Demystifying causal inference in randomised trials

Ian White[1], Sabine Landau[2], Graham Dunn[3,4] and Richard Emsley[3,4]

1 MRC Biostatistics Unit Cambridge

2 Department of Biostatistics, Institute of Psychiatry, Psychology and Neuroscience, King's College London

3 Centre for Biostatistics, The University of Manchester

4 MRC North West Hub for Trials Methodology Research

Methodology report

- Dunn G, Emsley RA, Liu H, Landau S, Green J, White I and Pickles A (2015). Evaluation and validation of social and psychological markers in randomised trials of complex interventions in mental health. *Health Technology Assessment* **19** (93).

  http://www.journalslibrary.nihr.ac.uk/__data/assets/pdf_file/0010/156592/FullReport-hta19930.pdf

- Non-technical introduction and summary of our work on analysing complex interventions:
  - Introduction to Complex Interventions
  - Mediation analysis
  - Process evaluation
  - Longitudinal extensions
  - Stratified medicine
  - Guidance and tips for trialists

Research Programme: Efficacy and Mechanisms Evaluation

Support for Methodology Research Group from Mental Health Research Network.

Funded by grants from MRC Methodology Research Programmes:

- "*Design and methods of explanatory (causal) analysis for randomised trials of complex interventions in mental health*" (2006-2009)
  - Graham Dunn (PI), Richard Emsley, Ian White, *et al.*
- "*Designs and analysis for the evaluation and validation of social and psychological markers in randomised trials of complex interventions in mental health*" (2010-12)
  - Graham Dunn (PI), Richard Emsley, Ian White, *et. al*.
- "*Developing methods for understanding mechanisms in complex interventions*" (2013-16)
  - Sabine Landau (PI), Richard Emsley, Ian White, Graham Dunn *et al*.

And MRC Methodology Hubs

- *"MRC NorthWest Hub for Trials Methodology Research (2013-2018)"*
  - Paula Williamson (PI), Richard Emsley, *et al*.
- "*MRC Biostatistics Unit Hub for Trials Methodology Research (2013-2018)"*
  - Adrian Mander (PI), Ian White, *et al*.

Aims of the workshop

We want to "demystify causal inference" by:

- introducing participants to the concepts of causal inference in randomised trials
- and to accessible statistical methods used to answer various causal questions;
- providing worked examples from different clinical areas;
- pointing out modelling issues and the key assumptions required;
- demonstrating how these methods can be implemented in standard statistical software.

ISCB Birmingham 2016 4

Timetable

9:30am – 11:00am	**Session 1:** Introduction to causal inference and non-compliance using Instrumental Variables (IV) (SL)
11:00am – 11:30am	Coffee break
11:30am – 1:00pm	**Session 2:** Non-compliance using Inverse Probability Weighting (IW)
1:00pm – 2:00pm	Lunch break
2:00pm – 3:30pm	**Session 3:** Introduction to mediation and mediation analysis using IV (GD)
3:30pm – 4:00pm	Coffee break
4:00pm – 5:30pm	**Session 4:** Mediation using IPW and other methods (RE)

ISCB Birmingham 2016 5

KING'S College LONDON

Lecture 1: Introduction to causal inference and non-compliance using Instrumental Variables

ISCB 2016

Sabine Landau

Department of Biostatistics, Institute of Psychiatry, Psychology and Neuroscience, King's College London

ISCB Birmingham 2016 6

Outline

1. **Causal estimands in clinical trials**
2. **ITT analysis for perfect trials**
3. **Challenges in imperfect trials**
4. **Commonly used efficacy estimators**
5. **Instrumental variables methods for complier average causal effect estimation**
6. **Assumptions trade-off**
7. **Practical: Analysis of ODIN trial**

ISCB Birmingham 2016 7

Learning objectives

By the end of this session you should be able to:

- understand the meaning of potential outcomes and the definition of individual and average causal effects;
- have become familiar with causal estimands for quantifying effectiveness and efficacy of treatments;
- understand instrumental variables (IV) approaches for estimating efficacy in trials;
- understand the assumptions made;
- understand the steps involved in IV estimation.

1. **Causal estimands in clinical trials**
2. ITT analysis for perfect trials
3. Challenges in imperfect trials
4. Commonly used efficacy estimators
5. Instrumental variables methods for complier average causal effect estimation
6. Assumptions trade-off
7. Practical: Analysis of ODIN trial

Clinical trials

- Randomised controlled trials (RCTs) are the gold standard for evaluating the benefits of interventions/treatments.
- Clinicians might be interested in:
  - The benefit of actually **receiving treatment** (effect of active incredient) =: **Efficacy**
  - The benefit of a **treatment policy** (effect of eligibility to receive treatment/effect of offer of treatment) =: **Effectiveness**
- Other possible interests include:
  - Effectiveness of a treatment policy outside of a trial setting.
  - Efficacy of a treatment only on those who start treatment.

Typical efficacy research questions

- Depressed patients are randomly allocated to receive either treatment as usual (TAU) from their GP or to attend a cognitive behavioural therapy (CBT) programme in addition to TAU.
- No one who is allocated to TAU gets access to CBT, but about half of those allocated to receive CBT do not take up the offer (therefore receiving only TAU).
- What is the effect of receiving CBT on recovery?
- Does the effect of therapy increase with the number of sessions attended?

Rubin's causal model

- We will now introduce Rubin's causal model (Rubin, 1974) for defining individual and population-average treatment effects.
- These population summaries do not always feature as parameters in statistical models and hence are more generically referred to as **estimands**.
- Whenever treatment effect heterogeneity is allowed we will assume that a subject's response to a treatment will be the same regardless of what treatments other subjects receive.
  - This is known as the **stable unit treatment value assumption (SUTVA)**

Formal definition of a causal effect

- The potential outcomes approach:
  - It is a comparison between **what is** and what might have been.
  - It is **counterfactual**.
  - We wish to estimate the difference between a patient's **observed outcome** and the outcome that **would have been observed** if, contrary to fact, the patient's treatment or care had been different (Neyman, 1923; Rubin, 1974).
- Without the possibility of comparison an individual treatment effect is not defined.
- (Note we cannot define an individual causal effect of non-modifiable variables such as gender.)

