS Elastic Net | -10.034 | 0.034 | 0.059 | 1.033 | 0.833 | 1.068 | 50.083 | -0.083 | 0.084 | 0.205 | 0.179 | 0.049 | 21.2 |
| OLS Elastic Net - m | -10.568 | 0.568 | | 0.964 | 0.902 | 1.252 | 50.996 | -0.996 | | 0.120 | 0.996 | 1.007 | |
| 2SLS Elastic Net - m | -10.043 | 0.043 | 0.075 | 1.029 | 0.828 | 1.061 | 50.108 | -0.108 | 0.108 | 0.202 | 0.186 | 0.052 | 16.5 |
| OLS Stepwise | -10.570 | 0.570 | | 0.957 | 0.899 | 1.240 | 50.997 | -0.997 | | 0.120 | 0.997 | 1.008 | |
| 2SLS Stepwise | -10.032 | 0.032 | 0.057 | 1.024 | 0.826 | **1.048** | 50.074 | **-0.074** | 0.074 | 0.208 | 0.179 | 0.049 | **23.2** |
| OLS no explanatory | -10.445 | 0.445 | | 0.984 | 0.869 | 1.164 | 50.776 | -0.776 | | 0.105 | 0.776 | 0.613 | |
| 2SLS no explanatory | -10.081 | 0.081 | 0.181 | 1.053 | 0.847 | 1.115 | 50.162 | -0.162 | 0.209 | 0.360 | 0.321 | 0.156 | 4.4 |

**A 36: Simulation study 1a, 10 explanatory variables of the continuous process variable, average F-statistic of relevant instruments=27, average F-statistic of individual instrument=13, N=400, simulations=1000, individual explanatory variables only**

| | Group Effect | | | | | Post randomisation mediator effect | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean coefficient | Mean bias | SD coefficent | Absolute Difference | MSE | Mean coefficient | Mean bias | SD coefficent | Absolute Difference | MSE | F-statistic of instrument |
| 2SLS X1 only | -9.969 | -0.031 | 1.156 | 0.909 | 1.335 | 49.943 | 0.057 | 0.817 | 0.573 | 0.670 | 13.447 |
| 2SLS X2 only | -9.956 | -0.044 | 1.212 | 0.927 | 1.470 | 49.938 | 0.062 | 0.831 | 0.587 | 0.693 | 12.915 |
| 2SLS X3 only | -9.925 | -0.075 | 1.196 | 0.928 | 1.434 | 49.885 | 0.115 | 0.812 | 0.584 | 0.672 | 13.050 |
| 2SLS X4 only | -9.960 | -0.040 | 1.184 | 0.920 | 1.401 | 49.921 | 0.079 | 0.878 | 0.601 | 0.776 | 13.323 |
| 2SLS X5 only | -10.074 | 0.074 | 4.396 | 1.047 | 19.313 | 50.272 | -0.272 | 12.190 | 0.997 | 148.515 | 13.062 |
| 2SLS X6 only | -9.920 | -0.080 | 1.221 | 0.942 | 1.496 | 49.890 | 0.110 | 0.849 | 0.623 | 0.732 | 12.700 |
| 2SLS X7 only | -9.925 | -0.075 | 1.188 | 0.922 | 1.416 | 49.916 | 0.084 | 0.834 | 0.572 | 0.702 | 13.240 |
| 2SLS X8 only | -9.941 | -0.059 | 1.202 | 0.932 | 1.447 | 49.921 | 0.079 | 0.816 | 0.589 | 0.671 | 13.405 |
| 2SLS X9 only | -9.932 | -0.068 | 1.226 | 0.943 | 1.506 | 49.930 | 0.070 | 0.821 | 0.595 | 0.679 | 13.260 |
| 2SLS X10 only | -9.936 | -0.064 | 1.227 | 0.936 | 1.508 | 49.897 | 0.103 | 0.892 | 0.616 | 0.806 | 12.853 |

**A 37: Full results simulation study 1a, 10 explanatory variables of the continuous process variable, f-statistic of relevant instruments=15, average F-statistic of individual instrument=9.7, N=400, simulations=1000**

| | Group effect | | | | | | Post randomisation mediator effect | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean Coef. | Mean Bias | bias 2SLS/ bias OLS | SD coef. | Absolute Difference | MSE | Mean Coef. | Mean Bias | bias 2SLS/ bias OLS | SD coef. | Absolute Difference | MSE | F-stat |
| OLS All | -10.604 | 0.604 | | 0.951 | 0.910 | 1.269 | 51.051 | -1.051 | | 0.108 | 1.051 | 1.117 | |
| 2SLS All | -10.082 | 0.082 | 0.136 | 1.026 | 0.824 | 1.059 | 50.175 | -0.175 | 0.167 | 0.228 | 0.237 | 0.083 | 7.3 |
| OLS relevant | -10.606 | 0.606 | | 0.938 | 0.902 | 1.247 | 51.051 | -1.051 | | 0.109 | 1.051 | 1.116 | |
| 2SLS relevant | -10.025 | 0.025 | 0.041 | 1.027 | 0.828 | 1.055 | 50.074 | -0.074 | 0.070 | 0.236 | 0.199 | 0.061 | 14.9 |
| OLS LASSO | -10.605 | 0.605 | | 0.943 | 0.904 | 1.255 | 51.045 | -1.045 | | 0.107 | 1.045 | 1.104 | |
| 2SLS LASSO | -10.062 | **0.062** | 0.103 | 1.022 | 0.824 | 1.048 | 50.133 | **-0.133** | 0.128 | 0.234 | 0.219 | 0.072 | 13.1 |
| OLS LASSO-m | -10.605 | 0.605 | | 0.943 | 0.903 | 1.253 | 51.049 | -1.049 | | 0.109 | 1.049 | 1.112 | |
| 2SLS LASSO - m | -10.077 | 0.077 | 0.127 | 1.018 | 0.820 | 1.041 | 50.164 | -0.164 | 0.156 | 0.229 | 0.231 | 0.079 | 9.9 |
| OLS Elastic Net | -10.606 | 0.606 | | 0.944 | 0.903 | 1.258 | 51.047 | -1.047 | | 0.108 | 1.047 | 1.107 | |
| 2SLS Elastic Net | -10.064 | 0.064 | 0.105 | 1.024 | 0.825 | 1.051 | 50.134 | -0.134 | 0.128 | 0.233 | 0.219 | **0.072** | 12.5 |
| OLS Elastic Net - m | -10.603 | 0.603 | | 0.944 | 0.904 | 1.255 | 51.049 | -1.049 | | 0.108 | 1.049 | 1.113 | |
| 2SLS Elastic Net - m | -10.076 | 0.076 | 0.126 | 1.019 | 0.821 | 1.043 | 50.164 | -0.164 | 0.156 | 0.229 | 0.232 | 0.079 | 9.6 |
| OLS Stepwise | -10.598 | 0.598 | | 0.939 | 0.897 | 1.238 | 51.046 | -1.046 | | 0.108 | 1.046 | 1.106 | |
| 2SLS Stepwise | -10.067 | 0.067 | 0.112 | 1.016 | 0.818 | **1.035** | 50.136 | -0.136 | 0.130 | 0.234 | 0.221 | 0.073 | **13.4** |
| OLS no explanatory | -10.510 | 0.510 | | 0.960 | 0.878 | 1.181 | 50.884 | -0.884 | | 0.100 | 0.884 | 0.791 | |
| 2SLS no explanatory | -10.129 | 0.129 | 0.254 | 1.046 | 0.843 | 1.109 | 50.244 | -0.244 | 0.276 | 0.402 | 0.385 | 0.221 | 3.4 |

**A 38: Simulation study 1a, 10 explanatory variables of the continuous process variable, average F-statistic of relevant instruments=15, average F-statistic of individual instrument=9.7, N=400, simulations=1000, individual explanatory variables only**

| | Group Effect | | | | | Post randomisation mediator effect | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean coefficient | Mean bias | SD coefficent | Absolute Difference | MSE | Mean coefficient | Mean bias | SD coefficent | Absolute Difference | MSE | F-statistic of instrument |
| 2SLS X1 only | -9.929 | -0.071 | 1.355 | 0.973 | 1.840 | 49.884 | 0.116 | 1.372 | 0.735 | 1.895 | 9.964 |
| 2SLS X2 only | -11.447 | 1.447 | 36.284 | 2.476 | 1317.300 | 51.803 | -1.803 | 47.366 | 2.636 | 2244.566 | 9.511 |
| 2SLS X3 only | -9.888 | -0.112 | 1.456 | 1.017 | 2.131 | 49.774 | 0.226 | 1.985 | 0.817 | 3.989 | 9.657 |
| 2SLS X4 only | -9.989 | -0.011 | 1.792 | 1.017 | 3.209 | 49.955 | 0.045 | 2.495 | 0.818 | 6.221 | 9.884 |
| 2SLS X5 only | -9.884 | -0.116 | 1.494 | 1.010 | 2.244 | 49.753 | 0.247 | 1.919 | 0.837 | 3.738 | 9.569 |
| 2SLS X6 only | -9.855 | -0.145 | 1.454 | 1.025 | 2.132 | 49.803 | 0.197 | 1.356 | 0.820 | 1.875 | 9.324 |
| 2SLS X7 only | -10.156 | 0.156 | 9.229 | 1.276 | 85.105 | 50.113 | -0.113 | 9.735 | 1.054 | 94.683 | 9.788 |
| 2SLS X8 only | -9.884 | -0.116 | 1.390 | 1.012 | 1.942 | 49.826 | 0.174 | 1.396 | 0.762 | 1.978 | 9.915 |
| 2SLS X9 only | -9.728 | -0.272 | 5.426 | 1.166 | 29.481 | 49.750 | 0.250 | 3.889 | 0.855 | 15.173 | 9.856 |
| 2SLS X10 only | -9.879 | -0.121 | 1.722 | 1.032 | 2.978 | 49.830 | 0.170 | 1.393 | 0.799 | 1.968 | 9.456 |

**A 39: Full results simulation study 1a, 10 explanatory variables of the continuous process variable, average F-statistic of relevant instruments=7.1, average F-statistic of individual instrument=5.9, N=400, simulations=1000**

| | Group effect | | | | | | Post randomisation mediator effect | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean Coef. | Mean Bias | bias 2SLS/ bias OLS | SD coef. | Absolute Difference | MSE | Mean Coef. | Mean Bias | bias 2SLS/ bias OLS | SD coef. | Absolute Difference | MSE | F-stat |
| OLS All | -10.625 | 0.625 | | 0.934 | 0.906 | 1.263 | 51.079 | -1.079 | | 0.097 | 1.079 | 1.175 | |
| 2SLS All | -10.159 | 0.159 | 0.255 | 1.007 | 0.816 | 1.039 | 50.306 | -0.306 | 0.284 | 0.275 | 0.349 | 0.169 | 3.81 |
| OLS relevant | -10.627 | 0.627 | | 0.919 | 0.899 | 1.237 | 51.079 | -1.079 | | 0.097 | 1.079 | 1.173 | |
| 2SLS relevant | -10.065 | 0.065 | 0.104 | 1.019 | 0.821 | 1.042 | 50.142 | -0.142 | 0.132 | 0.297 | 0.266 | 0.108 | 7.08 |
| OLS LASSO | -10.609 | 0.609 | | 0.922 | 0.890 | 1.220 | 51.043 | -1.043 | | 0.095 | 1.043 | 1.098 | |
| 2SLS LASSO | -6.248 | -3.752 | -6.157 | 13.035 | 4.876 | 183.826 | 50.307 | -0.307 | 0.294 | 0.298 | 0.361 | 0.183 | **9.19** |
| OLS LASSO-m | -10.625 | 0.625 | | 0.926 | 0.898 | 1.246 | 51.077 | -1.077 | | 0.097 | 1.077 | 1.170 | |
| 2SLS LASSO - m | -10.147 | 0.147 | 0.235 | 1.000 | 0.812 | 1.021 | 50.281 | -0.281 | 0.261 | 0.280 | 0.334 | **0.157** | 5.43 |
| OLS Elastic Net | -10.612 | 0.612 | | 0.926 | 0.895 | 1.232 | 51.048 | -1.048 | | 0.096 | 1.048 | 1.107 | |
| 2SLS Elastic Net | -7.060 | -2.940 | -4.804 | 12.308 | 4.125 | 159.981 | 50.307 | -0.307 | 0.293 | 0.303 | 0.363 | 0.186 | 8.72 |
| OLS Elastic Net - m | -10.623 | 0.623 | | 0.927 | 0.899 | 1.247 | 51.078 | -1.078 | | 0.097 | 1.078 | 1.171 | |
| 2SLS Elastic Net - m | -10.146 | **0.146** | 0.235 | 1.002 | 0.812 | 1.025 | 50.284 | -0.284 | 0.264 | 0.282 | 0.337 | 0.160 | 5.26 |
| OLS Stepwise | -10.611 | 0.611 | | 0.922 | 0.888 | 1.221 | 51.069 | -1.069 | | 0.096 | 1.069 | 1.153 | |
| 2SLS Stepwise | -10.148 | 0.148 | 0.242 | 0.997 | 0.806 | **1.015** | 50.278 | **-0.278** | 0.260 | 0.291 | 0.337 | 0.162 | 7.46 |
| OLS no explanatory | -10.572 | 0.572 | | 0.933 | 0.885 | 1.197 | 50.984 | -0.984 | | 0.093 | 0.984 | 0.977 | |
| 2SLS no explanatory | -10.228 | 0.228 | 0.398 | 1.033 | 0.845 | 1.118 | 50.410 | -0.410 | 0.417 | 0.475 | 0.521 | 0.394 | 2.37 |

