# The Stability-Controlled Trial and Quasi-Experiment? Identifying the effects of newly available treatments under self-selection

Chad Hazlett[*], Werner Maokola[†], David Ami Wulf[‡]

July 16, 2019

PRELIMINARY, NOT FOR CITATION

**Abstract**

Learning the effect of a new treatment through a randomized trial is often problematic due to ethical concerns, administrative constraints, or because those who are eligible for and consent to participation in a randomized trial differ from those who will later take the treatment. We discuss a recently proposed identification strategy that replaces the need for a randomized control group with an assumption on how the average non-treatment potential outcome for successive cohorts vary over time. This assumption — without ignorability — allows identification of the average treatment effect on the treated. A strong assumption of this kind may be defensible, for example, when there is no known change in the use other effective treatments for a given disease, or in the population getting that disease. More generally, if a sharp assumption of this kind is not defensible, a range of plausible assumptions can be used to show the corresponding range of plausible treatment effect estimates. We discuss this approach and use it to study the effectiveness of isoniazid preventive therapy (IPT) in reducing incidence of tuberculosis among HIV positive patients in Tanzania. At face value, IPT appears to have been highly effective: 16% of non-IPT takers developed active tuberculosis, while only 1% of IPT-takers developed it. By contrast, we show that the actual estimated effect could be either beneficial or harmful under a reasonable set of assumptions. These results protect against such compelling but misleading naive comparisons by requiring readers to actively select an assumption and see the corresponding result, and showing how easily one could have supported the opposite conclusion. Our results also reveal that those who received IPT were probably less likely to develop tuberculosis anyway. We concluded that this approach offers a step forward for implementing treatments and policies where randomized trials are impossible or undesirable.

Wordcount: 6637

[*]Corresponding author. UCLA Statistics and Political Science; chazlett@ucla.edu

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

[‡]UCLA Statistics; amiwulf@gmail.com

# 1 Introduction

Suppose a treatment becomes newly available or an effort is made to greatly increase access or uptake in a given population. Conventional clinical and statistical wisdom hold that only a randomized trial can reliably reveal the beneficial or harmful effects of that treatment. However, randomized trials may suffer due to concerns with feasibility, ethics, or representation. First, regarding feasibility, it may not be possible to restrict access to a randomly selected treatment group due to the way that a treatment is provided. The very nature of treatment dissemination may prohibit randomization, as in media campaigns. Or, the values and obligations of the organization implementing the treatment may be anathema to randomization. In other cases, the implementer may be open to randomization, but the project's implementation timeline may prevent a randomized evaluation from being planned and implemented in time. Second, randomization may pose ethical concerns. Denying a patient access to a potentially life-saving experimental drug may be argued to be ethically required if that is the only way to learn its effect. But if alternatives exist, such as the one proposed here, randomization may no longer be the most ethical choice. Relatedly, to decide who gets what treatment on the basis of an experimental design rather than a customized treatment plan may introduce harm or reduce benefit. Third and finally, even where RCTs are feasible and ethical, they may not provide the desired answer because of who is represented in them. Opportunities for randomization may dictate the location and population in which an experiment is run. In medical trials, eligibility for clinical trials is often restricted to those with minimal comorbidity and in otherwise good health. Moreover, the group willing to consent to a randomized trial knowing that they may be placed in the control condition can be a small proportion of those eligible. And, if treatment is not (successfully) blinded, those not receiving the treatment status they had hoped for may be resentful, which is both unfortunate and effectively changes the treatment being studied to include this.

Alternative designs that seek to address these concerns currently take on one of two forms. The first is a variety of designs that allow partial self-selection in an effort to address these concern, including "comprehensive cohort studies" (Olschewski and Scheurlen, 1985) and "patient preference trials" (Brewin and Bradley, 1989). These include a variety of research designs in which patients' preferences may be elicited, some individuals are randomized, and some receive

a treatment of their choosing. In newer patient-preference designs recently proposed by Knox et al. (2019), treatment preferences are elicited from all individuals, who are randomized first into a group that will have their treatment assigned at random, and a group that can choose its own treatment. This design allows for sharp-bound identification and sensitivity analysis for the average causal effects among those who would choose a given treatment. These seek to solve principally the representation problem just noted: those who select into a treatment differ from those who are randomly assigned to it. Such approaches involve the usual two arms of an RCT — randomized treatment and control groups — as well as other groups that are able to choose their treatment. While useful in many contexts to address the representation problem, they do not sidestep the need for also having randomized treatment and control groups, and thus do not solve the feasibility or ethical problems noted.

The second major set of alternatives to RCTs are simply observational studies, in which it is hoped that selection into treatment can be considered random conditionally on a set of observed covariates, i.e. "ignorability". This includes all the usual covariate adjustment approaches whether they are implemented through regression, weighting, matching, sub-classification, etc. Such approaches are, of course, susceptible to debilitating but unknowable biases due to unobserved confounding.

This paper is an early attempt to implement a recently proposed alternative (Hazlett, 2019) that side-steps the need for randomization without requiring the ignorability assumption: the stability-controlled trial (SCT) and stability-controlled quasi-experiment (SCQE). As described in detail below, this approach still caries a fundamental assumption: We must posit a numerical answer to the question of how the average outcome would have changed from one cohort to the next, had nobody taken the treatment. This "stability" assumption alone identifies the average treatment effect on the treated (ATT). Because this averages the treatment effect over a group who choose or were otherwise selected into treatment, this is is often a very desirable quantity, representationally. When a precise stability assumption is credible and defensible — as when experts can be certain there have been no other important changes in treatments or in the population of interest — so too will a precise result be credible. When an assumption on this difference can only be hypothesized or bounded, so too can a hypothetical or bounded result be provided in correspondence.

Many investigators will rightly remain skeptical of such an approach, and reluctant to rely

upon a research design without randomization. In this paper, we put the method to an early test, estimating the effectiveness of a treatment (isoniazid preventive therapy, IPT) on preventing tuberculosis among HIV positive patients visiting health clinics in Tanzania. This is a "tough case" for the method, owing to (i) the difficulty of establishing a strong assumption on how the non-treatment outcome would vary over time, and (ii) data limitations. In the course of the application, we also encounter methodological gaps that required further development of the method relative to Hazlett (2019), including the need for standard error estimators and different approaches to making assumptions on the shift in average non-treatment outcomes. In brief, we find that while naive comparisons would suggest a very beneficial effect of IPT on TB, our method shows that under a range of assumptions we cannot reject as plausible, the effect could have taken either sign. It thus clarifies exactly what must be argued (about the trend in non-treatment TB prevalence rates) in order to sustain a claim that IPT was either beneficial or harmful to those who took it. We hope this paves the way for future claims that are able to argue for restrictions on this assumption, but in the meantime, the clear relationship between these assumptions and results help to protect policymakers and investigators against unwarranted claims.

In what follows, Section 2 describes the proposed method. Section 3 introduces the application and the particulars of obtaining estimates. Section 4 produces the estimates. Section 5 discusses and concludes.

# 2 Proposed Method: Stability-controlled quasi-experiments

This section provides methodological details of the stability-controlled trial (SCT) and stability-controlled quasi-experiment (SCQE). We propose using the term SCT when the approach is proposed at the design stage, and SCQE when an opportunity is recognized retrospectively for this approach. In what remains, we use the term SCQE since the application here is of the latter type.

## 2.1 Setup

We use the potential outcomes framework (Splawa-Neyman, Dabrowska and Speed, 1990), with $Y_i(1)$ and $Y_i(0)$ representing the treatment and non-treatment potential outcomes, respectively for units $i = 1, \ldots, N$. The realized treatment status for unit $i$ is given by $D_i \in (0, 1)$. Two time periods are considered: $T = 0$ before the treatment is introduced, and $T = 1$ afterwards. Note that the cohorts observed at time $T = 0$ and time $T = 1$ in thie framework are assumed to be separate cohorts, not repeated measures as in a panel.

We use the operator $\hat{\mathbb{E}}[\cdot|\cdot]$ for sample averages and conditional averages. For notational ease we often suppress the index $i$ when referring to moments that do not depend on $i$, e.g. $\mathbb{E}[Y(0)|T = 1]$. Finally, we denote the proportion of individuals taking the treatment at time $T = 1$ as $\pi_1 = Pr(D = 1|T = 1)$ and the proportion taking it at time $T = 0$ is $\pi_0 = Pr(D = 1|T = 0)$. We consider first the simple case in which the treatment to be introduced is newly available, and thus nobody takes the treatment in the cohort for $T = 0$, i.e. $\pi_0 = 0$. We later generalize to the case with $\pi_0 \neq 0$.

