Principal Stratum estimands: Concepts, challenges and estimation

Prof Richard Emsley
NIHR Senior Investigator & Professor of Medical Statistics and Trials Methodology, IoPPN, King's College London
Academic Director, King's Clinical Trials Unit
Director, NIHR Research Support Service at King's College London and Partners
Theme Lead: Trials, Genomics & Prediction, NIHR Maudsley Biomedical Research Centre

Society for Clinical Trials Annual Meeting – 19th May 2026

NIHR | National Institute for Health and Care Research

NIHR | Maudsley Biomedical Research Centre

KING'S HEALTH PARTNERS

KING'S College LONDON

If you love SCT, come to ICTMC in September

ICTMC 2026
International Clinical Trials Methodology Conference
Birmingham, UK
14-17 September

ictmc.org

ictmc@-in-conference.org.uk

Disclosures

This represents independent research supported by the National Institute for Health and Care Research (**NIHR Research Professorship, NIHR300051**) and the National Institute for Health and Care Research (NIHR) Maudsley Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the National Institute for Health and Care Research or the Department of Health and Social Care.

No disclosures: Dr Emsley

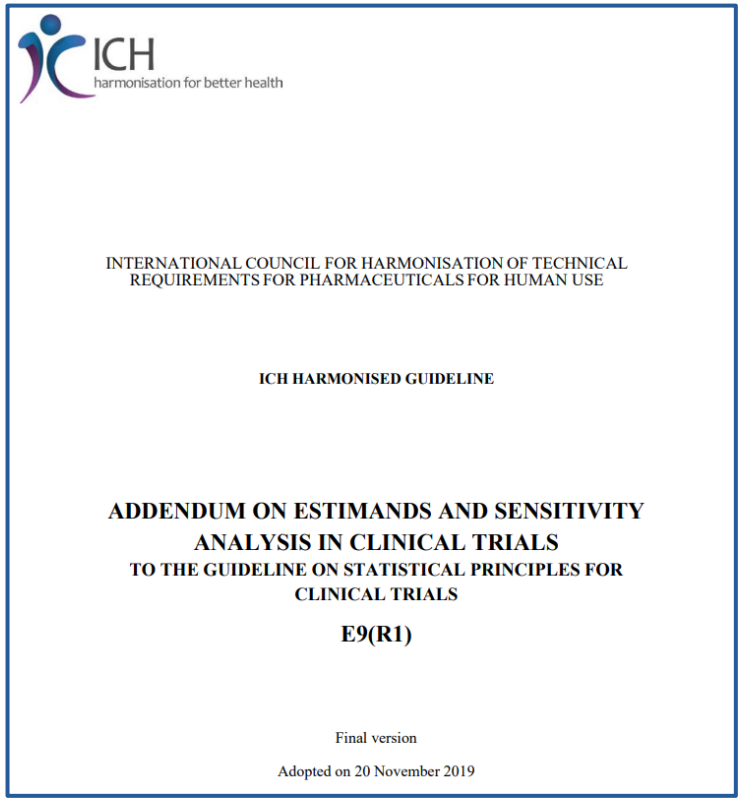
Some slides courtesy of previous joint presentation with Dr Brennan Kahan, UCL

FUNDED BY
NIHR | National Institute for Health and Care Research

Overview

- Concept of Principal Stratification
- Some estimation methods:
  - Re-weighting
  - Instrumental variables
- Summary

The estimand framework - ICH E9 (R1)



ICH
harmonisation for better health

INTERNATIONAL COUNCIL FOR HARMONISATION OF TECHNICAL REQUIREMENTS FOR PHARMACEUTICALS FOR HUMAN USE

ICH HARMONISED GUIDELINE

ADDENDUM ON ESTIMANDS AND SENSITIVITY ANALYSIS IN CLINICAL TRIALS
TO THE GUIDELINE ON STATISTICAL PRINCIPLES FOR CLINICAL TRIALS

E9(R1)

Final version

Adopted on 20 November 2019

ICH
harmonisation for better health

HOME | ABOUT ICH | WORK PRODUCTS | MEETINGS | TRAINING | NEWSROOM | Search...

Home \ News \

ICH E9(R1) Addendum reaches Step 4 of the ICH Process

4 December 2019

The ICH E9(R1) Addendum to Defining the Appropriate Estimand for a Clinical Trial/Sensitivity Analyses reached *Step 4* of the ICH Process at the ICH meeting in Singapore on 20 November 2019.

The ICH E9(R1) Addendum presents a structured framework to strengthen the dialogue between disciplines involved in the formulation of clinical trial objectives, design, conduct, analysis and interpretation, as well as between sponsor and regulator regarding the treatment effect (s) of interest that a clinical trial should address.

The ICH E9(R1) Addendum is available for download on the ICH website here.

Estimands

- Structured approach to defining the treatment effect, to make clear **what** is being estimated
  - Ensure everyone understands what is being estimated
  - Ensure what is being estimated is relevant
  - Ensure study design/data collection/analysis are aligned with the question

➢ Estimand = A precise description of exactly ***what*** *treatment effect* you **want (or *demand)*** to find out

Estimands framework

- Five key attributes that form the description of an estimand:

(A) The population; the patients targeted by the scientific question

(B) The treatment condition of interest and the alternative treatment condition(s) e.g. control or placebo

(C) The variables (or endpoint) to be obtained for each patient required to address the scientific question

(D) The specification of how to account for **intercurrent events** to reflect the scientific question of interest

(E) The population level summary for the variable which provides a basis for a comparison between treatment conditions

Intercurrent events

- Post-randomisation events which affect the **interpretation** or **occurrence** of outcome data

- Examples
  - Treatment discontinuation
  - Failure to initiate treatment
  - Treatment switching
  - Use of rescue medication
  - Death

- We can use different strategies for handling different intercurrent events

Strategies to address intercurrent events

Treatment policy

Hypothetical

Composite

While on treatment/alive

Principal stratum

Introducing notation and potential outcomes

- Y = outcome
- Z = treatment allocation

- $Y(1)$ = what Y would be if Z=1
- $Y(0)$ = what Y would be if Z=0

- $Y(1)$ and $Y(0)$ are potential outcomes

- Estimands are based on contrasts of **potential outcomes:** $E[Y(1)] - E[Y(0)]$

- When using potential outcomes we'll make the **consistency** assumption, which is that $Y(1) = Y$ if $Z = 1$ (and similar for $Z = 0$), i.e. each patients' observed outcome is equal to their potential outcome under their assigned treatment. We assume this is fulfilled under well-defined treatments and outcomes.

Principal stratification idea

- Classify subjects into latent classes which are defined by their joint potential responses of the intermediate variable (intercurrent event) to all possible random allocations
- These classes are known as principal strata which have the property that they are independent of treatment allocation and can be handled in the analysis in an analogous way to pre-randomisation variables e.g. via stratification
- Frangakis and Rubin (2002) introduced the concept of principal effects which compare treatments within principal strata (within-class or stratum-specific (ITT) effects)
- Can also fit more complicated models within principal strata (see Kim's talk…)

Emsley, R. and Dunn G. 2011: "Principal Stratification". In: Encyclopaedic Companion to Medical Statistics. Eds: Everitt B. Second Edition. Hodder Arnold.

Frangakis, C. E. and Rubin, D. B. 1999: "Addressing complications of intention-to-treat analysis in the combined presence of all-or-none treatment-noncompliance and subsequent missing outcomes", Biometrika, vol. 86, no. 2, pp. 365-379.

Frangakis, C. E. and Rubin, D. B. 2002: "Principal stratification in causal inference", Biometrics, vol. 58, no. 1, pp. 21-29.

Origin of Principal Stratification

BIOMETRICS 58, 21–29
March 2002

Principal Stratification in Causal Inference

Constan

Department of Biostat
615 N. Wolfe Street, Ba
email: cl

Don

Department of St
1 Oxford Street, Cambr
email: rub

Biometrika (1999), **86,** 2, *pp.* 365–379

Printed in Great Britain

Addressing complications of intention-to-treat analysis in the combined presence of all-or-none treatment-noncompliance and subsequent missing outcomes

BY CONSTANTINE E. FRANGAKIS AND DONALD B. RUBIN

Department of Statistics, Harvard University, 1 Oxford Street, Cambridge, Massachusetts 02138, U.S.A.

frangaki@hustat.harvard.edu rubin@stat.harvard.edu

Principal strata notation

- *Y* = outcome
- *Z* = treatment allocation
- *D* = intercurrent event (1=yes, no=0)

- $D(1)$ = what *D* would be if *Z*=1
- $D(0)$ = what *D* would be if *Z*=0

- Treatment effect in the subpopulation of participants who would not (or would) experience the intercurrent event under either arm:

$$E(Y(1) - Y(0) | \underbrace{D(1) = D(0) = 0}_{\text{Subset of patients who wouldn't experience IE under Z=1 or Z=0}})$$

Kahan BC, White IR, Edwards M, Harhay MO. Using modified intention-to-treat as a principal stratum estimator for failure to initiate treatment. Clin Trials. 2023 Jun;20(3):269-275.

Intercurrent event: Failure to initiate randomised treatment

- Four principal strata defined by binary intercurrent event ‘failure to initiate randomised treatment’ (*D*), which is fully observed in both arms:

	Initiate treatment (D=1)?	
	Control (Z=0)	Intervention (Z=1)
Always initiators	Yes	Yes
Intervention initiators	No	Yes
Control initiators	Yes	No
Never initiators	No	No

Intercurrent event: Failure to initiate randomised treatment

- Challenge: principal strata are unidentifiable without further assumptions

	Initiate treatment (D=1)?	
	Control (Z=0)	Intervention (Z=1)
Always initiators	Yes	Yes
Intervention initiators	No	Yes
Control initiators	Yes	No
Never initiators	No	No

- In the control arm, a participant who initiates treatment could be an ‘always initiator’ or a ‘control initiator’
- In the intervention arm, a participant who initiates treatment could be an ‘always initiator’ or an ‘intervention initiator’

Methods for estimating principal strata strategies

- Re-weighting
- Instrumental variables
- Modified intention-to-treat (not covered today)
- Finite mixture models (not covered today)
- Bayesian finite mixture models (not covered today)

Re-weighting method

- We know whether patients experience an intercurrent event under their assigned treatment
  - But not whether they would have under the alternate treatment
- Re-weighting uses baseline covariates, **X**, to predict the probability of the IE under the alternate treatment
- Then, using these predictions, we can infer which patients are mostly likely to belong to the principal stratum of interest

[1] Hayden D, Pauler DK and Schoenfeld D. An Estimator for Treatment Comparisons among Survivors in Randomized Trials. *Biometrics* 2005; 61: 305-310.

Re-weighting: Assumptions

- The re-weighting method makes two key assumptions:

1. Conditional independence between $D(Z = 0)$ and $D(Z = 1)$ given X
   - i.e. given baseline covariates, whether patients would experience the IE under one treatment assignment is independent of whether they would experience it under the other treatment assignment

2. Intercurrent event status under one assignment is independent of outcome under alternate assignment
   - E.g. $D(\mathrm{Z} = 0) \perp Y(Z = 1)|X, D(Z = 1) = 0$

[1] Hayden D, Pauler DK and Schoenfeld D. An Estimator for Treatment Comparisons among Survivors in Randomized Trials. *Biometrics* 2005; 61: 305-310.

Re-weighting: implementation

We focus on principal stratum of patients who would not experience the IE under either treatment assignment:

Steps:

1. Fit logistic model with outcome="No IE" (1=no IE, 0=experienced IE) and baseline covariates to intervention group patients

2. Obtain predictions using model in (1) for **control** arm patients

3. Fit logistic model with outcome="No IE" and baseline covariates to control arm patients

4. Obtain predictions using model in (3) for **intervention** arm patients

5. Generate new variable "weights", which is equal to the predictions Pr(no IE) from step (2) for usual care patients, and those from step (4) for intervention patients

6. Fit a weighted regression model which excludes patients who experienced the IE, with the weights from step (5)

Re-weighting: Example

Trial with:

- Outcome (Y)
- Intercurrent event status (D); and
- Binary baseline covariate (X)

Intervention arm (Z=1)					Control arm (Z=0)				
Y	D	X			Y	D	X		
3	0	1			7	0	1		
8	0	0			4	0	1		
12	1	0			6	0	0		
19	1	1			13	1	0		
…					…				

Re-weighting: Example

Step 1: fit logistic regression model with "no IE" as outcome and baseline covariates X to **intervention arm** patients

$$logit\big(P(D = 0|Z = 1)\big) = \alpha + \beta X$$

Intervention arm (Z=1)					Control arm (Z=0)				
Y	D	X			Y	D	X		
3	0	1			7	0	1		
8	0	0			4	0	1		
12	1	0			6	0	0		
19	1	1			13	1	0		
…					…				

Re-weighting: Example

Step 2: apply model in step (1) to control arm patients to get predicted probability of "no IE" if they had been assigned to intervention ($P\left(D^{(Z=1)}=0\right)$)

Intervention arm (Z=1)					Control arm (Z=0)				
Y	D	X			Y	D	X	$P\left(D^{(Z=1)}=0\right)$	
3	0	1			7	0	1	0.73	
8	0	0			4	0	1	0.73	
12	1	0			6	0	0	0.59	
19	1	1			13	1	0	0.59	

Re-weighting: Example

Step 3: same as step 1, but fit model to control arm patients

$$logit\big(P(D = 0|Z = 0)\big) = \alpha + \beta X$$

Intervention arm (Z=1)					Control arm (Z=0)				
Y	D	X			Y	D	X	$P\big(D^{(Z=1)} = 0\big)$	
3	0	1			7	0	1	0.73	
8	0	0			4	0	1	0.73	
12	1	0			6	0	0	0.59	
19	1	1			13	1	0	0.59	

Re-weighting: Example

Step 4: apply model in step (3) to intervention arm patients to get predicted probability of "no IE" if they had been assigned to control ($P\left(D^{(Z=0)}=0\right)$)

Intervention arm (Z=1)				Control arm (Z=0)			
Y	D	X	$P\left(D^{(Z=0)}=0\right)$	Y	D	X	$P\left(D^{(Z=1)}=0\right)$
3	0	1	0.65	7	0	1	0.73
8	0	0	0.45	4	0	1	0.73
12	1	0	0.45	6	0	0	0.59
19	1	1	0.65	13	1	0	0.59

Re-weighting: Example

Step 5: generate “weight” variable from the predicted probabilities

Intervention arm					Control arm				
Y	D	X	$P(D^{(Z=0)} = 0)$	Weight	Y	D	X	$P(D^{(Z=1)} = 0)$	Weight
3	0	1	0.65	0.65	7	0	1	0.73	0.73
8	0	0	0.45	0.45	4	0	1	0.73	0.73
12	1	0	0.45	0.45	6	0	0	0.59	0.59
19	1	1	0.65	0.65	13	1	0	0.59	0.59

Re-weighting: Example

Step 6: fit weighted regression model to outcome Y, excluding patients with intercurrent event

```
regress y z if d == 0,
[pweight=weight]
```

Intervention arm					Control arm				
Y	D	X	$P(D^{(Z=0)} = 0)$	Weight	Y	D	X	$P(D^{(Z=1)} = 0)$	Weight
3	0	1	0.65	0.65	7	0	1	0.73	0.73
8	0	0	0.45	0.45	4	0	1	0.73	0.73
~~12~~	~~1~~	~~0~~	~~0.45~~	~~0.45~~	6	0	0	0.59	0.59
~~19~~	~~1~~	~~1~~	~~0.65~~	~~0.65~~	~~13~~	~~1~~	~~0~~	~~0.59~~	~~0.59~~
…					…				

An alternative scenario: non-compliance

- Consider the setting of open-label intervention vs. usual care where the IE (*D*) is receipt of intervention being studied (1=yes, 0=no)
- Those randomised to usual care arm have no access to the intervention treatment
- Those randomised to receive the intervention treatment receive usual care if they do not take up the offer
- This is referred to as ‘all or none’ treatment compliance
- The IE is fully observed in the intervention arm but unknown in the usual care arm
- Use a principal stratification strategy to estimate the **Complier Average Causal Effect (CACE)**

Intercurrent event: Receipt of intervention treatment

- Four principal strata defined by binary intercurrent event ‘receipt of intervention treatment’ (*D*):

	Receipt of intervention treatment (D=1)?	
	Control (Z=0)	Intervention (Z=1)
Always-takers	Yes	Yes
Compliers	No	Yes
Defiers	Yes	No
Never-takers	No	No

Intercurrent event: Receipt of intervention treatment

- Four principal strata defined by binary intercurrent event 'receipt of intervention treatment' (*D*):

	Receipt of intervention treatment (D=1)?	
	Control (Z=0)	Intervention (Z=1)
~~Always-takers~~	~~Yes~~	~~Yes~~
Compliers	No	Yes
~~Defiers~~	~~Yes~~	~~No~~
Never-takers (non-compliers)	No	No

- Under 'all or none compliance', the controls have no access to receive the intervention treatment and so there are no *always-takers* or *defiers*.

- *Never-takers* often referred to as <u>non-compliers</u> in contrast to compliers:
  - Compliers get treatment if and only if allocated to the treatment condition.
  - Non-compliers never get the treatment, regardless of allocation

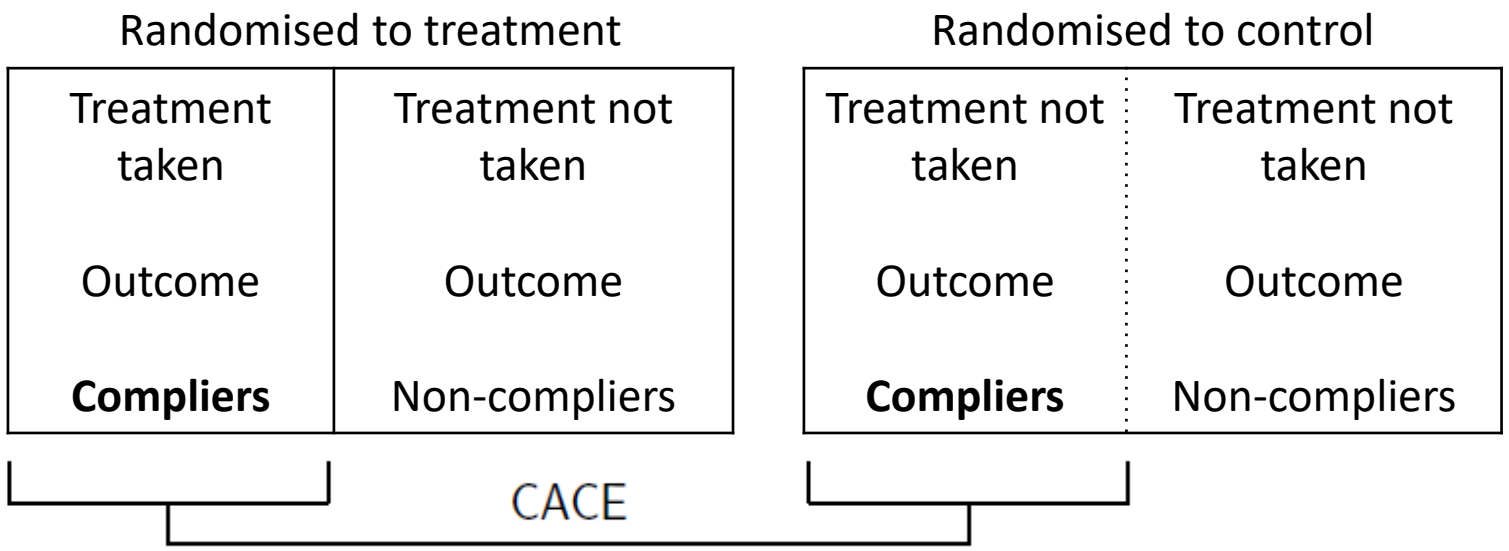
CACE estimation: assumptions

1. There are two latent classes of participants: Compliers and Non-compliers.
   - 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. In the absence of treatment (i.e. for the Non-compliers) there is no effect of randomisation (i.e. treatment arm) on outcome.
   - This assumption is often called an exclusion restriction

The Complier Average Causal Effect (CACE)



- 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

- This is a randomisation-respecting estimate $E(Y(1) - Y(0)|D(1) - D(0) = 1)$

- It is the ITT effect in the sub-group of participants who would always comply with their treatment allocation. It is not subject to confounding.

Hypothetical example: CACE analysis

# participants mean outcome	Compliers	Non-compliers	All
Treatment Z=1	180 14.25	55 13.10	235 13.98
Control Z=0	? ?	? ?	220 15.16

Hypothetical example: CACE analysis

# participants mean outcome	Compliers	Non-compliers	All
Treatment Z=1	180  14.25	55  13.10	235 13.98
Control Z=0	220*(180/235) = 168.51 ?	220*(55/235)= 51.49 ?	220 15.16

Use randomisation to fill in expected subgroup sizes in the control group (77% compliance).

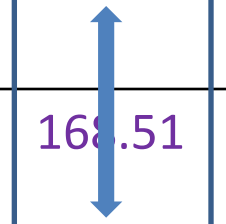
Hypothetical example: CACE analysis

# participants mean outcome	Compliers	Non-compliers	All
Treatment Z=1	180  14.25	55  13.10	235 13.98
Control Z=0	168.51 ?	51.49 13.10	220 15.16

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

Hypothetical example: CACE analysis

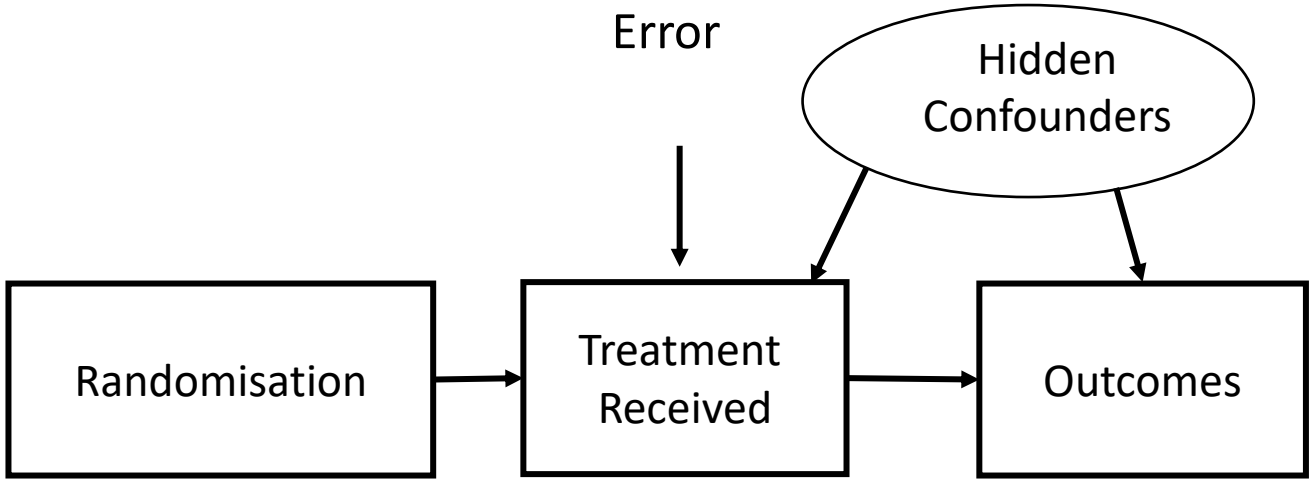
# participants mean outcome	Compliers	Non-compliers	All
Treatment Z=1	180 14.25	55 13.10	235 13.98
Control Z=0	168.51 15.79	51.49 13.10	220 15.16



Work out remaining entry from this information.

CACE = 14.25-15.79 = -1.54 and |CACE| > |-1.18| = |ITT|

The Instrumental Variable model



- Randomisation influences both treatment receipt and outcome. However, the influence on outcome is only through treatment received. Confounders are independent of randomisation.

- By assuming the absence of a direct path from randomisation to outcome, we assume the entire effect of randomisation acts through receipt of treatment.
  - **randomisation is an instrumental variable.**

- For continuous outcomes, there are several estimation approaches available

Two Stage Least Squares (2SLS) procedure

- The first stage involves:
  - Regress *D* on *Z* and *X* using OLS then save the predicted values of $\widehat{D}$.
  - $D = \alpha_0 + \alpha_1 Z + \alpha_2 X + \omega$

- Then at the second stage:
  - Regress *Y* on the predicted value of *D,* and *X* using OLS.
  - $Y = \beta_0 + \beta_1 \widehat{D} + \beta_2 X + \varepsilon$

- A correction needs to be made to the standard errors if performing this manually – standard software does this
  - e.g. **`ivregress y x (d = z)`**

- Note that same covariates are included in both stages

- Under normality 2SLS and LIML are asymptotically equivalent

- Imbens and Angrist (1994) have shown that if *Z* and *D* binary and provided that the target population contains no defiers (monotonicity assumption) then IV methods estimate CACE

Summary

- Principal stratification is an appropriate method to restrict the population to those who would (or would not) experience an IE
- There are a number of statistical approaches to estimating principal strata
- These rely on either strong assumptions or having good baseline predictors of the IE
- Baseline IE predictors should be considered at the design stage
- Some estimation methods e.g. reweighting, IVs can be straightforward to implement

Thank you for your attention

Email: richard.emsley@kcl.ac.uk

Twitter: @richardaemsley

NIHR | National Institute for Health and Care Research

NIHR | Maudsley Biomedical Research Centre

KING'S HEALTH PARTNERS

Trials Methodology TMRP Research Partnership

KING'S College LONDON

Modified Intention to Treat (mITT)

- Participants who do not initiate their randomised treatment are excluded
  - Analysis is performed only on those participants who initiate treatment in the respective groups

$$E(Y|Z = 1, D(1) = 1) - E(Y|Z = 0, D(0) = 1)$$

- Compare the subset of participants in the intervention group who initiate intervention treatment with the subset of participants in the control group who initiate the control treatment

- Example syntax: `regress y z if d == 1`

mITT estimator

- The mITT estimator excluding participants who do not initiate treatment compares the highlighted cells in the intervention group with the highlight cells in the control group

	Initiate treatment (D=1)?	
	Control (Z=0)	Intervention (Z=1)
Always initiators	Yes	Yes
Intervention initiators	No	Yes
Control initiators	Yes	No
Never initiators	No	No

- Problem: estimator is not based on a randomised comparison, as it compares different principal strata between the treatment arms (i.e. the analysis population differs between treatment groups).

mITT assumptions

- We need assumptions to identify the principal stratum membership of each participant

- The mITT estimator is unbiased for the PS estimand under assumption of no 'intervention initiators' and no 'control initiators'
  - If someone does not initiate treatment in one arm, they would also not initiate treatment in the other arm (and, conversely, if someone does initiate treatment in one arm, they would also initiate treatment in the other arm)

- This assumption implies there are only two principal strata in the trial, 'never initiators' and 'always initiators'

- If we observe a participant's initiation status in one treatment arm, we can infer their initiation status had they been allocated to the alternative treatment, and identify the principal stratum to which they belong

mITT estimator

- Assume there are no ‘intervention initiators’ and no ‘control initiators’
- The mITT estimator now compares the highlighted cells in a randomisation respecting analysis

	Initiate treatment (D=1)?	
	Control (Z=0)	Intervention (Z=1)
Always initiators	Yes	Yes
~~Intervention initiators~~	~~No~~	~~Yes~~
~~Control initiators~~	~~Yes~~	~~No~~
Never initiators	No	No

Can we check the assumption?

- The no ‘intervention or control initiators’ assumption can be partly assessed:
  - If true, then the proportion of non-initiators should, on average, be the same across randomised groups.
  - If there are large discrepancies in the proportion of non-initiators between groups, this may provide evidence that this assumption has been violated.
  - However, even if this proportion is the same across groups, this is not a guarantee the assumption is true.

When is mITT appropriate?

1. The intercurrent event of interest must be identifiable in each treatment arm
   - i.e. we must be able to measure which participants in each treatment group experience the intercurrent event.

2. The occurrence of the intercurrent event is not affected by treatment allocation
   - i.e. that if the intercurrent event occurs for a participant under one treatment condition, it would also occur under the other treatment condition.

➢ Kahan et al (2023) give detailed examples of trials where mITT principal stratum estimator is (and is not) appropriate:

mITT appropriate	mITT not appropriate
Double blind placebo trials where initiation of placebo arm is also monitored	Open label trials versus usual care where intercurrent event is failure to attend intervention – no way of identifying the IE in control arm
External events cause intercurrent event e.g. cancellation of planned surgery if surgery in both arms	Open label trials where IE is failure to attend allocated treatment but could be affected by allocated treatment

Estimating Intervention Effects in Post-Randomization Subgroups: A Longitudinal Mixed Effects Principal Stratification Model

Kimberley Goldsmith
Department of Biostatistics & Health Informatics
King's College London
Society for Clinical Trials Annual Meeting
May 19th, 2026

Disclosure

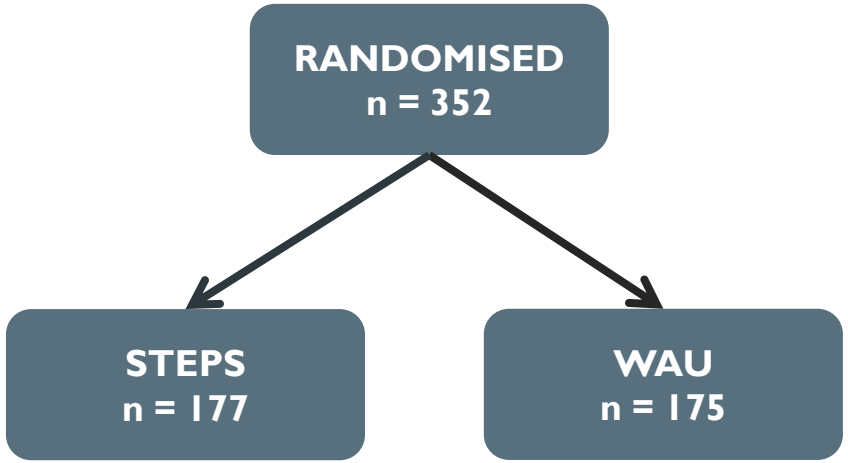
- The OPTIMA and SoCRATES studies were funded by the National Institute for Health and Care Research (NIHR), the SPARKLE study by UK Research and Innovation.

- KG reports funding from the NIHR Biomedical Research Centre and the South London and the Maudsley Applied Research Collaboration.

- This study was funded by the National Institute for Health and Care Research (NIHR). The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care

Acknowledgements: study participants, and colleagues involved in the studies or discussions about the work to be presented including Richard Emsley, Gordon Forbes, Ewan Carr, Nick Becksley-Hoelscher, Sam Norton, George Vamvakas, Oliver Pearson, and OPTIMA colleagues including the Chief Investigator Edmund Sonuga-Barke and the Project Manager Kasia Kostyrka-Allchorne.

OPTIMA (On-Line Parent Training for the Initial Management of ADHD Referrals)



A. STEPS home screen

STEPS

- Tools to support parents management of 5–11 year old children's behavioural problems. Delivery: modules incl short videos, audio clips, downloadable resources, space to record audio/written notes

WAU

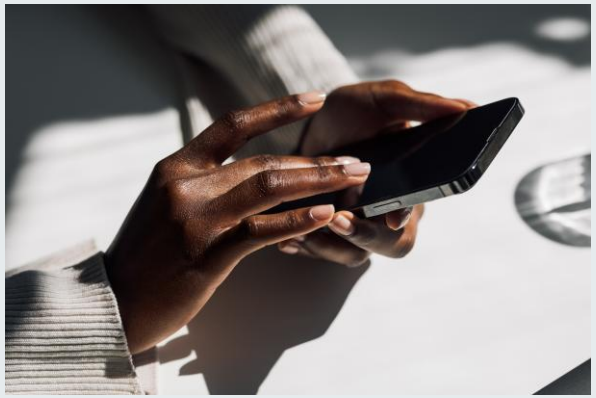
- Wait as usual

Primary outcome:
Parent reported oppositional defiant disorder (ODD) from Swanson, Nolan, and Pelham Rating Scale (SNAP-IV)
3-, 6-, 9-, 12-months post-randomisation

Protocol: Kostyrka-Allchorne K, …, Goldsmith K, …, Sonuga-Barke EJS. Trials. 2022 Dec 12;23(1):1003. doi: 10.1186/s13063-022-06952-z.

Digital interventions: what constitutes adequate dose/compliance?

- How do we measure engagement with an app-delivered intervention?
- What is "compliance" with respect to an app-delivered intervention?
- How is compliance handled in practice?



Elkes *et al. BMC Medical Research Methodology* (2024) 24:184
https://doi.org/10.1186/s12874-024-02308-0

BMC Medical Research Methodology

RESEARCH Open Access

User engagement in clinical trials of digital mental health interventions: a systematic review

Jack Elkes[1*], Suzie Cro[1], Rachel Batchelor[2], Siobhan O'Connor[3], Ly-Mee Yu[2], Lauren Bell[4], Victoria Harris[2], Jacqueline Sin[5†] and Victoria Cornelius[1†]

Discussion Many articles report user engagement metrics but few assessed the impact on the intervention effect missing opportunities to answer important patient centred questions for how well DMHIs work for engaged users. Defining engagement in this area is complex, more research is needed to obtain ways to categorise this into groups. However, the majority that considered engagement in analysis used approaches most likely to induce bias.

Parenting app engagement across the SPARKLE & OPTIMA trials

Across two trials:

- ~80 – 90% download/use app

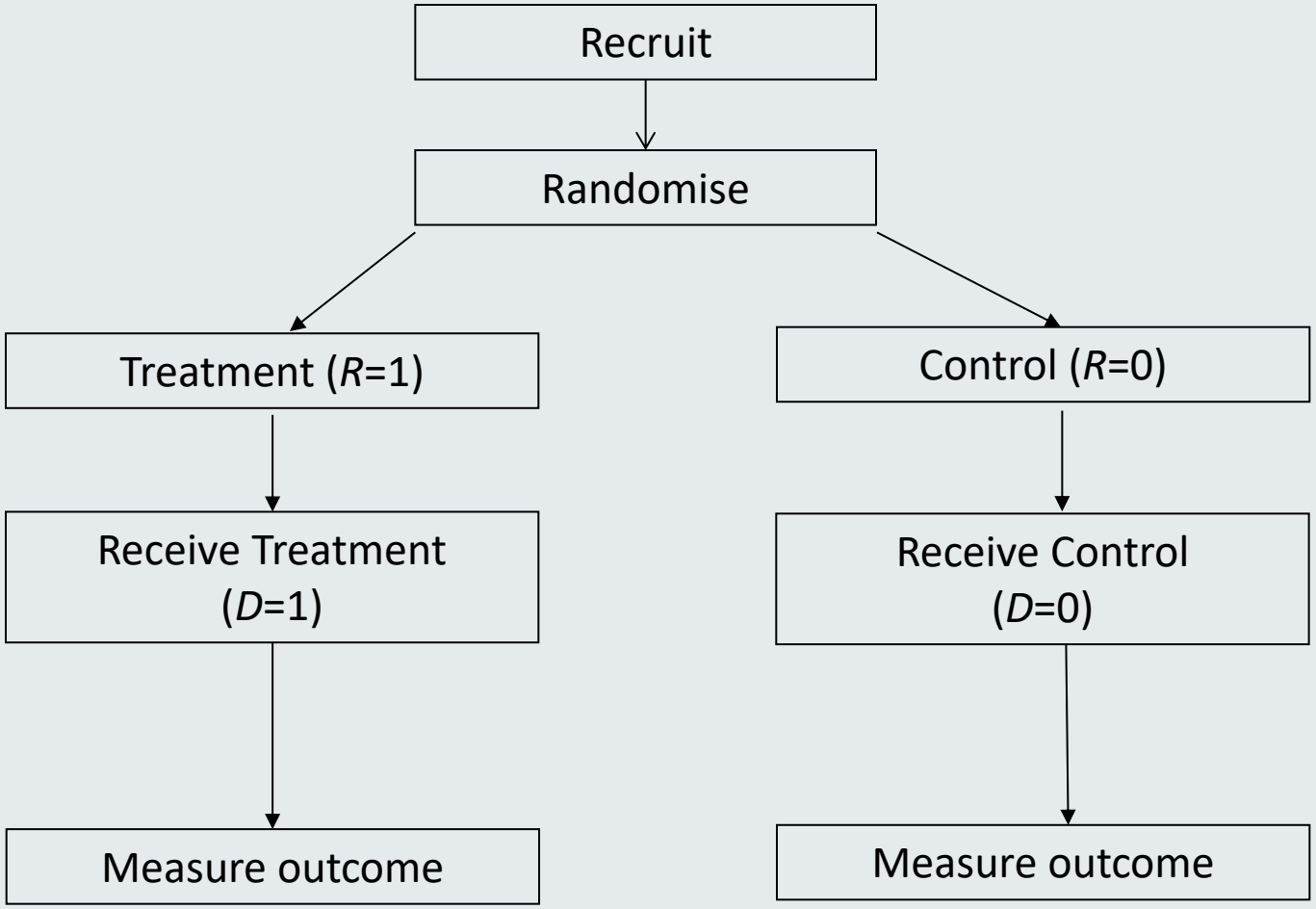
SPARKLE 45s or more in any one module

- 50% – 60%

OPTIMA: Completion of first two of eight (ordered) modules

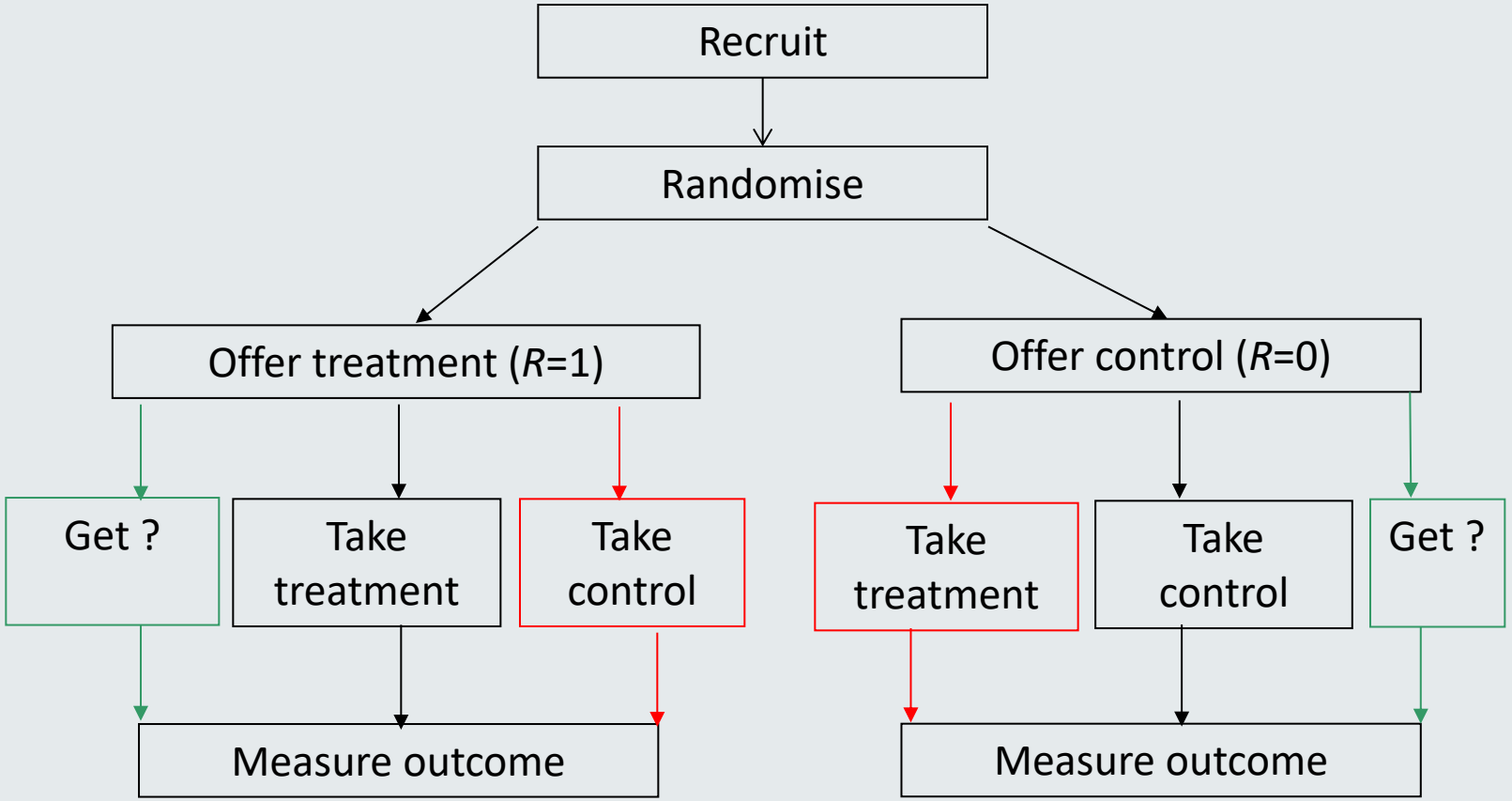
- ~50%

A ‘perfect’ randomised controlled trial



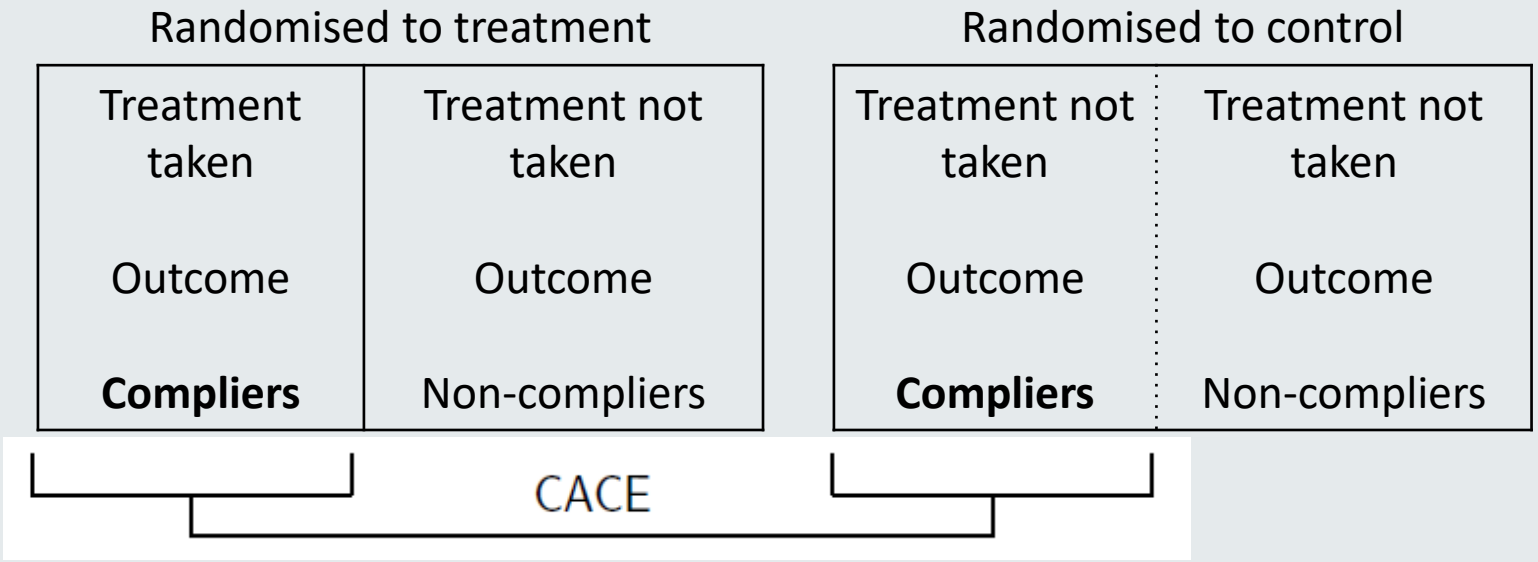
Slide courtesy of: Professor Richard Emsley

A more realistic RCT with treatment-related intercurrent events



Slide courtesy of: Professor Richard Emsley

The Complier Average Causal Effect (CACE)



Complier – accepts allocation, receives therapy as per protocol

Non-complier – would never receive therapy, whatever the allocation

Both can be identified in the Treatment Group, but **hidden or latent** in the controls

- ***Complier-Average Causal Effect (CACE) estimate*** compares average outcome of treatment arm compliers with average outcome of would-be compliers in control arm
- Randomisation-respecting estimate
- ITT effect in sub-group who would always comply with their treatment allocation – not subject to confounding

Effects/estimands of interest in OPTIMA

Main aim:

Single STEPS v WAU effect

Population: Intention to Treat

Treatment Policy Strategy

Key secondary aims:

Single STEPS v WAU effect

Population: those using the app/compliant

Principal Stratum Strategy
CACE estimate

Two STEPS v WAU effects

Population 1: those using the app more

Population 2: those using the app less

Principal Stratum Strategy
Relaxing CACE assumption of
0 effect in 'non-compliers'

https://www.ema.europa.eu/en/ich-e9-statistical-principles-clinical-trials-scientific-guideline

HEALTH TECHNOLOGY ASSESSMENT

VOLUME 19 ISSUE 93 NOVEMBER 2015
ISSN 1366-5278

Evaluation and validation of social and psychological markers in randomised trials of complex interventions in mental health: a methodological research programme

Graham Dunn, Richard Emsley, Hanhua Liu, Sabine Landau, Jonathan Green, Ian White and Andrew Pickles

Chapter 3

Colleagues have worked on applicable methods….

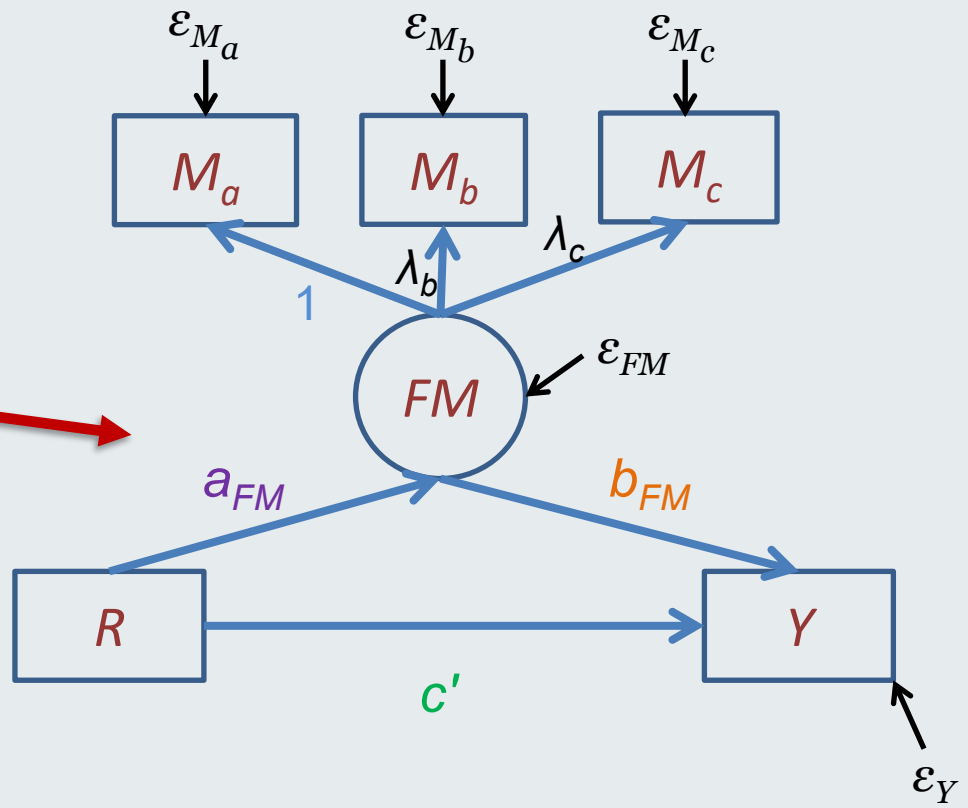
Statistical Methods in Medical Research 2010; **19**: 237–270

Mediation and moderation of treatment effects in randomised controlled trials of complex interventions

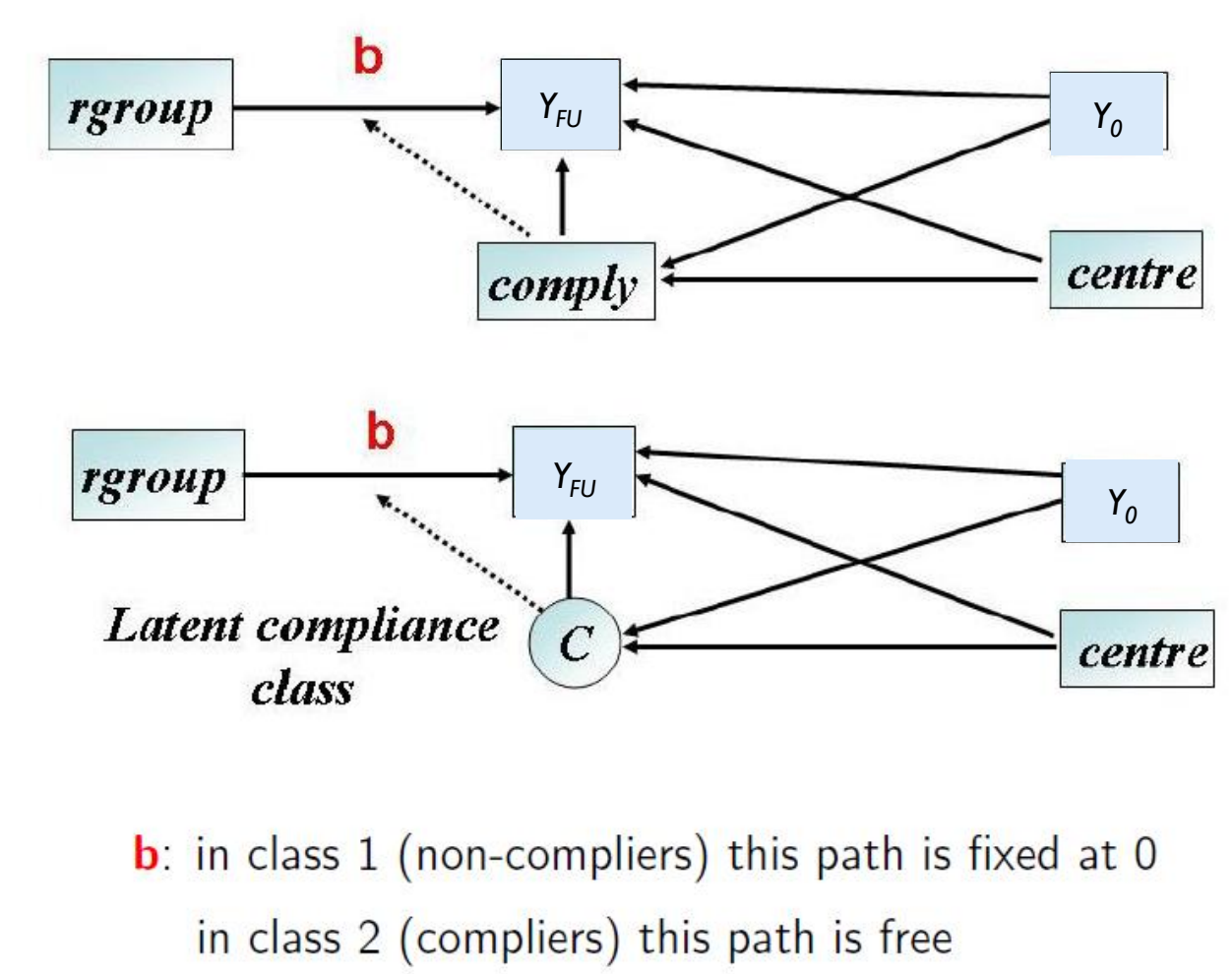
Richard Emsley, Graham Dunn Health Methodology Research Group, School of Community Based Medicine, University of Manchester, UK and **Ian R White** MRC Biostatistics Unit, Cambridge, UK

Structural Equation Modelling (SEM) Framework

- Fits multiple regression models simultaneously, using shared variance-covariance structure of the data
- Modelling of latent and observed variables
- Can also model variables that are partly observed/partly latent



