The Stability-Controlled Trial and Quasi-Experiment?
Learning effects of new treatments without randomization (or ignorability)

Chad Hazlett[1], Werner Maokola[2], David Ami Wulf[3]

[1]UCLA Statistics and Political Science; chazlett@ucla.edu

[2]Research Coordinator, National AIDS Control Program; Department of Epidemiology, Kilimanjaro Christian Medical University College, Moshi, Tanzania

[3]UCLA Statistics; amiwulf@gmail.com

Beginning in 2011, Tanzania's Ministry of Health has pushed for the use of isoniazid preventive therapy (IPT) to prevent tuberculosis among patients living with HIV.

Beginning in 2011, Tanzania's Ministry of Health has pushed for the use of isoniazid preventive therapy (IPT) to prevent tuberculosis among patients living with HIV.

Before offering IPT in clinics, 13% of patients with HIV developed TB.

Beginning in 2011, Tanzania's Ministry of Health has pushed for the use of isoniazid preventive therapy (IPT) to prevent tuberculosis among patients living with HIV.

Before offering IPT in clinics, 13% of patients with HIV developed TB.

After IPT was introduced, TB incidence was:

Beginning in 2011, Tanzania's Ministry of Health has pushed for the use of isoniazid preventive therapy (IPT) to prevent tuberculosis among patients living with HIV.

Before offering IPT in clinics, 13% of patients with HIV developed TB.

After IPT was introduced, TB incidence was:

- **16% among non-IPT users**, and

Beginning in 2011, Tanzania's Ministry of Health has pushed for the use of isoniazid preventive therapy (IPT) to prevent tuberculosis among patients living with HIV.

Before offering IPT in clinics, 13% of patients with HIV developed TB.

After IPT was introduced, TB incidence was:

- **16% among non-IPT users**, and
- **1% among IPT users**

Beginning in 2011, Tanzania's Ministry of Health has pushed for the use of isoniazid preventive therapy (IPT) to prevent tuberculosis among patients living with HIV.

Before offering IPT in clinics, 13% of patients with HIV developed TB.

After IPT was introduced, TB incidence was:

- **16% among non-IPT users**, and
- **1% among IPT users**

Can we do better than to just warn "this is only suggestive"?

Randomization is great, but...

- RCTs may be infeasible: Treatments that cannot be excluded; implementers who do not want to randomize.

Randomization is great, but...

- RCTs may be infeasible: Treatments that cannot be excluded; implementers who do not want to randomize.
- RCTs may not measure what you want: Stylized treatment; restrictive eligibility; different population than ultimate treatment-choosers.

Randomization is great, but...

- RCTs may be infeasible: Treatments that cannot be excluded; implementers who do not want to randomize.
- RCTs may not measure what you want: Stylized treatment; restrictive eligibility; different population than ultimate treatment-choosers.
- RCTs may be unethical: Assigns/denies treatment based on research aims, not what is best for individual. May no longer be ethical if credible alternative exists.

The SCT and SCQE (Hazlett, 2018)

The Stability-Controlled Trial (SCT) and Quasi-Experiment (SCQE) can estimate the ATT for newly available treatments

The SCT and SCQE (Hazlett, 2018)

The Stability-Controlled Trial (SCT) and Quasi-Experiment (SCQE) can estimate the ATT for newly available treatments

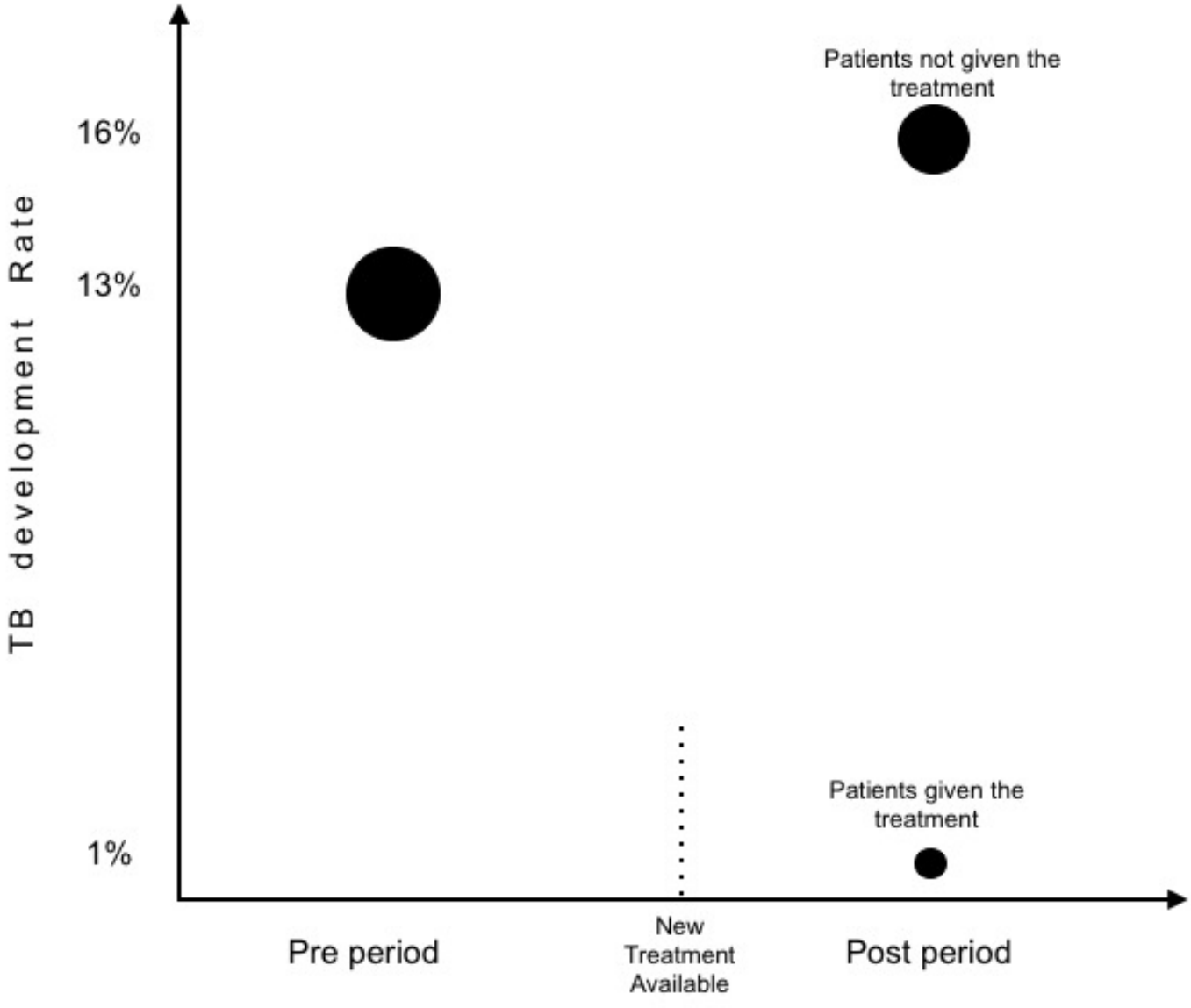
- Does not require: randomization, any knowledge of treatment assignment, conditional ignorability