The key assumption required is a postulated value for *the shift in the expected non-treatment potential outcome between the pre-treatment and post-treatment cohorts*, which we call $\delta$,

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

We require an assumption on $\delta$, or on a range of $\delta$ for which an ATT is to be determined. While various sources of information may be useful in informing a guess of $\delta$, as explored here, it fundamentally unknowable. If, for example, the outcome historically followed a stable and consistent trend, and subject matter experts agree that nothing else able to influence outcomes change over this time (besides the treatment introduction in question), then a $\delta$ representing a continuation of that trend may be a reasonable assumption. A choice of $\delta = 0$ states that the expected outcome would not be believed to change at all, in lieu of the new treatment. Such a case is most plausible when the (non-treatment) outcome is "hard to change", and so we can find no reason the observed outcome would have changed but for the new treatment. One example where this may be feasible (and where randomized trials would be ethically most troubling) includes the case of terminal illnesses where prior treatments have not proven effective.

With an assumed $\delta$, the average treatment effect on the treated (ATT) is identifiable as

follows. We have an unbiased estimate for the non-treatment outcome among the whole group in period one, using the mean observed (non-treatment) outcome in period zero, shifted by $\delta$. This group average is, in turn, a weighted combination of two other averages: the average non-treatment outcome among the untreated, which we observe, and the average non-treatment outcome among the treated, for which we can solve by applying the law of iterated expectations. That is,

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

which we can re-arrange to identify the important counterfactual quantity, $\mathbb{E}[Y(0)|D=1,T=1]$ in terms of observables,

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

Finally, the Average Treatment Effect on the Treated (ATT) is the difference between the treatment and non-treatment potential outcomes, taken solely among the treated, i.e. $\mathbb{E}[Y(1)|D=1,T=1] - \mathbb{E}[Y(0)|D=1,T=1]$. While we directly observe an estimate of the first quantity, the second term – the average outcome among the treated had they not taken the treatment – has now been given by the strategy above (Equation 2). The ATT is thus identifiable and given by

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

## 2.2 Instrumental variables interpretation

Despite appearing quite different in form, this estimator is actually equal to a version of the Wald estimator for instrumental variables, with an adjustment for $\delta$,

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

In this formulation, we interpret time as an instrument. The transition from period $T = 0$ to period $T = 1$ is an "encouragement", and any difference in non-treatment outcomes relating to this encouragement is a violation of the exclusion restriction, captured by $\delta$. Thus, $\delta$ must be subtracted out of the outcome for the encouraged (or out of the reduced form estimate itself) in order to correct for this. Since all of the "$\delta$-adjusted reduced form, i.e. $\mathbb{E}[Y|T=1] - \mathbb{E}[Y|T=0] - \delta$, is due to the treatment, dividing by the proportion treated ($\pi_1$) reveals the average treatment effect, "local to" the treated.

The idea of using time as an instrument has been exploited elsewhere (Johnston et al., 2008; Cain et al., 2009; Shetty, Vogt and Bhattacharya, 2009; Mack et al., 2015; Gokhale et al., 2018, see also Brookhart, Rassen and Schneeweiss, 2010 for discussion). The conceptualization of this approach in terms of a shift in the expected non-treatment outcome, though, seems not to have been offered previously and gives clarity on the underlying assumptions in the potential outcomes framework. Moreover, the ability to violate the exclusion restriction by postulating $\delta \neq 0$, either as a primary estimate or for sensitivity analysis, increases the usefulness, transparency, and credible debate around this approach relative to simply applying traditional IV techniques and making a claim that only if the exclusion restriction holds. Finally, we note that applications of IV in this context or more generally have attempted to buoy the required assumptions (i.e. ignorability of the encouragements and the exclusion restriction) by the addition of covariates. Our approach is generally skeptical of the conditional ignorability assumptions required of such covariate adjustment, instead relying only on $\delta$ to identify the effect.

## 2.3 Comparison to Difference-in-Differences

The method may superficially resemble the Difference-in-Difference (DID) approach. However, it operates in circumstances where DID is not possible, and provides a relaxation of DID in

cases where it is possible.

In the panel version of DID, we measure each unit before and after (some are exposed to) treatment. Alternatively, in the cross-sectional version, we must be able to place individuals into larger groupings that persist over time (such as states or schools), with treatment being assigned at the level of those larger units. It is this labeling or nesting that allows us to say whether an individual observed at time $T = 0$ "would have received treatment" had they been observed at $T = 1$.

Such labeling is necessary for DID to operate. In contrast, SCQE conceives of entirely separate cohorts at separate time points, and works even though there is no way to know if an individual observed at time $T = 0$ would have chosen treatment had they appeared at time $T = 1$. This is useful in cases such as new medical treatments: among those diagnosed with a given disease during $T = 0$, there is no way to say which would have taken the treatment one or more years later when a new treatment becomes available (at time $T = 1$).

The SCQE method is thus particularly useful where DID is not possible. When DID would have been possible in either the cross-sectional or panel form, DID is a special case of SCQE. Specifically, DID requires the parallel trends assumption,

$$\mathbb{E}\Big[Y(0)|T=1, D=1\Big] - \mathbb{E}\Big[Y(0)|T=0, D=1\Big] \;=\; \mathbb{E}\Big[Y(0)|T=1, D=0\Big] - \mathbb{E}\Big[Y(0)|T=0, D=0\Big],$$

whereas the present method instead assumes the overall average $Y(0)$ in the two periods differs by $\delta$. We could, for example, allow the trend in average non-treatment outcomes to be different for the would-be-treated group and the would-be-control group, thus violating parallel trends. Whatever trends we wish to assume for these two groups' average non-treatment outcomes, if we weight those trends by the population proportions, we would obtain the corresponding choice of $\delta$. DID, by comparison, is the special case of SCQE in which we (i) "learn $\delta$" from the over-time change in the units not eligible for treatment; and (ii) assumes the change in non-treatment outcomes is the same for both groups, i.e. parallel trends.

## 2.4 Inference

We consider here several approaches to constructing standard errors.

**Instrumental variables approach.** The first approach would be to use the "usual standard errors" for IV. That is, if $Z$ is a vector of ones for those in the post-treatment (encouraged) period and zeroes for those in the pre-treatment period, and $D$ is a vector indicating treatment status such as $(Z^\top D)^{-1} Z^\top \hat{\Sigma} Z (Z^\top D)^{-1}$ where $\hat{\Sigma}$ is an appropriate consistent estimator of the error variance, such as a simple spherical model, or the matrix with $\hat{\epsilon}_i^2$ on its diagonal for heteroskedastic standard errors (Cameron and Trivedi, 2005).[1] The downside of this approach is that it requires actual individual level data, whereas the point estimate itself can be constructed using just the three or four aggregate values noted above.[2]

**Boostrap.** Alternatively, as the "paired" (i.e. row-sampled) bootstrap is valid for instrumental variables (Freedman et al., 1984) it could be used here as well. In the application here we consider data from one policy but in which the implementation is spread across numerous clinics. We pool the data across clinics before constructing the ATT estimate (with the exception that $\delta$ is computed first at the individual level and used to produce the shifted expectation of the non-treatment outcome prior to pooling.In this case we use effectively a block-bootstrap, by re-sampling whole clinics on each iterate.

**Custom standard errors.** Computing any standard error invokes a number of assumptions including whether the proportion receiving treatment is fixed or random and choices as to which strata of the data can be assumed to have the same standard errors (in their potential outcomes) as well as covariances among these. While the IV estimator represents one set of choices, other choices could be made that are more or less appropriate depending upon the research design and assumptions. Moreover, conventional standard errors consider only "statistical uncertainty" and not any "identification uncertainty" we may wish to also propagate reflecting uncertainty over the choice of $\delta$. It may therefore be useful to explicate a set of standard error estimates under varying practical assumptions such as whether $\pi_1$ is fixed or random, and with the option of including uncertainty over $\delta$.[3]

[1] To implement this approach, we construct $\tilde{Y} = Y - \delta \mathbb{1}_{T=1}$ and use this as the outcome for IV/2SLS.

[2] We expect a simplified version can be constructed using only aggregate data estimating several the moments included in the Wald estimate, but have not yet done this.

[3] Developing these estimators is not currently a priority. We would be interested in any feedback on whether this would be useful to pursue.