**A 40: Simulation study 1a, 10 explanatory variables of the continuous process variable, average F-statistic of relevant instruments=7.1, average F-statistic of individual instrument=5.9, N=400, simulations=1000, individual explanatory variables only**

| | Group Effect | | | | | Post randomisation mediator effect | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean coefficient | Mean bias | SD coefficent | Absolute Difference | MSE | Mean coefficient | Mean bias | SD coefficent | Absolute Difference | MSE | F-statistic of instrument |
| 2SLS X1 only | -9.804 | -0.196 | 2.535 | 1.261 | 6.457 | 49.636 | 0.364 | 3.775 | 1.316 | 14.371 | 6.060 |
| 2SLS X2 only | -9.933 | -0.067 | 14.102 | 2.000 | 198.680 | 49.757 | 0.243 | 16.874 | 2.237 | 284.508 | 5.733 |
| 2SLS X3 only | -9.830 | -0.170 | 2.394 | 1.274 | 5.756 | 49.520 | 0.480 | 6.475 | 1.440 | 42.115 | 5.866 |
| 2SLS X4 only | -9.923 | -0.077 | 4.442 | 1.494 | 19.715 | 49.701 | 0.299 | 8.982 | 1.921 | 80.690 | 6.027 |
| 2SLS X5 only | -6.795 | -3.205 | 99.775 | 4.546 | 9955.275 | 46.077 | 3.923 | 116.800 | 5.317 | 13643.970 | 5.719 |
| 2SLS X6 only | -9.417 | -0.583 | 9.000 | 1.819 | 81.257 | 49.272 | 0.728 | 9.542 | 1.994 | 91.487 | 5.600 |
| 2SLS X7 only | -9.815 | -0.185 | 3.912 | 1.345 | 15.319 | 49.710 | 0.290 | 6.266 | 1.517 | 39.303 | 5.938 |
| 2SLS X8 only | -9.692 | -0.308 | 5.597 | 1.527 | 31.394 | 49.599 | 0.401 | 5.665 | 1.484 | 32.223 | 6.014 |
| 2SLS X9 only | -9.883 | -0.117 | 3.566 | 1.442 | 12.720 | 49.668 | 0.332 | 4.142 | 1.493 | 17.245 | 6.034 |
| 2SLS X10 only | -9.842 | -0.158 | 4.752 | 1.499 | 22.579 | 49.740 | 0.260 | 8.210 | 1.693 | 67.412 | 5.701 |

**CATEGORICAL PROCESS VARIABLES**

**A 41: Full results simulation study 1a, 5 explanatory variables of the categorical process variable, average F-statistic of relevant instruments=30.7, N=200, simulations=1000**

| | Group effect | | | | | | Post randomisation mediator effect | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean Coef. | Mean Bias | bias 2SLS/ bias OLS | SD coef. | Absolute Difference | MSE | Mean Coef. | Mean Bias | bias 2SLS/ bias OLS | SD coef. | Absolute Difference | MSE |
| OLS All | -11.739 | 1.739 | | 1.949 | 2.125 | 6.818 | 53.701 | -3.701 | | 2.551 | 3.872 | 20.201 |
| 2SLS All | -10.222 | 0.222 | 0.128 | 2.322 | 1.867 | 5.433 | 50.665 | -0.665 | 0.180 | 3.642 | 2.976 | 13.690 |
| OLS relevant | -11.770 | 1.770 | | 1.922 | 2.164 | 6.827 | 53.713 | -3.713 | | 2.486 | 3.855 | 19.962 |
| 2SLS relevant | -9.947 | -0.053 | -0.030 | 2.372 | 1.890 | 5.624 | 50.060 | -0.060 | 0.016 | 3.752 | 3.019 | 14.068 |
| OLS LASSO | -11.750 | 1.750 | | 1.944 | 2.161 | 6.837 | 53.667 | -3.667 | | 2.502 | 3.827 | 19.701 |
| 2SLS LASSO | -10.051 | **0.051** | 0.029 | 2.357 | 1.884 | 5.550 | 50.266 | **-0.266** | 0.073 | 3.728 | 3.005 | 13.955 |
| OLS LASSO-m | -11.723 | 1.723 | | 1.952 | 2.133 | 6.776 | 53.641 | -3.641 | | 2.532 | 3.822 | 19.659 |
| 2SLS LASSO - m | -10.126 | 0.126 | 0.073 | 2.330 | 1.867 | 5.438 | 50.445 | -0.445 | 0.122 | 3.648 | 2.922 | **13.489** |
| OLS Elastic Net | -11.758 | 1.758 | | 1.942 | 2.166 | 6.858 | 53.678 | -3.678 | | 2.503 | 3.843 | 19.788 |
| 2SLS Elastic Net | -10.075 | 0.075 | 0.043 | 2.360 | 1.887 | 5.572 | 50.310 | -0.310 | 0.084 | 3.722 | 2.987 | 13.937 |
| OLS Elastic Net - m | -11.734 | 1.734 | | 1.947 | 2.131 | 6.794 | 53.656 | -3.656 | | 2.516 | 3.824 | 19.694 |
| 2SLS Elastic Net - m | -10.171 | 0.171 | 0.099 | 2.326 | 1.861 | **5.436** | 50.528 | -0.528 | 0.145 | 3.647 | 2.933 | 13.568 |
| OLS Stepwise | -11.702 | 1.702 | | 1.916 | 2.111 | 6.562 | 53.627 | -3.627 | | 2.498 | 3.783 | 19.386 |
| 2SLS Stepwise | -10.103 | 0.103 | 0.061 | 2.367 | 1.899 | 5.607 | 50.320 | -0.320 | 0.088 | 3.866 | 3.104 | 15.031 |
| OLS no explanatory | -11.398 | 1.398 | | 1.879 | 1.895 | 5.480 | 53.018 | -3.018 | | 2.296 | 3.237 | 14.376 |
| 2SLS no explanatory | -10.205 | 0.205 | 0.147 | 2.862 | 2.279 | 8.226 | 50.628 | -0.628 | 0.208 | 4.946 | 3.968 | 24.833 |

**A 42: Full results simulation study 1a, 5 explanatory variables of the categorical process variable, average F-statistic of relevant instruments=30.7, N=200, simulations=1000, individual explanatory variabls only**

| | Group Effect | | | | | Post randomisation mediator effect | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Mean coefficient | Mean bias | SD coefficent | Absolute Difference | MSE | Mean coefficient | Mean bias | SD coefficent | Absolute Difference | MSE |
| 2SLS X1 only | -9.664 | -0.336 | 5.341 | 4.003 | 28.608 | 49.489 | 0.511 | 10.157 | 7.628 | 103.326 |
| 2SLS X2 only | -9.645 | -0.355 | 5.263 | 3.935 | 27.793 | 49.495 | 0.505 | 10.033 | 7.397 | 100.824 |
| 2SLS X3 only | -9.750 | -0.250 | 5.107 | 3.901 | 26.117 | 49.655 | 0.345 | 9.781 | 7.360 | 95.687 |
| 2SLS X4 only | -9.906 | -0.094 | 6.361 | 4.170 | 40.429 | 49.966 | 0.034 | 12.567 | 8.074 | 157.762 |
| 2SLS X5 only | -9.744 | -0.256 | 5.101 | 3.893 | 26.065 | 49.669 | 0.331 | 9.746 | 7.342 | 94.998 |

**A 43: Full results simulation study 1a, 5 explanatory variables of the categorical process variable, average F-statistic of relevant instruments=20.6, N=200, simulations=1000**

| | Group effect | | | | | | Post randomisation mediator effect | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean Coef. | Mean Bias | bias 2SLS/ bias OLS | SD coef. | Absolute Difference | MSE | Mean Coef. | Mean Bias | bias 2SLS/ bias OLS | SD coef. | Absolute Difference | MSE |
| OLS All | -13.350 | 3.350 | | 1.821 | 3.400 | 14.536 | 56.931 | -6.931 | | 2.276 | 6.938 | 53.209 |
| 2SLS All | -10.734 | 0.734 | 0.219 | 2.386 | 2.007 | 6.224 | 51.697 | -1.697 | 0.245 | 3.841 | 3.374 | 17.616 |
| OLS relevant | -13.379 | 3.379 | | 1.806 | 3.434 | 14.678 | 56.928 | -6.928 | | 2.222 | 6.935 | 52.928 |
| 2SLS relevant | -10.076 | 0.076 | 0.022 | 2.500 | 1.994 | 6.252 | 50.321 | -0.321 | 0.046 | 4.060 | 3.275 | 16.568 |
| OLS LASSO | -13.363 | 3.363 | | 1.807 | 3.418 | 14.576 | 56.898 | -6.898 | | 2.242 | 6.906 | 52.598 |
| 2SLS LASSO | -10.385 | **0.385** | 0.114 | 2.452 | 1.992 | 6.155 | 50.937 | **-0.937** | 0.136 | 4.010 | 3.281 | **16.944** |
| OLS LASSO-m | -13.351 | 3.351 | | 1.827 | 3.408 | 14.566 | 56.874 | -6.874 | | 2.274 | 6.885 | 52.422 |
| 2SLS LASSO - m | -10.573 | 0.573 | 0.171 | 2.441 | 2.006 | 6.279 | 51.316 | -1.316 | 0.191 | 3.934 | 3.334 | 17.192 |
| OLS Elastic Net | -13.371 | 3.371 | | 1.807 | 3.427 | 14.627 | 56.917 | -6.917 | | 2.239 | 6.924 | 52.850 |
| 2SLS Elastic Net | -10.420 | 0.420 | 0.124 | 2.467 | 2.004 | 6.255 | 51.009 | -1.009 | 0.146 | 4.022 | 3.306 | 17.178 |
| OLS Elastic Net - m | -13.356 | 3.356 | | 1.816 | 3.408 | 14.556 | 56.882 | -6.882 | | 2.274 | 6.891 | 52.525 |
| 2SLS Elastic Net - m | -10.589 | 0.589 | 0.175 | 2.430 | 2.008 | **6.246** | 51.347 | -1.347 | 0.196 | 3.925 | 3.322 | 17.207 |
| OLS Stepwise | -13.238 | 3.238 | | 1.799 | 3.308 | 13.717 | 56.788 | -6.788 | | 2.219 | 6.796 | 50.993 |
| 2SLS Stepwise | -10.453 | 0.453 | 0.140 | 2.498 | 2.029 | 6.439 | 51.025 | -1.025 | 0.151 | 4.280 | 3.519 | 19.352 |
| OLS no explanatory | -12.851 | 2.851 | | 1.796 | 2.946 | 11.348 | 55.924 | -5.924 | | 2.144 | 5.935 | 39.687 |
| 2SLS no explanatory | -10.755 | 0.755 | 0.265 | 2.897 | 2.396 | 8.952 | 51.734 | -1.734 | 0.293 | 5.102 | 4.321 | 29.008 |

**A 44: Full results simulation study 1a, 5 explanatory variables of the categorical process variable, average F-statistic of relevant instruments=20.6, N=200, simulations=1000, individual explanatory variables only**

| | Group Effect | | | | | Post randomisation mediator effect | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Mean coefficient | Mean bias | SD coefficent | Absolute Difference | MSE | Mean coefficient | Mean bias | SD coefficent | Absolute Difference | MSE |
| 2SLS X1 only | -9.455 | -0.545 | 5.978 | 4.431 | 36.000 | 49.087 | 0.913 | 11.423 | 8.487 | 131.185 |
| 2SLS X2 only | -9.403 | -0.597 | 6.061 | 4.350 | 37.052 | 49.041 | 0.959 | 11.557 | 8.225 | 134.350 |
| 2SLS X3 only | -9.579 | -0.421 | 6.202 | 4.384 | 38.610 | 49.314 | 0.686 | 12.092 | 8.377 | 146.541 |
| 2SLS X4 only | -9.770 | -0.230 | 9.654 | 4.811 | 93.156 | 49.647 | 0.353 | 18.395 | 9.350 | 338.178 |
| 2SLS X5 only | -9.478 | -0.522 | 6.473 | 4.460 | 42.129 | 49.174 | 0.826 | 12.419 | 8.481 | 154.752 |

**A 45: Full results simulation study 1a, 5 explanatory variables of the categorical process variable, average F-statistic of relevant instruments=10.4, N=200, simulations=1000**

| | Group effect | | | | | | Post randomisation mediator effect | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean Coef. | Mean Bias | bias 2SLS/ bias OLS | SD coef. | Absolute Difference | MSE | Mean Coef. | Mean Bias | bias 2SLS/ bias OLS | SD coef. | Absolute Difference | MSE |
| OLS All | -14.579 | 4.579 | | 1.725 | 4.587 | 23.944 | 59.396 | -9.396 | | 2.088 | 9.396 | 92.647 |
| 2SLS All | -11.598 | 1.598 | 0.349 | 2.573 | 2.461 | 9.166 | 53.432 | -3.432 | 0.365 | 4.279 | 4.487 | 30.069 |
| OLS relevant | -14.616 | 4.616 | | 1.704 | 4.622 | 24.209 | 59.412 | -9.412 | | 2.027 | 9.412 | 92.691 |
| 2SLS relevant | -10.290 | 0.290 | 0.063 | 2.878 | 2.309 | 8.358 | 50.754 | -0.754 | 0.080 | 4.955 | 4.040 | 25.096 |
| OLS LASSO | -14.522 | 4.522 | | 1.713 | 4.529 | 23.382 | 59.223 | -9.223 | | 2.057 | 9.223 | 89.293 |
| 2SLS LASSO | -10.664 | 0.664 | 0.147 | 4.304 | 2.771 | 18.948 | 52.333 | **-2.333** | 0.253 | 4.851 | 4.301 | 28.953 |
| OLS LASSO-m | -14.582 | 4.582 | | 1.735 | 4.586 | 24.000 | 59.354 | -9.354 | | 2.082 | 9.354 | 91.829 |
| 2SLS LASSO - m | -11.265 | 1.265 | 0.276 | 2.833 | 2.422 | 9.619 | 52.779 | -2.779 | 0.297 | 4.497 | 4.293 | **27.926** |
| OLS Elastic Net | -14.540 | 4.540 | | 1.722 | 4.547 | 23.572 | 59.252 | -9.252 | | 2.054 | 9.252 | 89.815 |
| 2SLS Elastic Net | -10.657 | **0.657** | 0.145 | 4.491 | 2.828 | 20.580 | 52.432 | -2.432 | 0.263 | 4.738 | 4.307 | 28.335 |
| OLS Elastic Net - m | -14.583 | 4.583 | | 1.727 | 4.588 | 23.987 | 59.365 | -9.365 | | 2.078 | 9.365 | 92.019 |
| 2SLS Elastic Net - m | -11.332 | 1.332 | 0.291 | 2.645 | 2.416 | **8.766** | 52.863 | -2.863 | 0.306 | 4.478 | 4.339 | 28.226 |
| OLS Stepwise | -14.395 | 4.395 | | 1.709 | 4.402 | 22.233 | 59.170 | -9.170 | | 2.049 | 9.170 | 88.283 |
| 2SLS Stepwise | -11.161 | 1.161 | 0.264 | 2.773 | 2.430 | 9.031 | 52.454 | -2.454 | 0.268 | 5.024 | 4.515 | 31.236 |
| OLS no explanatory | -14.179 | 4.179 | | 1.722 | 4.188 | 20.424 | 58.580 | -8.580 | | 2.023 | 8.580 | 77.709 |
| 2SLS no explanatory | -11.704 | 1.704 | 0.408 | 3.100 | 2.858 | 12.502 | 53.623 | -3.623 | 0.422 | 5.613 | 5.395 | 44.600 |