The SCT and SCQE (Hazlett, 2018)

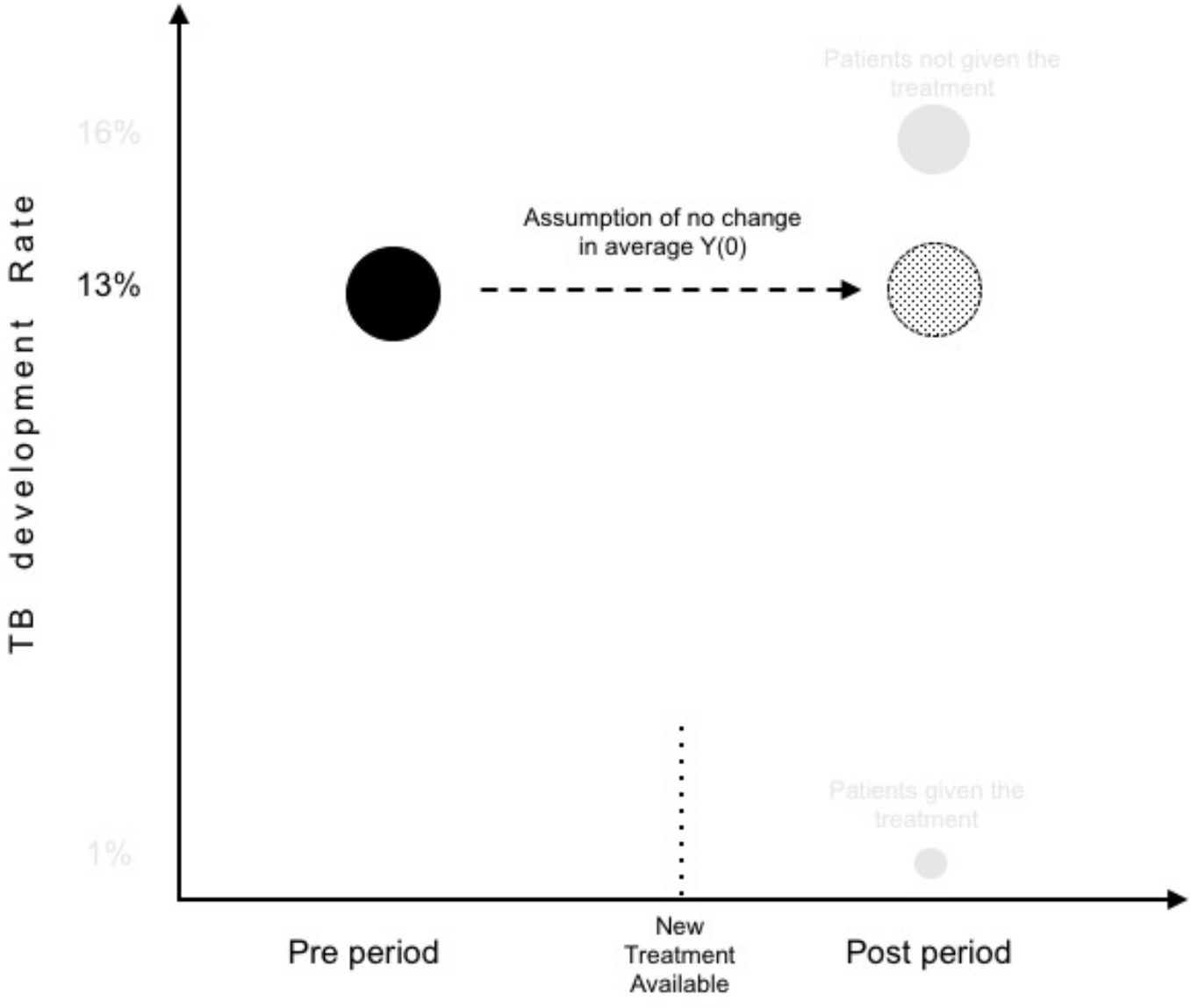
The Stability-Controlled Trial (SCT) and Quasi-Experiment (SCQE) can estimate the ATT for newly available treatments

- Does not require: randomization, any knowledge of treatment assignment, conditional ignorability
- Does require: One assumption, on how non-treatment outcome would have changed over time.

