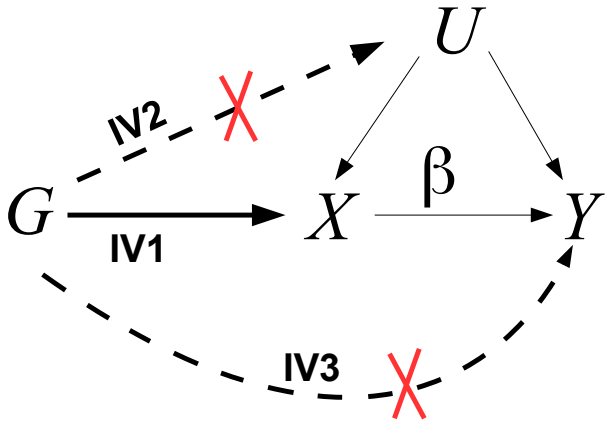
Triangulation Within a Single Study (TWIST): Combining Mendelian randomization, Confounder adjustment and Pharmacogenetics

Jack Bowden
@jack_bowdenjack,j.bowden2@exeter.ac.uk

July 2, 2021

MR uses genes as Instrumental Variables



1. IV1: G associated with X
2. IV2: G independent of U
3. IV3: G independent of Y given X, U
4. IV4: Homogeneity of effect

IV overcomes bias due to confounding (standard observational analysis)

X could equal modifiable exposure (BMI) or a treatment (drug)

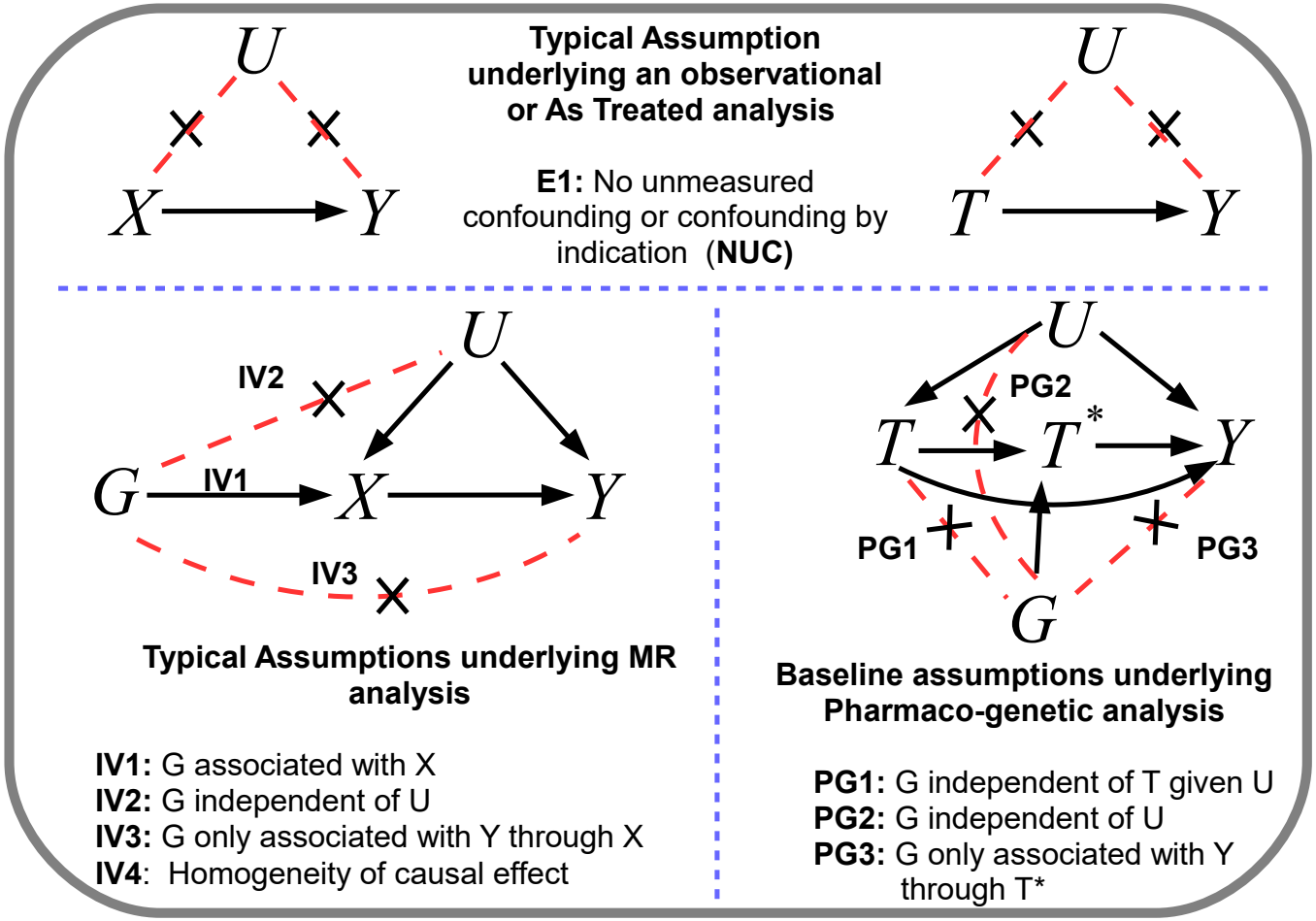
But what if we *a priori* do not believe IV4?