**A 46: Full results simulation study 1a, 5 explanatory variables of the categorical process variable, average F-statistic of relevant instruments=10.4, N=200, simulations=1000, individual explanatory variables only**

| | Group Effect | | | | | Post randomisation mediator effect | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Mean coefficient | Mean bias | SD coefficent | Absolute Difference | MSE | Mean coefficient | Mean bias | SD coefficent | Absolute Difference | MSE |
| 2SLS X1 only | -8.943 | -1.057 | 14.618 | 6.557 | 214.600 | 48.054 | 1.946 | 29.580 | 12.870 | 877.862 |
| 2SLS X2 only | 6.108 | -16.108 | 421.790 | 20.674 | 177988.100 | 19.780 | 30.220 | 781.721 | 39.179 | 611389.400 |
| 2SLS X3 only | -9.173 | -0.827 | 21.766 | 7.188 | 473.990 | 48.594 | 1.406 | 44.399 | 14.182 | 1971.278 |
| 2SLS X4 only | -9.156 | -0.844 | 14.631 | 6.879 | 214.562 | 48.536 | 1.464 | 29.788 | 13.651 | 888.576 |
| 2SLS X5 only | -7.629 | -2.371 | 26.752 | 7.511 | 720.595 | 45.382 | 4.618 | 52.287 | 14.624 | 2752.486 |

**A 47: Full results simulation study 1a, 5 explanatory variables of the categorical process variable, average F-statistic of relevant instruments=61.3, N=400, simulations=1000**

| | Group effect | | | | | | Post randomisation mediator effect | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean Coef. | Mean Bias | bias 2SLS/ bias OLS | SD coef. | Absolute Difference | MSE | Mean Coef. | Mean Bias | bias 2SLS/ bias OLS | SD coef. | Absolute Difference | MSE |
| OLS All | -11.833 | 1.833 | | 1.361 | 1.943 | 5.210 | 53.732 | -3.732 | | 1.734 | 3.759 | 16.935 |
| 2SLS All | -10.170 | 0.170 | 0.093 | 1.690 | 1.362 | 2.881 | 50.406 | -0.406 | 0.109 | 2.599 | 2.089 | 6.914 |
| OLS relevant | -11.819 | 1.819 | | 1.379 | 1.933 | 5.208 | 53.695 | -3.695 | | 1.779 | 3.728 | 16.813 |
| 2SLS relevant | -9.981 | -0.019 | -0.011 | 1.730 | 1.397 | 2.990 | 50.023 | -0.023 | 0.006 | 2.717 | 2.182 | 7.373 |
| OLS LASSO | -11.822 | 1.822 | | 1.377 | 1.935 | 5.214 | 53.707 | -3.707 | | 1.775 | 3.741 | 16.891 |
| 2SLS LASSO | -10.043 | **0.043** | 0.023 | 1.727 | 1.394 | 2.981 | 50.150 | **-0.150** | 0.040 | 2.701 | 2.177 | 7.313 |
| OLS LASSO-m | -11.820 | 1.820 | | 1.370 | 1.935 | 5.185 | 53.691 | -3.691 | | 1.742 | 3.717 | 16.657 |
| 2SLS LASSO - m | -10.123 | 0.123 | 0.068 | 1.705 | 1.379 | 2.918 | 50.299 | -0.299 | 0.081 | 2.629 | 2.100 | **6.995** |
| OLS Elastic Net | -11.817 | 1.817 | | 1.374 | 1.930 | 5.188 | 53.698 | -3.698 | | 1.772 | 3.732 | 16.813 |
| 2SLS Elastic Net | -10.062 | 0.062 | 0.034 | 1.722 | 1.391 | 2.965 | 50.188 | -0.188 | 0.051 | 2.684 | 2.159 | 7.232 |
| OLS Elastic Net - m | -11.820 | 1.820 | | 1.367 | 1.937 | 5.181 | 53.700 | -3.700 | | 1.744 | 3.729 | 16.730 |
| 2SLS Elastic Net - m | -10.131 | 0.131 | 0.072 | 1.702 | 1.378 | **2.911** | 50.322 | -0.322 | 0.087 | 2.627 | 2.102 | 7.000 |
| OLS Stepwise | -11.817 | 1.817 | | 1.369 | 1.928 | 5.173 | 53.704 | -3.704 | | 1.744 | 3.733 | 16.756 |
| 2SLS Stepwise | -10.061 | 0.061 | 0.034 | 1.722 | 1.387 | 2.966 | 50.155 | -0.155 | 0.042 | 2.697 | 2.148 | 7.290 |
| OLS no explanatory | -11.498 | 1.498 | | 1.347 | 1.675 | 4.059 | 53.071 | -3.071 | | 1.654 | 3.116 | 12.168 |
| 2SLS no explanatory | -10.134 | 0.134 | 0.090 | 2.090 | 1.682 | 4.382 | 50.339 | -0.339 | 0.110 | 3.531 | 2.842 | 12.572 |

**A 48: Full results simulation study 1a, 5 explanatory variables of the categorical process variable, average F-statistic of relevant instruments=61.3, N=400, simulations=1000, individual explanatory variables only**

| | Group Effect | | | | | Post randomisation mediator effect | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Mean coefficient | Mean bias | SD coef. | Absolute Difference | MSE | Mean coefficient | Mean bias | SD coef. | Absolute Difference | MSE |
| 2SLS X1 only | -9.674 | -0.326 | 3.323 | 2.622 | 11.136 | 49.425 | 0.575 | 6.305 | 4.979 | 40.048 |
| 2SLS X2 only | -10.069 | 0.069 | 3.396 | 2.671 | 11.529 | 50.190 | -0.190 | 6.427 | 5.016 | 41.295 |
| 2SLS X3 only | -9.982 | -0.018 | 3.433 | 2.704 | 11.775 | 50.017 | -0.017 | 6.480 | 5.108 | 41.948 |
| 2SLS X4 only | -9.840 | -0.160 | 3.326 | 2.609 | 11.075 | 49.746 | 0.254 | 6.311 | 4.913 | 39.851 |
| 2SLS X5 only | -9.968 | -0.032 | 3.427 | 2.700 | 11.731 | 49.996 | 0.004 | 6.565 | 5.182 | 43.053 |

**A 49: Full results simulation study 1a, 5 explanatory variables of the categorical process variable, average F-statistic of relevant instruments=40.4, N=400, simulations=1000**

| | Group effect | | | | | | Post randomisation mediator effect | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean Coef. | Mean Bias | bias 2SLS/ bias OLS | SD coef. | Absolute Difference | MSE | Mean Coef. | Mean Bias | bias 2SLS/ bias OLS | SD coef. | Absolute Difference | MSE |
| OLS All | -13.452 | 3.452 | | 1.296 | 3.458 | 13.597 | 56.976 | -6.976 | | 1.603 | 6.976 | 51.225 |
| 2SLS All | -10.443 | 0.443 | 0.128 | 1.769 | 1.467 | 3.321 | 50.951 | -0.951 | 0.136 | 2.813 | 2.372 | 8.807 |
| OLS relevant | -13.438 | 3.438 | | 1.307 | 3.443 | 13.528 | 56.941 | -6.941 | | 1.627 | 6.941 | 50.830 |
| 2SLS relevant | -10.043 | 0.043 | 0.012 | 1.826 | 1.472 | 3.332 | 50.142 | -0.142 | 0.021 | 2.971 | 2.395 | 8.838 |
| OLS LASSO | -13.434 | 3.434 | | 1.303 | 3.439 | 13.492 | 56.937 | -6.937 | | 1.619 | 6.937 | 50.738 |
| 2SLS LASSO | -10.170 | **0.170** | 0.050 | 1.822 | 1.477 | 3.346 | 50.404 | **-0.404** | 0.058 | 2.943 | 2.387 | 8.815 |
| OLS LASSO-m | -13.437 | 3.437 | | 1.294 | 3.442 | 13.488 | 56.941 | -6.941 | | 1.617 | 6.941 | 50.791 |
| 2SLS LASSO - m | -10.350 | 0.350 | 0.102 | 1.787 | 1.467 | 3.312 | 50.761 | -0.761 | 0.110 | 2.890 | 2.399 | 8.925 |
| OLS Elastic Net | -13.436 | 3.436 | | 1.303 | 3.441 | 13.502 | 56.942 | -6.942 | | 1.620 | 6.942 | 50.819 |
| 2SLS Elastic Net | -10.200 | 0.200 | 0.058 | 1.817 | 1.476 | 3.337 | 50.466 | -0.466 | 0.067 | 2.932 | 2.387 | **8.803** |
| OLS Elastic Net - m | -13.441 | 3.441 | | 1.299 | 3.446 | 13.527 | 56.945 | -6.945 | | 1.614 | 6.945 | 50.831 |
| 2SLS Elastic Net - m | -10.369 | 0.369 | 0.107 | 1.781 | 1.464 | **3.306** | 50.796 | -0.796 | 0.115 | 2.864 | 2.376 | 8.829 |
| OLS Stepwise | -13.419 | 3.419 | | 1.312 | 3.423 | 13.409 | 56.921 | -6.921 | | 1.634 | 6.921 | 50.569 |
| 2SLS Stepwise | -10.230 | 0.230 | 0.067 | 1.824 | 1.487 | 3.378 | 50.456 | -0.456 | 0.066 | 2.962 | 2.416 | 8.975 |
| OLS no explanatory | -12.958 | 2.958 | | 1.287 | 2.968 | 10.403 | 55.997 | -5.997 | | 1.534 | 5.996 | 38.309 |
| 2SLS no explanatory | -10.464 | 0.464 | 0.157 | 2.203 | 1.805 | 5.064 | 50.994 | -0.994 | 0.166 | 3.823 | 3.174 | 15.588 |

**A 50: Full results simulation study 1a, 5 explanatory variables of the categorical process variable, average F-statistic of relevant instruments=40.4, N=400, simulations=1000, individual explanatory variables only**

| | Group Effect | | | | | Post randomisation mediator effect | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Mean coefficient | Mean bias | SD coefficent | Absolute Difference | MSE | Mean coefficient | Mean bias | SD coefficent | Absolute Difference | MSE |
| 2SLS X1 only | -9.577 | -0.423 | 3.721 | 2.898 | 14.008 | 49.218 | 0.782 | 7.154 | 5.567 | 51.742 |
| 2SLS X2 only | -10.020 | 0.020 | 3.720 | 2.902 | 13.827 | 50.089 | -0.089 | 7.103 | 5.511 | 50.416 |
| 2SLS X3 only | -9.932 | -0.068 | 3.824 | 2.975 | 14.613 | 49.905 | 0.095 | 7.268 | 5.676 | 52.785 |
| 2SLS X4 only | -9.738 | -0.262 | 3.689 | 2.861 | 13.662 | 49.534 | 0.466 | 7.088 | 5.450 | 50.409 |
| 2SLS X5 only | -9.876 | -0.124 | 3.716 | 2.917 | 13.807 | 49.824 | 0.176 | 7.154 | 5.640 | 51.159 |

**A 51: Full results simulation study 1a, 5 explanatory variables of the categorical process variable, average F-statistic of relevant instruments=20.3, N=400, simulations=1000**

| | Group effect | | | | | | Post randomisation mediator effect | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean Coef. | Mean Bias | bias 2SLS/ bias OLS | SD coef. | Absolute Difference | MSE | Mean Coef. | Mean Bias | bias 2SLS/ bias OLS | SD coef. | Absolute Difference | MSE |
| OLS All | -14.656 | 4.656 | | 1.216 | 4.656 | 23.153 | 59.388 | -9.388 | | 1.437 | 9.388 | 90.202 |
| 2SLS All | -10.959 | 0.959 | 0.206 | 1.963 | 1.754 | 4.770 | 51.982 | -1.982 | 0.211 | 3.288 | 3.090 | 14.729 |
| OLS relevant | -14.648 | 4.648 | | 1.211 | 4.648 | 23.066 | 59.368 | -9.368 | | 1.433 | 9.368 | 89.814 |
| 2SLS relevant | -10.151 | 0.151 | 0.032 | 2.078 | 1.677 | 4.334 | 50.358 | -0.358 | 0.038 | 3.570 | 2.885 | 12.861 |
| OLS LASSO | -14.639 | 4.639 | | 1.212 | 4.639 | 22.985 | 59.348 | -9.348 | | 1.440 | 9.348 | 89.465 |
| 2SLS LASSO | -10.467 | **0.467** | 0.101 | 2.063 | 1.690 | 4.471 | 50.989 | **-0.989** | 0.106 | 3.542 | 2.935 | **13.512** |
| OLS LASSO-m | -14.646 | 4.646 | | 1.214 | 4.646 | 23.059 | 59.358 | -9.358 | | 1.444 | 9.358 | 89.655 |
| 2SLS LASSO - m | -10.777 | 0.777 | 0.167 | 2.010 | 1.733 | **4.639** | 51.607 | -1.607 | 0.172 | 3.413 | 3.041 | 14.218 |
| OLS Elastic Net | -14.640 | 4.640 | | 1.212 | 4.640 | 22.999 | 59.352 | -9.352 | | 1.439 | 9.352 | 89.532 |
| 2SLS Elastic Net | -10.501 | 0.501 | 0.108 | 2.059 | 1.697 | 4.486 | 51.060 | -1.060 | 0.113 | 3.540 | 2.948 | 13.639 |
| OLS Elastic Net - m | -14.646 | 4.646 | | 1.207 | 4.646 | 23.040 | 59.363 | -9.363 | | 1.438 | 9.363 | 89.740 |
| 2SLS Elastic Net - m | -10.796 | 0.796 | 0.171 | 2.005 | 1.740 | 4.651 | 51.649 | -1.649 | 0.176 | 3.400 | 3.042 | 14.270 |
| OLS Stepwise | -14.575 | 4.575 | | 1.210 | 4.575 | 22.393 | 59.288 | -9.288 | | 1.442 | 9.288 | 88.345 |
| 2SLS Stepwise | -10.567 | 0.567 | 0.124 | 2.108 | 1.758 | 4.762 | 51.139 | -1.139 | 0.123 | 3.635 | 3.036 | 14.499 |
| OLS no explanatory | -14.240 | 4.240 | | 1.222 | 4.240 | 19.472 | 58.566 | -8.566 | | 1.415 | 8.566 | 75.369 |
| 2SLS no explanatory | -11.053 | 1.053 | 0.248 | 2.492 | 2.157 | 7.315 | 52.172 | -2.172 | 0.254 | 4.484 | 4.006 | 24.804 |