ISCB Birmingham 2016 13

Notation

Participant/subject	i (i = 1 to N)
Measured baseline covariates	X_{ij} ($X_{i1}, X_{i2}, \ldots, X_{ip}$)
Random treatment offer	Z_i
Treatment received	D_i
Measured post-randomisation confounder	L_i
Unmeasured confounders	U_i
Intermediate outcome (mediator)	M_i (needed in afternoon lectures)
Continuous clinical outcome	Y_i
Time to event outcome	T_i

ISCB Birmingham 2016 14

Potential outcomes

- **Objective**: We want to define the causal effect of a binary variable, e.g. of exposure *D.*
- Prior to treatment allocation, there are two **potential outcomes** for each subject:
  - $Y_i(D = 1)$ or $Y_i(1)$, for short,
  - and $Y_i(D = 0)$ or $Y_i(0)$
- As a result of the allocation, however, **it is only ever possible to observe one of them** (the other is a counterfactual).
- The observed outcome is given by:

$$Y_i = D_i Y_i(1) + (1 - D_i) Y_i(0)$$

ISCB Birmingham 2016 15

Example: Potential outcomes

ID	*Y*(0)	*Y*(1)	*D*	*Y*
1	2	7	1	7
2	1	8	0	1
3	3	6	1	6
4	5	6	0	5
5	4	7	0	4
6	3	9	1	9
7	2	10	1	10

We can only observe the blue cells.

The pink *Y*s are counterfactual.

We can only ever observe one of *Y*(0) and *Y*(1) but never both.

ISCB Birmingham 2016

Individual treatment effects

- An **individual causal treatment effect (ITE)** for the *i*th subject with potential outcomes $Y_i(1)$ and $Y_i(0)$ can be defined by the contrast

$$\text{ITE} := Y_i(1) - Y_i(0)$$

- Note that treatment effects are allowed to be **heterogeneous**.
  - ITE is not constrained to be the same for each subject.
  - In principle it can vary from one subject to another.
  - And we might be able to investigate what characteristics of the subject influence the size of the treatment effect; that is look at effect moderation. (Not the subject of today's workshop – but note the use of such "predictive markers/moderators" as instruments in lecture 3.)

ISCB Birmingham 2016

Example: ITEs

ID	*Y*(0)	*Y*(1)	*D*	ITE
1	2	7	1	5
2	1	8	0	7
3	3	6	1	3
4	5	6	0	1
5	4	7	0	3
6	3	9	1	6
7	2	10	1	8

We can only observe the blue cells.

The pink *Y*s are counterfactual.

We cannot observe the ITEs.

The ITEs vary between individuals.

ISCB Birmingham 2016

Fundamental problem

- It is impossible to observe the value of $Y_i(1)$ and $Y_i(0)$ on the same subject and, therefore, it is impossible to observe ITE (the effect of treatment on subject *i*).
- This is the **fundamental problem of causal inference**.
- We can observe the marginal distributions under some designs but we can never observe the joint distribution.

ISCB Birmingham 2016

Average treatment effects

- We can now define an average treatment effect in the target populations or in the subpopulations of those who ultimately will (or will not) receive treatment:
  - **Average Treatment Effect (ATE):**

$$\text{ATE} := \text{E(ITE)} = \text{E}[Y(1)-Y(0)]$$

  - **Average Treatment effect on the Treated (ATT):**

$$\text{ATT} := \text{E(ITE} \mid D=1) = \text{E}[\, Y(1)-Y(0) \mid D=1]$$

  - **Average Treatment effect on the Untreated (ATU):**

$$\text{ATU} := \text{E(ITE} \mid D=0) = \text{E}[\, Y(1)-Y(0) \mid D=0]$$

- Clinicians are often interested in ATT.

Causal effect of treatment offer

- ATE, ATT, ATU quantify the causal effect of the **receipt of treatment (efficacy)**.
- In some research contexts the offer of the treatment – the **treatment policy** – is of interest.
- To quantify the causal effect of treatment offer (**effectiveness**) we can utilise potential outcomes in treatment offer Z.
- We define the **causal effect of treatment offer** by:

$$\text{ACE} := \text{E}[\, Y(Z=1) - Y(Z=0) \,]$$

(Note the index here is Z not D!)

ISCB Birmingham 2016 21

1. Causal estimands in clinical trials
2. **ITT analysis for perfect trials**
3. Challenges in imperfect trials
4. Commonly used efficacy estimators
5. Instrumental variables methods for complier average causal effect estimation
6. Assumptions trade-off
7. Practical: Analysis of ODIN trial

ISCB Birmingham 2016 22

Intention-to-treat (ITT) analysis in trials

- From Consolidated Standards of Reporting Trials (CONSORT) initiative, see http://www.consort-statement.org/resources/glossary:
  - *"A strategy for analyzing data in which all participants are included in the group to which they were assigned, whether or not they completed the intervention given to the group."*
  - *"Intention-to-treat analysis prevents bias caused by the loss of participants, which may disrupt the baseline equivalence established by random assignment and which may reflect non-adherence to the protocol."*
- ITT analyses compares (all) subjects as randomised, regardless of what intervention they actually received.

What does ITT analysis estimate?

- ITT analysis compares the outcome between the trial arms. Therefore it targets the **effect of the treatment offer (effectiveness)**.
- Under **full compliance** with treatment offer

  treatment offer = treatment receipt,

  and thus an ITT analyses **also targets efficacy**.
- Under **non-compliance**

  treatment offer ≠ treatment receipt,

  and thus an ITT analyses **no longer targets efficacy**.

Why does this work in a perfect trial?

- We can never observe the ITEs – so how can we estimate causal estimands?
- A causal estimand **is identified** if it can be expressed as a function of the distribution of the observed variables, e.g. first and second moments.
- The statistical answer to this problem is the estimation of such a function of first and second moments of the marginal distributions.
- In theory, once an estimand is identified, it can be estimated by respective observed moments in a random sample.
  - This gives an unbiased estimator for large samples.
  - More complicated if do not have a random sample.
  - (But without identifiability we can't go anywhere.)

Identifiability of ATE under random allocation and full compliance

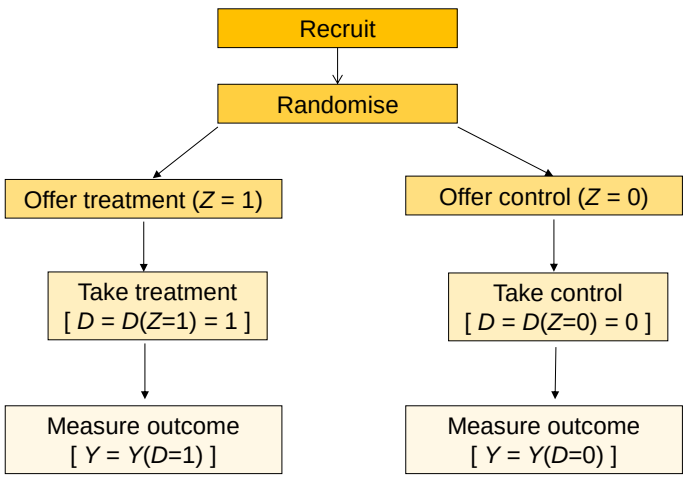
- Consider the average causal treatment effect (ATE).
- Under which designs can this be estimated from marginal distributions?
- For an RCT with random treatment offers and perfect compliance with allocated treatment (Z=D) we can write:

$$\begin{aligned} \text{ACE} = \text{ATE} &= \text{E}[Y(1)] - \text{E}[Y(0)] \\ &= \text{E}[Y(1) \mid Z=1] - \text{E}[Y(0) \mid Z=0] \\ &= \text{E}[Y(1) \mid D=1] - \text{E}[Y(0) \mid D=0] \\ &= \text{E}[Y \mid D=1] - \text{E}[Y \mid D=0] \\ (\ &= \ \text{E}[Y \mid Z=1] - \text{E}[Y \mid Z=0] =: \text{ITT}\ ) \end{aligned}$$

- Thus in a perfect RCT efficacy or effectiveness can be estimated by estimating the difference between the mean outcome for those receiving treatment and the mean outcome for those in the control condition (**ITT effect**).