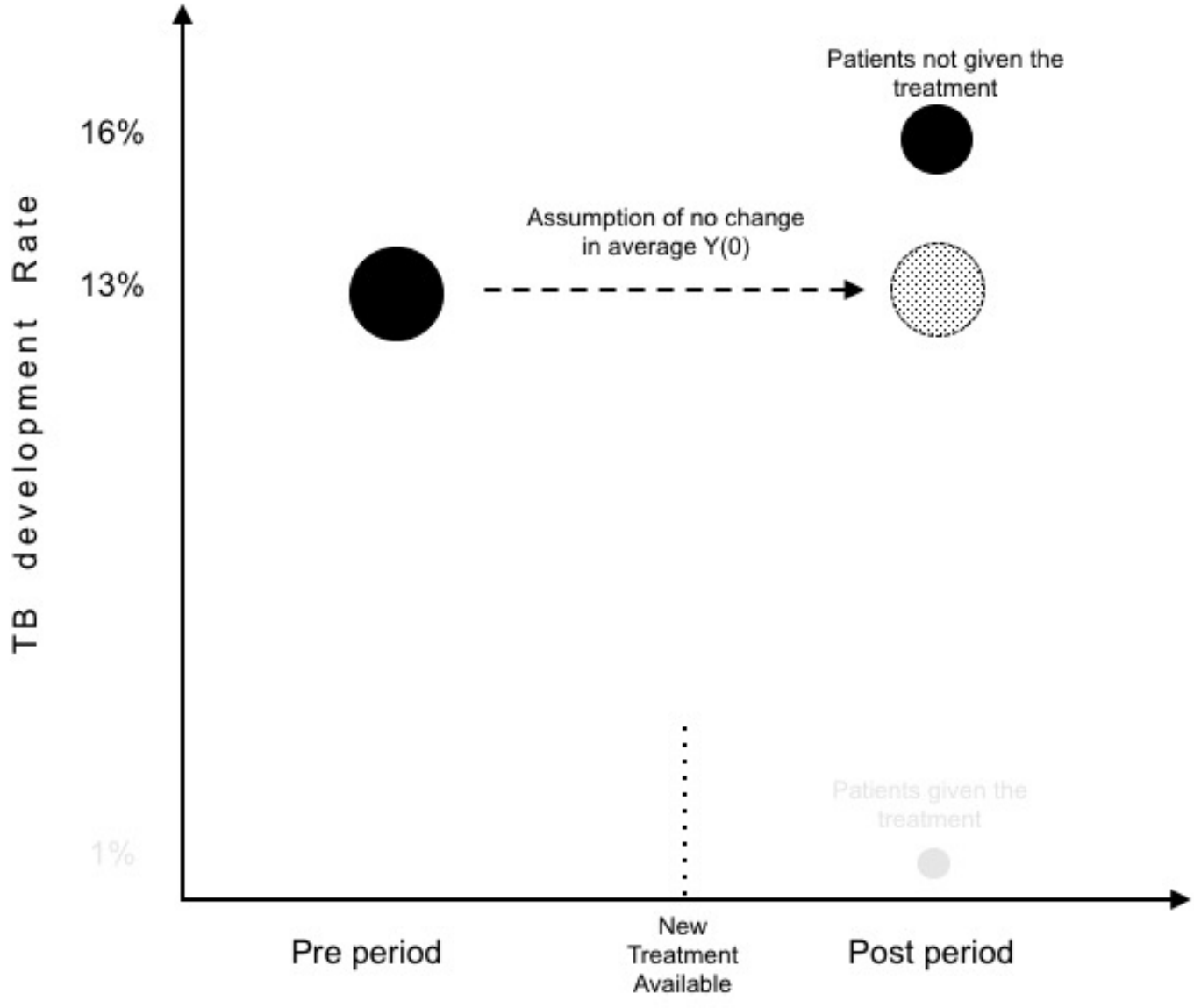
How it works: start with four observables