**A 52: Full results simulation study 1a, 5 explanatory variables of the categorical process variable, average F-statistic of relevant instruments=20.3, N=400, simulations=1000, individual explanatory variables only**

| | Group Effect | | | | | Post randomisation mediator effect | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Mean coefficient | Mean bias | SD coefficent | Absolute Difference | MSE | Mean coefficient | Mean bias | SD coefficent | Absolute Difference | MSE |
| 2SLS X1 only | -9.254 | -0.746 | 5.145 | 3.756 | 27.002 | 48.527 | 1.473 | 10.204 | 7.343 | 106.186 |
| 2SLS X2 only | -9.808 | -0.192 | 4.938 | 3.635 | 24.398 | 49.661 | 0.339 | 9.670 | 7.034 | 93.533 |
| 2SLS X3 only | -9.658 | -0.342 | 5.015 | 3.724 | 25.244 | 49.381 | 0.619 | 9.702 | 7.238 | 94.425 |
| 2SLS X4 only | -9.461 | -0.539 | 5.543 | 3.644 | 30.988 | 48.952 | 1.048 | 11.442 | 7.074 | 131.894 |
| 2SLS X5 only | -9.562 | -0.438 | 6.142 | 3.775 | 37.873 | 49.211 | 0.789 | 11.799 | 7.392 | 139.693 |

**A 53: Full results simulation study 1a, 10 explanatory variables of the categorical process variable, average F-statistic of relevant instruments=15.8, average F-statistic of individual instrument=7.4, N=200, simulations=1000**

| | Group effect | | | | | | Post randomisation mediator effect | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean Coef. | Mean Bias | bias 2SLS/ bias OLS | SD coef. | Absolute Difference | MSE | Mean Coef. | Mean Bias | bias 2SLS/ bias OLS | SD coef. | Absolute Difference | MSE |
| OLS All | -11.674 | 1.674 | | 2.012 | 2.134 | 6.848 | 53.507 | -3.507 | | 2.654 | 3.778 | 19.333 |
| 2SLS All | -10.286 | 0.286 | 0.171 | 2.426 | 1.955 | 5.960 | 50.729 | -0.729 | 0.208 | 3.681 | 3.034 | 14.069 |
| OLS relevant | -11.685 | 1.685 | | 1.983 | 2.123 | 6.766 | 53.486 | -3.486 | | 2.630 | 3.736 | 19.061 |
| 2SLS relevant | -10.076 | 0.076 | 0.045 | 2.490 | 1.982 | 6.201 | 50.270 | -0.270 | 0.077 | 3.841 | 3.092 | 14.811 |
| OLS LASSO | -11.668 | 1.668 | | 1.986 | 2.111 | 6.720 | 53.472 | -3.472 | | 2.605 | 3.723 | 18.834 |
| 2SLS LASSO | -10.196 | **0.196** | 0.118 | 2.445 | 1.963 | 6.010 | 50.531 | **-0.531** | 0.153 | 3.703 | 2.999 | 13.978 |
| OLS LASSO-m | -11.655 | 1.655 | | 1.991 | 2.106 | 6.697 | 53.464 | -3.464 | | 2.611 | 3.726 | 18.811 |
| 2SLS LASSO - m | -10.235 | 0.235 | 0.142 | 2.411 | 1.936 | 5.865 | 50.624 | -0.624 | 0.180 | 3.635 | 2.962 | **13.590** |
| OLS Elastic Net | -11.662 | 1.662 | | 1.981 | 2.107 | 6.683 | 53.463 | -3.463 | | 2.602 | 3.718 | 18.755 |
| 2SLS Elastic Net | -10.199 | 0.199 | 0.119 | 2.426 | 1.942 | 5.918 | 50.536 | -0.536 | 0.155 | 3.681 | 2.974 | 13.822 |
| OLS Elastic Net - m | -11.661 | 1.661 | | 1.990 | 2.106 | 6.715 | 53.474 | -3.474 | | 2.617 | 3.736 | 18.914 |
| 2SLS Elastic Net - m | -10.237 | 0.237 | 0.143 | 2.405 | 1.932 | **5.836** | 50.628 | -0.628 | 0.181 | 3.639 | 2.964 | 13.624 |
| OLS Stepwise | -11.529 | 1.529 | | 1.925 | 1.989 | 6.037 | 53.297 | -3.297 | | 2.549 | 3.557 | 17.361 |
| 2SLS Stepwise | -10.235 | 0.235 | 0.154 | 2.544 | 2.030 | 6.521 | 50.597 | -0.597 | 0.181 | 4.194 | 3.352 | 17.932 |
| OLS no explanatory | -11.255 | 1.255 | | 1.881 | 1.835 | 5.109 | 52.679 | -2.679 | | 2.294 | 2.960 | 12.437 |
| 2SLS no explanatory | -10.344 | 0.344 | 0.274 | 3.481 | 2.747 | 12.224 | 50.875 | -0.875 | 0.327 | 6.169 | 4.868 | 38.780 |

**A 54: Simulation study 1a, 10 explanatory variables of the categorical process variable, average F-statistic of relevant instruments=15.8, average F-statistic of individual instrument=7.4, N=200, simulations=1000, individual explanatory variables only**

| | Group Effect | | | | | Post randomisation mediator effect | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Mean coefficient | Mean bias | SD coefficent | Absolute Difference | MSE | Mean coefficient | Mean bias | SD coefficent | Absolute Difference | MSE |
| 2SLS X1 only | -9.98 | -0.02 | 11.07 | 6.58 | 122.36 | 50.09 | -0.09 | 22.06 | 13.00 | 486.00 |
| 2SLS X2 only | -11.06 | 1.06 | 38.02 | 8.29 | 1444.89 | 52.55 | -2.55 | 84.47 | 16.66 | 7133.74 |
| 2SLS X3 only | -8.03 | -1.97 | 70.38 | 8.54 | 4952.00 | 45.91 | 4.09 | 146.03 | 16.98 | 21321.06 |
| 2SLS X4 only | -9.20 | -0.80 | 14.78 | 7.12 | 218.88 | 48.51 | 1.49 | 28.26 | 13.76 | 800.19 |
| 2SLS X5 only | -1.44 | -8.56 | 252.99 | 14.26 | 64014.82 | 32.79 | 17.21 | 516.29 | 28.47 | 266584.40 |
| 2SLS X6 only | -5.62 | -4.38 | 88.23 | 10.50 | 7795.79 | 41.92 | 8.08 | 160.05 | 20.31 | 25655.91 |
| 2SLS X7 only | -9.23 | -0.77 | 11.85 | 6.59 | 140.94 | 48.61 | 1.39 | 23.49 | 12.91 | 553.38 |
| 2SLS X8 only | -9.81 | -0.19 | 14.09 | 6.96 | 198.48 | 49.74 | 0.26 | 28.04 | 13.67 | 785.29 |
| 2SLS X9 only | -11.88 | 1.88 | 39.05 | 8.33 | 1526.89 | 53.94 | -3.94 | 77.55 | 16.22 | 6024.23 |
| 2SLS X10 only | -6.72 | -3.28 | 78.76 | 9.19 | 6207.40 | 44.18 | 5.82 | 136.17 | 17.46 | 18558.62 |

**A 55: Full results simulation study 1a, 10 explanatory variables of the categorical process variable, average F-statistic of relevant instruments=11.2, average F-statistic of individual instrument=6.4, N=200, simulations=1000**

| | Group effect | | | | | | Post randomisation mediator effect | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean Coef. | Mean Bias | bias 2SLS/ bias OLS | SD coef. | Absolute Difference | MSE | Mean Coef. | Mean Bias | bias 2SLS/ bias OLS | SD coef. | Absolute Difference | MSE |
| OLS All | -13.209 | 3.209 | | 1.959 | 3.298 | 14.132 | 56.591 | -6.591 | | 2.438 | 6.598 | 49.381 |
| 2SLS All | -10.796 | 0.796 | 0.248 | 2.510 | 2.096 | 6.929 | 51.748 | -1.748 | 0.265 | 3.893 | 3.464 | 18.193 |
| OLS relevant | -13.219 | 3.219 | | 1.919 | 3.302 | 14.044 | 56.571 | -6.571 | | 2.403 | 6.574 | 48.941 |
| 2SLS relevant | -10.308 | 0.308 | 0.096 | 2.594 | 2.068 | 6.818 | 50.735 | -0.735 | 0.112 | 4.124 | 3.372 | 17.531 |
| OLS LASSO | -13.179 | 3.179 | | 1.923 | 3.263 | 13.799 | 56.493 | -6.493 | | 2.395 | 6.495 | 47.890 |
| 2SLS LASSO | -10.659 | 0.659 | 0.207 | 2.544 | 2.104 | 6.902 | 51.440 | -1.440 | 0.222 | 3.986 | 3.437 | 17.950 |
| OLS LASSO-m | -13.201 | 3.201 | | 1.930 | 3.284 | 13.965 | 56.530 | -6.530 | | 2.401 | 6.535 | 48.404 |
| 2SLS LASSO - m | -10.713 | 0.713 | 0.223 | 2.502 | 2.087 | **6.759** | 51.536 | -1.536 | 0.235 | 3.890 | 3.405 | **17.476** |
| OLS Elastic Net | -13.190 | 3.190 | | 1.920 | 3.273 | 13.862 | 56.513 | -6.513 | | 2.387 | 6.514 | 48.110 |
| 2SLS Elastic Net | -10.656 | **0.656** | 0.206 | 2.538 | 2.089 | 6.868 | 51.430 | **-1.430** | 0.220 | 3.964 | 3.418 | 17.744 |
| OLS Elastic Net - m | -13.203 | 3.203 | | 1.933 | 3.288 | 13.993 | 56.544 | -6.544 | | 2.410 | 6.550 | 48.626 |
| 2SLS Elastic Net - m | -10.726 | 0.726 | 0.227 | 2.499 | 2.087 | 6.766 | 51.572 | -1.572 | 0.240 | 3.889 | 3.414 | 17.581 |
| OLS Stepwise | -12.918 | 2.918 | | 1.870 | 3.025 | 12.008 | 56.199 | -6.199 | | 2.340 | 6.203 | 43.900 |
| 2SLS Stepwise | -10.678 | 0.678 | 0.232 | 2.697 | 2.177 | 7.726 | 51.495 | -1.495 | 0.241 | 4.531 | 3.815 | 22.745 |
| OLS no explanatory | -12.540 | 2.540 | | 1.852 | 2.698 | 9.877 | 55.264 | -5.264 | | 2.136 | 5.279 | 32.265 |
| 2SLS no explanatory | -10.995 | 0.995 | 0.392 | 3.643 | 2.978 | 14.247 | 52.167 | -2.167 | 0.412 | 6.523 | 5.410 | 47.203 |

**A 56: Simulation study 1a, 10 explanatory variables of the categorical process variable, average F-statistic of relevant instruments=11.2, average F-statistic of individual instrument=6.4, N=200, simulations=1000, individual explanatory variables only**

| | Group Effect | | | | | Post randomisation mediator effect | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Mean coefficient | Mean bias | SD coefficent | Absolute Difference | MSE | Mean coefficient | Mean bias | SD coefficent | Absolute Difference | MSE |
| 2SLS X1 only | -9.58 | -0.42 | 33.53 | 8.82 | 1123.28 | 49.11 | 0.89 | 70.45 | 17.60 | 4959.03 |
| 2SLS X2 only | -7.63 | -2.37 | 39.35 | 9.96 | 1552.78 | 45.75 | 4.25 | 73.35 | 19.39 | 5393.25 |
| 2SLS X3 only | -8.12 | -1.88 | 28.82 | 7.81 | 833.38 | 46.29 | 3.71 | 59.82 | 15.59 | 3588.53 |
| 2SLS X4 only | -21.43 | 11.43 | 259.09 | 19.86 | 67190.57 | 71.41 | -21.41 | 475.54 | 37.80 | 226368.70 |
| 2SLS X5 only | -9.46 | -0.54 | 29.47 | 8.79 | 867.66 | 49.25 | 0.75 | 59.31 | 17.20 | 3515.16 |
| 2SLS X6 only | -10.02 | 0.02 | 115.53 | 13.85 | 13334.25 | 50.79 | -0.79 | 237.46 | 27.70 | 56331.41 |
| 2SLS X7 only | -5.95 | -4.05 | 90.98 | 11.89 | 8284.88 | 42.37 | 7.63 | 178.91 | 23.41 | 32034.66 |
| 2SLS X8 only | -12.19 | 2.19 | 85.81 | 11.18 | 7360.11 | 54.77 | -4.77 | 182.08 | 22.62 | 33144.16 |
| 2SLS X9 only | -5.81 | -4.19 | 150.27 | 12.34 | 22577.51 | 42.50 | 7.50 | 278.89 | 23.60 | 77756.08 |
| 2SLS X10 only | -7.94 | -2.06 | 39.12 | 10.86 | 1533.28 | 46.15 | 3.85 | 76.92 | 21.37 | 5925.91 |

**A 57: Full results simulation study 1a, 10 explanatory variables of the categorical process variable, average F-statistic of relevant instruments=6.4, average F-statistic of individual instrument=4.8, N=200, simulations=1000**

| | Group effect | | | | | | Post randomisation mediator effect | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean Coef. | Mean Bias | bias 2SLS/ bias OLS | SD coef. | Absolute Difference | MSE | Mean Coef. | Mean Bias | bias 2SLS/ bias OLS | SD coef. | Absolute Difference | MSE |
| OLS All | -14.421 | 4.421 | | 1.849 | 4.441 | 22.960 | 58.992 | -8.992 | | 2.188 | 8.992 | 85.631 |
| 2SLS All | -11.571 | 1.571 | 0.355 | 2.651 | 2.459 | 9.490 | 53.287 | -3.287 | 0.366 | 4.219 | 4.384 | 28.586 |
| OLS relevant | -14.427 | 4.427 | | 1.810 | 4.443 | 22.871 | 58.967 | -8.967 | | 2.145 | 8.967 | 84.998 |
| 2SLS relevant | -10.708 | 0.708 | 0.160 | 2.821 | 2.310 | 8.452 | 51.528 | -1.528 | 0.170 | 4.675 | 3.953 | 24.164 |
| OLS LASSO | -14.265 | 4.265 | | 1.788 | 4.281 | 21.381 | 58.663 | -8.663 | | 2.127 | 8.663 | 79.565 |
| 2SLS LASSO | -10.349 | **0.349** | 0.082 | 5.924 | 3.430 | 35.177 | 52.987 | -2.987 | 0.345 | 4.693 | 4.539 | 30.922 |
| OLS LASSO-m | -14.399 | 4.399 | | 1.812 | 4.419 | 22.631 | 58.914 | -8.914 | | 2.168 | 8.914 | 84.160 |
| 2SLS LASSO - m | -11.392 | 1.392 | 0.316 | 2.679 | 2.434 | **9.106** | 52.900 | **-2.900** | 0.325 | 4.352 | 4.266 | **27.327** |
| OLS Elastic Net | -14.290 | 4.290 | | 1.798 | 4.309 | 21.634 | 58.728 | -8.728 | | 2.131 | 8.728 | 80.712 |
| 2SLS Elastic Net | -10.712 | 0.712 | 0.166 | 5.079 | 3.112 | 26.278 | 53.003 | -3.003 | 0.344 | 4.634 | 4.490 | 30.473 |
| OLS Elastic Net - m | -14.400 | 4.400 | | 1.823 | 4.419 | 22.676 | 58.939 | -8.939 | | 2.173 | 8.939 | 84.619 |
| 2SLS Elastic Net - m | -11.364 | 1.364 | 0.310 | 2.971 | 2.482 | 10.680 | 52.973 | -2.973 | 0.333 | 4.370 | 4.310 | 27.913 |
| OLS Stepwise | -14.057 | 4.057 | | 1.777 | 4.079 | 19.613 | 58.527 | -8.527 | | 2.098 | 8.527 | 77.103 |
| 2SLS Stepwise | -11.381 | 1.381 | 0.341 | 2.882 | 2.560 | 10.205 | 52.925 | -2.925 | 0.343 | 5.047 | 4.748 | 34.001 |
| OLS no explanatory | -13.800 | 3.800 | | 1.798 | 3.835 | 17.668 | 57.760 | -7.760 | | 2.019 | 7.760 | 64.294 |
| 2SLS no explanatory | -11.912 | 1.912 | 0.503 | 3.880 | 3.446 | 18.696 | 54.005 | -4.005 | 0.516 | 7.005 | 6.389 | 65.063 |