# 3 Application: Isoniazid preventive therapy

## 3.1 Isoniazid preventive therapy

Tanzania is experiencing a major health crisis in TB prevalence and mortality. For those already immunocompromised by HIV, devloping an active TB infection is both more likely and more lethal. Isoniazid, an antibiotic, is often used in the treatment of active TB. The prophylactic use of isoniazid to prevent TB is refered to as isoniazid preventative therapy (IPT). Randomized trials have demonstrated isoniazid as an effective treatmnet for active TB (Smieja et al., 1999), and more importantly here, the effectiveness of IPT in preventing latent TB from developing into active TB (Churchyard et al., 2007; Akolo et al., 2010). As a result, the World Health Organization strongly encourages the use of IPT to prevent active TB in those immunocompromised by HIV, even in settings where testing for latent TB cannot be provided (WHO, 2008). However, we are not aware of any evidence as to the actual effectiveness of IPT-promoting policies in developing country settings.

Beginning in 2011, Tanzania has been making IPT available in HIV clinics, and encouraging its use through a nationwide clinician education program for the prevention of active TB development. Groups of individual clinics are selected in waves, and clinicians from these clinics receive educational training encouraging the use of IPT in *all* HIV patients not yet diagnosed with active TB. Prior to these trainings, although isoniazid was formally a standard part of care in these clinics for patients with active TB, use of IPT remained at or near zero. Following the trainings, while IPT should be widely used, we find that it is still only used on occasion (25% of the time, in the clinics that adopt it at all). New clinics were enrolled in the program through 2017, incrementally increasing the number of facilities using IPT and the number of patients given the treatment nationwide. By the end of that period, more than a third of the 318 HIV clinics have been enrolled in the program. Table 1 shows the times at which IPT use effectively began, the TB development rates before this, and the subsequent levels of IPT uptake, in each of the 21 clinics included in the analysis below.

We know little about the process by which certain clinics are chosen rather than others. More importantly, we also know little about the process by which certain patients receive IPT while others do not. The latter feature in particular suggests an approach such as SCQE because, knowing so little about the treatment assignment process, we see little hope for a defensible claim

that conditioning on any set of observed covariates would render the treatment unconfounded.

When estimating the effect of IPT, in principle each clinic provides an opportunity for a clinic-specific estimate. Alternatively, we can take clinics as merely sampling units and aim to construct a single, nationwide effect estimate pooling together patients across facilities. We discuss both estimates below though we rely mainly on the pooled estimator as a consequence of limited sample sizes. In either case, the unit of interest is a patient, the treatment is the prescription of IPT, and the outcome is whether or not the patient was eventually diagnosed with active TB.[4]

Finally, recall that the SCQE approach will estimate an average effect of IPT on TB incidence, among those who opted to take it. This is to be distinguished from efforts to estimate the *efficacy* of IPT in preventing TB (as in a randomized trial), or alternatively about the effectiveness of the program as a whole on all HIV positive patients. The estimand from SCQE is thus a highly relevant one from a policy perspective or out of interest in a retrospective analysis seeking to see what the actual effect of IPT has been on those prescribed it.

## 3.2 Patient and facility inclusion criteria

The differences in expected non-treatment outcomes for the pre- and post-treatment cohorts would be affected by any compositional differences between the cohorts. It is thus preferable when possible to construct cohorts in a way that does not generate any such compositional differences. We limit the population of interest to (a) the first year of clinic visits of "new" patients, who (b) show up in the data for at least that year, (c) that whole year of which was contained within either the pre-IPT or post-IPT period. These cohorts are thus limited to "new patients". The first part of this definition avoids systematically differing visit histories between the two cohorts. The second part was shaped by the need to accurately establish the treatment and outcome status of patients in the analysis, which requires the demonstrated presence in the data. The final part was to ensure that patients are exposed to IPT encouragement for either none or all of their first year. We note that by limiting our sample to "new patients", we limit our estimand to the "ATT among new patients".

At the clinic level, to focus on clinics where an estimate is feasible we required that there

[4]Critically, because of the lack of latent TB testing in Tanzania, TB development refers to *active* TB development throughout this paper

be 100 patients in each cohort, and that at least 10% of patients in the post-treatment period received IPT. We thus required that at least 10 patients were treated in the post-IPT period.[5]

## 3.3 Inferring timing of IPT

Some data-related difficulties complicated the measurement of IPT exposure. We transformed the visit-level dataset available to us into patient-level summaries containing an initial visit date and whether the patient received IPT or developed TB within that year. The source of information for both IPT prescription and TB development was a single variable in the medical record documenting isoniazid use. Using TB treatment status to establish the date of TB development carries with it any the delays between the two events. Unfortunately, no comprehensive record of TB test results was available to us.

There was also no record of the exact dates that each facility was trained in IPT use. Instead, we imputed these facility-specific dates by observing the time series of individual IPT prescriptions at each given facility. We chose the $2^{\text{nd}}$ percentile of IPT prescription dates as our indicator for when IPT began, which aligned well with what stood out visually as the major initial spike of uses at each facility. We contend that IPT records before this imputed date were often data entry errors, and removed them accordingly. It should be noted that an incorrect imputation date could cause attenuation bias. If any of the ignored IPT uses before that date were genuine, some effect of the treatment could be improperly experienced in the $T = 0$ period. Likewise, there can be coding errors whereby a patient is coded as treated in $T = 1$ when they should not have been. Both of these problems would lead to an understatement of the ATT. However, a sensitivity analysis shows that the causal effect estimates were robust to other choices of percentile; comparison of the ATTs from SCQE when using the $1^{\text{st}}$ or $5^{\text{th}}$ percentiles showed no substantive differences.

## 3.4 Considering $\delta$

Where should our beliefs about $\delta$ come from? Ideally, domain knowledge would provide a strong claim as to the appropriate value or range of values. One of the authors (Dr. Maokola) is an

[5]Conveniently, under this arrangement, the minimum F-statistic one would get for a first-stage regression (of treatment status on the post-treatment indicator) would be 11, favorably comparing to the traditional guideline of 10 (Stock and Yogo, 2002). In practice our clinic-level F-statistics ranged from 26 to 1356.

Table 1: Clinic Implementation of IPT

| Clinic Number | Total Patients | Pre-Implementation TB Rate | IPT Implementation Date | Post-Implementation IPT Rate |
|---|---|---|---|---|
| 1 | 3,031 | 0.15 | 2014-06-06 | 0.16 |
| 2 | 2,457 | 0.19 | 2014-06-10 | 0.25 |
| 3 | 2,448 | 0.14 | 2014-05-27 | 0.26 |
| 4 | 2,053 | 0.12 | 2014-06-26 | 0.28 |
| 5 | 1,610 | 0.04 | 2014-09-15 | 0.23 |
| 6 | 1,602 | 0.12 | 2015-09-03 | 0.20 |
| 7 | 1,569 | 0.18 | 2015-05-13 | 0.26 |
| 8 | 1,406 | 0.01 | 2015-01-19 | 0.39 |
| 9 | 1,382 | 0.13 | 2014-06-23 | 0.23 |
| 10 | 1,186 | 0.27 | 2015-09-22 | 0.23 |
| 11 | 1,035 | 0.00 | 2015-09-14 | 0.14 |
| 12 | 962 | 0.06 | 2015-03-16 | 0.16 |
| 13 | 946 | 0.17 | 2015-03-23 | 0.22 |
| 14 | 895 | 0.21 | 2015-11-17 | 0.33 |
| 15 | 818 | 0.03 | 2015-12-18 | 0.15 |
| 16 | 688 | 0.12 | 2015-03-23 | 0.51 |
| 17 | 638 | 0.06 | 2016-01-04 | 0.73 |
| 18 | 591 | 0.30 | 2015-03-13 | 0.24 |
| 19 | 490 | 0.10 | 2015-03-10 | 0.32 |
| 20 | 485 | 0.13 | 2015-09-14 | 0.19 |
| 21 | 423 | 0.02 | 2015-01-19 | 0.11 |

*Note:* Implementation details for the 21 clinics that qualify for an ATT estimate, as defined in Section 3.2. Total patients is the number of pre- and post-implementation patients used, following the same criteria.

expert on this topic and we thus documented his beliefs about $\delta$ prior to examining the data. He noted that there were no known changes in TB development rates in recent years or any medical or epidemiological reasons to expect a change, but that ongoing efforts to increase the amount of testing and treatment for TB in the population we studied could have increased reporting rates slightly. His best guess for the non-treatment outcome trend was an absolute increase of 0.5 to 1 percentage points per year. He indicated that he was not confident in the coverage of this range, though, and encouraged a data-driven approach. Notably, because he suggested no major changes over these years other than a potential increase in reporting, it becomes much more reasonable to trust that trends found in TB rates elsewhere in the data might continue on at a steady pace.