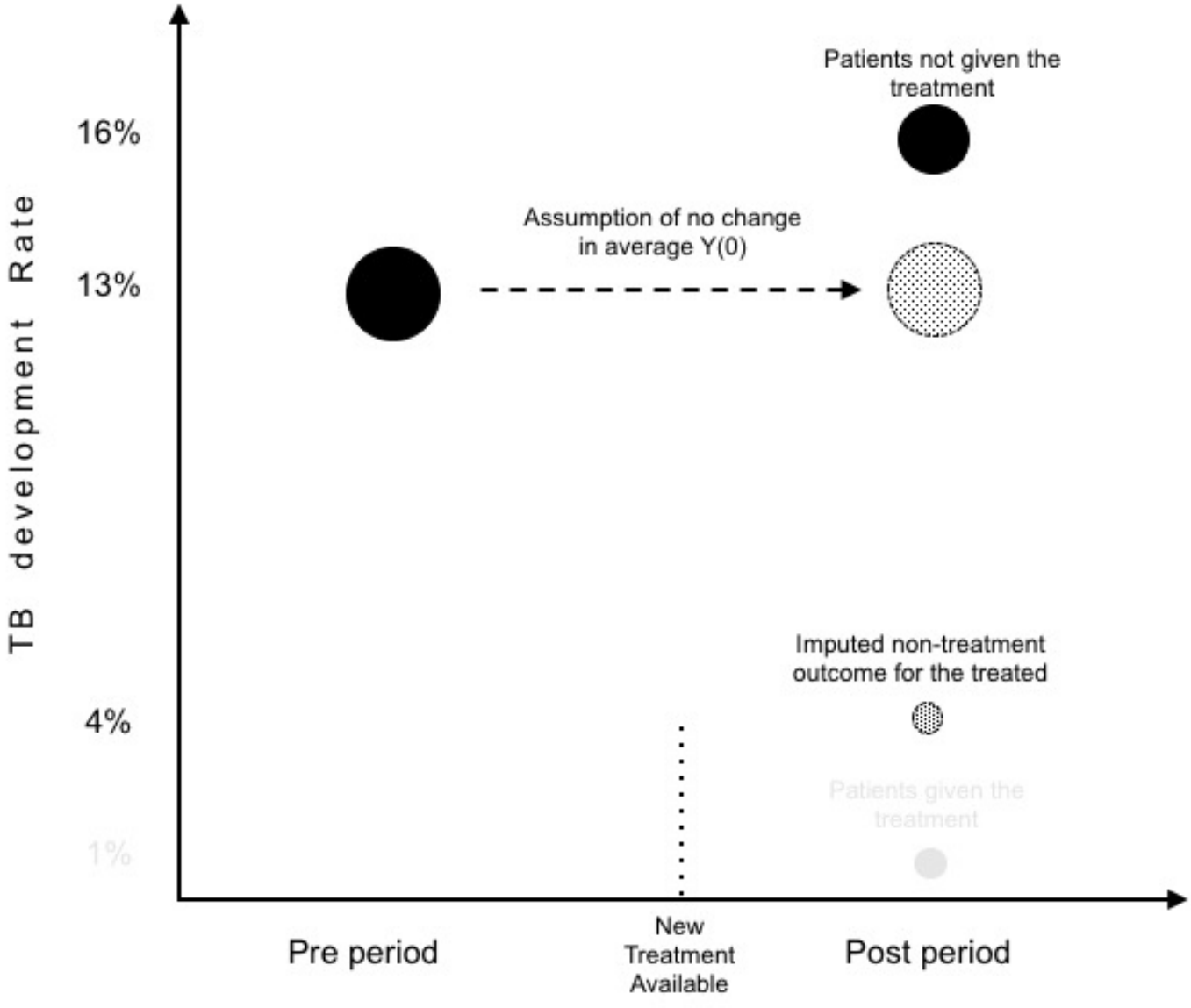
How it works: 1. Add assumption on change in $\mathbb{E}[Y(0)]$ over time