**A 58: Simulation study 1a, 10 explanatory variables of the categorical process variable, average F-statistic of relevant instruments=6.4, average F-statistic of individual instrument=4.8, N=200, simulations=1000, individual explanatory variables only**

| | Group Effect | | | | | Post randomisation mediator effect | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Mean coefficient | Mean bias | SD coefficent | Absolute Difference | MSE | Mean coefficient | Mean bias | SD coefficent | Absolute Difference | MSE |
| 2SLS X1 only | -14.36 | 4.36 | 279.81 | 22.02 | 78234.96 | 60.37 | -10.37 | 604.88 | 45.02 | 365619.30 |
| 2SLS X2 only | -7.18 | -2.82 | 117.62 | 18.91 | 13828.51 | 44.83 | 5.17 | 232.27 | 37.42 | 53924.30 |
| 2SLS X3 only | -6.63 | -3.37 | 49.58 | 12.11 | 2467.09 | 43.22 | 6.78 | 99.61 | 24.06 | 9958.92 |
| 2SLS X4 only | -1.33 | -8.67 | 204.87 | 22.59 | 42004.92 | 33.50 | 16.50 | 385.66 | 44.98 | 148859.50 |
| 2SLS X5 only | -6.11 | -3.89 | 104.39 | 17.10 | 10900.70 | 42.96 | 7.04 | 195.66 | 33.33 | 38295.57 |
| 2SLS X6 only | -7.02 | -2.98 | 59.02 | 13.34 | 3488.43 | 44.42 | 5.58 | 112.96 | 26.19 | 12777.79 |
| 2SLS X7 only | -12.74 | 2.74 | 95.79 | 17.29 | 9174.10 | 56.13 | -6.13 | 197.87 | 34.62 | 39152.78 |
| 2SLS X8 only | -15.07 | 5.07 | 240.96 | 20.25 | 58029.92 | 61.39 | -11.39 | 511.58 | 41.07 | 261577.00 |
| 2SLS X9 only | -7.84 | -2.16 | 38.26 | 11.81 | 1466.95 | 45.58 | 4.42 | 78.64 | 23.41 | 6198.02 |
| 2SLS X10 only | 1.29 | -11.29 | 229.69 | 23.57 | 52833.78 | 28.72 | 21.28 | 430.56 | 45.78 | 185646.30 |

**A 59: Full results simulation study 1a, 10 explanatory variables of the categorical process variable, average F-statistic of relevant instruments=31.2, average F-statistic of individual instrument=14.0, N=400, simulations=1000**

| | Group effect | | | | | | Post randomisation mediator effect | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean Coef. | Mean Bias | bias 2SLS/ bias OLS | SD coef. | Absolute Difference | MSE | Mean Coef. | Mean Bias | bias 2SLS/ bias OLS | SD coef. | Absolute Difference | MSE |
| OLS All | -11.723 | 1.723 | | 1.334 | 1.843 | 4.747 | 53.480 | -3.480 | | 1.745 | 3.504 | 15.151 |
| 2SLS All | -10.151 | 0.151 | 0.088 | 1.634 | 1.312 | 2.691 | 50.342 | -0.342 | 0.098 | 2.526 | 2.054 | 6.490 |
| OLS relevant | -11.734 | 1.734 | | 1.339 | 1.854 | 4.797 | 53.498 | -3.498 | | 1.763 | 3.527 | 15.342 |
| 2SLS relevant | -10.045 | 0.045 | 0.026 | 1.662 | 1.322 | 2.762 | 50.125 | -0.125 | 0.036 | 2.590 | 2.080 | 6.717 |
| OLS LASSO | -11.732 | 1.732 | | 1.339 | 1.849 | 4.790 | 53.473 | -3.473 | | 1.774 | 3.499 | 15.205 |
| 2SLS LASSO | -10.115 | **0.115** | 0.066 | 1.650 | 1.321 | 2.733 | 50.243 | **-0.243** | 0.070 | 2.557 | 2.071 | 6.589 |
| OLS LASSO-m | -11.725 | 1.725 | | 1.348 | 1.845 | 4.791 | 53.465 | -3.465 | | 1.749 | 3.488 | 15.063 |
| 2SLS LASSO - m | -10.146 | 0.146 | 0.085 | 1.648 | 1.319 | 2.734 | 50.312 | -0.312 | 0.090 | 2.523 | 2.057 | 6.457 |
| OLS Elastic Net | -11.729 | 1.729 | | 1.340 | 1.846 | 4.784 | 53.471 | -3.471 | | 1.764 | 3.495 | 15.159 |
| 2SLS Elastic Net | -10.118 | 0.118 | 0.068 | 1.651 | 1.322 | 2.737 | 50.253 | -0.253 | 0.073 | 2.553 | 2.066 | 6.575 |
| OLS Elastic Net - m | -11.723 | 1.723 | | 1.341 | 1.841 | 4.763 | 53.466 | -3.466 | | 1.745 | 3.488 | 15.053 |
| 2SLS Elastic Net - m | -10.141 | 0.141 | 0.082 | 1.640 | 1.310 | **2.707** | 50.309 | -0.309 | 0.089 | 2.517 | 2.046 | **6.422** |
| OLS Stepwise | -11.690 | 1.690 | | 1.342 | 1.816 | 4.655 | 53.435 | -3.435 | | 1.757 | 3.465 | 14.884 |
| 2SLS Stepwise | -10.123 | 0.123 | 0.073 | 1.672 | 1.328 | 2.808 | 50.259 | -0.259 | 0.075 | 2.625 | 2.103 | 6.950 |
| OLS no explanatory | -11.333 | 1.333 | | 1.279 | 1.524 | 3.411 | 52.689 | -2.689 | | 1.507 | 2.738 | 9.501 |
| 2SLS no explanatory | -10.213 | 0.213 | 0.159 | 2.576 | 2.037 | 6.674 | 50.453 | -0.453 | 0.168 | 4.675 | 3.717 | 22.042 |

**A 60: Simulation study 1a, 10 explanatory variables of the categorical process variable, average F-statistic of relevant instruments=31.2, average F-statistic of individual instrument=14.0, N=400, simulations=1000, individual explanatory variables only**

| | Group Effect | | | | | Post randomisation mediator effect | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Mean coefficient | Mean bias | SD coefficent | Absolute Difference | MSE | Mean coefficient | Mean bias | SD coefficent | Absolute Difference | MSE |
| 2SLS X1 only | -9.90 | -0.10 | 4.80 | 3.62 | 23.05 | 49.82 | 0.18 | 9.41 | 7.11 | 88.50 |
| 2SLS X2 only | -9.87 | -0.13 | 4.95 | 3.68 | 24.45 | 49.78 | 0.22 | 9.68 | 7.18 | 93.73 |
| 2SLS X3 only | -9.66 | -0.34 | 4.86 | 3.64 | 23.67 | 49.34 | 0.66 | 9.49 | 7.09 | 90.36 |
| 2SLS X4 only | -9.92 | -0.08 | 5.10 | 3.76 | 25.94 | 49.81 | 0.19 | 9.95 | 7.30 | 98.87 |
| 2SLS X5 only | -9.74 | -0.26 | 5.11 | 3.80 | 26.11 | 49.52 | 0.48 | 9.96 | 7.41 | 99.28 |
| 2SLS X6 only | -9.73 | -0.27 | 5.06 | 3.81 | 25.61 | 49.48 | 0.52 | 9.83 | 7.43 | 96.78 |
| 2SLS X7 only | -9.92 | -0.08 | 5.15 | 3.73 | 26.50 | 49.86 | 0.14 | 10.03 | 7.15 | 100.52 |
| 2SLS X8 only | -9.93 | -0.07 | 4.94 | 3.79 | 24.43 | 49.86 | 0.14 | 9.54 | 7.30 | 91.01 |
| 2SLS X9 only | -10.00 | 0.00 | 5.18 | 3.79 | 26.83 | 50.01 | -0.01 | 10.05 | 7.34 | 100.84 |
| 2SLS X10 only | -9.84 | -0.16 | 4.89 | 3.78 | 23.87 | 49.70 | 0.30 | 9.51 | 7.40 | 90.45 |

**A 61: Full results simulation study 1a, 10 explanatory variables of the categorical process variable, average F-statistic of relevant instruments=21.7, average F-statistic of individual instrument=11.9, N=400, simulations=1000**

| | Group effect | | | | | | Post randomisation mediator effect | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean Coef. | Mean Bias | bias 2SLS/ bias OLS | SD coef. | Absolute Difference | MSE | Mean Coef. | Mean Bias | bias 2SLS/ bias OLS | SD coef. | Absolute Difference | MSE |
| OLS All | -13.301 | 3.301 | | 1.281 | 3.303 | 12.538 | 56.633 | -6.633 | | 1.655 | 6.633 | 46.730 |
| 2SLS All | -10.416 | 0.416 | 0.126 | 1.685 | 1.399 | 3.008 | 50.875 | -0.875 | 0.132 | 2.713 | 2.311 | 8.118 |
| OLS relevant | -13.306 | 3.306 | | 1.281 | 3.306 | 12.568 | 56.640 | -6.640 | | 1.666 | 6.640 | 46.861 |
| 2SLS relevant | -10.161 | 0.161 | 0.049 | 1.730 | 1.389 | 3.018 | 50.360 | -0.360 | 0.054 | 2.787 | 2.256 | 7.886 |
| OLS LASSO | -13.304 | 3.304 | | 1.280 | 3.304 | 12.551 | 56.627 | -6.627 | | 1.658 | 6.627 | 46.661 |
| 2SLS LASSO | -10.307 | **0.307** | 0.093 | 1.716 | 1.402 | 3.037 | 50.645 | **-0.645** | 0.097 | 2.749 | 2.284 | 7.966 |
| OLS LASSO-m | -13.294 | 3.294 | | 1.283 | 3.296 | 12.497 | 56.612 | -6.612 | | 1.656 | 6.612 | 46.457 |
| 2SLS LASSO - m | -10.382 | 0.382 | 0.116 | 1.692 | 1.402 | **3.006** | 50.800 | -0.800 | 0.121 | 2.715 | 2.294 | 8.003 |
| OLS Elastic Net | -13.297 | 3.297 | | 1.281 | 3.298 | 12.511 | 56.620 | -6.620 | | 1.657 | 6.620 | 46.571 |
| 2SLS Elastic Net | -10.313 | 0.313 | 0.095 | 1.712 | 1.400 | 3.027 | 50.662 | -0.662 | 0.100 | 2.745 | 2.283 | **7.965** |
| OLS Elastic Net - m | -13.300 | 3.300 | | 1.283 | 3.301 | 12.533 | 56.614 | -6.614 | | 1.659 | 6.614 | 46.491 |
| 2SLS Elastic Net - m | -10.391 | 0.391 | 0.119 | 1.697 | 1.404 | 3.029 | 50.809 | -0.809 | 0.122 | 2.722 | 2.299 | 8.059 |
| OLS Stepwise | -13.205 | 3.205 | | 1.290 | 3.207 | 11.935 | 56.517 | -6.517 | | 1.654 | 6.517 | 45.202 |
| 2SLS Stepwise | -10.356 | 0.356 | 0.111 | 1.735 | 1.422 | 3.133 | 50.730 | -0.730 | 0.112 | 2.837 | 2.351 | 8.573 |
| OLS no explanatory | -12.658 | 2.658 | | 1.243 | 2.669 | 8.611 | 55.339 | -5.339 | | 1.478 | 5.339 | 30.688 |
| 2SLS no explanatory | -10.587 | 0.587 | 0.221 | 2.707 | 2.193 | 7.664 | 51.199 | -1.199 | 0.225 | 4.980 | 4.065 | 26.213 |

**A 62: Simulation study 1a, 10 explanatory variables of the categorical process variable, average F-statistic of relevant instruments=21.7, average F-statistic of individual instrument=11.9, N=400, simulations=1000, individual explanatory variables only**

| | Group Effect | | | | | Post randomisation mediator effect | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Mean coefficient | Mean bias | SD coefficent | Absolute Difference | MSE | Mean coefficient | Mean bias | SD coefficent | Absolute Difference | MSE |
| 2SLS X1 only | -9.78 | -0.22 | 5.37 | 3.94 | 28.84 | 49.57 | 0.43 | 10.55 | 7.78 | 111.43 |
| 2SLS X2 only | -9.69 | -0.31 | 5.99 | 4.11 | 35.89 | 49.41 | 0.59 | 11.97 | 8.06 | 143.39 |
| 2SLS X3 only | -9.62 | -0.38 | 6.95 | 4.16 | 48.33 | 49.30 | 0.70 | 14.12 | 8.17 | 199.62 |
| 2SLS X4 only | -9.71 | -0.29 | 5.83 | 4.16 | 34.07 | 49.39 | 0.61 | 11.44 | 8.11 | 131.05 |
| 2SLS X5 only | -9.49 | -0.51 | 6.80 | 4.28 | 46.49 | 49.00 | 1.00 | 13.19 | 8.39 | 174.85 |
| 2SLS X6 only | -9.32 | -0.68 | 8.19 | 4.46 | 67.54 | 48.66 | 1.34 | 16.31 | 8.74 | 267.63 |
| 2SLS X7 only | -9.67 | -0.33 | 5.82 | 4.12 | 34.00 | 49.35 | 0.65 | 11.47 | 7.94 | 131.77 |
| 2SLS X8 only | -9.64 | -0.36 | 6.43 | 4.26 | 41.38 | 49.27 | 0.73 | 12.82 | 8.28 | 164.69 |
| 2SLS X9 only | -9.29 | -0.71 | 19.53 | 4.75 | 381.35 | 48.52 | 1.48 | 41.66 | 9.35 | 1736.39 |
| 2SLS X10 only | -9.57 | -0.43 | 5.79 | 4.29 | 33.63 | 49.17 | 0.83 | 11.31 | 8.42 | 128.46 |