Path diagram for CACE latent class model



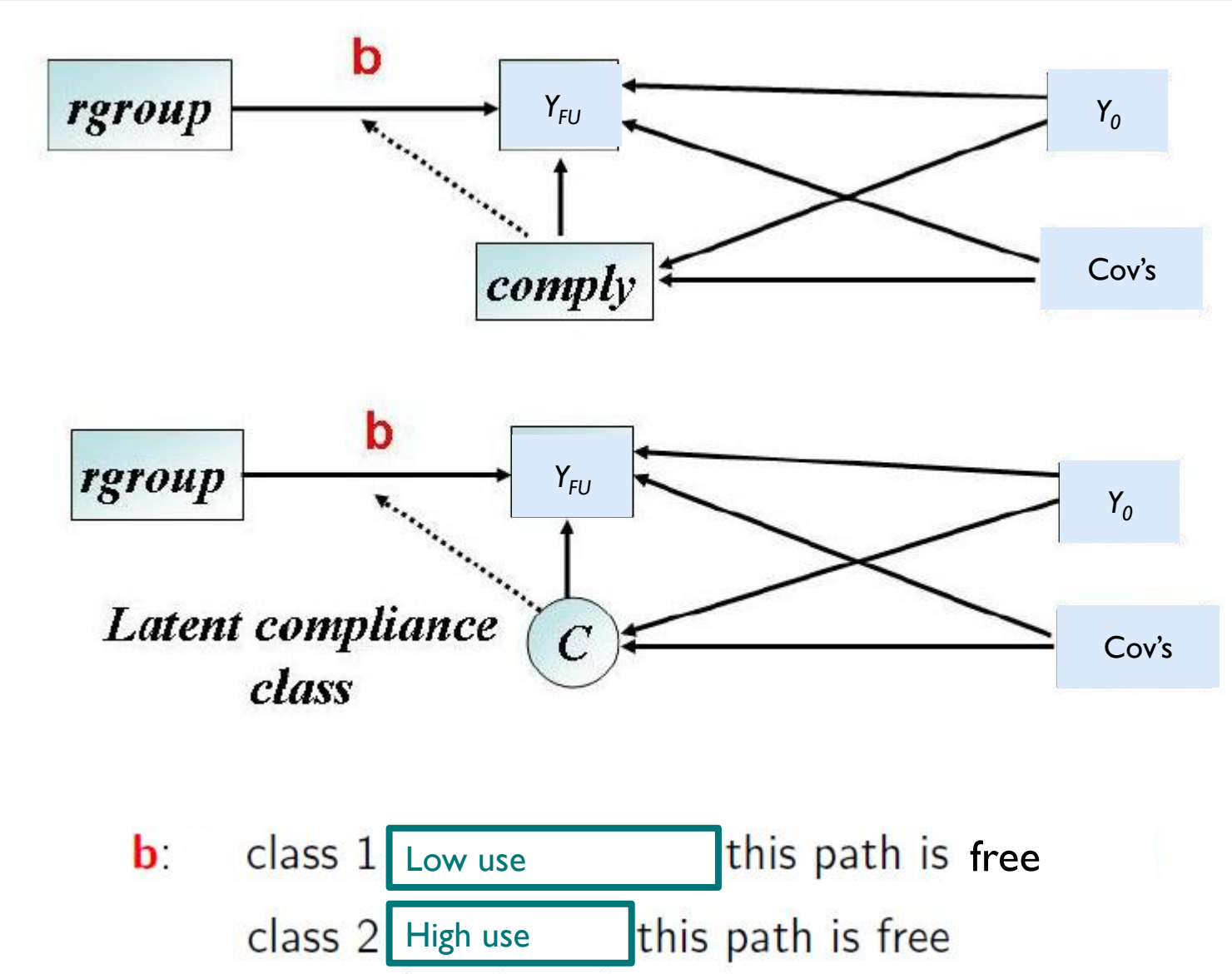
- Formalise the CACE as a discrete latent variable model
- Compliance observed in the treatment arm and latent in the control arm
- Allows for predictors of class membership that might not be included in the model for the outcome
- Allows for exclusion restriction to be relaxed

NEED GOOD PREDICTORS OF COMPLIANCE TO BE ABLE TO DO THIS

- Can extend theory to non-linear outcomes

Slide courtesy of: Professor Richard Emsley

Path diagram for PREM* latent class model - OPTIMA



*Post-randomisation effect modification

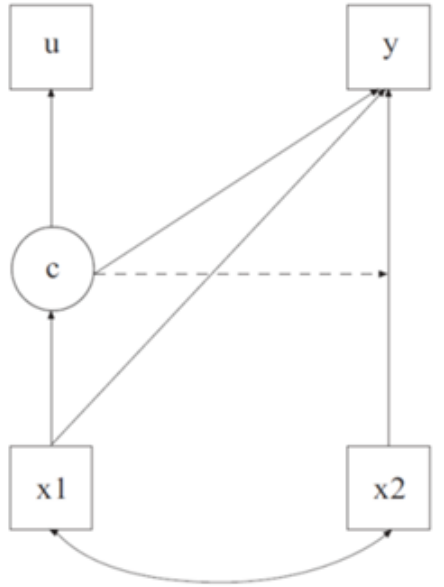
For whom compliance is observed / unobserved: software

Mplus (manual)

```
TITLE:     this is an example of mixture randomized
           trials modeling using CACE estimation with
           missing data on the latent class indicator
DATA:      FILE IS ex7.24.dat;
VARIABLE:  NAMES ARE u y x1 x2;
           CLASSES = c (2);
           CATEGORICAL = u;
           MISSING = u (999);
ANALYSIS:  TYPE = MIXTURE;
```

Ex 7.24

```
MODEL:
           %OVERALL%
           y ON x1 x2;
           c ON x1;

           %c#1%      non-complier
           [u$1@15];
           [y];
           y;
           y ON x2@0;

           %c#2%      complier
           [u$1@-15];
           [y*.5];
           y;
OUTPUT:    TECH1 TECH8;
```



Single measure of outcome

Stata

```
gsem ///
(1: depvar <- i.treatment@0 [indepvars] [, family(familyname) ] ) ///
(2: depvar <- i.treatment [indepvars] [, family(familyname) ] ) ///
(C <- [varlist] ) ///
(1: comp <- _cons@-15, logit) ///
(2: comp <- _cons@15, logit), ///
lclass(C 2)
```

u and comp:

Treatment complier = 1
Treatment non-complier = 0
Control = missing

Troncoso & Morales-Gomez, 2022, doi: 10.1177/1536867X221106416

Primary analysis model for OPTIMA trial

In OPTIMA, **outcomes were measured repeatedly post-randomisation**

So we fitted a mixed effects model with a random intercept for participant
(to account for repeated measures, a model we commonly use):

```
mixed odd time##randomisation_arm odd_baseline || optima_id:
```

`odd` = outcomes – long form data with outcomes at all post-randomisation time points

`time` = time point

`Randomisation_arm` = binary allocation variable

`odd_baseline` = baseline measure of conduct outcome (ANCOVA)

`|| optima_id:` = random intercept

Discrete latent variable mixed effects models - Stata

gsem formulation of mixed effects model:

```
gsem (y <- i.arm i.time y0 M[id])
```

Not currently possible in Stata (as far as I know):

```
gsem  (1.C: y <- i.arm@0 i.time y0 M1[id]) ///
        (2.C: y <- i.arm i.time y0 M1[id]) ///
        (C <- ) ///
        (1: compliance <- _cons@-15, logit) ///
        (2: compliance <- _cons@15, logit), ///
        lclass(C 2) latent(M1)
```

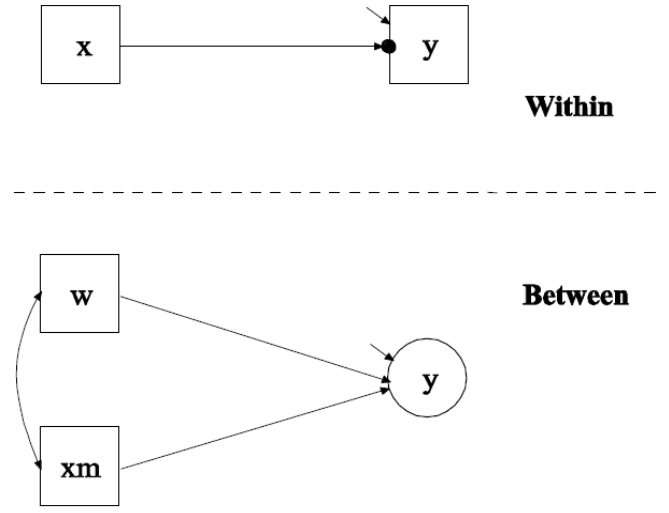
```
*option lclass() is not allowed with models specified with
continuous latent variables
```

Discrete latent variable mixed effects models - Mplus

Is possible here, requires combining CACE type model shown earlier with multilevel model

```
TITLE:     this is an example of a two-level
           regression analysis for a continuous
           dependent variable with a random intercept
           and an observed covariate
DATA:      FILE = ex9.1a.dat;
VARIABLE:  NAMES = y x w xm clus;
           WITHIN = x;
           BETWEEN = w xm;
           CLUSTER = clus;
DEFINE:    CENTER x (GRANDMEAN);
ANALYSIS:  TYPE = TWOLEVEL;
MODEL:
           %WITHIN%
           y ON x;
           %BETWEEN%
           y ON w xm;
```

Ex 9.1



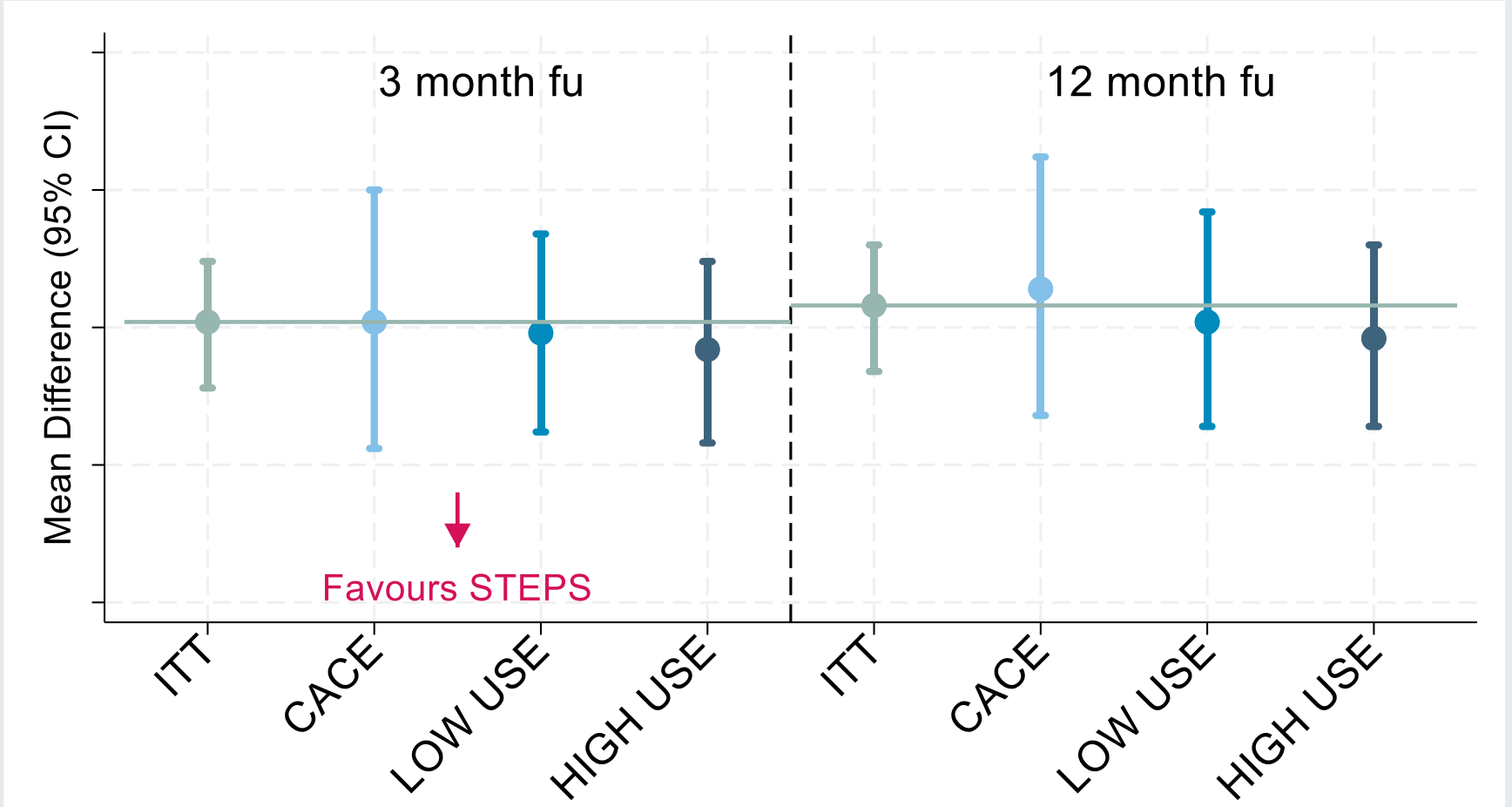

```
MODEL:
           %OVERALL%
           y ON x1 x2;
           c ON x1;

           %c#1%
           [u$1@15];
           [y];
           y;
           y ON x2@0;

           %c#2%
           [u$1@-15];
           [y*.5];
           y;
OUTPUT:    TECH1 TECH8;
```

Ex 7.24

OPTIMA outcome, Scores higher = worse

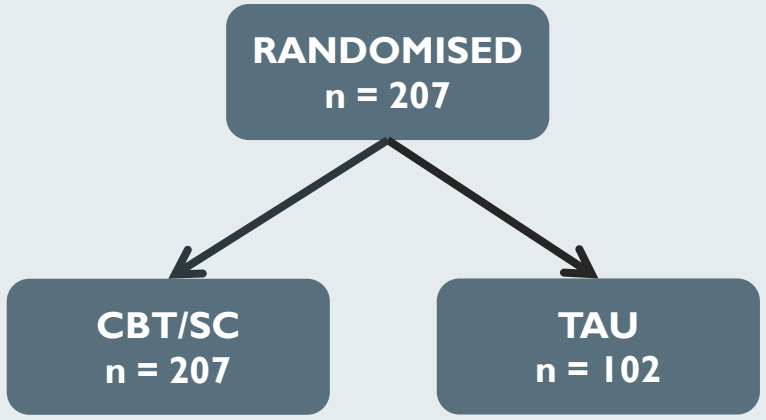
Predictors:

- centre (stratifier)
- baseline measure of outcome
- education
- autism features
- other neurodivergent children
- employment status
- income
- marital status
- completed objective rating task



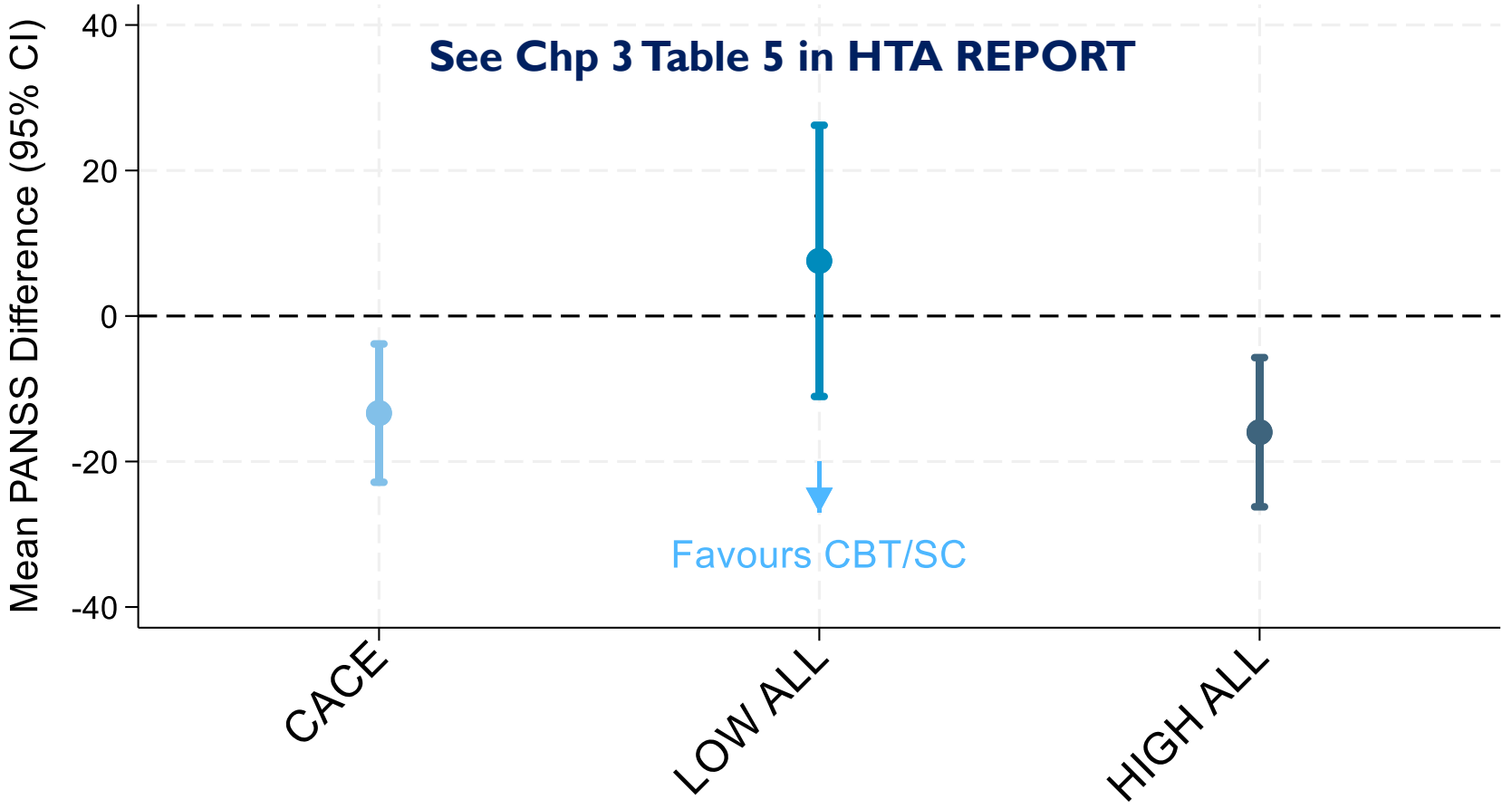
STEPS (n = 177 / 154) v WAU (n = 175 / 155)
mean differences (95% CI)

SoCRATES study, PANSS scores higher = worse



Predictors:

- centre
- baseline PANSS score
- years of education
- logarithm of duration of untreated illness



CBT/Supportive Counselling + TAU (n = 207 / 92)
v TAU (n = 102 / 69)
mean differences (95% bootstrapped CIs)

Discussion / Future work

- Further checks on models

- Restricted to complete cases to get models working
  - Properly incorporate missing data / latent ignorability

- Ongoing issues with meaningful measures of engagement with digital interventions

- Good predictors of compliance – and gathering these in trials
  - Especially for digital interventions?
  - Prespecify?

kimberley.goldsmith@kcl.ac.uk

www.linkedin.com/in/kimberley-goldsmith-a787a029b

NIHR | National Institute for Health and Care Research

NIHR | Maudsley Biomedical Research Centre

KING'S HEALTH PARTNERS

KING'S College LONDON

Combatting assumptions in handling intercurrent events: Semiparametric principal stratification analysis beyond monotonicity

Fan Li

Department of Biostatistics
Yale School of Public Health

https://lifan90.com/

The Society for Clinical Trials (SCT) 46th Annual Meeting
May 19, 2026

Discloure

- This talk is primarily based on a manuscript to appear in *Statistica Sinica*
  - Tong J, Kahan B, Harhay MO, Li F (2028). Semiparametric principal stratification analysis beyond monotonicity. *Statistica Sinica*. In Press.
- This work was supported by the United States National Institutes of Health (NIH), National Heart, Lung, and Blood Institute (NHLBI, grant numbers R01-HL168202 and 1R01HL178513). All statements in this report, including its findings and conclusions, are solely those of the authors and do not necessarily represent the views of the NIH.
- The authors also acknowledge the Yale University-Mayo Clinic Center of Excellence in Regulatory Science and Innovation (CERSI) for supporting this study.

Intercurrent events

- Intercurrent events: post-treatment events (or intermediate outcomes) that affect either the interpretation or existence of the final outcome
  - Direct regression adjustment for an observed intercurrent event, often fails to produce causally interpretable treatment effects due to post-randomization selection bias
- Examples
  - Truncation by death: mortality occurring post-randomization can lead to ambiguously defined non-mortality outcomes
  - Treatment noncompliance or discontinuation
- The ICH E9 (R1) addendum suggested the Estimands Framework, which recognized the challenges due to intercurrent events and provided handling strategies

Handling intercurrent events

Kahan et al. (2023)[1]

Strategy for Handling Intercurrent Events in Estimand Definition	Description
Treatment policy	The outcome regardless of the occurrence of the intercurrent event is of interest (i.e., the intercurrent event is considered part of the treatment strategy). Cannot be used for truncating events which preclude the occurrence of the outcome.
Composite	The intercurrent event is included in the endpoint definition—for example, by assigning a particular value of the endpoint to patients who experience the intercurrent event. Different composite estimands could be defined on the basis of the particular value assigned to the endpoint.
While on treatment/while alive	The endpoint prior to the occurrence of the intercurrent event is of interest.
Hypothetical	The effect of treatment in a hypothetical scenario where the intercurrent event would not occur is of interest. There can be multiple hypothetical settings that would apply to any particular intercurrent event, so it is necessary to describe the precise hypothetical setting envisaged.
Principal stratum	The treatment effect in the subset of participants who would (or would not) experience the intercurrent event is of interest. Different principal strata could be defined (e.g., the set of participants who would not discontinue treatment under either treatment vs. the set who would not discontinue if assigned to the intervention).

[1] Kahan BC, Cro S, Li F, Harhay MO (2023). Eliminating ambiguous treatment effects using estimands. *American Journal of Epidemiology*.

Principal stratification

- Among all five strategies, principal stratification targets the finest estimands
- A rich literature since Frangakis and Rubin (2002)[2]
  - nonparametric bounds
  - model-based identification
  - nonparametric point identification
- Nonparametric point identification transparently separates structural assumptions from model assumptions
  - monotonicity
  - principal ignorability
  - exclusion restriction
  - ...

[2] Frangakis CE, Rubin DB. (2002). Principal stratification in causal inference. *Biometrics*.

Principal stratification - cont'd

- Among all five strategies, principal stratification targets the finest estimands **BUT** come with strongest assumptions
- When adopting the principal stratum strategy in critical care trials
  - Acute Respiratory Distress Syndrome (ARDS) trial
  - 549 patients with acute lung injury and ARDS who were on mechanical ventilation were randomly assigned to either lower or higher positive end-expiratory pressure (PEEP) levels (NHLBI ARDS Clinical Trials Network 2004)
  - interested in long-term outcomes, which may be truncated by death
  - neither the high nor low PEEP levels are considered standard care and their relative impact on survival remains controversial
- "*I don't trust the principal stratum strategy as the monotonicity assumption is questionable. . .*"

Monotonicity or non-monotonicity?

- **Monotonicity** as a mainstream assumption for nonparametric identification
  - highly questionable in the presence of two active treatments
- **Counterfactual intermediate independence** offers another solution (Hayden et al. 2005)
  - capturing all relevant baseline predictors of a post-baseline variable is challenging (Vansteelandt and Van Lancker 2024)[3]
- **Question**: Is there a middle ground between these two extremes? Can we construct optimal estimators beyond monotonicity?

[3] Vansteelandt S, Van Lancker K. (2024). Chasing shadows: how implausible assumptions skew our understanding of causal estimands. *Statistics in Biopharmaceutical Research*.

Notation

Consider a study with treatment assignment $Z_i \in \{0, 1\}$, a binary intermediate outcome $D_i \in \{0, 1\}$ and a final outcome Y_i

- N: sample size
- Z_i: refer to $Z_i = 1$ and $Z_i = 0$ as the treatment and control for simplicity
- $\{Y_i(0), Y_i(1)\}$: a pair of potential final outcomes; observed final outcome is $Y_i = Y_i(Z_i)$
- $\{D_i(0), D_i(1)\}$: a pair of potential intercurrent event (IE); observed value $D_i = D_i(Z_i)$
- Assuming treatment ignorability $Z_i \perp \{D_i(0), D_i(1), Y_i(0), Y_i(1)\}|\mathbf{X}$
  - handles randomized assignment as a special case

Principal causal effects

- The joint potential IE—$\{D(0), D(1)\}$—can be treated as a pre-treatment covariate; classify patient subgroups (principal strata) $\{D(0), D(1)\} = d_0 d_1$ if $D(0) = d_0$ and $D(1) = d_1$

$d_0 d_1$	$D(0) = d_0, D(1) = d_1$	*Truncation by Death*	*Noncompliance*
01	$D(0) = 0, D(1) = 1$	Protected	Compliers
11	$D(0) = 1, D(1) = 1$	Always-survivors	Always-takers
10	$D(0) = 1, D(1) = 0$	Harmed	Defiers
00	$D(0) = 0, D(1) = 0$	Never-survivors	Never-takers

- Principal causal effects are defined as

$$\mu_{d_0 d_1} = E\{Y(1) - Y(0) | D(0) = d_0, D(1) = d_1\}.$$

  - survivor average causal effect (SACE) (stratum 11)
  - complier average causal effect (CACE) (stratum 01)

Two "competing" assumptions

Assumption 1 (*Monotonicity*)

$D(1) \geq D(0)$.

Assumption 2 (*Counterfactual intermediate independence*)

$D(0) \perp D(1)|\mathbf{X}$.

- Assumption 1 $\Rightarrow$ 3 strata (no harmed or defier stratum), but requires strong content knowledge
- Assumption 2 $\Rightarrow$ 4 strata, but requires all relevant baseline predictors of a post-baseline variable
- Which one to consider?

A sensitivity assumption

Assumption 3 (*Conditional odds ratio*)

Defining $e_{d_0 d_1}(\mathbf{X}) \equiv \Pr(D(0) = d_0, D(1) = d_1|\mathbf{X})$ *as the principal score, there exists a known function* $\theta(\bullet) : \mathcal{X} \to [0, \infty]$, *such that*

$$\frac{e_{11}(\mathbf{X})e_{00}(\mathbf{X})}{e_{10}(\mathbf{X})e_{01}(\mathbf{X})} = \theta(\mathbf{X}).$$

- leveraging odds ratio (OR) as an association measure
  - $\theta(\mathbf{X}) > 1$: the odds of $D(1) = 1|D(0) = 1$ greater than $D(1) = 1|D(0) = 0$
  - $\theta(\mathbf{X}) < 1$: the odds of $D(1) = 1|D(0) = 1$ smaller than $D(1) = 1|D(0) = 0$
  - bears the flavor of stochastic monotonicity
- **margin-free property**: the range of the dependence measure is not constrained by the unknown margins

Special cases

- $\theta(\mathbf{X}) = 1 \Leftrightarrow D(0) \perp D(1)|\mathbf{X}$
- Monotonicity is more nuanced
  - $\theta(\mathbf{X}) \to \infty$ is implied by either $e_{10}(\mathbf{X}) = 0$ or $e_{01}(\mathbf{X}) = 0$
  - $\theta(\mathbf{X}) \to \infty$ only refers to oscillating monotonicity
  - see example distribution of $\{D(0), D(1)\}$ given a binary X

		$D(0) = 0$	$D(0) = 1$	$\theta(X)$	$D(1) \geq D(0)$?
$X = 0$	$D(1) = 0$ $D(1) = 1$	$e_{00}(X = 0) = 1/3$ $e_{01}(X = 0) = 1/3$	$e_{10}(X = 0) = 0$ $e_{11}(X = 0) = 1/3$	∞	✓
$X = 1$	$D(1) = 0$ $D(1) = 1$	$e_{00}(X = 1) = 1/3$ $e_{01}(X = 1) = 0$	$e_{10}(X = 1) = 1/3$ $e_{11}(X = 1) = 1/3$	∞	×

- Define $p_z(\mathbf{X}) = \Pr(D|\mathbf{X}, Z = z), \quad z = 0, 1$
- $\{\theta(\mathbf{X}) \to \infty\} + \{p_1(\mathbf{X}) > p_0(\mathbf{X})\} \Leftrightarrow D(1) \geq D(0)$
- Recommending use existing methods assuming monotonicity (Assumption 1 holds)

Ignorability assumption

Assumption 4 (*Mean Principal ignorability*)

4(a): $E\{Y(1)|D(0)=1, D(1)=1, \mathbf{X}\} = E\{Y(1)|D(0)=0, D(1)=1, \mathbf{X}\}$

4(b): $E\{Y(1)|D(0)=1, D(1)=0, \mathbf{X}\} = E\{Y(1)|D(0)=0, D(1)=0, \mathbf{X}\}$

4(c): $E\{Y(0)|D(0)=1, D(1)=1, \mathbf{X}\} = E\{Y(0)|D(0)=1, D(1)=0, \mathbf{X}\}$

4(d): $E\{Y(0)|D(0)=0, D(1)=1, \mathbf{X}\} = E\{Y(0)|D(0)=0, D(1)=0, \mathbf{X}\}$

- weaker than $\{Y(1), Y(0)\} \perp \{D(1), D(0)\}|\mathbf{X}$
- only 2 out of 4 needed for a single principal causal estimand

$d_0 d_1$	Estimand	Nonparametric identification
01 (CACE)	μ_{01}	Assumptions 3 + 4(a) + 4(d)
11 (SACE)	μ_{11}	Assumptions 3 + 4(a) + 4(c)
10	μ_{10}	Assumptions 3 + 4(b) + 4(c)
00	μ_{00}	Assumptions 3 + 4(b) + 4(d)

Nonparametric point identification

- For simplicity, mainly focus on the truncation by death and SACE
  - full results can be found in Tong et al. (2025)[4]
- Under treatment ignorability, and Assumptions 3, 4(a) and 4(c)

$$\mu_{11} = \frac{E\{e_{11}(\mathbf{X})ZDY/[\pi(\mathbf{X})p_1(\mathbf{X})]\}}{E\{e_{11}(\mathbf{X})ZD/[\pi(\mathbf{X})p_1(\mathbf{X})]\}} - \frac{E\{e_{11}(\mathbf{X})(1-Z)DY/[(1-\pi(\mathbf{X}))p_0(\mathbf{X})]\}}{E\{e_{11}(\mathbf{X})(1-Z)D/[(1-\pi(\mathbf{X}))p_0(\mathbf{X})]\}},$$

and, equivalently, with

$$\mu_{11} = \frac{E\{e_{11}(\mathbf{X})(m_{11}(\mathbf{X}) - m_{01}(\mathbf{X}))\}}{E\{e_{11}(\mathbf{X})\}},$$

where $m_{zd}(\mathbf{X}) = E\{Y|Z=z, D=d, \mathbf{X}\}$ is the conditional outcome mean, and

$$e_{11}(\mathbf{X}) = \frac{1 + (\theta(\mathbf{X})-1)(p_0(\mathbf{X}) + p_1(\mathbf{X})) - \sqrt{\delta(\mathbf{X})}}{2(\theta(\mathbf{X})-1)},$$

$$\delta(\mathbf{X}) = [1 + (\theta(\mathbf{X})-1)(p_0(\mathbf{X}) + p_1(\mathbf{X}))]^2 - 4\theta(\mathbf{X})(\theta(\mathbf{X})-1)p_0(\mathbf{X})p_1(\mathbf{X}).$$

[4] Tong J, Kahan B, Harhay MO, Li F (2025). Semiparametric principal stratification analysis beyond monotonicity. *Statistica Sinica*. In Press.

Conditionally doubly robust estimator

- To improve robustness and efficiency, we also derive the efficient influence function (EIF) motivate an improved estimator—combining principal & propensity score weighting and outcome regression
  - complicated exact form given in Tong et al. (2025)
  - **Property 1**: as long as the principal score model is correctly specified, $\widehat{\mu}_{11}^{\mathrm{CDR}}$ is consistent if either the propensity score model or the outcome mean model is correctly specified.
  - **Property 2**: when all the working models are correctly specified, $\widehat{\mu}_{11}^{\mathrm{CDR}}$ is efficient and attains the variance lower bound.
- Further enhancement via debiased machine learning
  - estimating all nuisance models via machine learning, under cross-fitting
  - if nuisance learned at a rate $\geq n^{1/4}$, $\widehat{\mu}_{11}^{\mathrm{ML}}$ is $\sqrt{n}$-consistency and efficient—addressing model misspecification bias

Simulation study

- Conducted extensive simulations to confirm the properties of the conditionally doubly robust estimators (machine learning extensions)
- The new class of estimators naturally allow us to explore the impact of incorrect assumptions about $\{D(1), D(0)\}$
  - **design 1:** assuming monotonicity holds (loosely $\theta_{\text{true}} = \infty$), fit estimators under $\theta \in \{0.5, 1, 2, \infty\}$
  - **design 2:** assuming non-monotonicity holds (e.g., $\theta_{\text{true}} = 0.5$), fit estimators under $\theta \in \{0.5, 1, 2, \infty\}$

θ-specification	θ_{true} in the data-generating process	
	Monotonicity	Non-monotonicity
Monotonicity	The scenario considered in Jiang et al. (2022), where we extend the DML under cross-fitting.	1. Largest bias and undercoverage of CI; 2. Unstable estimates for CACE; 3. Unable to estimate PCE for defiers.
Non-monotonicity	Mild bias and undercoverage, which vanishes with increasing θ.	1. Our approaches produce valid inference with correctly specified θ; 2. Some bias and undercoverage with moderate misspecification of θ.

Design 1: under monotonicity $\theta_{\text{true}} = \infty$

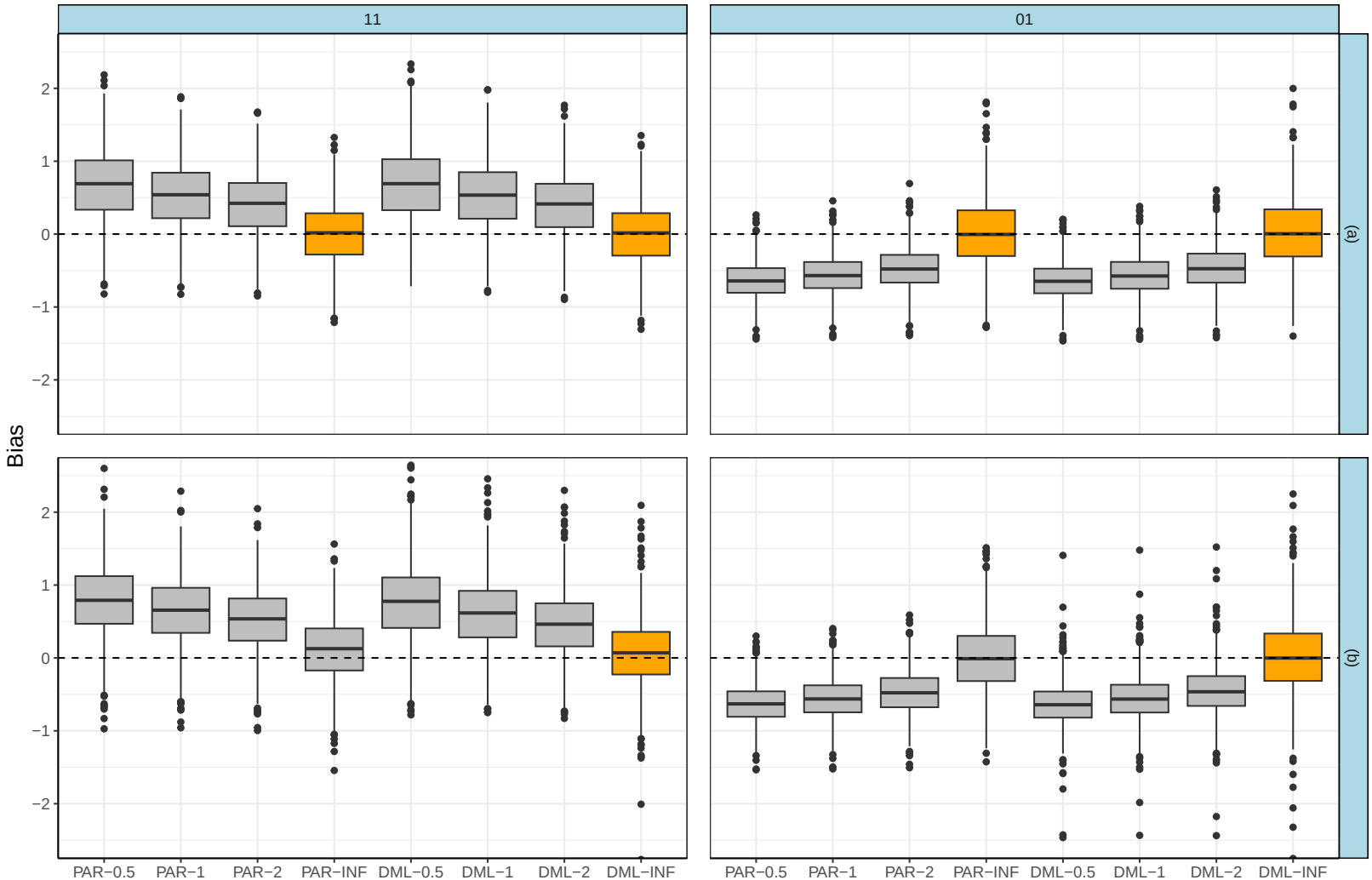


Figure 1: Scenarios (a) and (b) correspond to cases where all working models are either correctly specified or all incorrectly specified.

Design 2: under non-monotonicity $\theta_{\text{true}} = 0.5$

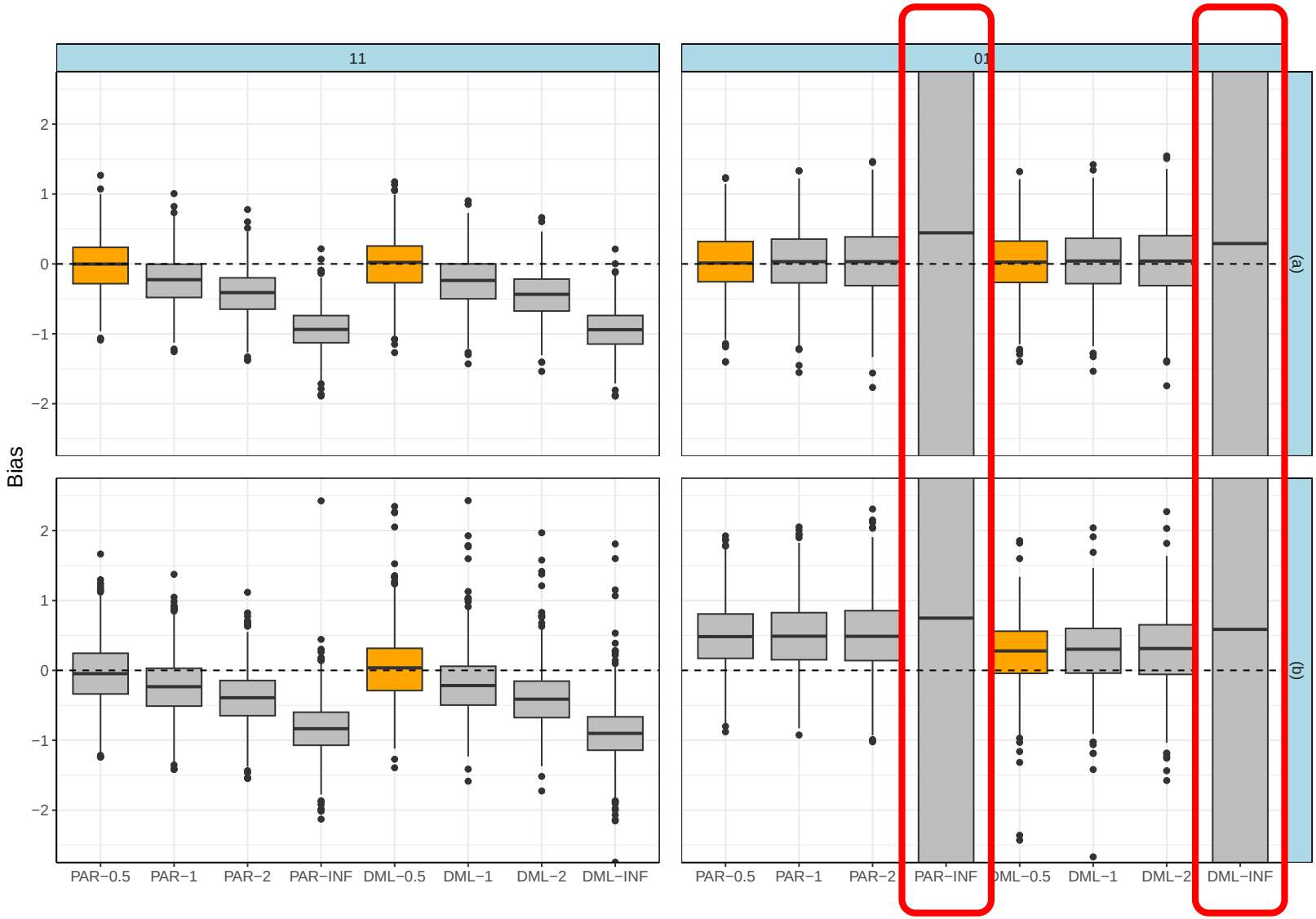


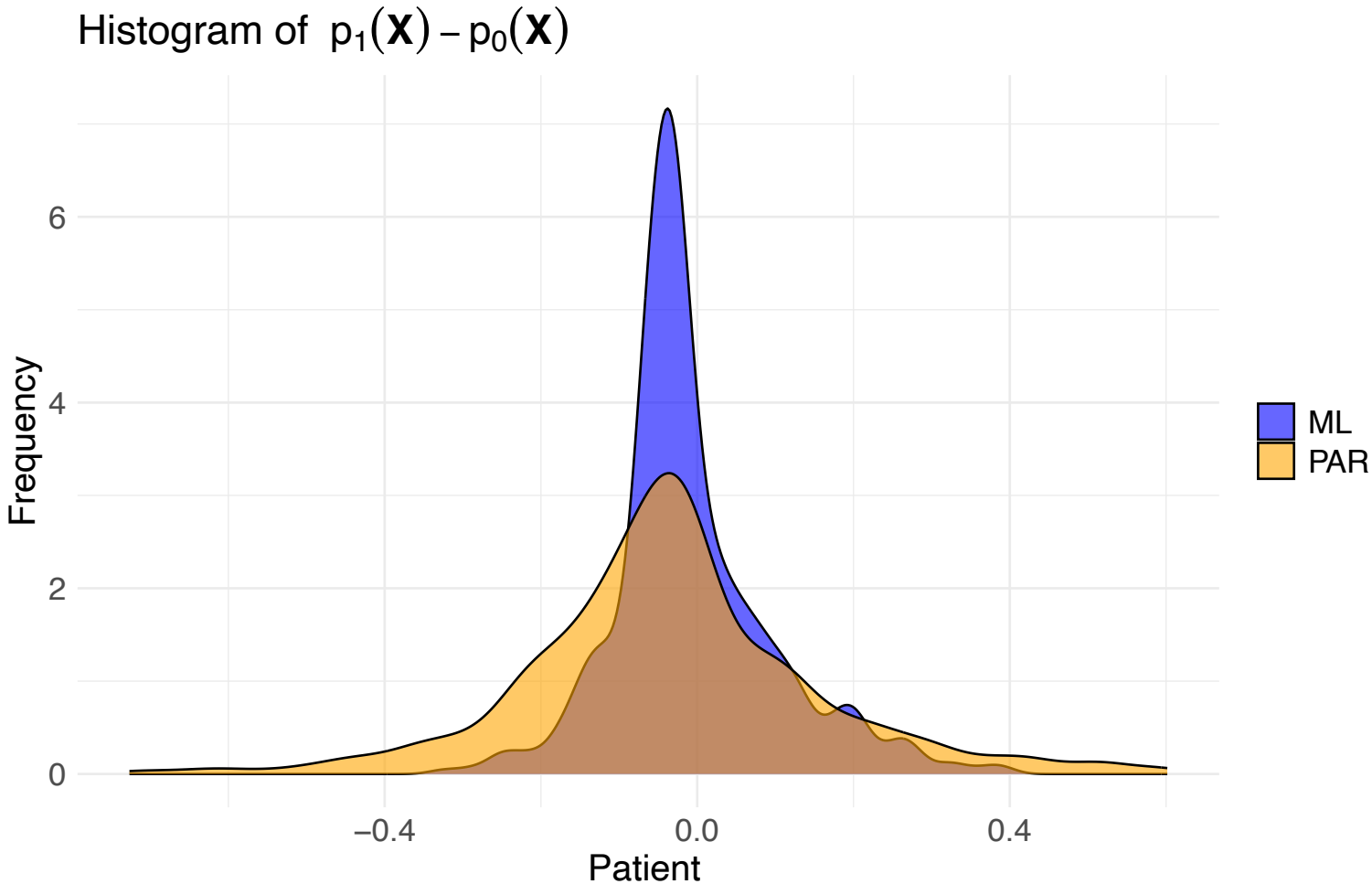
Figure 2: Scenarios (a) and (b) correspond to cases where all working models are either correctly specified or all incorrectly specified.

Applications to the ARDS trial

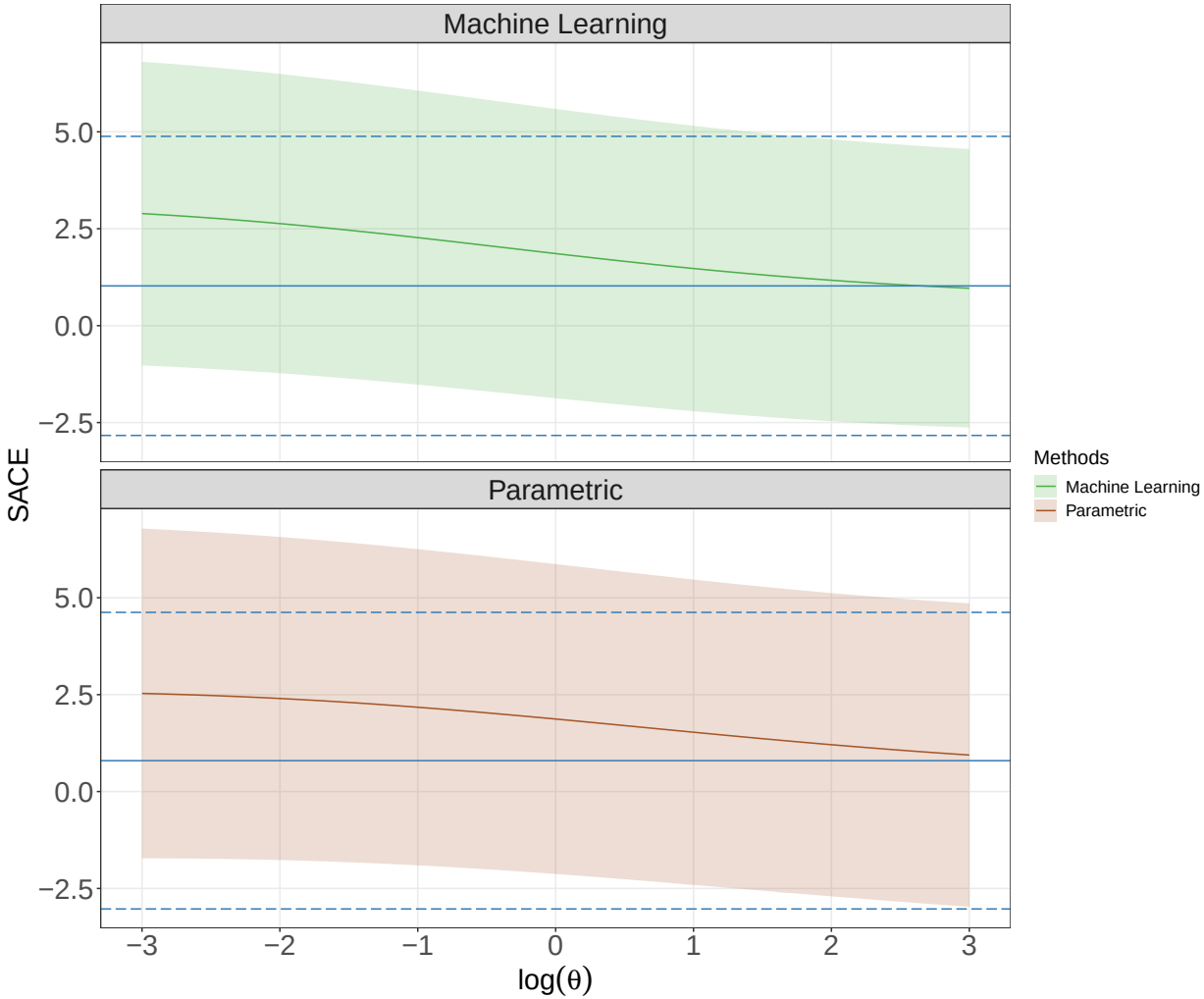
- **Context:** comparative effectiveness trial of PEEP levels in patients with acute lung injury and ARDS who are receiving mechanical ventilation
- **Outcome:** days to returning home (DTRH)—truncated by death
  - 549 patients in total; 68 patients in the lower PEEP arm and 76 patients in the higher PEEP arm died
  - interested in survivor average causal effect
- **Covariates:** an elaborate approach including 25 baseline covariates
  - attempt to capture $D - Y$ confounding and satisfy Assumption 4
- **Methods:** both parametric conditionally doubly robust and debiased machine learning estimators
  - assuming $\theta \in [0.05, 20]$; also compute estimates under monotonicity
  - constant θ as a practical strategy (simulations provide some support)

An empirical "check" of monotonicity

- $\widehat{e}_{01}(\mathbf{X}) = \widehat{p}_1(\mathbf{X}) - \widehat{p}_0(\mathbf{X})$ across all individuals, calculated using Super Learner ('ML') and parametric logistic regression ('PAR')
- if monotonicity holds, $e_{01}(\mathbf{X}) = p_1(\mathbf{X}) - p_0(\mathbf{X}) \geq 0$



Survivor average causal effects under $\log(\theta) \in [-3, 3]$



- Blue solid and dashed lines $\Rightarrow$ point estimates and CI assuming monotonicity
- Assumption on IEs affects point estimates not direction in this example

Discussion

- Going beyond monotonicity
  - require an association measure for $\{D(1), D(0)\}$
  - odds ratio is a strong candidate—margin free and relatively familiar
  - a constant assumption often suffices, but needs more empirical exploration
  - provide an R package `PSor` for implementation (continuing to develop)
- How to go beyond monotonicity is not the end of the story
  - we still require mean principal ignorability—an assumption that mimics what we need for analyzing observational data
  - we developed a confounding function approach for sensitivity analysis (Tong et al. 2025)
  - **ongoing work**: can we make it easier? E-value for principal ignorability

References I


Frangakis, Constantine E and Donald B Rubin (2002). "Principal stratification in causal inference". In: *Biometrics* 58.1, pp. 21–29.


Hayden, Douglas, Donna K Pauler, and David Schoenfeld (2005). "An estimator for treatment comparisons among survivors in randomized trials". In: *Biometrics* 61.1, pp. 305–310.


Jiang, Zhichao, Shu Yang, and Peng Ding (2022). "Multiply robust estimation of causal effects under principal ignorability". In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 84.4, pp. 1423–1445.


Kahan, Brennan C, Suzie Cro, Fan Li, and Michael O Harhay (2023). "Eliminating ambiguous treatment effects using estimands". In: *American Journal of Epidemiology* 192.6, pp. 987–994.


NHLBI ARDS Clinical Trials Network (2004). "Higher versus lower positive end-expiratory pressures in patients with the acute respiratory distress syndrome". In: *New England Journal of Medicine* 351.4, pp. 327–336.


Tong, Jiaqi, Brennan Kahan, Michael O Harhay, and Fan Li (2025). "Semiparametric principal stratification analysis beyond monotonicity". In: *Statistica Sinica*.


Vansteelandt, Stijn and Kelly Van Lancker (2024). "Chasing shadows: how implausible assumptions skew our understanding of causal estimands". In: *Statistics in Biopharmaceutical Research*, pp. 1–7.

Extra: Summary of properties

Table 1: A quick summary of multiple robustness properties of the semiparametric estimator across θ. $\mathcal{M}_\pi$ represents the propensity score model $\pi(\mathbf{X})$; $\mathcal{M}_{p_z}$ represents the principal score model $p_z(\mathbf{X})$; $\mathcal{M}_{m_{zd}}$ represents the outcome mean model $m_{zd}(\mathbf{X})$. ✓ means the correctly specified model.

odds ratio θ	$\mathcal{M}_\pi$	$\mathcal{M}_{p_0}$	$\mathcal{M}_{p_1}$	$\mathcal{M}_{m_{01}}$	$\mathcal{M}_{m_{11}}$
Monotonicity $\theta = \infty$; triply robust	✓	✓	✓		
		✓	✓	✓	✓
	✓			✓	✓
Independence $\theta = 1$; quadruply robust	✓	✓	✓		
		✓	✓	✓	✓
	✓	✓		✓	
	✓		✓		✓
$\theta \in (0, 1) \cup (1, \infty)$; conditionally doubly robust	✓	✓	✓		
		✓	✓	✓	✓

Extra: `PSor` R package

- For implementation, the `PSor` GitHub R package can be installed via `devtools::install_github("deckardt98/PSor")`.
- Below is a simple code demonstration assuming $\theta = 1.5$:

```
PSor.fit(
  out.formula = Y~X1+X2+X3+X4,
  ps.formula = D~X1+X2+X3+X4,
  pro.formula = Z~X1+X2+X3+X4,
  df = df,
  out.name = "Y",
  int.name = "D",
  trt.name = "Z",
  cov.names = c("X1","X2","X3","X4"),
  or = 1.5,
  SLmethods = c("SL.glm", "SL.rpart", "SL.nnet"),
  n.fold = 5,
  scale = "RD",
  alpha = 0.05
)
```

```
>                CDR.Est CDR.SE CDR.ci.low CDR.ci.up DML.Est DML.SE DML.ci.low DML.ci.up
> Always-Takers    2.137  0.317      1.515     2.759   2.136  0.332      1.485     2.786
> Compliers       -0.183  0.468     -1.101     0.735  -0.186  0.481     -1.128     0.757
> Never-Takers    -3.098  0.377     -3.837    -2.359  -3.053  0.368     -3.774    -2.331
> Defiers          0.023  0.424     -0.808     0.853  -0.088  0.442     -0.955     0.778
```