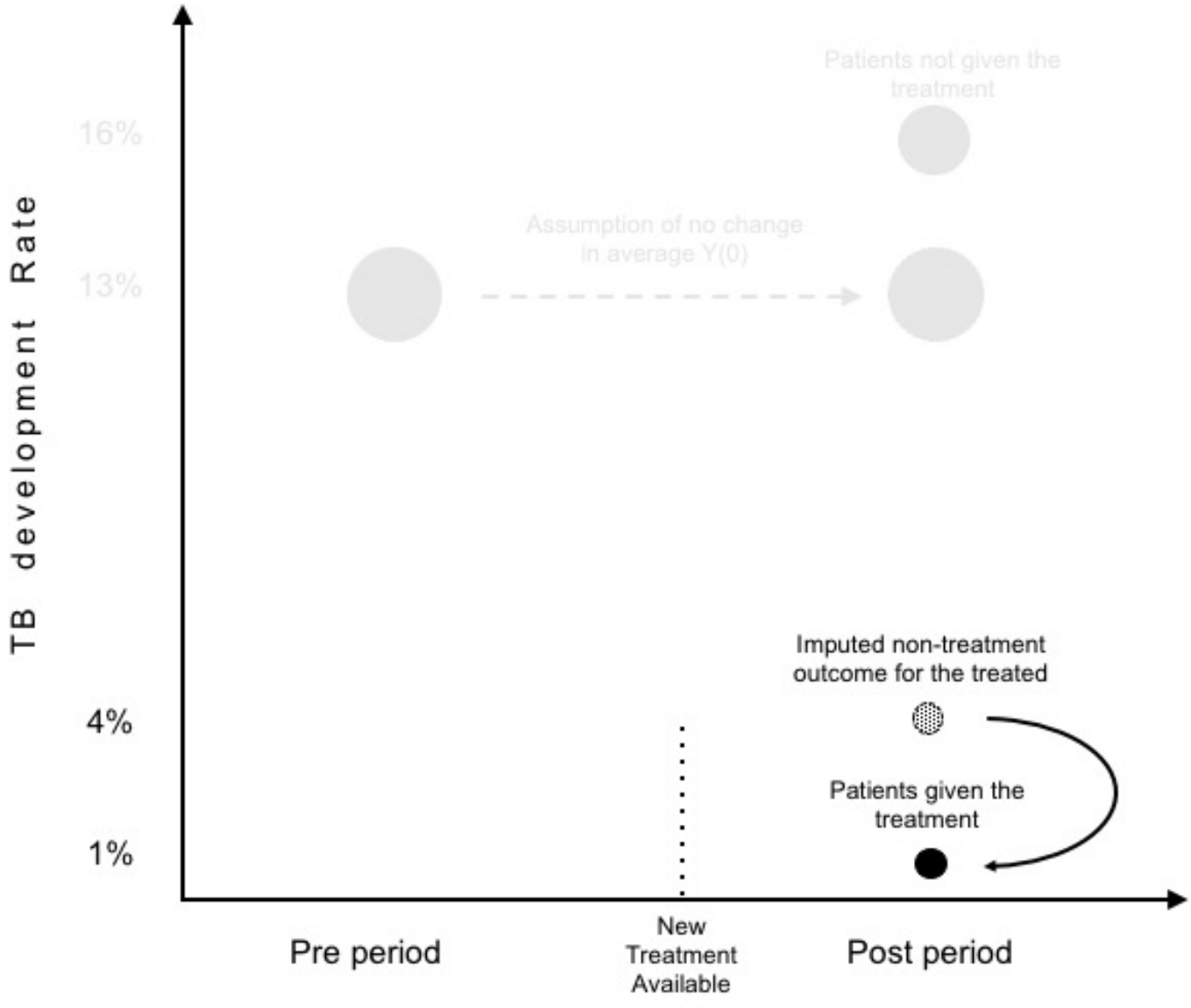
How it works: 2. The LIE doesn't lie



How it works: 2. the LIE doesn't lie



How it works: 3. Compare



In other words,

Define the change in non-treatment average outcomes,

$$\delta \equiv \mathbb{E}[Y(0)|T=1] - \mathbb{E}[Y(0)|T=0], \tag{1}$$

In other words,

Define the change in non-treatment average outcomes,

$$\delta \equiv \mathbb{E}[Y(0)|T=1] - \mathbb{E}[Y(0)|T=0], \tag{1}$$

Which identifies $\mathbb{E}[Y(0)|D=1,T=1]$,

$$\begin{aligned}\mathbb{E}[Y(0)|D=1,T=1] &= \frac{\mathbb{E}[Y(0)|T=0] - \mathbb{E}[Y(0)|D=0,T=1](1-\pi_1)+\delta}{\pi_1} \\ &= \frac{\mathbb{E}[Y|T=0] - \mathbb{E}[Y|D=0,T=1](1-\pi_1)+\delta}{\pi_1}\end{aligned} \tag{2}$$

In other words,

Define the change in non-treatment average outcomes,

$$\delta \equiv \mathbb{E}[Y(0)|T=1] - \mathbb{E}[Y(0)|T=0], \tag{1}$$

Which identifies $\mathbb{E}[Y(0)|D=1,T=1]$,

$$\begin{aligned}\mathbb{E}[Y(0)|D=1,T=1] &= \frac{\mathbb{E}[Y(0)|T=0] - \mathbb{E}[Y(0)|D=0,T=1](1-\pi_1)+\delta}{\pi_1} \\ &= \frac{\mathbb{E}[Y|T=0] - \mathbb{E}[Y|D=0,T=1](1-\pi_1)+\delta}{\pi_1}\end{aligned} \tag{2}$$

Yielding the ATT,

$$\begin{aligned}ATT &= \mathbb{E}[Y(1)|D=1,T=1] - \mathbb{E}[Y(0)|D=1,T=1] \\ &= \mathbb{E}[Y|D=1,T=1] - \left(\frac{\mathbb{E}[Y|T=0] - \mathbb{E}[Y|D=0,T=1](1-\pi_1)+\delta}{\pi_1}\right).\end{aligned} \tag{3}$$

Comparison to IV

This is IV with "time as the instrument", and some twists:

- δ allows a prescribed deviation from the exclusion restriction
- encouraged to give answer conditionally on δ, rather than producing estimate that is correct only if exclusion true
- we don't want covariates to buoy assumptions, just δ

Comparison to IV

This is IV with "time as the instrument", and some twists:

- δ allows a prescribed deviation from the exclusion restriction
- encouraged to give answer conditionally on δ, rather than producing estimate that is correct only if exclusion true
- we don't want covariates to buoy assumptions, just δ

The IV equivalence usefully reminds us:

- only need a shift in probability of treatment, not totally new treatment (for LATE instead of ATT).
- existing standard error estimators

Comparison to Difference-in-Difference (DID)

Biggest difference compared to DID is where you can use it:

- Both cross-sectional and panel versions of DID require labeling each individual as "would be treated" or not.
- However SCQE works where you have a pre-treatment cohort for whom you cannot say who would have later been treated. E.g. a new medication or policy or media treatment.

Comparison to Difference-in-Difference (DID)

Biggest difference compared to DID is where you can use it:

- Both cross-sectional and panel versions of DID require labeling each individual as "would be treated" or not.
- However SCQE works where you have a pre-treatment cohort for whom you cannot say who would have later been treated. E.g. a new medication or policy or media treatment.