**A 63: Full results simulation study 1a, 10 explanatory variables of the categorical process variable, average F-statistic of relevant instruments=12.1, average F-statistic of individual instrument=8.5, N=400, simulations=1000**

| | Group effect | | | | | | Post randomisation mediator effect | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean Coef. | Mean Bias | bias 2SLS/ bias OLS | SD coef. | Absolute Difference | MSE | Mean Coef. | Mean Bias | bias 2SLS/ bias OLS | SD coef. | Absolute Difference | MSE |
| OLS All | -14.517 | 4.517 | | 1.230 | 4.517 | 21.913 | 59.070 | -9.070 | | 1.507 | 9.070 | 84.533 |
| 2SLS All | -10.862 | 0.862 | 0.191 | 1.836 | 1.633 | 4.111 | 51.767 | -1.767 | 0.195 | 3.137 | 2.939 | 12.954 |
| OLS relevant | -14.518 | 4.518 | | 1.225 | 4.518 | 21.906 | 59.066 | -9.066 | | 1.504 | 9.066 | 84.449 |
| 2SLS relevant | -10.367 | 0.367 | 0.081 | 1.908 | 1.556 | 3.772 | 50.775 | -0.775 | 0.086 | 3.259 | 2.701 | 11.212 |
| OLS LASSO | -14.507 | 4.507 | | 1.229 | 4.507 | 21.820 | 59.032 | -9.032 | | 1.496 | 9.032 | 83.814 |
| 2SLS LASSO | -10.697 | **0.697** | 0.155 | 1.864 | 1.597 | 3.958 | 51.419 | **-1.419** | 0.157 | 3.182 | 2.817 | 12.128 |
| OLS LASSO-m | -14.512 | 4.512 | | 1.228 | 4.512 | 21.867 | 59.052 | -9.052 | | 1.509 | 9.052 | 84.207 |
| 2SLS LASSO - m | -10.800 | 0.800 | 0.177 | 1.844 | 1.618 | 4.038 | 51.635 | -1.635 | 0.181 | 3.154 | 2.891 | **12.611** |
| OLS Elastic Net | -14.509 | 4.509 | | 1.226 | 4.509 | 21.834 | 59.038 | -9.038 | | 1.493 | 9.038 | 83.913 |
| 2SLS Elastic Net | -10.704 | 0.704 | 0.156 | 1.857 | 1.596 | **3.939** | 51.433 | -1.433 | 0.159 | 3.174 | 2.811 | 12.116 |
| OLS Elastic Net - m | -14.519 | 4.519 | | 1.225 | 4.519 | 21.922 | 59.058 | -9.058 | | 1.506 | 9.058 | 84.310 |
| 2SLS Elastic Net - m | -10.818 | 0.818 | 0.181 | 1.840 | 1.621 | 4.050 | 51.663 | -1.663 | 0.184 | 3.156 | 2.895 | 12.714 |
| OLS Stepwise | -14.338 | 4.338 | | 1.220 | 4.338 | 20.307 | 58.852 | -8.852 | | 1.480 | 8.852 | 80.552 |
| 2SLS Stepwise | -10.751 | 0.751 | 0.173 | 1.919 | 1.664 | 4.242 | 51.539 | -1.539 | 0.174 | 3.400 | 3.017 | 13.918 |
| OLS no explanatory | -13.910 | 3.910 | | 1.209 | 3.911 | 16.748 | 57.845 | -7.845 | | 1.403 | 7.845 | 63.503 |
| 2SLS no explanatory | -11.215 | 1.215 | 0.311 | 2.976 | 2.593 | 10.322 | 52.451 | -2.451 | 0.313 | 5.649 | 4.974 | 37.889 |

**A 64: Simulation study 1a, 10 explanatory variables of the categorical process variable, average F-statistic of relevant instruments=12.1, average F-statistic of individual instrument=8.5, N=400, simulations=1000, individual explanatory variables only**

| | Group Effect | | | | | Post randomisation mediator effect | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Mean coefficient | Mean bias | SD coefficent | Absolute Difference | MSE | Mean coefficient | Mean bias | SD coefficent | Absolute Difference | MSE |
| 2SLS X1 only | -10.26 | 0.26 | 29.56 | 5.91 | 872.99 | 50.40 | -0.40 | 54.97 | 11.58 | 3019.08 |
| 2SLS X2 only | -9.88 | -0.12 | 13.78 | 5.74 | 189.66 | 49.74 | 0.26 | 26.70 | 11.31 | 712.51 |
| 2SLS X3 only | -8.73 | -1.27 | 11.67 | 5.68 | 137.66 | 47.52 | 2.48 | 23.66 | 11.23 | 565.37 |
| 2SLS X4 only | -8.11 | -1.89 | 23.95 | 6.62 | 576.43 | 46.18 | 3.82 | 48.28 | 13.06 | 2343.12 |
| 2SLS X5 only | -8.91 | -1.09 | 27.56 | 6.53 | 760.13 | 47.85 | 2.15 | 53.37 | 12.89 | 2850.11 |
| 2SLS X6 only | -8.51 | -1.49 | 13.18 | 6.11 | 175.76 | 47.08 | 2.92 | 26.13 | 12.08 | 690.70 |
| 2SLS X7 only | -9.76 | -0.24 | 15.36 | 5.94 | 235.75 | 49.62 | 0.38 | 31.35 | 11.65 | 981.87 |
| 2SLS X8 only | -9.44 | -0.56 | 17.95 | 5.91 | 322.21 | 48.88 | 1.12 | 36.72 | 11.67 | 1348.31 |
| 2SLS X9 only | -10.17 | 0.17 | 27.70 | 6.56 | 766.69 | 50.29 | -0.29 | 53.53 | 12.87 | 2862.36 |
| 2SLS X10 only | -9.04 | -0.96 | 13.09 | 6.18 | 172.22 | 48.02 | 1.98 | 26.32 | 12.23 | 696.17 |

**A 65: MEDIAN ABSOLUTE BIAS of estimates by estimation method with increasing uncertainty in post-randomisation process variable: 1000 simulations, CONTINUOUS process variable, covariance of error terms =0.69, sample size=200**

| Variance of mediator error | Average first stage f-statistic | | | OLS | ALL | | | | LASSO | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | All relevant variable | ALL variables | Effect | All | 2SLS | LIML | Fuller | GMM | 2SLS | LIML | Fuller | GMM |
| 5 explanatory variables | | | | | | | | | | | | |
| 1 | 140.7 | 46.2 | Group | 1.012 | 0.992 | 0.981 | **0.981** | 1.038 | 0.976 | 0.976 | **0.976** | 1.041 |
| | | | Mediator | 0.508 | 0.171 | 0.173 | 0.171 | **0.168** | 0.168 | **0.170** | 0.170 | 0.173 |
| 2 | 30.7 | 10.6 | Group | 1.032 | 0.979 | 0.976 | **0.973** | 1.059 | 0.971 | 0.970 | **0.967** | 1.021 |
| | | | Mediator | 0.879 | 0.204 | 0.193 | **0.189** | 0.200 | 0.197 | **0.184** | 0.185 | 0.197 |
| 3 | 10.3 | 4 | Group | 1.026 | 1.000 | 1.007 | **0.993** | 1.037 | 0.983 | **0.982** | 0.984 | 1.001 |
| | | | Mediator | 1.083 | 0.292 | 0.244 | **0.240** | 0.290 | 0.271 | **0.231** | 0.232 | 0.263 |
| 4 | 3.3 | 1.8 | Group | 1.051 | **0.983** | 1.195 | 1.134 | 1.053 | **0.978** | 1.019 | 1.000 | 0.987 |
| | | | Mediator | 1.154 | 0.706 | 0.636 | **0.594** | 0.703 | 0.638 | **0.522** | 0.543 | 0.636 |
| 10 explanatory variables | | | | | | | | | | | | |
| 1 | 161.4 | 87.4 | Group | 1.034 | 0.980 | 0.979 | **0.977** | 1.051 | 0.962 | 0.959 | **0.958** | 1.062 |
| | | | Mediator | 0.450 | 0.167 | **0.165** | 0.166 | 0.174 | 0.165 | **0.163** | 0.163 | 0.178 |
| 2 | 35 | 19.3 | Group | 1.021 | 0.986 | **0.967** | 0.971 | 1.036 | 0.965 | 0.967 | **0.963** | 1.056 |
| | | | Mediator | 0.798 | 0.192 | **0.177** | 0.178 | 0.196 | 0.188 | 0.176 | **0.174** | 0.195 |
| 3 | 11.7 | 6.7 | Group | 1.042 | 0.991 | 0.984 | **0.982** | 1.031 | **0.968** | 0.975 | 0.974 | 1.059 |
| | | | Mediator | 0.997 | 0.259 | 0.223 | **0.218** | 0.269 | 0.252 | **0.213** | 0.214 | 0.266 |
| 4 | 3.5 | 2.4 | Group | 1.007 | **0.970** | 1.026 | 1.006 | 1.027 | **0.957** | 0.970 | 0.967 | 1.006 |
| | | | Mediator | 1.087 | 0.525 | 0.404 | **0.387** | 0.527 | 0.498 | **0.377** | 0.385 | 0.497 |

**A 66: MEDIAN ABSOLUTE BIAS of estimates by estimation method with increasing uncertainty in post-randomisation process variable: 1000 simulations, CATEGORICAL process variable, covariance of error terms =0.69, sample size=200**

| Variance of mediator error | Effect | OLS | ALL | | | | LASSO | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | All | 2SLS | LIML | Fuller | GMM | 2SLS | LIML | Fuller | GMM |
| 5 explanatory variables | | | | | | | | | | |
| 1 | Group | 2.031 | **1.591** | 1.784 | 1.766 | 1.692 | **1.632** | 1.685 | 1.674 | 1.633 |
| | Mediator | 3.850 | **2.429** | 2.734 | 2.685 | 2.555 | **2.480** | 2.603 | 2.587 | 2.565 |
| 2 | Group | 3.521 | **1.707** | 1.904 | 1.889 | 1.803 | 1.677 | 1.756 | 1.752 | **1.662** |
| | Mediator | 6.975 | **2.813** | 2.998 | 2.983 | 2.869 | 2.768 | 2.744 | 2.732 | **2.708** |
| 3 | Group | 4.786 | **2.243** | 2.425 | 2.391 | 2.290 | **2.061** | 2.134 | 2.112 | 2.101 |
| | Mediator | 9.511 | **3.909** | 4.289 | 4.198 | 3.899 | 3.724 | 3.479 | **3.458** | 3.650 |
| 10 explanatory variables | | | | | | | | | | |
| 1 | Group | 1.949 | **1.560** | 1.705 | 1.677 | 1.685 | **1.604** | 1.694 | 1.693 | 1.756 |
| | Mediator | 3.591 | **2.585** | 2.806 | 2.808 | 2.755 | **2.591** | 2.798 | 2.776 | 2.810 |
| 2 | Group | 3.310 | **1.729** | 1.954 | 1.912 | 1.874 | **1.756** | 1.874 | 1.853 | 1.851 |
| | Mediator | 6.664 | **2.900** | 3.137 | 3.084 | 3.055 | **2.816** | 2.991 | 2.958 | 3.014 |
| 3 | Group | 4.585 | **2.176** | 2.310 | 2.282 | 2.263 | **2.085** | 2.139 | 2.110 | 2.148 |
| | Mediator | 9.067 | **3.688** | 4.276 | 4.224 | 3.834 | **3.671** | 3.894 | 3.771 | 3.756 |

**Appendix 2: Full EDIE-II data results**

**2.1. Impact of attendance at therapy sessions on secondary outcomes by analysis method**

**A 67: Attendance as a continuous measure, complete case analysis**

| | | OLS | | | IV - 2SLS | | | IV in 2 stages - all observations | | | G-estimation | | | Mplus - all obs | | | Mplus - complete case | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Outcome | N | Coef. | Std. Err. | p-value | Coef. | Std. Err. | p-value | Coef | Boot S.E. | p-value | Coef | Boot S.E. | p-value | Coef | std.err | p-value | Coef | std.err | p-value |
| Severity | 187 | -0.47 | 0.17 | 0.008 | -0.66 | 0.22 | 0.002 | -0.82 | 0.22 | <0.001 | -0.85 | 0.26 | 0.001 | -0.79 | 0.28 | 0.005 | -0.67 | 0.22 | 0.002 |
| Distress | 166 | -0.33 | 0.19 | 0.085 | -0.40 | 0.24 | 0.090 | -0.50 | 0.23 | 0.030 | -0.54 | 0.29 | 0.062 | -0.45 | 0.31 | 0.143 | -0.40 | 0.24 | 0.090 |
| BDI | 183 | -0.04 | 0.05 | 0.431 | -0.03 | 0.06 | 0.598 | -0.04 | 0.07 | 0.552 | -0.05 | 0.08 | 0.531 | -0.03 | 0.08 | 0.722 | -0.03 | 0.06 | 0.598 |
| SIAS | 178 | 0.01 | 0.16 | 0.947 | -0.07 | 0.20 | 0.744 | -0.08 | 0.20 | 0.680 | -0.10 | 0.24 | 0.670 | -0.12 | 0.26 | 0.646 | -0.07 | 0.20 | 0.743 |
| MANSA | 173 | 0.00 | 0.10 | 0.964 | 0.04 | 0.12 | 0.739 | 0.05 | 0.11 | 0.659 | 0.03 | 0.13 | 0.814 | 0.07 | 0.16 | 0.657 | 0.04 | 0.12 | 0.739 |