Specifically, what if there is genetically driven effect heterogeneity?

Canonical Example: Clopidogrel, CYPC219 and stroke

- Clopidogrel primary drug for stroke prevention in the UK
- Requires *CYP2C19* enzyme activation in order to be properly metabolised and thus work to its fullest extent
- Common loss-of-function variant within the *CYP2C19* gene massively impacts ability to metabolise the drug
- We may want to know *'How much would the population benefit if everyone could receive the full effect of treatment?'*
- Call this the genetically mediated treatment effect (GMTE)
- Treatment T therefore our exposure

Observational, MR and Pharmacogenetic causal inference



- We want to estimate the $T^* \rightarrow Y$ path
- $T \rightarrow Y$ path accounts for treatment effect heterogeneity

The standard pharmacogenetic approach

- Suppose assumptions PG1-PG3 hold
- Assume binary treatment and genotype
- The GMTE can be estimated using **only** people who are treated with Clopidogrel
- The GMTE(1) estimate is

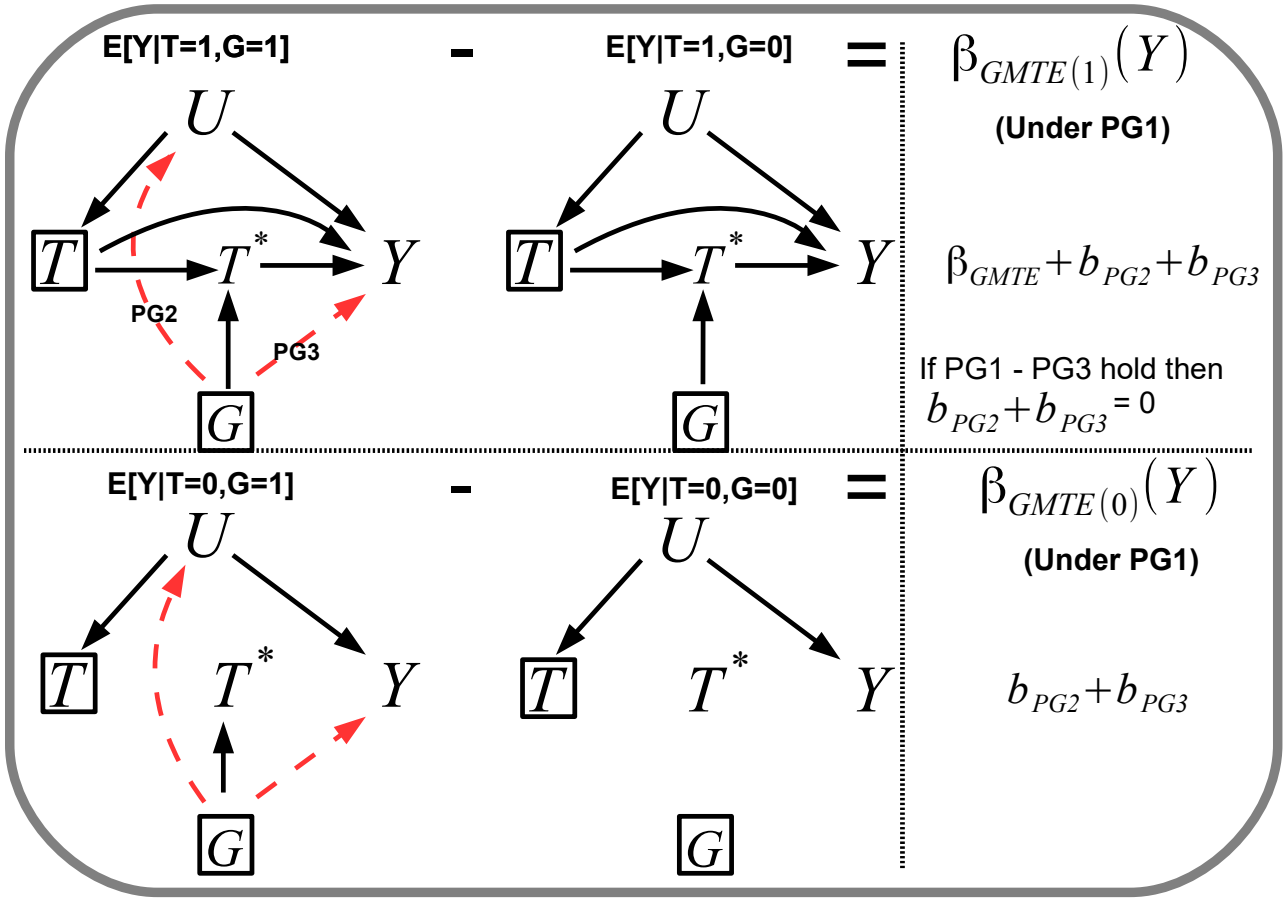
$$\hat{\beta}_{GMTE(1)}(Y) = \hat{E}[Y|T = 1, G = 1] - \hat{E}[Y|T = 1, G = 0]$$

- The mean/risk/hazard difference between genetic groups in the treated population

Problems with the basic approach

- Suppose that assumptions PG1-PG3 are violated
- **Poss. PG1 violation**: If LoF carriers have an increased risk of side effects on treatment and choose to immediately come off the drug..
- **Poss PG2 violation**: LoF variant associated with a risk factor for Stroke, which then increases their likelihood of being treated
- **Poss: PG3 violation**: The LoF variant affects stroke through independent pathway
- Standard estimate will reflect the genetically mediated effect of treatment plus the bias due to PG1-3 violation

The Robust GMTE (RGMTE) under violation of PG2-PG3



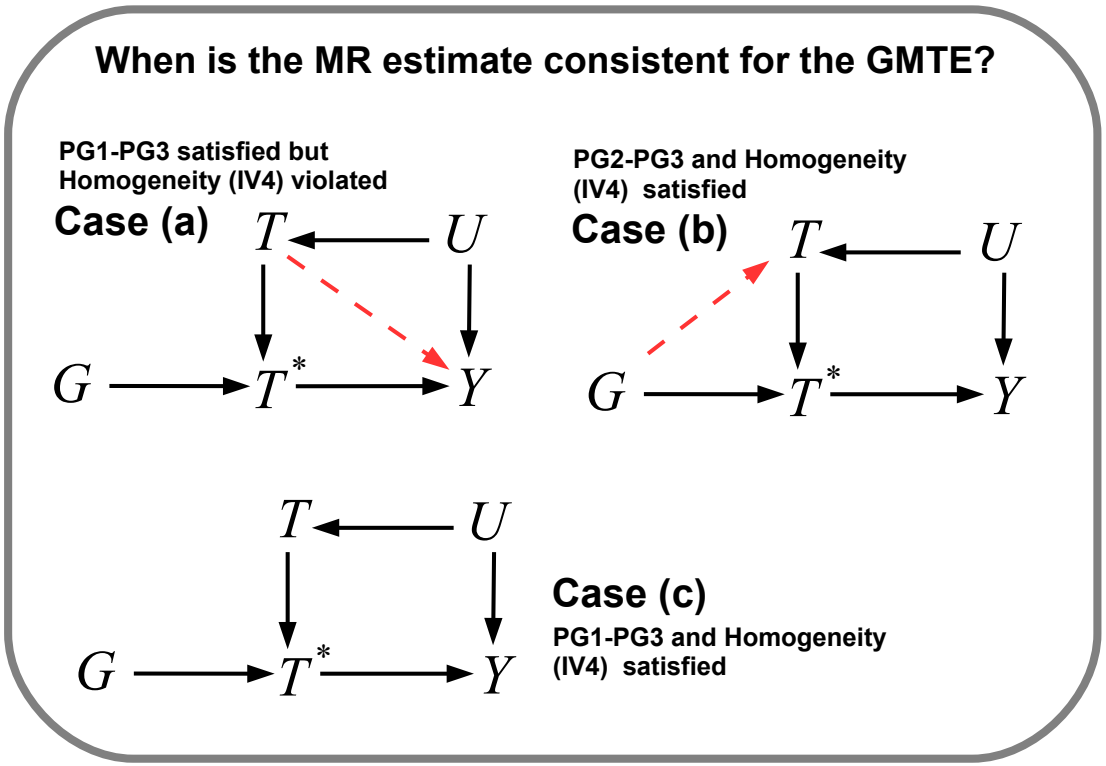
$$\hat{\beta}_{RGMTE}(Y) = \hat{\beta}_{GMTE(1)}(Y) - \hat{\beta}_{GMTE(0)}(Y)$$

The GMTE(0) estimate should be zero if PG2-3 satisfied

This is like a Difference-in-Difference estimator

When is the MR estimate consistent for the GMTE?

$$\hat{\beta}_{MR}(Y) = \frac{E[Y|G=1] - E[Y|G=0]}{E[T^*|G=1] - E[T^*|G=0]}$$

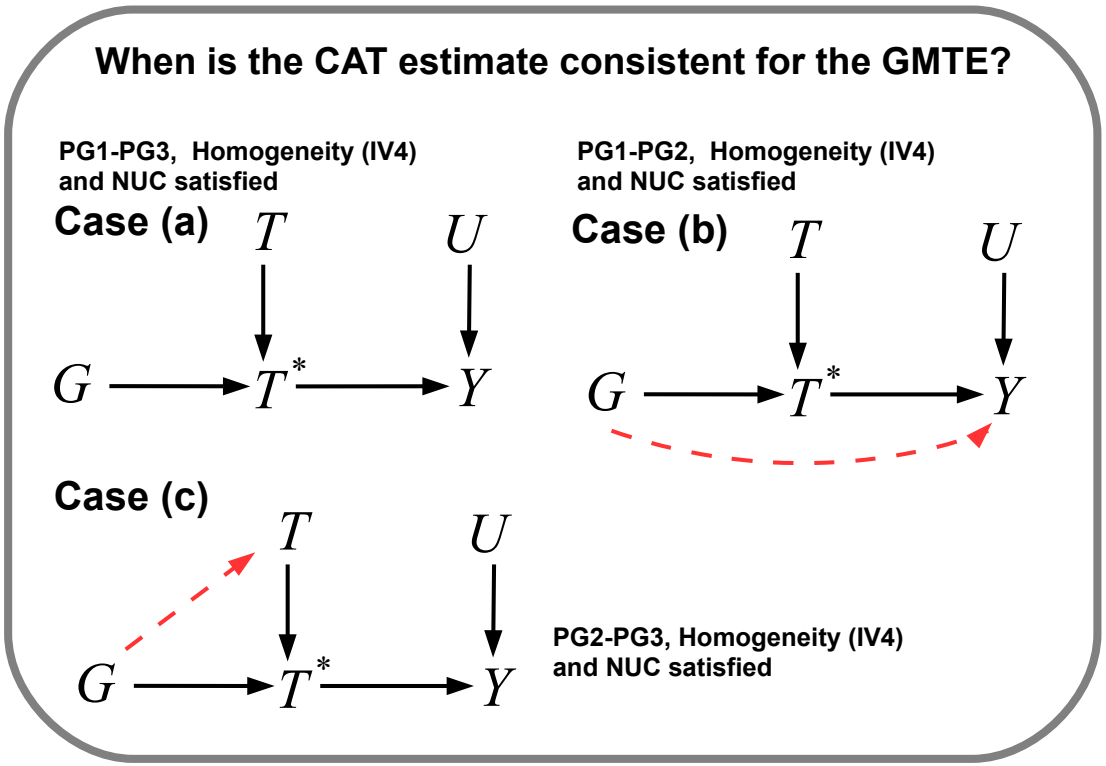


- PG1 violation not a problem because we don't condition on T !
- PG1 & Homogeneity violation is

When is the Corrected 'As Treated' estimate consistent for the GMTE?

- The 'As Treated' estimate will be strongly confounded
- But a scaled version of the estimate can in principle be unbiased

$$\hat{\beta}_{CAT}(Y) = \frac{\hat{E}[Y|T=1] - \hat{E}[Y|T=0]}{E[G|T=1]}$$



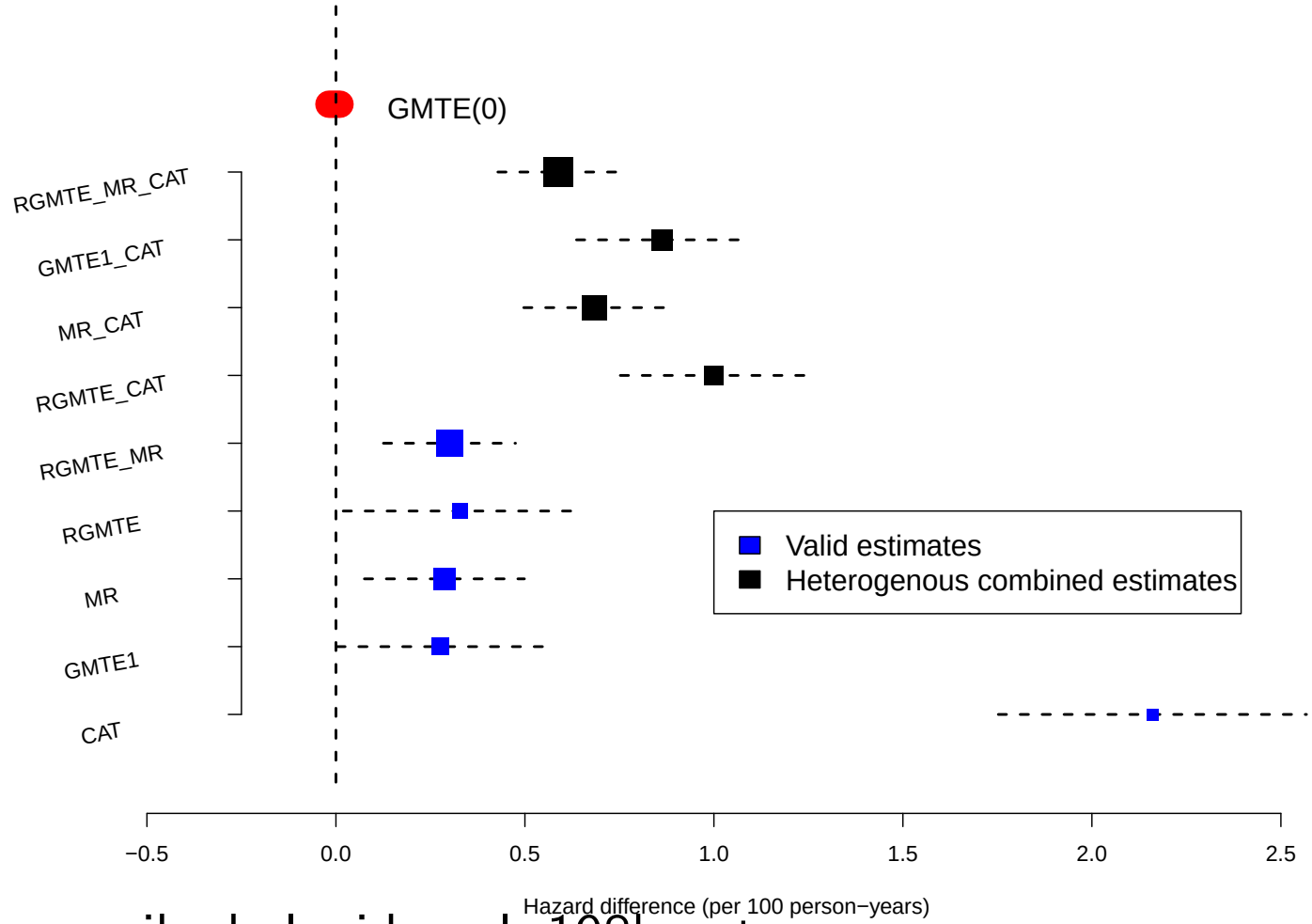
- Requires NUC & Homogeneity assumptions

The most suprising thing of all!

$$Cov(CAT, GMTE(1), RGMTE, MR) = \begin{pmatrix} \sigma^2_{CAT} & 0 & 0 & 0 \\ 0 & \sigma^2_{GMTE(1)} & a & b \\ 0 & a & \sigma^2_{RGMTE} & 0 \\ 0 & b & 0 & \sigma^2_{MR} \end{pmatrix}$$

- Many estimates are independent of each other!
- Pairs or triples of estimates can be combined if sufficiently 'similar'
- Heterogeneity test statistic used to assess this
- Valid estimates should be similar, invalid estimates are hopefully different
- 4 single, 4 two-way & 1 three-way estimates for the GMTE

Application to CPRD-UKB data (n=207,000)

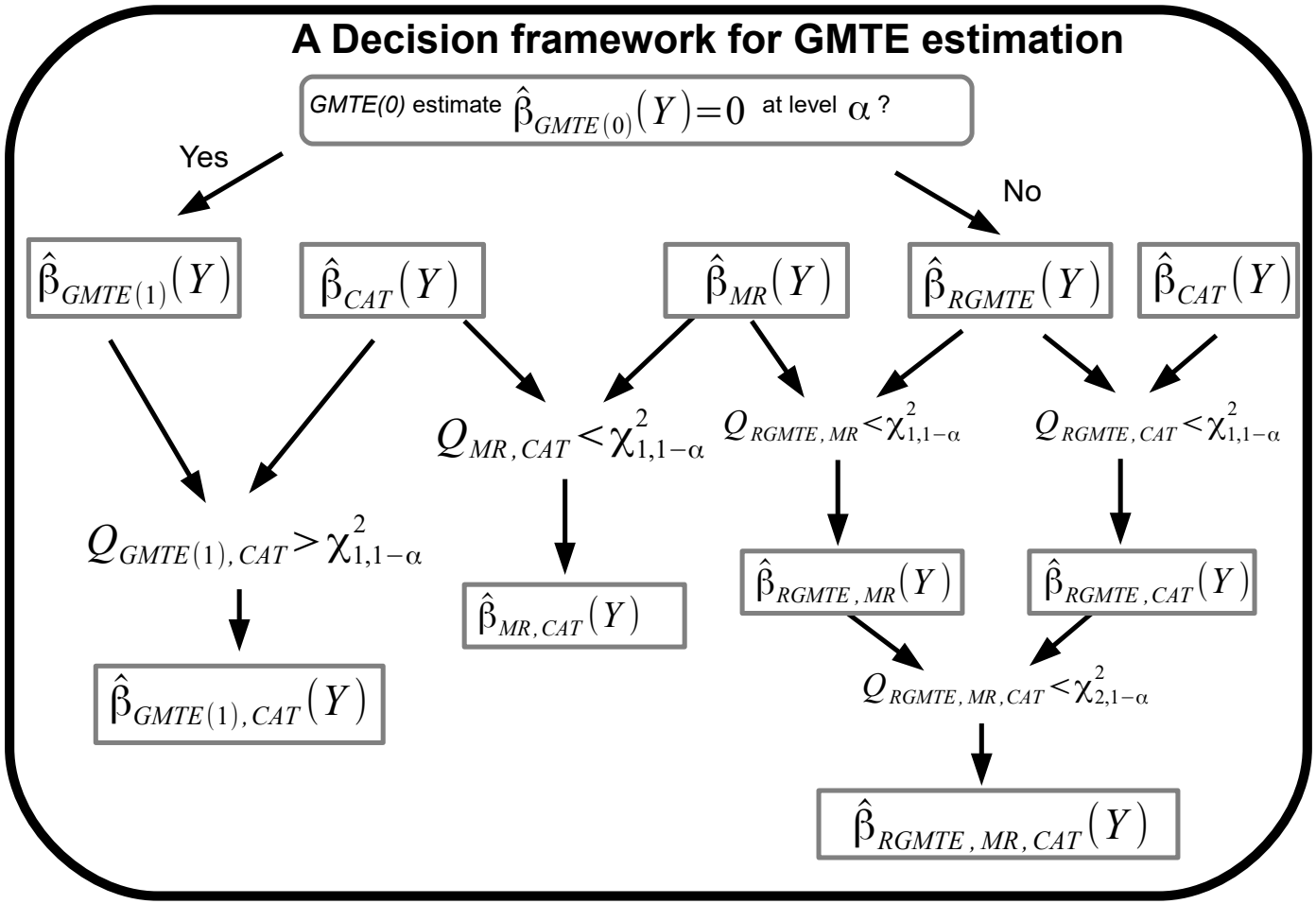


- 8k prescribed clopidogrel, 198k not
- GMTE(0) estimate centred strongly on zero (reassuring)
- Most efficient estimate is the combined MR/RGMTE estimate
- Results suggests 13.7% fewer strokes if G=0 group switched

Summary and Future work

- Triangulation of evidence from different data sources is becoming a well established principle for strengthening causal analysis
- We call this framework '**T**riangulation **WI**thin a **ST**udy' (TWIST)' in order to emphasise that an analysis in this spirit is also possible within a single data set, using causal estimators that are statistically independent, but reliant on a different sets of assumptions.
- Comments or suggestions welcome
- Thanks for listening! Pre-print out there!

9 potentially valid single and combined estimates for the GMTE



- 4 single, 4 two-way & 1 three-way estimates for the GMTE
- Valid estimates should be similar, invalid estimates are hopefully different

The GMTE(1) estimate applied to CPRD-UKB data

Variable	Prescribed Clopidogrel (n=7,483)	Never Prescribed Clopidogrel n = 198,868
Age at recruitment	61.4± 6.2	56.5±8.0
Age at first prescription	64.1±7.3	-
Sex(Female%)	2,559(34.2%)	110,569(55.6%)
CYPC219 LoF carrier	2,145(28.7%)	56,043(28.2%)
Incident Ischemic Stroke diagnosis*	110 (1.5%)	2,078(1.0%)
Incident MI diagnosis*	1,822 (24.8%)	13,796(6.9%)

- Time to stroke modelled using an Aalen additive hazards model, adjusted for age, sex and 10 Genetic PCs
- Being a CYP2C19 LoF carrier increases the risk of stroke by 0.28% per year (p=0.048)
- If we could reduce the LoF carrier's risk by this amount, then summed over the 5264 patient years in the data, it would lead to a 13.2% reduction in the number of strokes (from 110 to 95)
- But what if PG1-PG3 are not satisfied?