When you can do DID, it is a special case of SCQE:

- Parallel trends: the two groups have the same trend in $\mathbb{E}[Y(0)]$
- SCQE: There exists an average trend over the two groups, δ
- The connection: SCQE is DID if you (i) learn the trend from the controls, and (ii) assume parallel trends
- SCQE thus gives alternative route to identification or sensitivity of DID.

Coming up with δ?

Returning to the effect of IPT on TB,

Coming up with δ?

Returning to the effect of IPT on TB,

1. Domain knowledge. Prior to data analysis, Dr. Maokola registered:
   - no known other reasons for change in TB incidence
   - except that reporting may be increasing: guesses rise of 0.5 to 1 percentage point per year due to improved reporting.

Coming up with δ?

Returning to the effect of IPT on TB,

1. Domain knowledge. Prior to data analysis, Dr. Maokola registered:
   - no known other reasons for change in TB incidence
   - except that reporting may be increasing: guesses rise of 0.5 to 1 percentage point per year due to improved reporting.

2. Informed by data. δ not identifiable, but if no other major changes, the existing trends may be informative
   - linear model on non-treated periods/clinics. -0.3 [-.1, -0.5].
   - exponential model; annual decay rate of 0.93 [0.89, 0.97]).

Results

If you believe in a specific δ or range, great, use it.

- E.g. At $\delta = 0$, then the ATT is -3 [-12, 8] percentage points

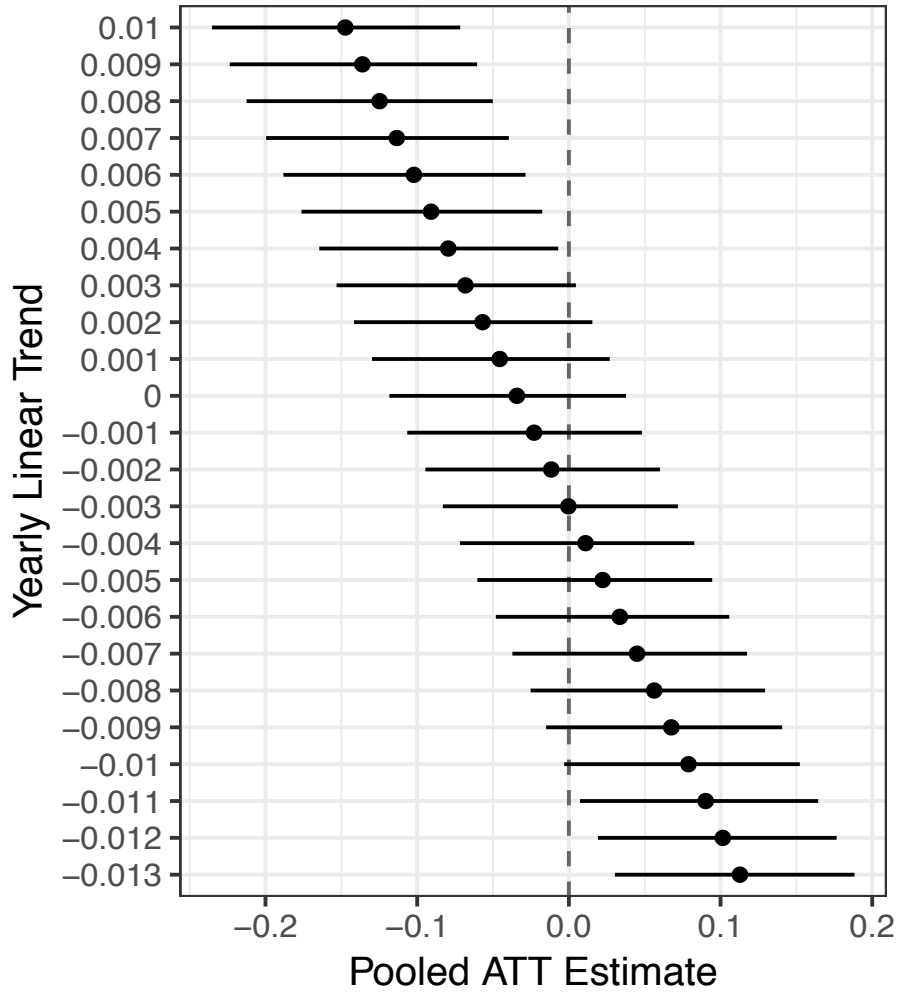
Results

If you believe in a specific δ or range, great, use it.