1. Causal estimands in clinical trials
2. ITT analysis for perfect trials
3. **Challenges in imperfect trials**
4. Commonly used efficacy estimators
5. Instrumental variables methods for complier average causal effect estimation
6. Assumptions trade-off
7. Practical: Analysis of ODIN trial

A 'perfect' randomised controlled trial

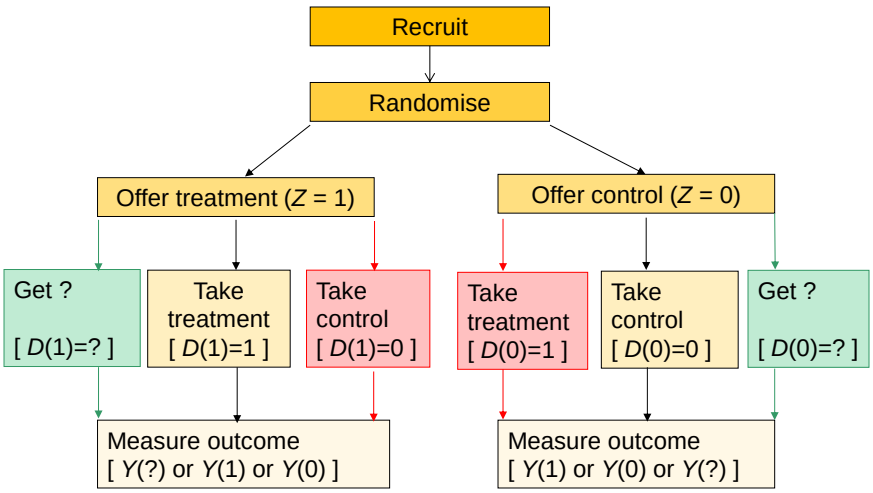


An 'imperfect' RCT

"…*There could not be worse experimental animals on earth than human beings; they complain, they go on vacations, they take things they are not supposed to take, they lead incredibly complicated lives, and, sometimes, they do not take their Medicine*…."

Efron B. Foreword. *Statistics in Medicine* 1998; 17: 249–50.

A more realistic RCT



ISCB Birmingham 2016 30

RCT example: ODIN

- Here we use data from the ODIN trial - an evaluation of the effect of psychotherapy on the severity of depression.
  - Depression measured pre-randomisation and six months after randomisation.
- RCT with two parallel arms.
  - 427 patients were randomised (236 treatment : 191 controls).
- There was non-adherence (non-compliance):
  - Only 128 of those offered treatment actually took up the offer.
  - None of the controls had access to treatment.
- There were lots of missing outcome data.
  - For the moment we assume a very simple missing at random data generating process – MAR, that is, probability of missing outcome dependent only on randomised group – and carry out complete case (CC) analysis.

ISCB Birmingham 2016 31

ODIN trial data (Dunn *et al.*, 2003)

		# at baseline	# (%) at 6 months	BDI mean at 6 months
Treatment Arm	Non-compliers	108	59 (55%)	13.22
	Compliers	128	118 (92%)	13.32
Control Arm		191	140 (73%)	15.16
	Total	427	317	

There was a trial arm difference (estimated ITT effect = -1.87) suggesting that the therapy was effective.

But ODIN was not a perfect trial ($Z \neq D$) ; so we can't use this as an efficacy estimate.

32

1. Causal estimands in clinical trials
2. ITT analysis for perfect trials
3. Challenges in imperfect trials
4. **Commonly used efficacy estimators**
5. Instrumental variables methods for complier average causal effect estimation
6. Assumptions trade-off
7. Practical: Analysis of ODIN trial

Commonly used efficacy estimators

- Clinicians have long tried to address this.
- Three commonly used strategies to evaluating treatment efficacy:
  - **ACCEPT - ITT analysis:** Comparison of all those allocated to the treatment group with all those allocated to the control group.
  - **EXCLUDE - Per-protocol analysis:** Comparison of those who complied with random allocation in the treatment group with all of the controls.
  - **MODEL - As-treated analysis:** Comparison of those who received the treatment in question with those who did not receive it, regardless of random allocation.

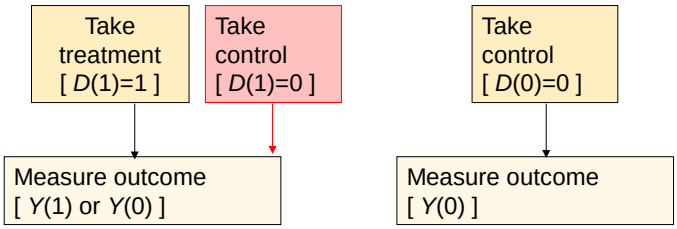
Special case: no treatment access

- Consider the special case where:

  I. those randomised to control have no access to treatments other than those forming part of the control condition
     - e.g. novel therapy

  II. and those randomised to receive the experimental treatment receive the control condition if they do not comply

- Then the RCT simplifies to:



ISCB Birmingham 2016 35

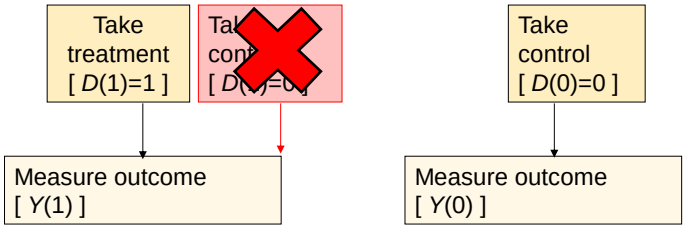
ITT approach

- Clinicians sometimes suggest simply **accepting** that the ITT estimate will be biased for the purpose of efficacy estimation when there is non-adherence with randomly allocated treatment.
- It has been suggested that is an "conservative approach" since ITT = ACE < ATE.
  - That is the ITT estimator underestimates ATE.
- However, consider that such an approach would not be "conservative" in the context of a non-inferiority trial.
  - We might be suggesting non-inferiority of a new (say brief and cheaper) treatment compared to a gold standard simply because the gold standard is not taken up.