**A 68: Attendance as a categorical, four or more sessions, complete case analysis**

| | | OLS | | | IV | | | IV in 2 stages - all observations | | | G-estimation | | | Mplus - all obs | | | Mplus - complete case | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Outcome | N | Coef. | Std. Err. | p-value | Coef. | Std. Err. | p-value | Coef | Boot S.E. | p-value | Coef | Boot S.E. | p-value | Coef | std.err | p-value | Coef | std.err | p-value |
| Severity | 187 | -7.07 | 2.46 | 0.005 | -8.49 | 2.77 | 0.002 | -9.98 | 2.84 | <0.001 | -10.43 | 3.34 | 0.002 | -9.55 | 3.38 | 0.005 | -8.49 | 2.77 | 0.002 |
| Distress | 166 | -3.74 | 2.73 | 0.172 | -5.13 | 3.04 | 0.091 | -6.05 | 2.96 | 0.041 | -6.44 | 3.67 | 0.080 | -6.23 | 3.73 | 0.094 | -5.13 | 3.04 | 0.091 |
| BDI | 169 | -0.27 | 0.71 | 0.702 | -0.42 | 0.79 | 0.599 | -0.49 | 0.87 | 0.571 | -0.56 | 0.95 | 0.556 | -0.53 | 0.98 | 0.586 | -0.42 | 0.79 | 0.599 |
| SIAS | 154 | -0.88 | 2.27 | 0.700 | -0.82 | 2.49 | 0.743 | -0.98 | 2.52 | 0.696 | -1.39 | 2.98 | 0.642 | -0.76 | 3.16 | 0.810 | -0.82 | 2.50 | 0.743 |
| MANSA | 156 | 0.60 | 1.36 | 0.660 | 0.51 | 1.53 | 0.739 | 0.61 | 1.44 | 0.672 | 0.52 | 1.67 | 0.756 | 0.44 | 1.91 | 0.819 | 0.51 | 1.53 | 0.739 |

## 2.2. Full results of process variable analysis for each type of estimator

**A 69: Agreement of problems and goals and randomisation on symptom severity at 12 months, imputed data, comparison of estimators**

| Effect | Coefficient | Std.Error | 95% CI | | p-value | Coefficient | Std.Error | 95% CI | | p-value |
|---|---|---|---|---|---|---|---|---|---|---|
| | All instruments - 2SLS | | | | | LASSO instruments - 2SLS | | | | |
| Randomisation | -2.45 | 6.58 | -15.41 | 10.51 | 0.710 | 0.56 | 6.07 | -11.33 | 12.44 | 0.927 |
| Problem agreement | -7.36 | 8.76 | -24.65 | 9.93 | 0.402 | -12.35 | 8.15 | -28.34 | 3.63 | 0.130 |
| | All instruments - LIML | | | | | LASSO instruments - LIML | | | | |
| andomisation | 0.30 | 8.67 | -16.88 | 17.49 | 0.972 | 4.10 | 8.73 | -13.02 | 21.22 | 0.639 |
| Problem agreement | -11.26 | 11.87 | -34.83 | 12.32 | 0.345 | -17.33 | 12.05 | -40.96 | 6.29 | 0.151 |
| | All instruments - GMM | | | | | LASSO instruments - GMM | | | | |
| Randomisation | -0.15 | 5.40 | -10.78 | 10.47 | 0.978 | 0.64 | 5.94 | -11.00 | 12.29 | 0.914 |
| Problem agreement | -10.63 | 7.34 | -25.09 | 3.82 | 0.149 | -13.91 | 7.87 | -29.33 | 1.51 | 0.077 |
| | All instruments - FIML | | | | | LASSO instruments - FIML | | | | |
| Randomisation | 0.06 | 8.49 | -16.76 | 16.87 | 0.995 | 3.72 | 8.44 | -12.81 | 20.26 | 0.659 |
| Problem agreement | -10.90 | 11.60 | -33.93 | 12.12 | 0.349 | -16.80 | 11.62 | -39.58 | 5.98 | 0.148 |

**A 70: Formulation and randomisation on symptom severity at 12 months, imputed data, comparison of estimators**

| Effect | Coefficient | Std.Error | 95% CI | | p-value | Coefficient | Std.Error | 95% CI | | p-value |
|---|---|---|---|---|---|---|---|---|---|---|
| | All instruments - 2SLS | | | | | LASSO instruments - 2SLS | | | | |
| Randomisation | 1.58 | 6.41 | -11.04 | 14.19 | 0.806 | 4.13 | 6.21 | -8.04 | 16.29 | 0.506 |
| Formulation | -14.84 | 9.54 | -33.64 | 3.95 | 0.121 | -19.39 | 8.58 | -36.21 | -2.56 | 0.024 |
| | All instruments - LIML | | | | | LASSO instruments - LIML | | | | |
| Randomisation | 10.32 | 9.06 | -7.53 | 28.17 | 0.256 | 10.57 | 9.07 | -7.20 | 28.35 | 0.244 |
| Formulation | -28.91 | 13.95 | -56.44 | -1.38 | 0.040 | -29.81 | 13.20 | -55.69 | -3.93 | 0.024 |
| | All instruments - GMM | | | | | LASSO instruments - GMM | | | | |
| Randomisation | 2.07 | 4.95 | -7.69 | 11.83 | 0.677 | 2.66 | 5.97 | -9.04 | 14.36 | 0.656 |
| Formulation | -15.90 | 6.86 | -29.41 | -2.39 | 0.021 | -18.39 | 9.11 | -36.24 | -0.54 | 0.047 |
| | All instruments - FIML | | | | | LASSO instruments - FIML | | | | |
| Randomisation | 9.20 | 8.68 | -7.89 | 26.29 | 0.290 | 11.75 | 9.78 | -7.41 | 30.91 | 0.230 |
| Formulation | -27.11 | 13.32 | -53.39 | -0.83 | 0.043 | -31.28 | 14.25 | -59.22 | -3.34 | 0.029 |

**A 71: Proportion of sessions involving homework and randomisation on symptom severity at 12 months, imputed data, comparison of estimators**

| Effect | Coefficient | Std.Error | 95% CI | | p-value | Coefficient | Std.Error | 95% CI | | p-value |
|---|---|---|---|---|---|---|---|---|---|---|
| | All instruments - 2SLS | | | | | LASSO instruments - 2SLS | | | | |
| Randomisation | 0.106 | 6.166 | -12.093 | 12.306 | 0.986 | 0.063 | 5.261 | -10.248 | 10.374 | 0.990 |
| % homework | -0.182 | 0.132 | -0.442 | 0.078 | 0.168 | -0.186 | 0.113 | -0.407 | 0.035 | 0.100 |
| | All instruments - LIML | | | | | LASSO instruments - LIML | | | | |
| Randomisation | 4.714 | 8.362 | -11.940 | 21.368 | 0.575 | 4.250 | 8.499 | -12.408 | 20.909 | 0.618 |
| % homework | -0.291 | 0.185 | -0.660 | 0.078 | 0.121 | -0.285 | 0.192 | -0.661 | 0.091 | 0.140 |
| | All instruments - GMM | | | | | LASSO instruments - GMM | | | | |
| Randomisation | 3.096 | 5.347 | -7.610 | 13.802 | 0.565 | 3.241 | 5.198 | -6.948 | 13.430 | 0.534 |
| % homework | -0.233 | 0.110 | -0.452 | -0.015 | 0.037 | -0.241 | 0.108 | -0.453 | -0.030 | 0.027 |
| | All instruments - FIML | | | | | LASSO instruments - FIML | | | | |
| Randomisation | 4.220 | 8.105 | -11.902 | 20.343 | 0.604 | 3.795 | 8.100 | -12.081 | 19.671 | 0.640 |
| % homework | -0.279 | 0.179 | -0.635 | 0.077 | 0.123 | -0.274 | 0.182 | -0.632 | 0.083 | 0.135 |

**A 72: More than half of sessions involving homework and randomisation on symptom severity at 12 months, imputed data, comparison of estimators**

| Effect | Coefficient | Std.Error | 95% CI | | p-value | Coefficient | Std.Error | 95% CI | | p-value |
|---|---|---|---|---|---|---|---|---|---|---|
| | All instruments - 2SLS | | | | | LASSO instruments - 2SLS | | | | |
| Randomisation | -2.12 | 5.14 | -12.31 | 8.08 | 0.682 | -1.62 | 4.73 | -10.88 | 7.64 | 0.733 |
| >50% homework | -10.88 | 8.65 | -27.97 | 6.21 | 0.210 | -12.17 | 7.75 | -27.35 | 3.02 | 0.120 |
| | All instruments - LIML | | | | | LASSO instruments - LIML | | | | |
| Randomisation | 1.65 | 7.89 | -14.24 | 17.53 | 0.836 | 1.99 | 8.54 | -14.75 | 18.73 | 0.817 |
| >50% homework | -18.29 | 14.38 | -47.18 | 10.59 | 0.209 | -19.29 | 15.71 | -50.08 | 11.51 | 0.225 |
| | All instruments - GMM | | | | | LASSO instruments - GMM | | | | |
| Randomisation | 0.66 | 4.46 | -8.36 | 9.69 | 0.883 | 1.10 | 4.72 | -8.15 | 10.34 | 0.818 |
| >50% homework | -14.52 | 7.28 | -29.16 | 0.13 | 0.052 | -15.79 | 7.75 | -30.99 | -0.60 | 0.049 |
| | All instruments - FIML | | | | | LASSO instruments - FIML | | | | |
| Randomisation | 1.17 | 7.52 | -13.94 | 16.28 | 0.877 | 1.48 | 7.83 | -13.87 | 16.83 | 0.851 |
| >50% homework | -17.35 | 13.63 | -44.66 | 9.96 | 0.208 | -18.30 | 14.32 | -46.37 | 9.76 | 0.206 |

**A 73: Proportion of sessions involving active change strategies and randomisation on symptom severity at 12 months, imputed data, comparison of estimators**

| Effect | Coefficient | Std.Error | 95% CI | | p-value | Coefficient | Std.Error | 95% CI | | p-value |
|---|---|---|---|---|---|---|---|---|---|---|
| | All instruments - 2SLS | | | | | LASSO instruments - 2SLS | | | | |
| Randomisation | 3.279 | 6.760 | -10.095 | 16.653 | 0.628 | 3.203 | 5.979 | -8.516 | 14.923 | 0.594 |
| % change strategies | -0.224 | 0.128 | -0.476 | 0.029 | 0.082 | -0.225 | 0.113 | -0.446 | -0.004 | 0.050 |
| | All instruments - LIML | | | | | LASSO instruments - LIML | | | | |
| Randomisation | 7.823 | 7.989 | -7.979 | 23.625 | 0.329 | 8.084 | 8.421 | -8.420 | 24.589 | 0.341 |
| % change strategies | -0.317 | 0.154 | -0.622 | -0.012 | 0.041 | -0.325 | 0.164 | -0.647 | -0.004 | 0.052 |
| | All instruments - GMM | | | | | LASSO instruments - GMM | | | | |
| Randomisation | 4.341 | 4.960 | -5.434 | 14.116 | 0.382 | 3.318 | 5.494 | -7.450 | 14.086 | 0.548 |
| % change strategies | -0.245 | 0.092 | -0.426 | -0.064 | 0.008 | -0.236 | 0.100 | -0.433 | -0.040 | 0.020 |
| | All instruments - FIML | | | | | LASSO instruments - FIML | | | | |
| Randomisation | 7.372 | 7.847 | -8.142 | 22.887 | 0.349 | 7.622 | 8.171 | -8.393 | 23.637 | 0.355 |
| % change strategies | -0.308 | 0.151 | -0.606 | -0.009 | 0.043 | -0.316 | 0.159 | -0.628 | -0.004 | 0.052 |

**A 74: More than half of sessions involving active change strategies and randomisation on symptom severity at 12 months, imputed data, comparison of estimators**

| Effect | Coefficient | Std.Error | 95% CI | | p-value | Coefficient | Std.Error | 95% CI | | p-value |
|---|---|---|---|---|---|---|---|---|---|---|
| | All instruments - 2SLS | | | | | LASSO instruments - 2SLS | | | | |
| Randomisation | 0.05 | 5.99 | -11.94 | 12.04 | 0.993 | 0.22 | 5.23 | -10.03 | 10.47 | 0.967 |
| >50% change strategies | -13.64 | 9.44 | -32.55 | 5.26 | 0.154 | -14.46 | 7.88 | -29.92 | 0.99 | 0.069 |
| | All instruments - LIML | | | | | LASSO instruments - LIML | | | | |
| Randomisation | 5.68 | 8.00 | -10.41 | 21.77 | 0.481 | 5.44 | 7.75 | -9.75 | 20.64 | 0.484 |
| >50% change strategies | -23.65 | 13.23 | -50.31 | 3.01 | 0.081 | -23.86 | 12.63 | -48.62 | 0.91 | 0.061 |
| | All instruments - GMM | | | | | LASSO instruments - GMM | | | | |
| Randomisation | 0.96 | 4.50 | -7.96 | 9.88 | 0.831 | 0.39 | 4.59 | -8.60 | 9.38 | 0.932 |
| >50% change strategies | -15.17 | 6.74 | -28.48 | -1.87 | 0.026 | -15.12 | 7.00 | -28.84 | -1.40 | 0.031 |
| | All instruments - FIML | | | | | LASSO instruments - FIML | | | | |
| Randomisation | 5.03 | 7.70 | -10.43 | 20.48 | 0.517 | 5.67 | 7.48 | -8.98 | 20.33 | 0.448 |
| >50% change strategies | -22.49 | 12.68 | -47.98 | 3.01 | 0.082 | -24.04 | 12.09 | -47.73 | -0.35 | 0.048 |

**A 75: All components of therapy and randomisation on symptom severity at 12 months, imputed data, comparison of estimators**

| Effect | Coefficient | Std.Error | 95% CI | | p-value | Coefficient | Std.Error | 95% CI | | p-value |
|---|---|---|---|---|---|---|---|---|---|---|
| | All instruments - 2SLS | | | | | LASSO instruments - 2SLS | | | | |
| Randomisation | -3.31 | 3.53 | -10.25 | 3.62 | 0.348 | -3.13 | 3.34 | -9.67 | 3.41 | 0.350 |
| All Components | -13.98 | 8.13 | -29.97 | 2.01 | 0.086 | -15.19 | 8.05 | -30.96 | 0.59 | 0.065 |
| | All instruments - LIML | | | | | LASSO instruments - LIML | | | | |
| Randomisation | -1.15 | 4.11 | -9.24 | 6.94 | 0.780 | -0.72 | 4.32 | -9.19 | 7.75 | 0.868 |
| All Components | -20.96 | 10.34 | -41.34 | -0.57 | 0.044 | -23.02 | 12.24 | -47.01 | 0.97 | 0.067 |
| | All instruments - GMM | | | | | LASSO instruments - GMM | | | | |
| Randomisation | -2.21 | 2.96 | -8.14 | 3.71 | 0.458 | -2.48 | 3.24 | -8.83 | 3.87 | 0.447 |
| All Components | -15.73 | 6.30 | -28.16 | -3.30 | 0.013 | -16.90 | 7.59 | -31.78 | -2.02 | 0.031 |
| | All instruments - FIML | | | | | LASSO instruments - FIML | | | | |
| Randomisation | -1.36 | 4.05 | -9.33 | 6.60 | 0.737 | -0.98 | 4.19 | -9.19 | 7.24 | 0.816 |
| All Components | -20.27 | 10.12 | -40.21 | -0.33 | 0.046 | -22.19 | 11.72 | -45.15 | 0.78 | 0.064 |

**A 76: All and some components of therapy on symptom severity at 12 months, imputed data, comparison of estimators**

| Effect | Coefficient | Std.Error | 95% CI | | p-value | Coefficient | Std.Error | 95% CI | | p-value |
|---|---|---|---|---|---|---|---|---|---|---|
| | All instruments - 2SLS | | | | | LASSO instruments - 2SLS | | | | |
| Some therapy | -4.37 | 4.63 | -13.47 | 4.73 | 0.346 | 17.13 | 8.08 | 1.29 | 32.97 | 0.045 |
| Full therapy | -17.05 | 6.14 | -29.11 | -4.98 | 0.006 | -1.42 | 5.49 | -12.18 | 9.33 | 0.797 |
| | All instruments - LIML | | | | | LASSO instruments - LIML | | | | |
| Some therapy | -0.89 | 6.04 | -12.79 | 11.01 | 0.883 | 32.30 | 23.35 | -13.48 | 78.07 | 0.176 |
| Full therapy | -22.48 | 8.20 | -38.66 | -6.30 | 0.007 | -11.08 | 15.26 | -40.99 | 18.84 | 0.473 |
| | All instruments - GMM | | | | | LASSO instruments - GMM | | | | |
| Some therapy | -3.13 | 3.91 | -10.96 | 4.70 | 0.427 | 17.97 | 7.11 | 4.04 | 31.90 | 0.016 |
| Full therapy | -17.72 | 4.88 | -27.31 | -8.14 | 0.000 | -2.43 | 5.23 | -12.68 | 7.82 | 0.647 |
| | All instruments - FIML | | | | | LASSO instruments - FIML | | | | |
| Some therapy | -1.29 | 5.88 | -12.86 | 10.28 | 0.827 | 29.53 | 18.87 | -7.45 | 66.51 | 0.127 |
| Full therapy | -21.88 | 7.96 | -37.57 | -6.19 | 0.006 | -9.30 | 12.34 | -33.48 | 14.88 | 0.456 |

**2.3 Results of session and session*process interaction on belief mediators**

**A 77: Attendance and attendance by post-randomisation process variable interactions on belief mediators at 6 months, imputed data results**

| | | Agreement of problems and goals | | | Formulation | | | >50% homework | | | >50% active change strategies | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Outcome | Effect | Coef. | Std. Err. | p-value | Coef. | Std. Err. | p-value | Coef. | Std. Err. | p-value | Coef. | Std. Err. | p-value |
| MCQ Cog Con | sessions | 0.189 | 0.471 | 0.689 | 0.156 | 0.430 | 0.718 | 0.085 | 0.215 | 0.695 | -0.127 | 0.338 | 0.708 |
| | process*sessions | -0.241 | 0.516 | 0.640 | -0.211 | 0.515 | 0.683 | -0.174 | 0.313 | 0.578 | 0.160 | 0.425 | 0.707 |
| MCQ Cog Self | sessions | -0.358 | 0.492 | 0.467 | -0.322 | 0.441 | 0.466 | -0.036 | 0.235 | 0.878 | -0.297 | 0.297 | 0.317 |
| | process*sessions | 0.391 | 0.549 | 0.476 | 0.379 | 0.532 | 0.476 | 0.008 | 0.351 | 0.983 | 0.374 | 0.390 | 0.338 |
| MCQ Neg Thoughts | sessions | 0.342 | 0.536 | 0.525 | 0.643 | 0.508 | 0.210 | 0.085 | 0.247 | 0.733 | 0.216 | 0.337 | 0.522 |
| | process*sessions | -0.477 | 0.593 | 0.423 | -0.871 | 0.616 | 0.163 | -0.305 | 0.382 | 0.427 | -0.409 | 0.440 | 0.354 |
| MCQ Neg Control | sessions | -0.312 | 0.390 | 0.424 | -0.027 | 0.389 | 0.945 | -0.159 | 0.205 | 0.439 | -0.174 | 0.278 | 0.532 |
| | process*sessions | 0.258 | 0.440 | 0.558 | -0.071 | 0.465 | 0.879 | 0.117 | 0.319 | 0.714 | 0.113 | 0.358 | 0.752 |
| BCSS Neg Self | sessions | 0.391 | 0.547 | 0.474 | 0.699 | 0.507 | 0.170 | 0.062 | 0.250 | 0.805 | 0.487 | 0.373 | 0.198 |
| | process*sessions | -0.516 | 0.622 | 0.407 | -0.939 | 0.632 | 0.141 | -0.228 | 0.385 | 0.557 | -0.768 | 0.512 | 0.143 |
| BCSS Pos Self | sessions | -0.368 | 0.700 | 0.599 | -0.361 | 0.571 | 0.528 | -0.117 | 0.277 | 0.673 | -0.438 | 0.428 | 0.307 |
| | process*sessions | 0.268 | 0.771 | 0.728 | 0.270 | 0.698 | 0.700 | -0.019 | 0.438 | 0.965 | 0.435 | 0.561 | 0.440 |
| BAPS negative | sessions | 0.275 | 0.616 | 0.659 | 0.550 | 0.498 | 0.275 | 0.164 | 0.254 | 0.520 | 0.411 | 0.345 | 0.235 |
| | process*sessions | -0.325 | 0.673 | 0.633 | -0.687 | 0.605 | 0.263 | -0.252 | 0.380 | 0.508 | -0.553 | 0.435 | 0.205 |
| BAPS normal | sessions | -0.354 | 0.528 | 0.502 | 0.048 | 0.424 | 0.911 | -0.060 | 0.223 | 0.788 | 0.147 | 0.347 | 0.673 |
| | process*sessions | 0.481 | 0.591 | 0.416 | 0.019 | 0.512 | 0.971 | 0.244 | 0.356 | 0.493 | -0.106 | 0.461 | 0.818 |
| PBEQ NAE | sessions | 0.065 | 0.623 | 0.919 | 0.037 | 0.485 | 0.940 | -0.173 | 0.248 | 0.497 | -0.055 | 0.461 | 0.908 |
| | process*sessions | -0.139 | 0.671 | 0.840 | -0.113 | 0.542 | 0.837 | 0.181 | 0.336 | 0.595 | -0.018 | 0.553 | 0.975 |
| PBEQ SAE | sessions | 0.096 | 0.325 | 0.774 | -0.367 | 0.235 | 0.134 | -0.010 | 0.174 | 0.957 | -0.058 | 0.233 | 0.809 |
| | process*sessions | -0.101 | 0.372 | 0.791 | 0.466 | 0.269 | 0.095 | 0.041 | 0.237 | 0.869 | 0.095 | 0.275 | 0.737 |

**Appendix 3: Full COMMAND trial results**

**3.1 Compliance modelled as a continuous variable**

**A 78: Instrumental variables analysis of the voice power mediator: compliance outcome modelled as a continuous measure, comparison of instrument used**

| Model | Effect | Coefficient | Bootstrap Std. Err. | Bootstrap Normal 95% CI | |
|---|---|---|---|---|---|
| All variable by group interactions as instruments | | | | | |
| OLS | Randomisation | -0.127 | 0.094 | -0.308 | 0.061 |
| | Power of voice | 0.053 | 0.035 | -0.017 | 0.122 |
| 2SLS | Randomisation | -0.090 | 0.100 | -0.264 | 0.127 |
| | Power of voice | 0.120 | 0.066 | 0.019 | 0.280 |
| VPD total by group interaction as instrument | | | | | |
| OLS | Randomisation | -0.135 | 0.082 | -0.299 | 0.025 |
| | Power of voice | 0.068 | 0.032 | 0.004 | 0.130 |
| 2SLS | Randomisation | -0.080 | 0.810 | -1.705 | 1.470 |
| | Power of voice | 0.174 | 1.085 | -2.014 | 2.240 |
| LASSO variables by group interactions as instruments | | | | | |
| OLS | Randomisation | -0.123 | 0.086 | -0.279 | 0.057 |
| | Power of voice | 0.064 | 0.033 | -0.001 | 0.129 |
| 2SLS | Randomisation | -0.095 | 0.104 | -0.271 | 0.135 |
| | Power of voice | 0.112 | 0.162 | -0.187 | 0.447 |
| Stepwise variables by group interactions as instruments | | | | | |
| OLS | Randomisation | -0.113 | 0.087 | -0.267 | 0.074 |
| | Power of voice | 0.054 | 0.034 | -0.014 | 0.118 |
| 2SLS | Randomisation | -0.107 | 0.092 | -0.276 | 0.084 |
| | Power of voice | 0.061 | 0.068 | -0.096 | 0.171 |

**A 79: Instrumental variables analysis of the voice power mediator: compliance outcome modelled as a continuous measure, all variable by group interactions as instruments, comparison of estimation methods, N=140**

| Model | Effect | Coefficient | Std. Err. | 95% CI | | p-value |
|---|---|---|---|---|---|---|
| OLS linear | Randomisation | -0.127 | 0.087 | -0.300 | 0.046 | 0.147 |
| | Power of voice | 0.053 | 0.036 | -0.017 | 0.124 | 0.138 |
| 2SLS linear | Randomisation | -0.102 | 0.084 | -0.267 | 0.063 | 0.224 |
| | Power of voice | 0.099 | 0.062 | -0.021 | 0.220 | 0.107 |
| GMM linear | Randomisation | -0.060 | 0.096 | -0.248 | 0.129 | 0.535 |
| | Power of voice | 0.178 | 0.096 | -0.010 | 0.367 | 0.064 |
| LIML linear | Randomisation | -0.138 | 0.070 | -0.276 | 0.000 | 0.050 |
| | Power of voice | 0.107 | 0.055 | -0.001 | 0.215 | 0.051 |
| LIML-F linear | Randomisation | -0.064 | 0.095 | -0.250 | 0.121 | 0.498 |
| | Power of voice | 0.170 | 0.093 | -0.012 | 0.352 | 0.067 |

**A 80: Instrumental variables analysis of the voice power mediator: compliance outcome modelled as a continuous measure, VPD total score by group interaction as instrument, comparison of estimation methods, N=150**

| Model | Effect | Coefficient | Std. Err. | 95% CI | | p-value |
|---|---|---|---|---|---|---|
| OLS linear | Randomisation | -0.132 | 0.076 | -0.282 | 0.018 | 0.085 |
| | Power of voice | 0.067 | 0.031 | 0.005 | 0.129 | 0.034 |
| 2SLS linear | Randomisation | -0.088 | 0.098 | -0.279 | 0.104 | 0.369 |
| | Power of voice | 0.171 | 0.148 | -0.118 | 0.461 | 0.246 |
| GMM linear | Randomisation | -0.088 | 0.091 | -0.267 | 0.092 | 0.337 |
| | Power of voice | 0.171 | 0.146 | -0.115 | 0.458 | 0.241 |
| LIML linear | Randomisation | -0.088 | 0.098 | -0.279 | 0.104 | 0.369 |
| | Power of voice | 0.171 | 0.148 | -0.118 | 0.461 | 0.246 |
| LIML-F linear | Randomisation | -0.094 | 0.094 | -0.279 | 0.091 | 0.321 |
| | Power of voice | 0.157 | 0.137 | -0.111 | 0.426 | 0.250 |

**A 81: Instrumental variables analysis of the voice power mediator: compliance outcome modelled as a continuous measure, LASSO selected variable by group interactions as instrument (cohabiting, gender and compliance), comparison of estimation methods, N=154**

| Model | Effect | Coefficient | Std. Err. | 95% CI | | p-value |
|---|---|---|---|---|---|---|
| B&K linear | Randomisation | -0.147 | 0.076 | -0.297 | 0.003 | 0.054 |
| | Power of voice | 0.076 | 0.031 | 0.015 | 0.138 | 0.014 |
| 2SLS linear | Randomisation | -0.152 | 0.093 | -0.334 | 0.029 | 0.100 |
| | Power of voice | 0.066 | 0.120 | -0.169 | 0.301 | 0.583 |
| GMM linear | Randomisation | -0.158 | 0.091 | -0.338 | 0.021 | 0.083 |
| | Power of voice | 0.068 | 0.117 | -0.162 | 0.298 | 0.562 |
| LIML linear | Randomisation | -0.153 | 0.094 | -0.338 | 0.032 | 0.105 |
| | Power of voice | 0.065 | 0.125 | -0.180 | 0.310 | 0.603 |
| LIML-F linear | Randomisation | -0.152 | 0.092 | -0.333 | 0.028 | 0.099 |
| | Power of voice | 0.066 | 0.119 | -0.166 | 0.299 | 0.576 |

**A 82: Instrumental variables analysis of the voice power mediator: compliance outcome modelled as a categorical measure, comparison of instruments used**

| Model | N | Effect | Coefficient | Standard error | 95% CI | | p-value | average marginal effect | 95% CI of marginal effect | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | All variable by group interactions as instruments | | | | | | | |
| B&K probit | 140 | Randomisation | -0.466 | 0.278 | -1.011 | 0.079 | 0.094 | -0.131 | -0.279 | 0.017 |
| | | Power of voice | 0.168 | 0.118 | -0.062 | 0.399 | 0.153 | 0.047 | -0.016 | 0.111 |
| IV probit | 140 | Randomisation | -0.088 | 0.334 | -0.743 | 0.567 | 0.793 | -0.023 | -0.197 | 0.150 |
| | | Power of voice | 0.625 | 0.220 | 0.194 | 1.056 | 0.005 | 0.165 | 0.059 | 0.271 |
| | | | VPD total score by group interaction as instrument | | | | | | | |
| B&K probit | 150 | Randomisation | -0.377 | 0.226 | -0.819 | 0.065 | 0.095 | -0.122 | -0.262 | 0.018 |
| | | Power of voice | 0.192 | 0.095 | 0.006 | 0.378 | 0.043 | 0.062 | 0.005 | 0.120 |
| IV probit | 150 | Randomisation | -0.228 | 0.324 | -0.863 | 0.408 | 0.483 | -0.068 | -0.270 | 0.133 |
| | | Power of voice | 0.459 | 0.365 | -0.257 | 1.174 | 0.209 | 0.138 | -0.046 | 0.322 |
| | | | LASSO variables by group interaction as instruments | | | | | | | |
| B&K probit | 154 | Randomisation | -0.422 | 0.225 | -0.864 | 0.020 | 0.061 | -0.136 | -0.274 | 0.002 |
| | | Power of voice | 0.229 | 0.094 | 0.045 | 0.413 | 0.015 | 0.074 | 0.018 | 0.130 |
| IV probit | 154 | Randomisation | -0.449 | 0.291 | -1.019 | 0.122 | 0.123 | -0.147 | -0.346 | 0.053 |
| | | Power of voice | 0.175 | 0.401 | -0.612 | 0.961 | 0.663 | 0.057 | -0.190 | 0.304 |