- E.g. At $\delta = 0$, then the ATT is -3 [-12, 8] percentage points

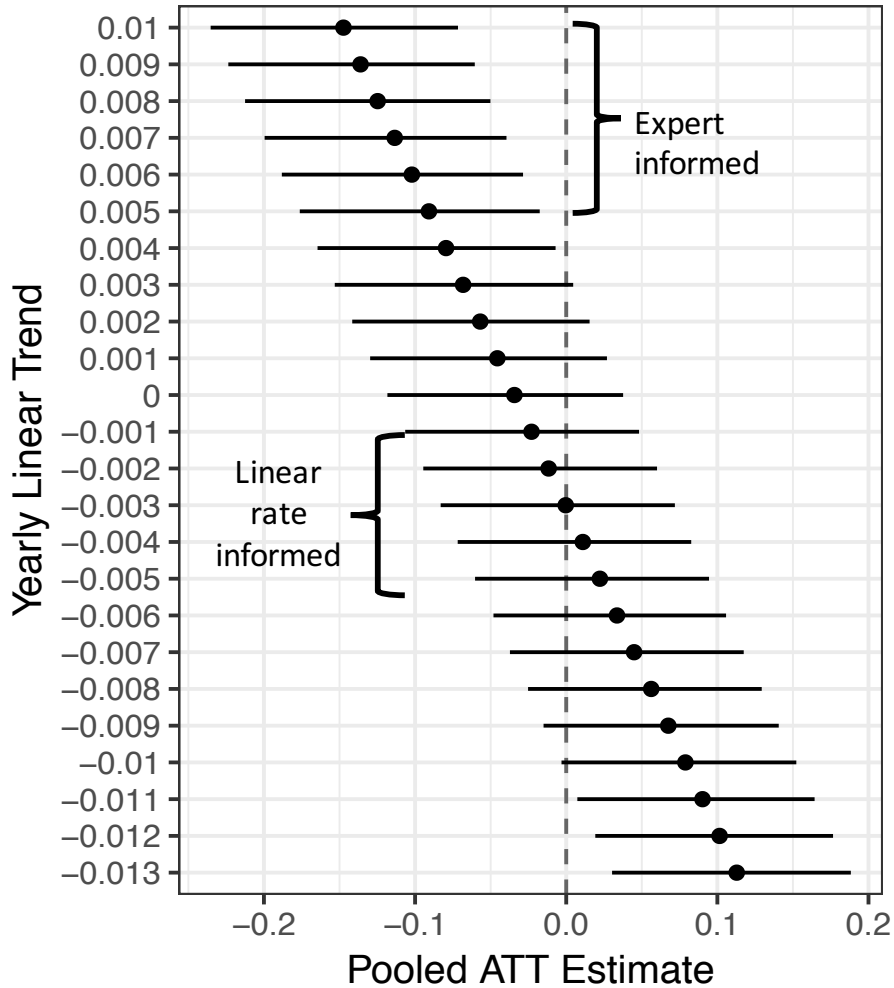
Or if wide range of believable δ, just look:

ATTs as function of δ



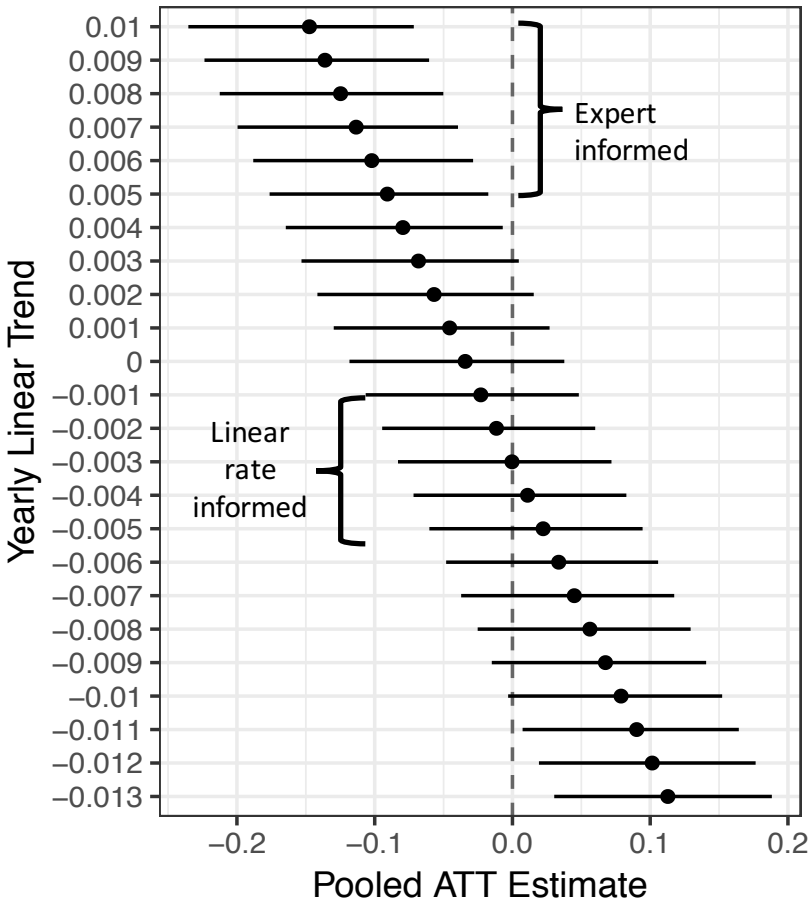
95% CI by block boostrap over clinics

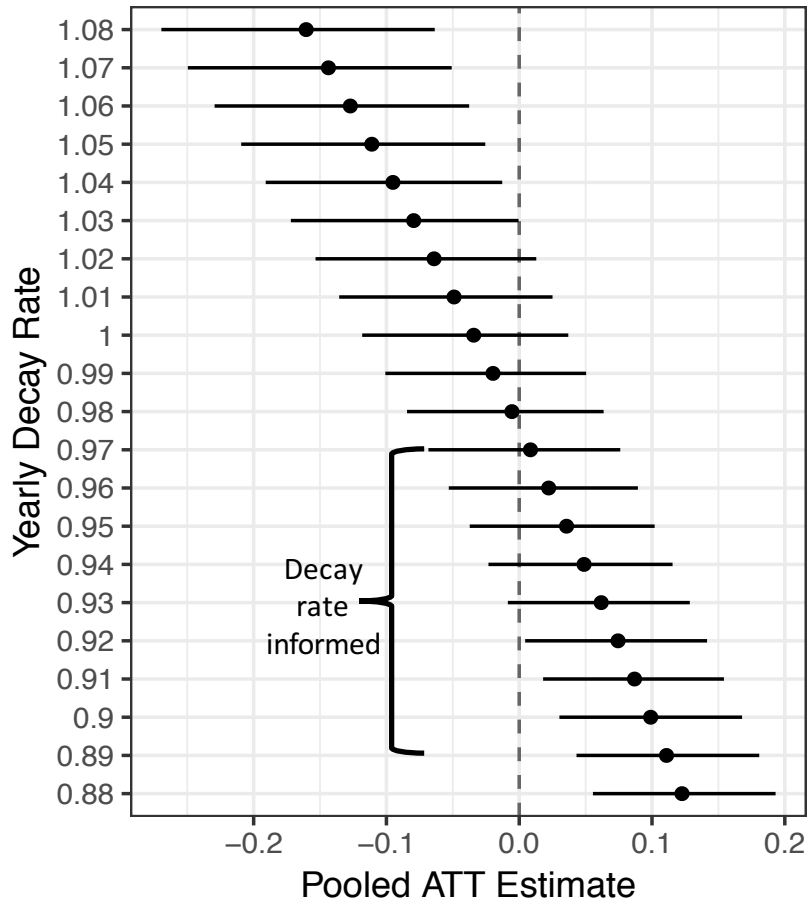
ATTs as function of δ



95% CI by block boostrap over clinics

ATTs as function of δ





A side benefit: Who takes IPT?

This also sheds light on who takes treatment, not in terms of their X but their $Y(0)$.

- pre-IPT TB average $Y(0)$ was 13%, compared to 16% for controls after IPT
- this signals that, as long $\delta \leq 3\%$ per year, those taking IPT were already "better off'
- may be useful in understanding selection and improving policy

Conclusions 1: SCQE in General

Identification of ATT for the price of an assumption, δ, without randomization or conditional ignorability.

Conclusions 1: SCQE in General

Identification of ATT for the price of an assumption, δ, without randomization or conditional ignorability.

Assumption on δ is easy to understand: "baseline trend"

- Where narrow δ is defensible, can produce sharp result.
- Where δ hard to know, broader possible results – but so be it!

Conclusions 1: SCQE in General

Identification of ATT for the price of an assumption, δ, without randomization or conditional ignorability.

Assumption on δ is easy to understand: "baseline trend"

- Where narrow δ is defensible, can produce sharp result.
- Where δ hard to know, broader possible results – but so be it!

The SCQE and SCT may be useful in many cases to go beyond observational claims while being rigorous.

Conclusions 2: IPT through the SCQE lens

Recall: **16% TB without IPT vs. 1% with it.**

Conclusions 2: IPT through the SCQE lens

Recall: **16% TB without IPT vs. 1% with it**.

- For the statisticians to warn "this is only suggestive" fails to tell the whole story,

Conclusions 2: IPT through the SCQE lens

Recall: **16% TB without IPT vs. 1% with it**.

- For the statisticians to warn "this is only suggestive" fails to tell the whole story,
- And is probably ineffective as a warning

Conclusions 2: IPT through the SCQE lens

Recall: **16% TB without IPT vs. 1% with it**.

- For the statisticians to warn "this is only suggestive" fails to tell the whole story,
- And is probably ineffective as a warning

But showing the effect estimate over range of δ communicates a very different picture:

Conclusions 2: IPT through the SCQE lens

Recall: **16% TB without IPT vs. 1% with it.**

- For the statisticians to warn "this is only suggestive" fails to tell the whole story,
- And is probably ineffective as a warning

But showing the effect estimate over range of δ communicates a very different picture:

- The result is frail and depends upon your assumptions

Conclusions 2: IPT through the SCQE lens

Recall: **16% TB without IPT vs. 1% with it.**

- For the statisticians to warn "this is only suggestive" fails to tell the whole story,
- And is probably ineffective as a warning

But showing the effect estimate over range of δ communicates a very different picture:

- The result is frail and depends upon your assumptions
- Even if you like $\delta = 0$, ambiguous impact.

Conclusions 2: IPT through the SCQE lens

Recall: **16% TB without IPT vs. 1% with it.**

- For the statisticians to warn "this is only suggestive" fails to tell the whole story,
- And is probably ineffective as a warning

But showing the effect estimate over range of δ communicates a very different picture:

- The result is frail and depends upon your assumptions
- Even if you like $\delta = 0$, ambiguous impact.
- Ruling out that it is harmful, too, requires an argument (baseline trend did not drop by more than 1%).

Conclusions 2: IPT through the SCQE lens

Recall: **16% TB without IPT vs. 1% with it.**

- For the statisticians to warn "this is only suggestive" fails to tell the whole story,
- And is probably ineffective as a warning

But showing the effect estimate over range of δ communicates a very different picture:

- The result is frail and depends upon your assumptions
- Even if you like $\delta = 0$, ambiguous impact.
- Ruling out that it is harmful, too, requires an argument (baseline trend did not drop by more than 1%).
- In short: You must defend an assumption to advocate for a result.

Extra slides

Results: Clinic level

Suppose we use the range of δ offered by the linear estimate above: