A proposal for estimand and endpoint development, using the REMoxTB and STREAM trials as case studies

Tra My Pham

MRC Clinical Trials Unit at UCL

SCT session “The What and the How: Practical challenges in aligning estimands and estimation using tuberculosis as a case study”

22/05/2023 | tra.pham.09@ucl.ac.uk

Smarter Studies
Global Impact
Better Health

Disclosure

No relevant disclosures

Outline

1. Tuberculosis (TB) treatment trials
2. Motivation for using estimands in TB trials
3. The estimand framework
4. Application in TB trials
5. Thinking through the intercurrent events (ICEs)
6. Conclusions

Based on paper: Pham TM, Tweed CD, Carpenter JR, et al. Rethinking intercurrent events in defining estimands for tuberculosis trials. Clin Trials 2022; 19: 522–533.

TB treatment trials

Current treatment regimens for TB are effective but burdensome

- 6 months for drug-sensitive TB
- 9 months for rifampin-resistant TB

The MRC CTU at UCL runs a number of TB trials

- Our main aim is to develop less onerous but similarly effective regimens
- Two examples (non-inferiority trials)
  - REMoXTB: 4-month regimens vs 6-month control
  - STREAM Stage 1: 9-month regimen vs 20-month control

MRC Clinical Trials Unit

UCL

Gillespie SH, Crook AM, McHugh TD, et al. Four-month moxifloxacin-based regimens for drug-sensitive tuberculosis. *N Engl J Med* 2014; 371: 1577–1587

Nunn AJ, Phillips PPJ, Meredith SK, et al. A trial of a shorter regimen for rifampin-resistant tuberculosis. *N Engl J Med* 2019; 380: 1201–1213

Motivation for using estimands in TB trials

Primary outcome in TB treatment trials is often a binary (composite) outcome

- Several components, but primarily based on culture results
- **Favourable**: infection is cleared (negative culture results) and does not return
- **Unfavourable**: infection is not clear (positive culture results) or participant experiences some components

Some components were inevitably included as 'unfavourable'

- Actual outcome was missing and 'unfavourable' was most likely
- Chosen method of analysis required a binary classification

The estimand framework provides a systematic way for thinking about TB outcomes

- Many aspects have already been addressed in previous outcome definitions
- But not clear whether intercurrent events were considered / handled appropriately
- Difficult to identify precise research question

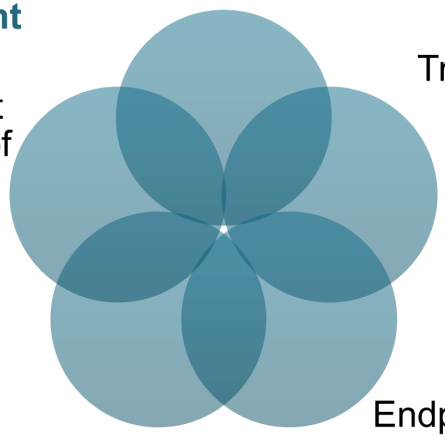
The estimand framework

Population

Population of patients

Handling of InterCurrent Events (ICEs)

Identification of relevant ICEs and specification of how they are handled



Treatment conditions

Treatment conditions being compared

Population-level summary

Summary measure used to compare treatment conditions (e.g. risk difference, odds ratio)

Endpoint

Endpoint to be obtained for each patient

The estimand is a precise description of the treatment effect that the trial aims to estimate

STRATEGIES FOR HANDLING ICEs

Treatment policy
- Treatment effect regardless of occurrence of ICE

Composite
- Treatment effect on composite of outcome or occurrence of ICE

Hypothetical
- Treatment effect if ICE did not occur

Principal stratum
- Treatment effect in sub-population for whom ICE would not occur

While on treatment
- Treatment effect up to point of ICE occurring

Application in TB trials

Discussed between trial methodologists, TB clinicians, TB statisticians

Identified common intercurrent events (ICEs)

- Used REMoxTB and STREAM Stage 1 as motivating examples

Explained why each event was an ICE

- Events that affect interpretation of or preclude measurement of the intended final outcome