In addition to domain knowledge, data may indirectly inform beliefs about $\delta$, though we

hasten to emphasize that $\delta$ is never identified by observed data. In our case, informative data can be gleaned from the trends in values of $\hat{E}[Y_0]$ that were not affected by IPT implementation, taken either from clinics that did not receive IPT training or before such training in clinics that did receive it.

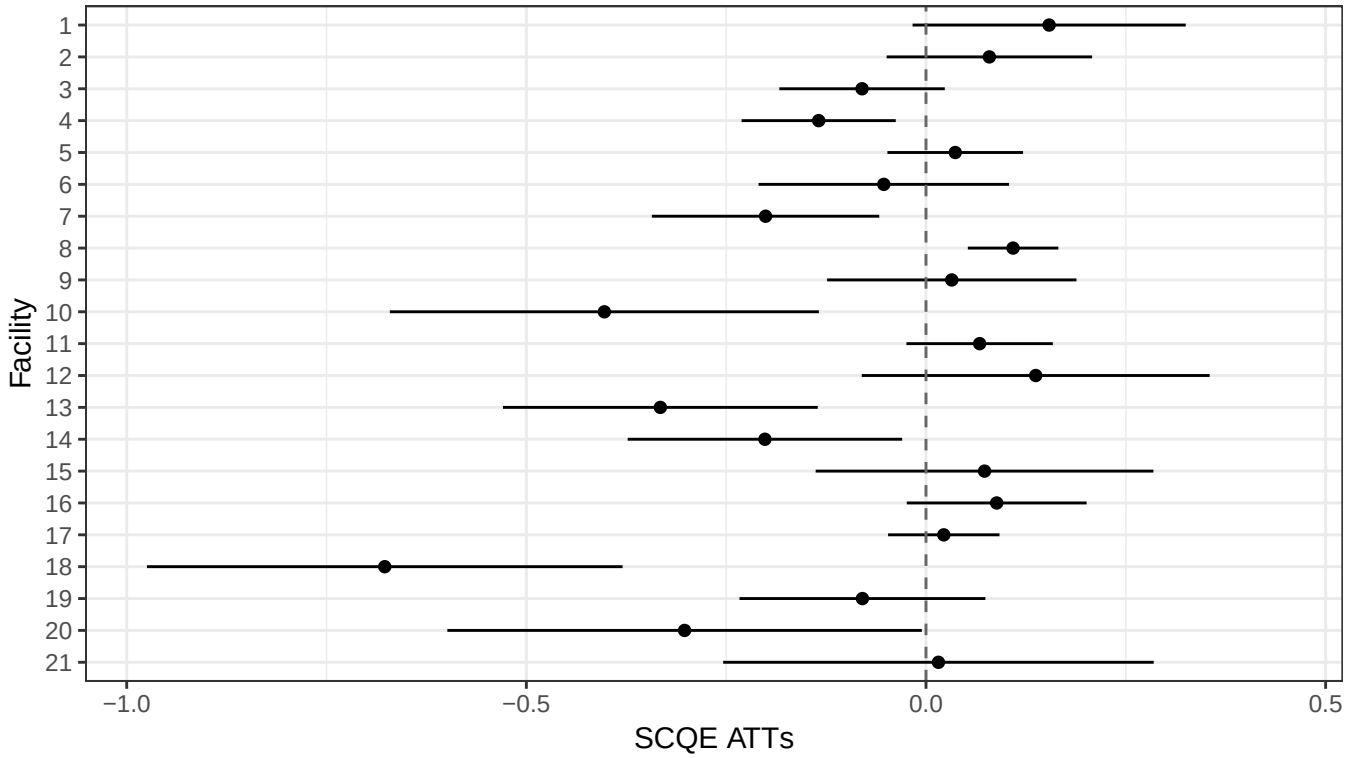
To inform $\delta$ empirically, we must first choose whether to estimate a single, system-wide trend in $\mathbb{E}[Y_0]$, or separate clinic-wise estimates. We choose the former for two reasons. First, even if the $\delta$ varies by clinic, there exists an average $\delta$ over cinics, and this is all we need in order to construct the across-clinic etimate of the ATT that we focus on. Second, the data are so noisy at the clinic-level that even the statistical uncertainty in these estimates remains large.

Second, we can conceive the trend over time as an additive one that grows linearly with the passage of time, or as an exponential or other non-linear rate (though the latter is then tranlsated back into an absolute shift to apply the appropriate $\delta$). This is particularly reasonable given that the TB incidence rate tends to be near zero in some clinics.

For the linear estimate, we regressed the binary outcome on the first visit date for all pre-implementation and non-implementing patients using a linear regression, and included intercept terms for each facility. The resulting estimate was a daily shift in $\mathbb{E}[Y_0]$ of -7.98e-06, or, as a yearly shift, of -0.00291 (95% CI [-0.00068, -0.00515]). By multiplying the linear estimate by the time between the pre-implementation and post-implementation periods, we obtain a value of $\delta$ to be used in Equation 3. For the exponential decay estimate, we ran a binary regression with a log link using the same terms as the linear regression, which produced a daily decay rate of 0.99980, or a yearly decay rate of 0.9309 (95% CI [0.8933, 0.9701]).[6] By exponentiating the decay rate by the pre-to-post time, we get a relative shift of $\mathbb{E}[Y_0]$, which can be combined with $\mathbb{E}[Y_0|T = 0]$ to obtain the corresponding value of $\delta$ for including in Equation 3. For example, given the average of the TB development rate in the pre-implementation periods of .1277, the yearly decay maps to an absolute change in $\mathbb{E}[Y_0]$ of -0.0090 (95% CI [-0.0046, -0.0128]).

[6]We call these multipliers "decay" rates because the data produced estimates of less than 1, but they could have represented "growth" rates had they been above 1.

Figure 1: Clinic-Specific ATTs for $\delta = 0$




*Note:* ATT estimates for each facility given $\delta$ is assumed to be 0. The whiskers represent the 95% CI from the IV estimator for the standard errors.

# 4 Results

## 4.1 Clinic level results

We begin with clinic-level estimates that, while relying on smaller samples, are useful for their simplicity and ability to show variability across clinics. We first generate an ATT at each clinic from Equation 3 with the appropriately-scaled values of $\delta$. We then obtain standard errors using the IV approach. Note that the standard errors constructed this way account for only statistical uncertainty at a given choice of $\delta$ and not uncertainty in $\delta$ itself. For illustration, we show the clinic-level estimates at $\delta = 0$. At this value, seven of the 21 clinics show a negative (beneficial) estimate with 95% confidence intervals excluding zero (i.e. two-sided $p < 0.05$); one clinic shows a statistically significant positive (harmful) effect, and the remaining 13 having confidence intervals that include zero.

Next, we consider clinic-level ATTs based on other choices of $\delta$, or better, various intervals for $\delta$. Without constructing specialized standard errors that directly integrate uncertainty over