ISCB Birmingham 2016 36

Per-protocol approach

- Per-protocol analysis = comparison of outcome in those who take the treatment versus outcome in all those randomised to control.
- An attempt to estimate the average effect of treatment on the treated (ATT) by **excluding** those who do not start their active treatment (observed non-compliers) from the analysis.



ISCB Birmingham 2016 37

Per-protocol assumptions

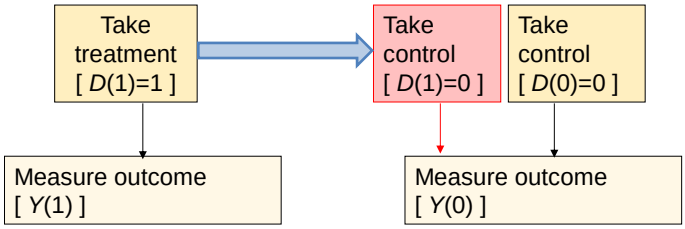
- For simplicity assuming no treatment access in the control arm.
- Per-protocol analysis is estimating

$$\begin{aligned} &E[\,Y(1) \mid Z=1 \,\&\, D(Z=1)=1] - E[\,Y(0) \mid Z=0] \\ &= E[\,Y(1) \mid Z=1 \,\&\, D=1] - E[\,Y(0)\,] \\ &= E[\,Y(1) \mid D=1] - E[\,Y(0)\,] \\ &\neq \text{ATT} \end{aligned}$$

  - Only estimating ATT if $E[\,Y(0) \mid D=1] = E[\,Y(0) \mid D=0]$
  - I.e. only if treatment-free prognosis is independent of treatment receipt
  - (Otherwise there is confounding of the effect of D on Y.)

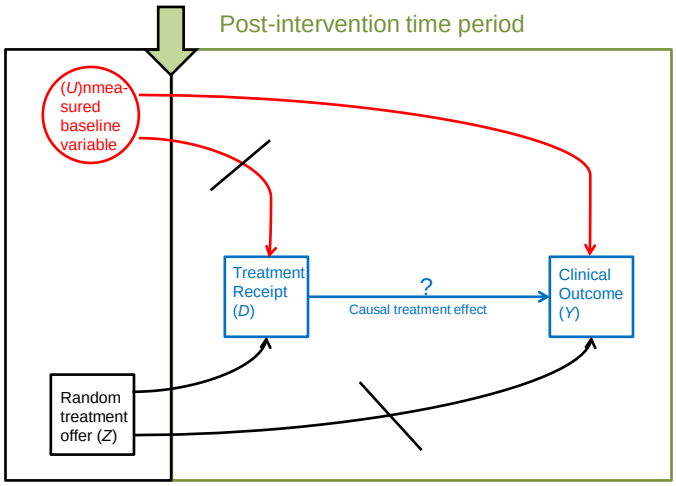
As-treated approach

- As-treated analysis = comparison of outcome in those who take the treatment versus outcome in those who take control, regardless of randomisation.
- An attempt to estimate ATE by **modelling** the effect of treatment receipt.
- Non-compliers in the active treatment "moved" into the control group.



ISCB Birmingham 2016 39

Confounding of the as-treated estimator



CC analysis for ODIN trial

Estimators	Missing Data	Difference in BDI6 (std error)
ITT – unadjusted	CC	-1.87 (1.14)
As-treated – unadjusted	CC	-1.26 (1.18)
Per-protocol – unadjusted	CC	-1.84 (1.29)

Interpretation of ODIN results

- Efficacy would appear to be less than effectiveness.
  - This is strange.
  - Considering that not taking a treatment should make the randomisation group outcomes more similar.
- How much less depends on whether we use as-treated or per-protocol approaches.
- **Something is wrong!**

1. Causal estimands in clinical trials
2. ITT analysis for perfect trials
3. Challenges in imperfect trials
4. Commonly used efficacy estimators
5. **Instrumental variables methods for complier average causal effect estimation**
6. Assumptions trade-off
7. Practical: Analysis of ODIN trial

What might be a better approach?

- For the moment we are assuming:
  - **"all-or-nothing" compliance**, which means that everyone gets either the treatment or the control (i.e. *D* binary), and
  - **no contamination**, which means that participants do not have access to treatment other than being offered it.
- In the intervention arm, we observe:
  - $Y(D=1)$ in Compliers
  - $Y(D=0)$ in Non-compliers
- In the control arm, we observe $Y(0)$ in everyone, but really we have two subgroups again:
  - $Y(D=0)$ in **potential** Compliers
  - $Y(D=0)$ in **potential** Non-compliers

Complier Average Causal Effect (CACE)

- Thus when comparing those who comply when offered treatment with controls we are comparing outcomes between different populations.
- We may consider the average treatment effect in the (sub)population of compliers with either treatment offer.
- Remember for such **compliers**:
  - $D(Z=1) = 1$ and $D(Z=0) = 0$
  - and therefore $D(1)-D(0) = 1$
- This estimand is known as the **Complier Average Causal Effect (CACE)**.
- Formally:

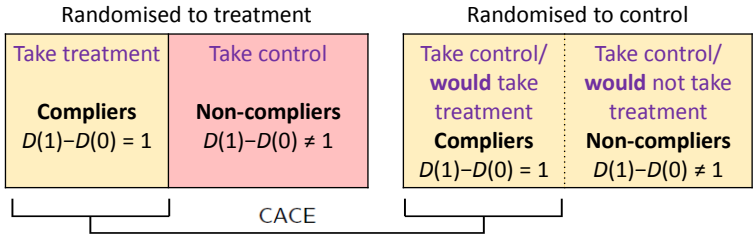
$$\text{CACE} := \text{E}[\ \text{ITE} \mid D(1)-D(0)=1\ ] = \text{E}[\ Y(1)-Y(0) \mid D(1)-D(0)=1\ ]$$

CACE estimation

- Need to be clear for which population we wish to evaluate the effect of receiving the treatment.
- The Complier-Average Causal Effect (CACE) estimate is the comparison of the average outcome of the compliers in the treatment arm with the average outcome of the comparable group of would-be compliers in the control arm.
- A CACE estimator provides a **randomisation-respecting estimate** (White, 2005).
  - We are estimating the ITT effect in the sub-group of participants who would always comply with their treatment allocation.

CACE estimation cont'd

- CACE compares a subgroup of those randomised to treatment with the same subgroup of those randomised to control:



47

CACE estimation cont'd

- How can we estimate CACE?
- First, we make some explicit assumptions:

1. **There are two latent classes of participants:**
   - Compliers get treatment if and only if randomly allocated to the treatment.
   - Non-compliers never get the treatment, regardless of allocation.
   - We can identify these two groups in the treatment arm, but they remain hidden (unobserved) in the control arm.
2. As a consequence of randomisation, on average, **the proportion of compliers is the same in the two arms of the trial**.
3. We need to assume that **for the non-compliers the offer of treatment in itself does not influence outcome.** This assumption is often called an **exclusion restriction**.

Problems with CACE

- We need to be able to define (binary) compliance.
  - Can be problematic, e.g. participants undergoing therapy might attend a certain number of sessions. How many are enough to be called a "complier"?
- In practice, we may want to know what will be observed…
  - …if compliance is worse than in the trial (e.g. if rolled out in clinical practice)
  - …if compliance is better than in the trial (e.g. because intervention is well publicised/ marketed)
  - This means we may want to know the average causal effect in a different subgroup.
- Compliance with a treatment offer could be setting-specific and thus different in the target population.
  - We might assume this is the CACE – but it is an assumption.

ODIN CACE analysis (CC)

# participants mean BDI	Compliers	Non-compliers	All
Treatment	118 13.32	59 13.22	177 13.29
Control	? ?	? ?	140 15.16

ODIN CACE analysis (CC) cont'd

# participants mean BDI	Compliers	Non-compliers	All
Treatment	118 13.32	59 13.22	177 13.29
Control	93.3=140x(118/177) ?	46.7=140x(59/177) ?	140 15.16

Use randomisation to fill in expected subgroup sizes in the control group (66.7% compliance).

ODIN CACE analysis (CC)

# participants mean BDI	Compliers	Non-compliers	All
Treatment	118 13.32	59 13.22	177 13.29
Control	93.3 ?	46.7 13.22	140 15.16

Use exclusion restriction to fill in mean outcome of non-compliers in controls.

ODIN CACE analysis (CC) cont'd

# participants mean BDI	Compliers	Non-compliers	All
Treatment	118 13.32	59 13.22	177 13.29
Control	93.3 16.13	46.7 13.22	140 15.16

Work out remaining entry from this information.
CACE = 13.32-16.13 = -2.81 and |CACE| > |-1.87| = |ITT|

ODIN per-protocol analysis (CC)

# participants mean BDI	Compliers	Non-compliers	All
Treatment	118 13.32	59	177 13.29
Control	93.3 15.16	46.7 15.16	140 15.16

Within controls per-protocol works out the outcome for compliers by assuming that it is the same as for non-compliers ("random non-compliance").

CC analysis for ODIN trial

Estimators	Missing Data	Difference in BDI6 (std error)
ITT – unadjusted	CC	-1.87 (1.14)
As-treated – unadjusted	CC	-1.26 (1.18)
Per-protocol – unadjusted	CC	-1.84 (1.29)
CACE – unadjusted	CC	-2.81 (1.72)

CACE under more general scenario

- So far we have concentrated on the no contamination scenario.
- More generally, a population splits into four principal strata:
  - **Compliers** (take what is offered)
  - **Never-takers** (take the control treatment under either offer)
  - **Always-takers** (take the active treatment under either offer)
  - **Defiers** (take the opposite treatment to what is offered)
- In the no contamination scenario, there were no Always-takers nor Defiers and therefore all the Non-compliers were Never-takers.
- We can relax this to allow for the presence of Always-takers.
  - However, we still have to assume the absence of Defiers. This is known as the **monotonicity assumption**.

ISCB Birmingham 2016 56

CACE under more general scenario cont'd

- The **(extended) exclusion restriction** then stipulates that the offer of treatment alone does not have an effect on outcome in the Never-takers nor the Always-takers.
- CACE can still be estimated under this exclusion restriction although the procedure now involves more steps than previously demonstrated.

CACE estimation "by-hand"

- We have demonstrated the operations that need to be performed when we have only Compliers and Never-takers to estimate CACE.
- To get an even quicker answer consider the following relationships between the overall ITT effect and the three principal strata:

$$
\begin{aligned}
ITT = \quad & p_{Complier} && * ITT_{Compliers} \\
+ \quad & p_{Never\text{-}taker} && * ITT_{Never\text{-}taker} \\
+ \quad & p_{Always\text{-}taker} && * ITT_{Always\text{-}taker} \\
= \quad & p_{Complier} && * CACE
\end{aligned}
$$

(from the exclusion restriction)

Then:

estimate of CACE (CC) = estimate of ITT / estimate of $p_{Complier}$

= -1.87 / 0.667 = -2.80 

(SE can be generated by bootstrapping.)

Instrumental variables approach

- Instrumental variables theory for **structural linear models** (continuous *Y*):
  - An explanatory variable is **endogenous** if is correlated with the model's error term.
  - A least squares estimator will be biased for the causal effect of endogenous variable. (It is confounded.)
- An instrumental variable (IV) is a variable not included in the model for *Y* that:
  - (IV1) **Relevance condition**: Predicts the endogenous variable.
  - (IV2) Is not correlated with the model's error term.
- Theoretical result: An estimator that employs the instrumental variable (**IV estimator) is consistent** for the causal effect of the endogenous variable.

Application to CACE estimation

- In trials:
  - **Causal effect of interest:** Causal effect of treatment receipt (*D*) on outcome (*Y*).
  - **Endogenous variable:** Treatment receipt (*D*).
    - This is a post-randomisation variable.
    - So can be affected by unobserved baseline variables (*U*) that determine both treatment receipt (*D*) and outcome (*Y*).
    - Latent *U* forms part of the error term of the linear model.

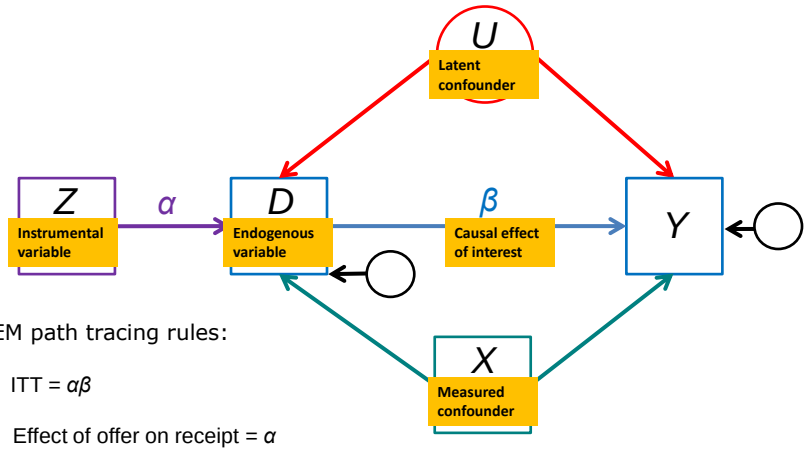
Application to CACE estimation cont'd

- In trials:
  - **Instrumental variable:** Random treatment offer (*Z*).
    - Should meet (IV1) as treatment offer should predict treatment receipt.
    - Meeting (IV2) implies that:-
      - (a) there is no direct effect of Z on *Y* (other than via *D*) – the **exclusion restriction;**
        - » This remains an assumption.
      - (b) that there are no common causes of Z and *Y*.
        - » This holds in a trial due to randomisation.
  - So IV estimator using *Z* is consistent for causal effect of *D*.

ISCB Birmingham 2016 61

SEM fitted to construct IV estimator

Structural equation model (SEM) diagram:



Linear SEM path tracing rules:

- $Z \rightarrow Y$: ITT = $\alpha\beta$
- $Z \rightarrow D$: Effect of offer on receipt = α
- Estimator of β = ITT estimator / estimator of α

62

Treatment effect heterogeneity

- Which causal effect of *D* does an IV estimator estimate?
- IV theory as applied in econometrics assumes effect homogeneity.
  - Were we to believe this then we would be estimating ATE = CACE.
- We wish to allow treatment effects to vary between individuals.
- Important result: Under treatment heterogeneity an IV estimator estimates CACE (Imbens and Angrist, 1994) assuming that:
  - Randomisation is an instrument for treatment receipt (**exclusion restriction**).
  - There are no defiers in the target population (**monotonicity**).

More on CACE estimation

- CACE estimation using IVs is equivalent to CACE estimation by fitting **structural mean models** (see e.g. Fischer-Lapp and Goetghebeur, 1999; Dunn and Bentall, 2007)
  - See also later lecture 3 for a demonstration in the context of mediation.
- The IV approach can be extended to:-
  - deal with continuous adherence measures/dose-response relationships (Dunn and Bentall, 2007).
  - to cover non-continuous outcomes,
    - such as binary outcomes (e.g. Clarke, Palmer and Windmeijer, 2015)
    - or censored survival times (e.g. Tchetgen Tchetgen *et al*., 2015).

More on CACE cont'd

- A **principal stratification** approach (Frangakis and Rubin, 2002) can be used to estimate CACE.
  - Needs specialist SEM software for fitting.
  - Can be extended to allow for missing data generating mechanisms that are **latent ignorable** (Pickles and Croudace, 2009).
- All CACE estimators rely on some form of exclusion restriction assumption.

1. Causal estimands in clinical trials
2. ITT analysis for perfect trials
3. Challenges in imperfect trials
4. Commonly used efficacy estimators
5. Instrumental variables methods for complier average causal effect estimation
6. **Assumptions trade-off**
7. Practical: Analysis of ODIN trial

CACE estimators and assumptions

- The estimators are valid if and only if the assumptions are true.
- Which assumptions seem to be more realistic?
  - The **commonly used estimators rely on the assumption of no residual confounding** (no prognostic variables driving treatment receipt).
  - The **vital assumption for the CACE estimator is the exclusion restriction** (no effect of randomisation on outcome except through treatment received).

CACE estimators and assumptions cont'd

- The exclusion restriction is likely to hold in a double-blind placebo-controlled drug trial, but what about the case of an **unblinded trial** of psychotherapy?
  - Resentful demoralisation.
- The exclusion restriction might be unrealistic when a **cut-off** is applied to define adherence
  - Is there really no outcome difference between those receiving a few sessions and those receiving none?
- In that case better to use continuous adherence measure, see slide 64.

1. Causal estimands in clinical trials
2. ITT analysis for perfect trials
3. Challenges in imperfect trials
4. Commonly used efficacy estimators
5. Instrumental variables methods for complier average causal effect estimation
6. Assumptions trade-off
7. **Practical: Analysis of ODIN trial**

ISCB Birmingham 2016 69

ODIN trial variable names

`rgroup`: random allocation (1=treatment, 0=control)

`treat`: treatment received (1=treated, 0=not treated; "0" for control arm)

`bdi6`: six month score on the Beck Depression Inventory (BDI)

`bdi0`: pre-randomisation (baseline) BDI score

`centre`: code for each of 8 centres (randomisation stratifier)

`sessions`: number of therapy session received (missing in control arm)

`resp`: response at six months ("1"=BDI observed at six months, "0"=BDI missing)

(Using Stata version 14 here.)

Trial data set (CC only)

	subjectn	centre	bdi0	rgroup	typeint	sessions	complian	resp6	bdi6	miss	wm	treat	subjectid
1	23	7	17	1	indivi...	3	attended	0	.	1	1.084746	1	70023
2	26	1	19	1	group	5	discon...	1	0	0	1.830508	0	10026
3	28	7	18	1	indivi...	6	attended	0	.	1	1.084746	1	70028
4	42	1	23	0	.	.	.	0	.	1	1.364286	0	10042
5	45	3	33	0	.	.	.	0	.	1	1.364286	0	30045
6	50	1	19	1	group	0	dna	0	.	1	1.830508	0	10050
7	60	7	39	1	indivi...	2	attended	1	40	0	1.084746	1	70060
8	65	1	18	0	.	.	.	1	18	0	1.364286	0	10065
9	67	1	17	0	.	.	.	1	5	0	1.364286	0	10067
10	68	2	25	1	indivi...	6	attended	1	7	0	1.084746	1	20068
11	72	2	13	1	indivi...	6	attended	1	11	0	1.084746	1	20072
12	73	7	20	1	indivi...	4	attended	1	8	0	1.084746	1	70073
13	90	3	13	1	indivi...	6	attended	1	6	0	1.084746	1	30090
14	92	2	24	1	indivi...	.	refused	0	.	1	1.830508	0	20092
15	93	1	15	0	.	.	.	1	9	0	1.364286	0	10093
16	101	3	20	1	indivi...	6	attended	1	0	0	1.084746	1	30101
17	106	1	15	1	group	7	attended	1	5	0	1.084746	1	10106
18	110	3	22	1	indivi...	6	attended	1	18	0	1.084746	1	30110
19	119	1	18	0	.	.	.	0	.	1	1.364286	0	10119
20	121	1	14	1	group	8	attended	1	11	0	1.084746	1	10121
21	122	7	14	0	.	.	.	1	3	0	1.364286	0	70122
22	123	1	21	1	group	8	attended	1	10	0	1.084746	1	10123
23	126	1	27	0	.	.	.	1	11	0	1.364286	0	10126

Some descriptives

```
tab2 rgroup resp6, row

-> tabulation of rgroup by resp6

+----------------+
| Key            |
|----------------|
|   frequency    |
| row percentage |
+----------------+

   offered |   non-missing bdi6
 treatment |         0          1 |     Total
-----------+----------------------+----------
         0 |        51        140 |       191
           |     26.70      73.30 |    100.00
-----------+----------------------+----------
         1 |        59        177 |       236
           |     25.00      75.00 |    100.00
-----------+----------------------+----------
     Total |       110        317 |       427
           |     25.76      74.24 |    100.00
```

"191 participants were randomised to the control condition.
236 participants were randomised to treatment.
The response rate in each trial arm was approx. 75%."

Baseline balance

```
univar bdi0, by(rgroup)

-> rgroup=0
                                                  ------------- Quantiles -------------
Variable       n      Mean     S.D.       Min       .25       Mdn       .75       Max
-------------------------------------------------------------------------------------
    bdi0     191     22.58     8.06     13.00     16.00     21.00     27.00     51.00
-------------------------------------------------------------------------------------

-> rgroup=1
                                                  ------------- Quantiles -------------
Variable       n      Mean     S.D.       Min       .25       Mdn       .75       Max
-------------------------------------------------------------------------------------
    bdi0     236     22.79     8.32      6.00     16.50     21.00     27.00     49.00
-------------------------------------------------------------------------------------
```

"Baseline BDI was similar between the two trial arms."

(As was the centre distribution due to randomisation, not shown.)

Compliance with allocated treatment

```
tab2 treat complian if rgroup==1, row

-> tabulation of treat by complian if rgroup==1

+-----------------+
| Key             |
|-----------------|
|   frequency     |
| row percentage  |
+-----------------+

           |            treatment compliance
     treat |  attended    refused  discontin        dna |      Total
-----------+--------------------------------------------+----------
         0 |         0         52         36         20 |       108
           |      0.00      48.15      33.33      18.52 |    100.00
-----------+--------------------------------------------+----------
         1 |       128          0          0          0 |       128
           |    100.00       0.00       0.00       0.00 |    100.00
-----------+--------------------------------------------+----------
     Total |       128         52         36         20 |       236
           |     54.24      22.03      15.25       8.47 |    100.00
```

"Of the 236 participants randomised to therapy, 108 did not complete the treatment that was offered to them while 128 did."

Producing commonly used estimators for ODIN

- Need to specify appropriate explanatory variables:
- **ITT estimator**
  - Unadjusted: `regress bdi6 rgroup`
  - Adjusted for baseline covariates:
    `regress bdi6 rgroup bdi0 i.centre`
- **As-treated estimator**
  - Unadjusted: `regress bdi6 treat`
  - Adjusted: `regress bdi6 treat bdi0 i.centre`
- **Per-protocol estimator**
  - Unadjusted:
```
preserve
drop if rgroup==1 & treat==0
regress bdi6 treat
restore
```
  - Adjusted:
```
preserve
drop if rgroup==1 & treat==0
regress bdi6 treat bdi0 i.centre
restore
```

ISCB Birmingham 2016 75

ITT estimates in Stata

- ITT estimate under simple MAR assumption

```
regress bdi6 rgroup
------------------------------------------------------------------------------
        bdi6 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      rgroup |  -1.869007   1.143556    -1.63   0.103    -4.11898     .3809657
       _cons |   15.15714   .8545045    17.74   0.000     13.47589     16.8384
------------------------------------------------------------------------------
```

- ITT estimate with conditioning on centre and baseline values of outcome to gain extra precision

```
regress bdi6 rgroup i.centre bdi0
------------------------------------------------------------------------------
        bdi6 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      rgroup |  -2.341451   .9620675    -2.43   0.016    -4.234532   -.4483707
    centre 2 |   1.927232   2.602716     0.74   0.460    -3.194189    7.048652
    centre 3 |   3.667262   2.342902     1.57   0.119    -.9429163    8.277441
  …
        bdi0 |   .5302925   .0608799     8.71   0.000     .4104978    .6500873
       _cons |  -2.095002   2.336221    -0.90   0.371    -6.692033    2.502029
------------------------------------------------------------------------------
```

76

As-treated estimates in Stata

- As-treated estimate under simple MAR assumption

```
regress bdi6 treat
```

```
------------------------------------------------------------------------------
        bdi6 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       treat |  -1.260881   1.177566    -1.07   0.285    -3.577769    1.056008
       _cons |   14.58291     .71845    20.30   0.000     13.16935    15.99648
------------------------------------------------------------------------------
```

- As-treated estimate adjusting for centre and baseline BDI

```
regress bdi6 treat bdi0 i.centre
```

```
------------------------------------------------------------------------------
        bdi6 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       treat |  -1.111169   1.002084    -1.11   0.268    -3.082991    .8606528
        bdi0 |   .5267403   .0615095     8.56   0.000     .4057067    .6477739
             |
      centre |
          2  |   1.544469   2.619601     0.59   0.556    -3.610176    6.699115
          3  |   3.552439   2.364203     1.50   0.134    -1.099654    8.204532
…            |
       _cons |  -2.618306   2.341458    -1.12   0.264    -7.225644    1.989032
------------------------------------------------------------------------------
```


Two-stage least squares estimator

- There are various estimators that employ instrumental variables to estimate the effect of endogenous variables without bias.
- A commonly used estimator is the **two-stage least squares estimator (2SLS):**
  - **Stage 1:** regress endogenous variable on IV and covariates and then use this fitted linear model to generate a prediction of the endogenous variable
  - **Stage 2:** fit linear model of interest after replacing the endogenous variable by this prediction
- Assumptions:
  - (i) Linear second stage model
  - (ii) Existence of IVs for endogenous variables
- The 2SLS is **consistent** for the causal parameters of the linear model.
- Theory provides an SE, but note that this is not the SE that would be produced by a regression on predictions.

2SLS commands

- Most general purpose statistics packages include a 2SLS and other IV estimators.
  - SAS: PROC SYSLIN
  - R: ivreg command, package systemfit and steinIV
  - (SPSS menu: Analyze – Regression – 2-Stage Least Squares)
- Stata's 2SLS command structure:

type of IV estimator — Response variable and exogenous covariates

```
ivregress 2sls y x1 x2 … (endo = iv)
```

(Set of) endogenous variable(s) — (Set of) instrumental variable(s)

  - Note: The exogenous covariates x1 x2 ... are automatically included in the first stage regression.

ISCB Birmingham 2016 79

2SLS unadjusted

```
ivregress 2sls bdi6 (treat = rgroup), first

First-stage regressions
                                                  Number of obs     =        317
                                                  F(   1,    315)   =     278.23
                                                  Prob > F          =     0.0000
                                                  R-squared         =     0.4690
                                                  Adj R-squared     =     0.4673
                                                  Root MSE          =     0.3534

------------------------------------------------------------------------------
       treat |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      rgroup |   .6666667   .0399672    16.68   0.000     .5880302    .7453031
       _cons |  -1.11e-15   .0298649    -0.00   1.000    -.0587599    .0587599
------------------------------------------------------------------------------


Instrumental variables (2SLS) regression          Number of obs   =        317
                                                  Wald chi2(1)    =       2.66
                                                  Prob > chi2     =     0.1029
                                                  R-squared       =          .
                                                  Root MSE        =      10.13

------------------------------------------------------------------------------
        bdi6 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       treat |  -2.803511   1.718695    -1.63   0.103    -6.172091    .5650694
       _cons |   15.15714    .856179    17.70   0.000     13.47906    16.83522
------------------------------------------------------------------------------
Instrumented:  treat
Instruments:   rgroup
```

note CC approach

ISCB Birmingham 2016 80

2SLS adjusted

```
ivregress 2sls bdi6 bdi0 i.centre (treat = rgroup), first

First-stage regressions
------------------------------------------------------------------------------
       treat |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        bdi0 |    .003287   .0025343     1.30   0.196    -.0016998    .0082738
             |
      centre |
          2  |   .0537707   .1083458     0.50   0.620    -.1594236     .266965
          3  |   .0936235   .0975303     0.96   0.338    -.0982889    .2855359
…
      rgroup |   .6658542   .0400489    16.63   0.000     .5870491    .7446593
       _cons |   -.040299   .0972521    -0.41   0.679     -.231664    .1510661
------------------------------------------------------------------------------


Instrumental variables (2SLS) regression          Number of obs   =        317
------------------------------------------------------------------------------
        bdi6 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       treat |  -3.516463   1.446056    -2.43   0.015    -6.35068    -.682245
        bdi0 |   .5418511   .0614513     8.82   0.000     .4214088    .6622935
             |
      centre |
          2  |   2.116314   2.613953     0.81   0.418    -3.006939    7.239568
          3  |   3.996486   2.356315     1.70   0.090     -.621807    8.614779
…
       _cons |  -2.236712   2.331698    -0.96   0.337    -6.806756    2.333333
------------------------------------------------------------------------------
Instrumented:  treat
Instruments:   bdi0 2.centre 3.centre 4.centre 5.centre 6.centre 7.centre
               8.centre rgroup
```

better "predictors in first stage regression (IVs and measured confounders)"

ISCB Birmingham 2016

Results summary (CC)

Efficacy estimators	Missing Data	Difference in BDI6 (std error)
ITT – unadjusted	CC	-1.87 (1.14)
ITT – adjusted	CC	-2.34 (0.96)
As-treated – unadjusted	CC	-1.26 (1.18)
As-treated – adjusted	CC	-1.11 (1.00)
Per-protocol – unadjusted	CC	-1.84 (1.29)
Per-protocol – adjusted	CC	-1.74 (1.09)
CACE – unadjusted	CC	-2.80 (1.72)
CACE – adjusted	CC	-3.52 (1.45)

Discussion (1)

It appears 2SLS estimates the efficacy of psychotherapy for depression.

But for whom?

Discussion (2)

Some participants characterised as "non-compliers" in the therapy received a few therapy sessions.

```
tab1 sessions if (rgroup==1 & treat==0), miss

-> tabulation of sessions if (rgroup==1 & treat==0)

        no. |
   sessions |
interventio |
n completed |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |         56       51.85       51.85
          1 |         19       17.59       69.44
          2 |          7        6.48       75.93
          3 |          7        6.48       82.41
          4 |          2        1.85       84.26
          5 |          3        2.78       87.04
          . |         14       12.96      100.00
------------+-----------------------------------
      Total |        108      100.00
```

What are the implications of this for the exclusion restriction?

What would be a better approach?

Discussion (3)

Non-compliance with therapy offer in the active arm predicted loss to follow-up at 6 months. (This is typical for trials of psychological interventions.)

```
tab2 treat resp if rgroup==1, row exact


           |   non-missing bdi6
     treat |         0          1 |     Total
-----------+----------------------+----------
         0 |        49         59 |       108
           |     45.37      54.63 |    100.00
-----------+----------------------+----------
         1 |        10        118 |       128
           |      7.81      92.19 |    100.00
-----------+----------------------+----------
     Total |        59        177 |       236
           |     25.00      75.00 |    100.00

          Fisher's exact =                 0.000
```

Does our CC analysis allow for such a missing data generating process?

What would be a better approach?

Summary: Assessing efficacy in the presence of non-compliance

- We can consider causal effects of treatment offer (effectiveness) or treatment receipt (efficacy).
- Individual causal treatment effects can vary between individuals (effect heterogeneity).
- Statistical inference is concerned with unbiased estimation of average causal treatment effects.
- Under non-compliance with randomly allocated treatments ITT, per-protocol and as-treated approaches suffer biases when estimating ATE or ATT respectively.
- If we are willing to focus on the subpopulation of compliers and believe the exclusion restriction holds then we can estimate CACE in order to assess treatment efficacy.
- The most versatile method for doing this is an IV approach.

Key message

Does the treatment work?

And if so, by how much?

References

- Clarke P., Palmer T. and Windmeijer F. (2015). Estimating structural mean models with multiple instrumental variables using the generalised method of moments. *Statistical Science* **30** (1); 96-117
- Dunn G., Maracy M., Dowrisk C., Ayuso-Mateos J.L, Dalgard O.S., Page H., Lehtinen V., Casey P., Wilkinson C., Vaquez-Barquero J.L and Wilkinson G. (2003) Estimating psychological treatment effects from a randomised controlled trial with both non-compliance and loss to follow-up, *British Journal of Psychiatry* **183** (4): 323-331
- Dunn G. and Bentall R. (2007). Modelling treatment-effect heterogeneity in randomized controlled trials of complex interventions (psychological treatments). *Statistics in Medicine* **26**: 4719–4745
- Fischer-Lapp K. and Goetghebeur E. (1999). Practical properties of structural mean models for the analysis of noncompliance data. *Controlled Clinical Trials* **20**: 531-546
- Frangakis C.E. and Rubin D. (2002). Principal stratification in causal inference. *Biometrics* **58** (1): 21-29.
- Imbens G.W. and Angrist J.D. (1994) Identification and Estimation of Local Average Treatment Effects. *Econometrica* **62** (2): 467-475
- Pickles A. and Croudace T. (2010). Latent mixture models for multivariate ad longitudinal outcomes. *Statistical Methods in Medical Research* **19**(3):271-89
- Rubin, D.B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* **66:** 688-701.
- Tchetgen-Tchetgen E.J., Walter S., Vansteelandt S., Martinussen T. and Glymour M. (2015) Instrumental variables estimation in a survival context. *Epidemiology* **26**: 402-410.
- White I.R. (2005). Uses and limitations of randomization-based efficacy estimators. *Statistical Methods in Medical Research* **14**(4):327-47