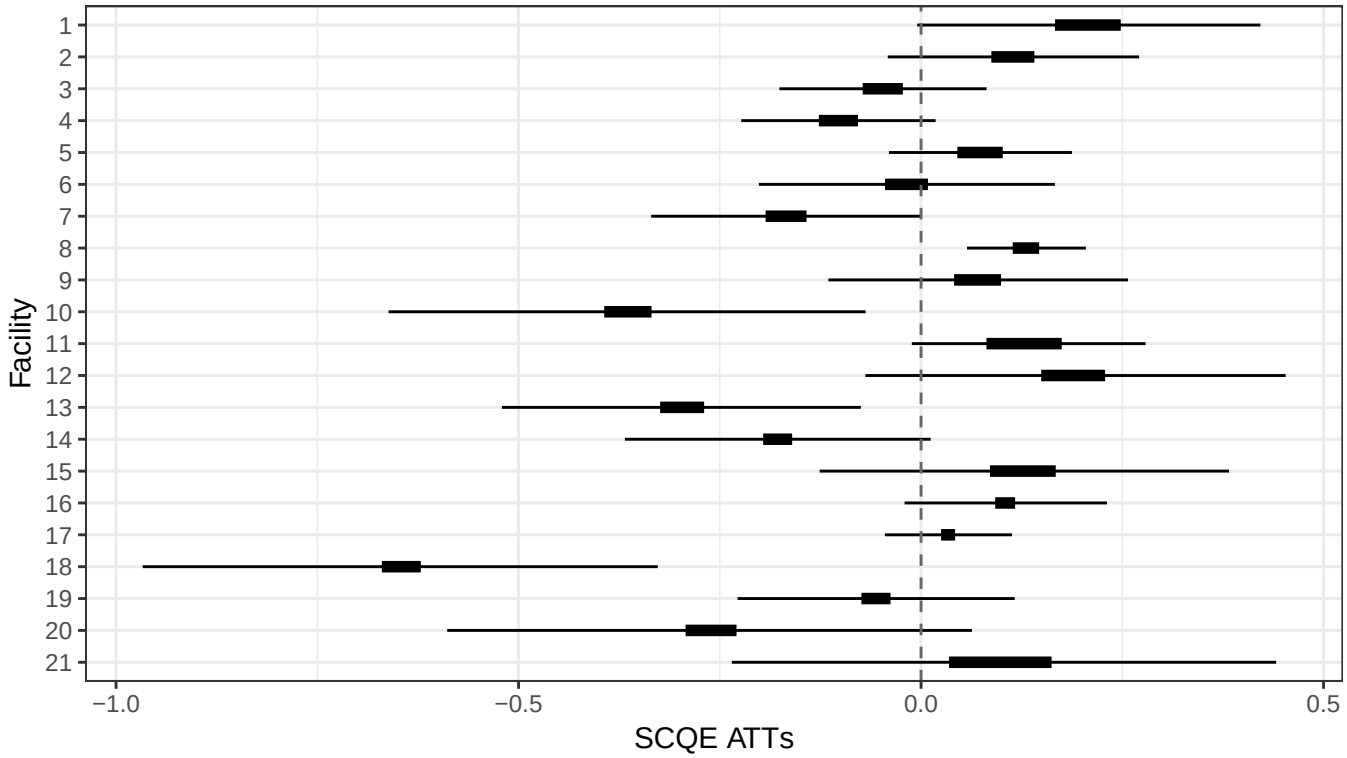
$\delta$, we can use the IV-based standard errors by repeating the analysis at the high- and low-ends of an interval over $\delta$. Augmented confidence intervals can then be built by taking the lower CI from the choice of $\delta$ that produces the lowest point estimate and the upper CI from the choice that produces the higher point estimate. This results in a "range-and-whisker" rather than a "dot-and-whisker" visualization of clinic-level ATT estimates.

Figure 2 shows such ATT estimates using a range of $\delta$ based on the linear trend and its statistical uncertainty. The thick bands show the range between the point estimates at the higher and lower choices of $\delta$. The whiskers then show the lower or upper portions of the confidence intervals extending from these. In four clinics, the effect estimate is in the beneficial direction with the augmented 95% confidence interval excluding zero; in one it is significant in the opposite direction and in the remaining 16 the augmented confidence interval includes zero.

In the Appendix, we show similar plots with both the range of $\delta$ suggest by Dr. Maokola of 0.5-1 percentage points increase per year (Figure 4), or the range obtained by using an exponential decay rate to learn from the prior trends (Figure 5). The first are more sanguine, with 8 clinics having augmented CIs that exclude zero in the beneficial direction and one showing significant estimates in the other direction. The latter is the most pessimistic: three clinics appear to have positive (harmful) effects of IPT with augmented confidence intervals excluding zero and only one shows evidence of a significant beneficial effect.

Several finite-sample considerations may arise using this approach, particularly at lower levels of aggregation such as the clinic-level estimates. First, while we do not face the equivalent of a "weak instrument" problem (too small a $\pi_1$), that could be an issue in other applications or if we included other clinics where uptake of IPT was lower. Second, we also came to find that small samples can give rise to imputed intermediate quantities that can take on impossible values: Both the estimate for $\mathbb{E}[Y(0)|T=1]$ and the estimate of the counterfactual, $\mathbb{E}[Y(0)|T=1, D=1]$ can be *negative* for some clinics. This can occur under certain constellations of values such as when the pre-treatment incidence rate was already low, and a negative $\delta$ was then applied (resulting in a negative $\hat{\mathbb{E}}[Y(0)|T=1]$ for that clinic), or when the observed outcome $\hat{\mathbb{E}}[Y(0)|T=1, D=0]$ is higher than the imputed $\hat{\mathbb{E}}[Y(0)|T=1]$, forcing $\hat{\mathbb{E}}[Y(0)|T=1, D=0]$ to be below it, and possibly below zero. While disconcerting at first, these are not really "impossible" or problematic: they arise due to finite sample error. Relatedly, other methods can also produce such "impossible" implied values when applied to smaller sampling units. For

Figure 2: Clinic-Specific ATTs for data driven linear $\delta$




*Note:* ATT estimates for each facility, using the range of $\delta$ implied by learning the linear trend over untreated periods, and constructing estimates using the upper and lower 95% confidence interval of that $\delta$, together with the 95% confidence interval around the ATT from each of those. The results appear to be significantly and substantively beneficial in four clinics (those to the left of zero line, with the augmented 95% confidence interval excluding zero); in one it is significant in the opposite direction; and in the remaining 16 the augmented confidence interval includes zero. See Appendix for similar results but using choices of $\delta$ arrived at by different assumptions.

example, in a difference-in-difference approach, one could selects a particular unit's outcome in the pre-treatment period, and shifts it by the common "over-time" shift (i.e. that learned from the difference between the average of the control group outcomes before and after treatment). That result, too, could easily take on an impossible value (e.g. an employment level below zero), despite representing an imputed non-treatment value for that unit in the second time period. And yet, this is a consequence of idiosyncratic sampling error and averages out to yield a valid treatment effect estimate. The same is true here.[7]

[7]Such situations may provide opportunities to improve upon our estimates: when applying $\delta$ produces an impossible second period outcome, it tells us something about the finite sample error in the first period, i.e. that it was drawn from a truncated distribution. This information could be used to update our estimate of the pre-treatment value. We do not pursue this here.

## 4.2 Pooled estimates

Our primary estimate of interest pools across clinics so as to attain a single estimate that is as precise as possible. Regardless of heterogeneity that may exist — and without assuming it away — at any level of aggregation, there exists a choice of $\delta$ applicable to the population in question, and for every choice of $\delta$ there is a consequent ATT estimate. In this pooling, we obtain a system-wide estimate for each of the moments estimators involved in Equation 3 through weighted averages of clinic-level conditional averages. Note that because the time gap between $T = 0$ and $T = 1$ can differ by clinic, the $\delta$ must be applied to each clinic's data in order to arrive at that clinic's estimate for $\mathbb{E}[Y(0)|T = 1]$. Once the four key sample moments have been constructed in this pooled fashion, the final ATT is estimated using these and the weighted-average $\pi_1$ as per Equation 3. Bootstrap confidence intervals are obtained by resampling over clinics when constructing the weighted averages and then recomputing the ATT.

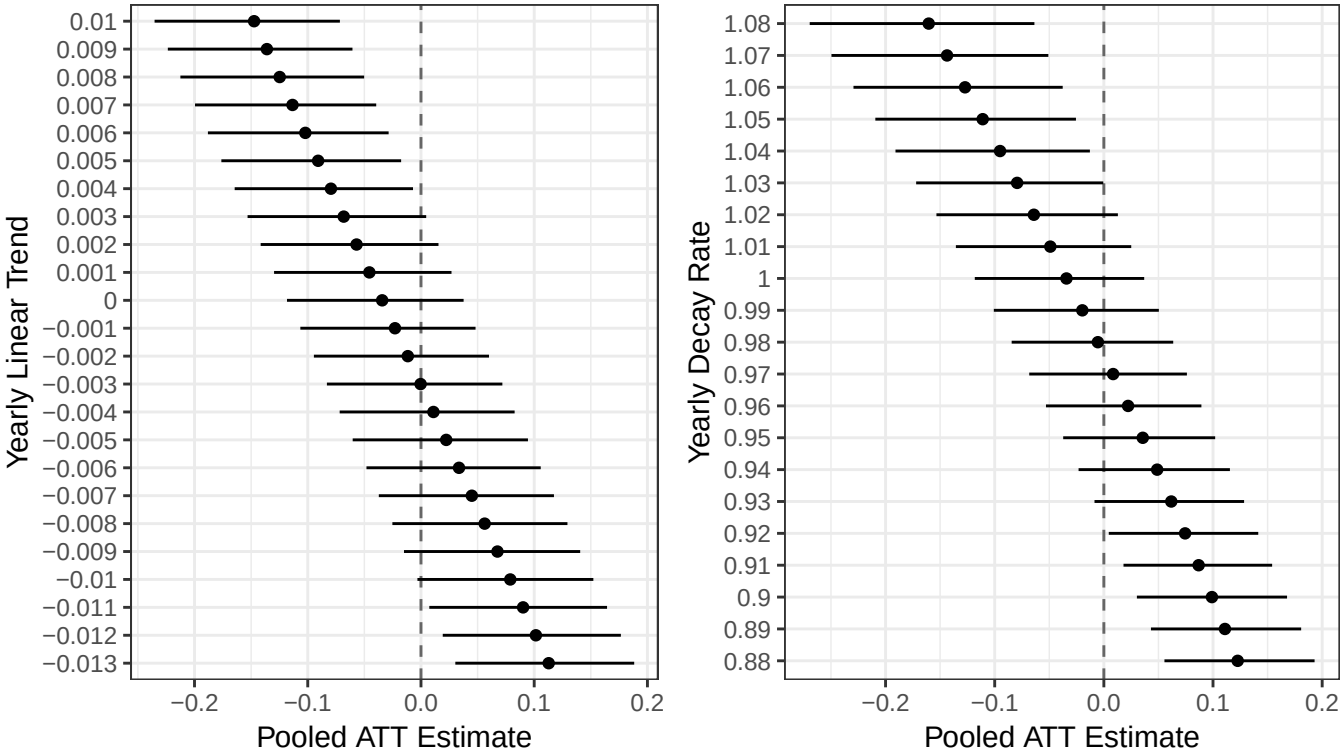
We find that prior to IPT, the pooled average TB incidence rate was 13%. After IPT, 25% of patients were prescribed IPT. The observed average TB incidence rate for those who did not take IPT was 16%, slightly higher than the 13% per-IPT. But the incidence was radically lower for those who took IPT at just 1%. These are the only four values we require from the data. The question to be answered is whether the large difference in TB rates between those taking IPT (1%) and those who did not (16%) in the second period s due to an effect of IPT, or a selection process.

We can answer this, for a choice of $\delta$. Suppose we employ $\delta = 0$, i.e. we believe that nothing important has changed over time in the non-treatment expected TB rate. Informally, the small shift from 13% getting TB pre-IPT, to 16% of non-IPT-takers getting it in the second period suggests that those choosing not to take IPT once available are similar to but slightly "worse off" than the average patient. This means that those choosing/selected to take IPT are slightly "better off". Thus, the naive difference in TB rates in the second period (16%-1% = 15%) is partly due to this selection since the treated would have already have lower TB rates.

The SCQE method adjusts for this through Equation 3. Under a $\delta$ of zero, the estimated average non-treatment rate of TB over everybody in the post-IPT cohort is then imputed to be 13%. For treated and non-treated groups to collectively reach the average non-treatment outcome of 13% prevalence, the treated group must have had a non-treatment outcome of 4%.

Comparing this to the observed, treatment outcome for the treated gives the ATT of 1% - 4% = -3%. That is, IPT appeared to reduce the TB incidence rate by 3 percentage points.

Figure 3: Pooled ATTs, by $\delta$




*Note:* Pooled-estimates for the ATT under varying assumptions on the linear trend generating $\delta$ (*left*), or an exponential decay rate giving rise to $\delta$ (*right*). Confidence intervals were generated using the block bootstrap method described in section 2.4. These plots help to demonstrate the mapping between a choice of $\delta$ or of an unerlying parameter that implies an effective $\delta$ and the consequent ATT estimate. Under our prior assumption that the non-treatment average TB incidence would rise by 0.005 to 0.01 per year, we see negative (beneficial) ATT estimates (*left*), distinguishable from zero and ranging from 9 to 15 percentage points lower TB incidence. The data-assisted choice of $\delta$ under a linear model suggested annualized trends of -0.001 to -0.005, still on the *left*, which corresponds to non-significant pooled ATTs. By contrast, a data-assisted choice of $\delta$ under an exponential decay model suggested annualized decay rates of 0.97 to 0.89. On the *right*, we see these decay rates correspond to a combination of non-significant and significant positive (harmful) estimates.

We have thus dealt with selection concerns not through assuming the observability of all confounders and adjusting for them, but simply through an assumption on $\delta$. Ideally, we would have a strong argument for our assumption on $\delta$, and would take such a result as the final estimate of the ATT. However, these estimates are better understood simply as "the logical implication of an assumption about $\delta$ on the estimate of the ATT." Figure 3 shows this more comprehensively by visualizing estimates under varying choices of $\delta$. The left panel of Figure 3 shows how as-

sumptions on a linear trend in the non-treatment outcome generate varying estimates.[8] Those values produce the ATT estimates plotted, with 95% confidence intervals produced by the IV method. Under the "domain knowledge" assumption that reported TB rates would have risen by 0.5 to 1 percentage point over per year, the consequent ATT estimates would range from a 9 to 15 percentage point reduction in the TB incidence rate. By contrast, the data-informed choice of $\delta$ based on linear trends in the non-IPT data suggests a range of -0.001 to -0.005. These correspond to small and non-significant estimated ATTs. Finally, the right panel of Figure 3 indexes estimates by the annualized decay rate used to formulate $\delta$. The data-driven assumption that decay rates vary from 0.97 to 0.89 produces ATT estimates ranging from a 12 percentage point harmful (and significant) effect of IPT down to an estimate of approximately zero.

# 5 Conclusions

Randomized trials are not always possible or preferable. Does the SCQE approach provide a feasible alternative? We argue that it provides useful, credible information and protects against misleading results — though like any method, the answer is not always as definitive as we would like. On the one hand, the SCQE method allows a clear understanding of how one's assumptions on the non-treatment average outcome trends translate into ATT estimates, free of concerns over endogeneity or selection into treatment. On the other hand, in this case the range of $\delta$ values we find to be plausible produces ambiguous results that can change in sign and significance. Without stronger assumptions to limit the range of defensible $\delta$ values, we cannot yet be certain that IPT is beneficial, harmful, or neutral. Of course, applied to different situations and with different abilities to narrow assumptions on $\delta$, the results may be more or less definitive in other cases.

Though such uncertain results may be frustrating to the investigator or policymaker, they are no less important and constitute an important step forward in program evaluation. First, the approach provides a set of simple "what-if" analyses that correctly characterizes what we can say about the effect of IPT, and tells us what further restrictions to our assumptions (on $\delta$) would be required to arrive at a definitive claim. Second, these results stand in stark contrast to naive comparisons that may be employed in the absence of a randomized trial: With just 1%

[8]The time gap between the average pre- and post-IPT patients was close to three years for all clinics, however we took the actual average and multiplied by this annual rate to get an effective $\delta$ as an additive shift.

of patients on IPT developing active TB as compared to 16% not on IPT at the same time, it is easy for policymakers to enthusiastically conclude that this program is working and call for its expansion. The SCQE approach provides a degree of protection against such potentially false and even harmful conclusions, going beyond the usual practice of simply reminding policymakers that naive comparisons are "only suggestive". Our approach turn this problem around, pointing not to an invisible threat of confounding and hoping it is taken seriously, but rather requiring the reader to actively choose and defend an assumption on the trend ($\delta$) if they wish to argue for a given result. In the process, it also shows how easily one could have drawn the opposite conclusion, encouraging skepticism. In this case, we show that the program is beneficial only if we can defend a claim that the (non-treatment) TB incidence rate was climbing by at 0.4 percentage points or more per year over this period. Further, unless the policymaker can rule out a 1 percentage point or more decrease in the (non-treatment) TB incidence rate per year, we cannot reject the possibility that the program was actually harmful. We believe policymakers are better off with this type of information than without it.

An additional benefit of this approach is that it tells us something about who was selected into the treatment, not in terms of their covariates, but directly in terms of their non-treatment potential outcomes. Those in the second cohort who did not take the treatment had higher TB prevalence than the average of the first cohort. As long as $\delta \leq 3\%$, which we judge to be quite plausible, we can conclude that those who chose to take ITP had lower chances of getting TB anyway. Knowing that this treatment is often assigned to those who are already "healthier" is likely useful information for assessing and redesigning the national effort to increase IPT use.

In conclusion, the SCQE and SCT approaches have much to offer where randomized trials are infeasible, unethical, or otherwise not preferred. They may produce sharp and definite conclusions, particularly where effects are strong and/or narrow assumptions on $\delta$ can be supported due to the nature of the application. In other cases, as here, the assumptions we can make on $\delta$ are may not be sharp enough to produce a narrow range of credible effect estimates. Even in these cases, however, this method aids in protecting against false conclusions, describes what assumptions on trends have to be ruled out (or in) to get find a beneficial or harmful effect, and tell us about the nature of those who are selected into treatment. It thus provides a rigorous alternative to both naive comparisons that may be otherwise attempted, and randomized trials when they are not feasible or desirable.

# References


Akolo, C, I Adetifa, S Shepperd and J Volmink. 2010. "Treatment of latent tuberculosis infection in HIV infected persons." *Cochrane Database of Systematic Reviews* .

Brewin, Chris R and Clare Bradley. 1989. "Patient preferences and randomised clinical trials." *BMJ: British Medical Journal* 299(6694):313.

Brookhart, M Alan, Jeremy A Rassen and Sebastian Schneeweiss. 2010. "Instrumental variable methods in comparative safety and effectiveness research." *Pharmacoepidemiology and drug safety* 19(6):537–554.

Cain, Lauren E, Stephen R Cole, Sander Greenland, Todd T Brown, Joan S Chmiel, Lawrence Kingsley and Roger Detels. 2009. "Effect of highly active antiretroviral therapy on incident AIDS using calendar period as an instrumental variable." *American journal of epidemiology* 169(9):1124–1132.

Cameron, A Colin and Pravin K Trivedi. 2005. *Microeconometrics: methods and applications*. Cambridge university press.

Churchyard, Gavin J, Fabio Scano, Alison D Grant and Richard E Chaisson. 2007. "Tuberculosis Preventive Therapy in the Era of HIV Infection: Overview and Research Priorities." *The Journal of Infectious Diseases* 196:S52–S62.

Freedman, David et al. 1984. "On bootstrapping two-stage least-squares estimates in stationary linear models." *The Annals of Statistics* 12(3):827–842.

Gokhale, Mugdha, John B Buse, Christina DeFilippo Mack, Michele Jonsson Funk, Jennifer Lund, Ross J Simpson and Til Stürmer. 2018. "Calendar time as an instrumental variable in assessing the risk of heart failure with antihyperglycemic drugs." *Pharmacoepidemiology and drug safety* 27(8):857–866.

Hazlett, Chad. 2019. "Estimating causal effects of new treatments despite self-selection: The case of experimental medical treatments." *Journal of Causal Inference* 7(1).

Johnston, KM, P Gustafson, AR Levy and P Grootendorst. 2008. "Use of instrumental variables in the analysis of generalized linear models in the presence of unmeasured confounding with applications to epidemiological research." *Statistics in medicine* 27(9):1539–1556.

Knox, Dean, Teppei Yamamoto, Matthew A Baum and Adam J Berinsky. 2019. "Design, identification, and sensitivity analysis for patient preference trials." *Journal of the American Statistical Association* pp. 1–27.

Mack, Christina DeFilippo, M Alan Brookhart, Robert J Glynn, Anne Marie Meyer, William R Carpenter, Robert S Sandler and Til Stürmer. 2015. "Comparative effectiveness of oxaliplatin vs. 5-flourouricil in older adults: an instrumental variable analysis." *Epidemiology (Cambridge, Mass.)* 26(5):690.

Olschewski, M and H Scheurlen. 1985. "Comprehensive cohort study: an alternative to randomized consent design in a breast preservation trial." *Methods of information in medicine* 24(03):131–134.

Shetty, Kanaka D, William B Vogt and Jayanta Bhattacharya. 2009. "Hormone replacement therapy and cardiovascular health in the United States." *Medical Care* pp. 600–605.

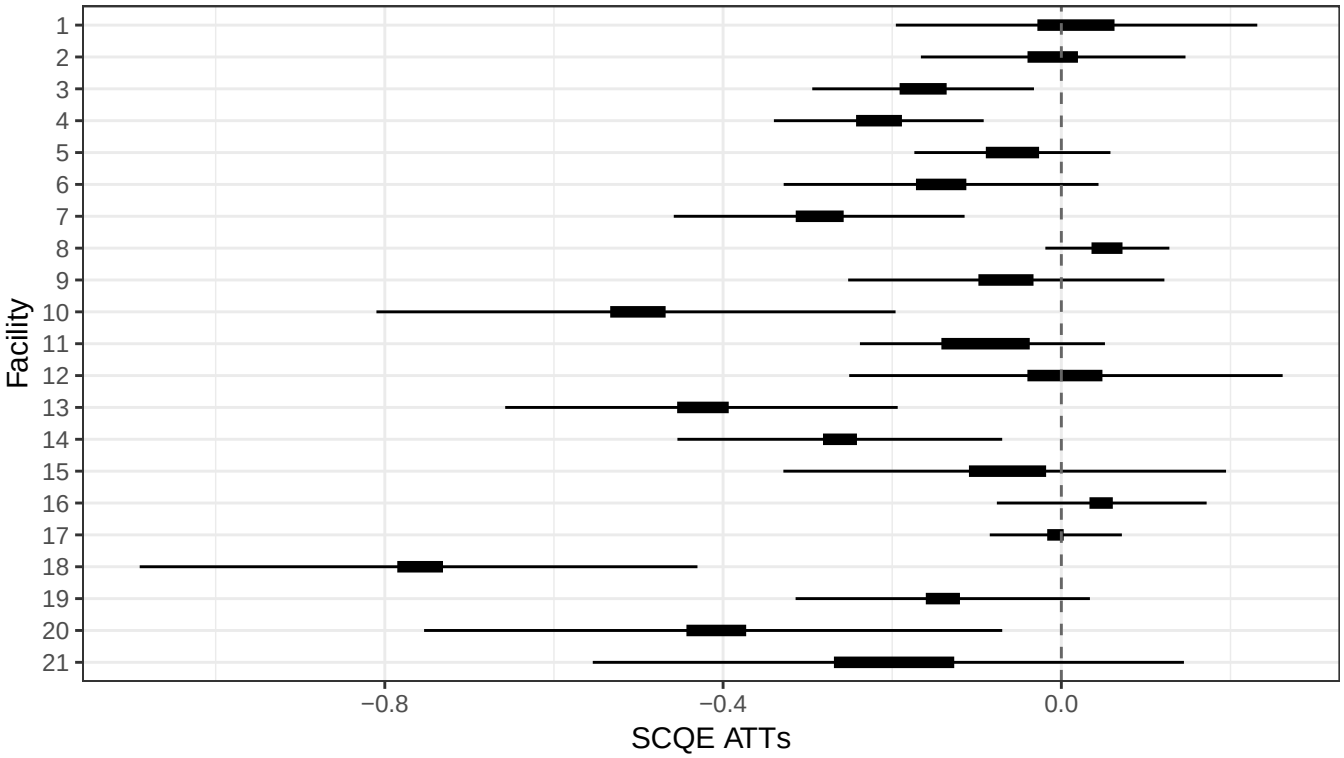
Smieja, Marek, Catherine Marchetti, Deborah Cook and Fiona M Smaill. 1999. "Isoniazid for preventing tuberculosis in non-HIV infected persons." *Cochrane Database of Systematic Reviews* .

Splawa-Neyman, Jerzy, Dorota M Dabrowska and TP Speed. 1990. "On the application of probability theory to agricultural experiments. Essay on principles. Section 9." *Statistical Science* pp. 465–472.

Stock, James H and Motohiro Yogo. 2002. "Testing for weak instruments in linear IV regression.".

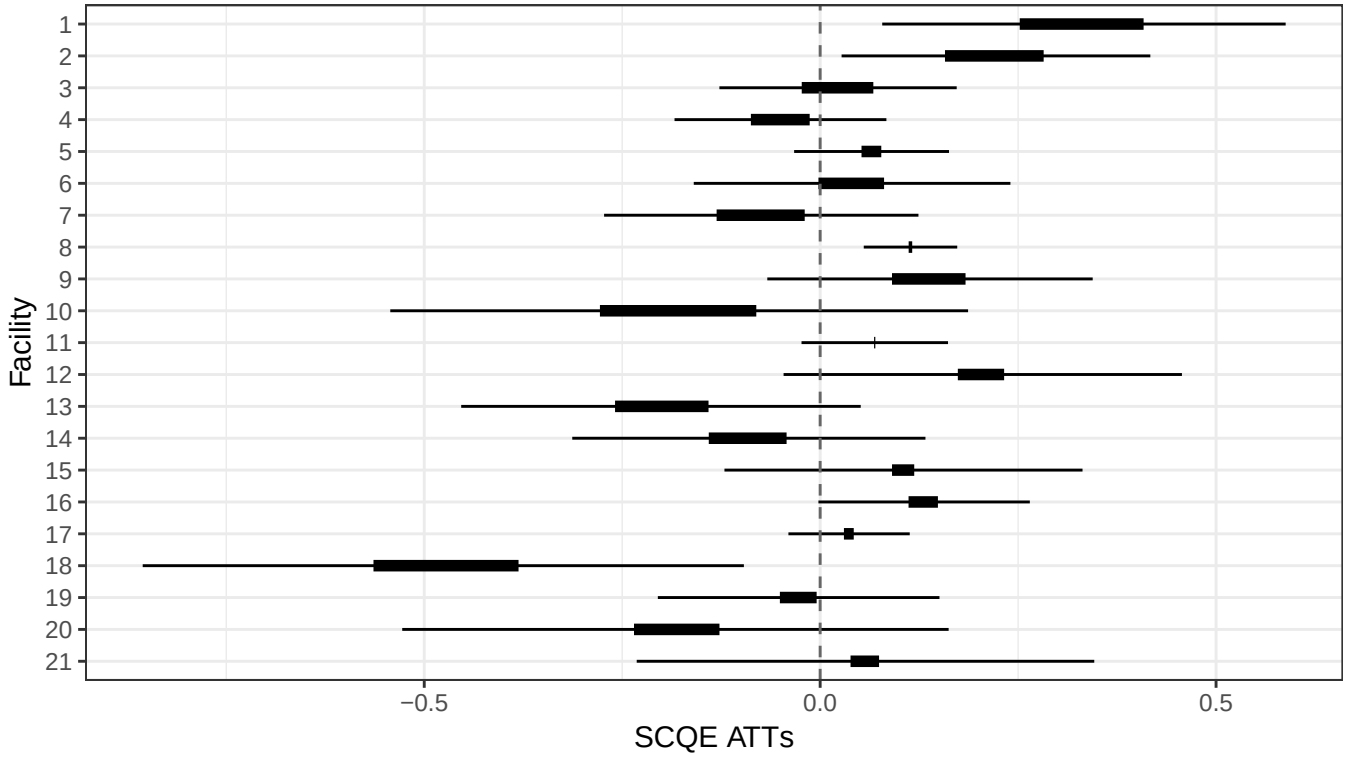
WHO. 2008. Isoniazid Preventive Therapy. In *Implementing the WHO Stop TB Strategy: A Handbook for National Tuberculosis Control Programmes*. chapter 7.

Figure 4: Clinic-Specific ATTs for domain-knowledge-based $\delta$




*Note:* ATT estimates for each facility, using the $\delta$ range provided from domain knowledge, together with the 95% confidence interval around each end of that range. The results appear to be significantly and substantively beneficial in eight clinics (those to the left of zero line, with the augmented 95% confidence interval excluding zero); in none are they significant in the opposite direction; and in the remaining 13 the augmented confidence interval includes zero. These results are the most optimistic of the three sets of clinic-level estimates we generated.

Figure 5: Clinic-Specific ATTs for data driven exponential $\delta$




*Note:* ATT estimates for each facility, using the $\delta$ implied by learning the exponential decay trend over untreated periods, and constructing estimates using the upper and lower 95% confidence interval of that $\delta$, together with the 95% confidence interval around the ATT from each of those. The results appear to be significantly and substantively beneficial in one clinic (that clinic to the left of zero line, with the augmented 95% confidence interval excluding zero); in three they are significant in the opposite direction; and in the remaining 17 the augmented confidence interval includes zero. These results are the most pessimistic of the three sets of clinic-level estimates we generated.

Table 2: Linear $\delta$ results

| Yearly Linear Trend | Effective Used $\delta$ | Pooled Resultant ATT | Bootstrapped 2.5% CI | Bootstrapped 97.5% CI |
|---:|---:|---:|---:|---:|
| 0.010 | 0.029 | -0.147 | -0.235 | -0.072 |
| 0.009 | 0.026 | -0.136 | -0.224 | -0.061 |
| 0.008 | 0.023 | -0.125 | -0.212 | -0.050 |
| 0.007 | 0.020 | -0.113 | -0.200 | -0.040 |
| 0.006 | 0.017 | -0.102 | -0.188 | -0.029 |
| 0.005 | 0.014 | -0.091 | -0.176 | -0.018 |
| 0.004 | 0.011 | -0.080 | -0.165 | -0.007 |
| 0.003 | 0.009 | -0.068 | -0.153 | 0.005 |
| 0.002 | 0.006 | -0.057 | -0.142 | 0.015 |
| 0.001 | 0.003 | -0.046 | -0.130 | 0.027 |
| 0.000 | 0.000 | -0.034 | -0.118 | 0.038 |
| -0.001 | -0.003 | -0.023 | -0.107 | 0.048 |
| -0.002 | -0.006 | -0.012 | -0.095 | 0.060 |
| -0.003 | -0.009 | -0.000 | -0.083 | 0.072 |
| -0.004 | -0.011 | 0.011 | -0.072 | 0.083 |
| -0.005 | -0.014 | 0.022 | -0.060 | 0.094 |
| -0.006 | -0.017 | 0.034 | -0.048 | 0.106 |
| -0.007 | -0.020 | 0.045 | -0.037 | 0.117 |
| -0.008 | -0.023 | 0.056 | -0.025 | 0.129 |
| -0.009 | -0.026 | 0.068 | -0.015 | 0.141 |
| -0.010 | -0.029 | 0.079 | -0.003 | 0.152 |
| -0.011 | -0.031 | 0.090 | 0.007 | 0.164 |
| -0.012 | -0.034 | 0.101 | 0.019 | 0.177 |
| -0.013 | -0.037 | 0.113 | 0.030 | 0.188 |

*Note:* Effective $\delta$ values used, and the resultant ATT estimates, for a range of yearly counterfactual TB development linear trends across the population. Those estimates, as well as the confidence intervals around them generated by bootstrapped resampling at the clinic level, produce the left half of Figure 3. The data-assisted choice of $\delta$ under this linear model ranged in yearly trends from -0.001 to -0.005, while the best guess at a yearly trend using domain knowledge ranged from 0.005 to 0.01

Table 3: Exponential $\delta$ results

| Yearly Decay Rate | Effective Used $\delta$ | Pooled Resultant ATT | Bootstrapped 2.5% CI | Bootstrapped 97.5% CI |
|---|---|---|---|---|
| 1.08 | 0.032 | -0.160 | -0.269 | -0.064 |
| 1.07 | 0.028 | -0.144 | -0.250 | -0.051 |
| 1.06 | 0.024 | -0.127 | -0.229 | -0.038 |
| 1.05 | 0.019 | -0.111 | -0.209 | -0.026 |
| 1.04 | 0.015 | -0.095 | -0.191 | -0.013 |
| 1.03 | 0.011 | -0.079 | -0.172 | -0.001 |
| 1.02 | 0.008 | -0.064 | -0.153 | 0.013 |
| 1.01 | 0.004 | -0.049 | -0.136 | 0.025 |
| 1.00 | 0.000 | -0.034 | -0.118 | 0.037 |
| 0.99 | -0.004 | -0.020 | -0.101 | 0.050 |
| 0.98 | -0.007 | -0.005 | -0.084 | 0.064 |
| 0.97 | -0.011 | 0.008 | -0.068 | 0.076 |
| 0.96 | -0.014 | 0.022 | -0.053 | 0.089 |
| 0.95 | -0.018 | 0.036 | -0.037 | 0.102 |
| 0.94 | -0.021 | 0.049 | -0.023 | 0.115 |
| 0.93 | -0.024 | 0.062 | -0.009 | 0.128 |
| 0.92 | -0.028 | 0.074 | 0.005 | 0.141 |
| 0.91 | -0.031 | 0.087 | 0.018 | 0.154 |
| 0.90 | -0.034 | 0.099 | 0.030 | 0.168 |
| 0.89 | -0.037 | 0.111 | 0.043 | 0.181 |
| 0.88 | -0.040 | 0.123 | 0.056 | 0.193 |

*Note:* Effective $\delta$ values used, and the resultant ATT estimates, for a range of yearly counterfactual TB development decay rates across the population. The Those estimates, as well as the confidence intervals around them generated by bootstrapped resampling at the clinic level, produce the right half of Figure 3. The data-assisted choice of $\delta$ under this exponential decay model ranged in yearly decay rates from 0.89 to 0.97