Asked TB clinicians & statisticians how each ICE was handled and discussed the rationale

- This was often hard to disentangle

Suggested possible ways for handling ICE that follow the estimand framework

- Patient and healthcare provider perspectives were considered

Events considered and why they are ICEs

Treatment changes & treatment extension	Data collected after a treatment change no longer reflect the patient's status while receiving original treatment
Poor adherence	Data collected after poor adherence no longer reflect the patient's status while receiving original treatment
Re-infection	It hinders our ability to confirm whether the patient has been durably cured from their original strain of TB
Death	For those who die before completing the trial, their clinical data and culture results no longer exist and are truncated by death
Loss to follow-up	Not an ICE, but often associated with discontinuation of treatment which is an ICE

1. Treatment changes & treatment extension

How has it been handled?	• Unfavourable composite for a change of one drug (REMoxTB) / ≥ two drugs (STREAM Stage 1) • Treatment policy for a change of one drug (STREAM Stage 1) or treatment extension to make up for missed doses
Rationale	• Treatment changes reflect treatment failure • Fewer treatment options available in a national TB programme • New regimen often longer / more toxic
How might it be handled?	• From a patient perspective, a treatment change is not inherently a poor outcome → treatment policy? • For treatment changes not available in practice (e.g. experimental to standard) → hypothetical?

2. Poor adherence

How has it been handled?	• Poor adherence was ignored in main (*modified* intention-to-treat) analysis → treatment policy • Poor adherers were excluded from per-protocol analysis → not clear (hypothetical)?
Rationale	• Poor adherence is unavoidable in practice • But in a NI trial it tends to make the arms more similar
How might it be handled?	• Patient perspective: might be interested in outcomes if they adhere → hypothetical for non-adherence *except* for toxicity? • Healthcare perspective: treatment policy?

3. Re-infection with a new strain of TB

How has it been handled?	Re-infected patients were excluded from analysis
Rationale	“Not assessable”
How might it be handled?	• Patient perspective: can the re-infection also be cured? → treatment policy? • Healthcare perspective: aim is to cure the original strain → hypothetical?

4. Death

How has it been handled?	• Unfavourable composite for any death • Except: patients dying from violent/accidental causes were excluded from analysis (REMoxTB)
Rationale	• TB-related death is a bad outcome • But cause of death may be hard to determine
How might it be handled?	• Patient perspective: any death is bad → unfavourable composite strategy? • Healthcare perspective: regimens not designed to prevent non-TB related deaths → hypothetical?

5. Loss to follow-up (not an ICE)

How has it been handled?	• Unfavourable composite strategy for LTFU during treatment phase • LTFU after treatment phase is not an ICE
Rationale	• LTFU during treatment phase taken to predict lack of cure • Same definition used in arm with shorter treatment, to avoid bias
How might it be handled?	• Discontinuation of treatment as an ICE → treatment policy? unfavourable composite? • LTFU as a missing data problem (Rehal et al, submitted)

Conclusions

Although commonly used in TB trials, the current primary outcome definition implies a number of estimand choices that are hard to justify

- Excluding participants from analysis is also hard to justify

The estimand framework provides a common language for discussion and comparison across trials

- Challenges TB trialists to justify and improve their outcome definitions
- Implemented in our recent TB trials (e.g. STREAM Stage 2 and PARADIGM4TB)

Using two motivating trials as examples, we hope to stimulate and inform discussions among trial teams on how estimands should be defined for their trials

Further work is needed to identify practical estimation methods that align with chosen estimand

Acknowledgements

James Carpenter
Angela Crook
Hanif Esmail
Ruth Goodall
Brennan Kahan
Andrew Nunn
Patrick Phillips
Conor Tweed
Ian White

Study with us online for an MSc Clinical Trials or an MSc Statistics in Clinical Trials.
Find out more: https://bit.ly/3EwrvJK

MRC Clinical Trials Unit

UCL

@MRCCTU

www.mrcctu.ucl.ac.uk

Smarter Studies
Global Impact
Better Health

Practical Challenges in Aligning Estimands & Estimation

A MDR-TB Case Study: STREAM Stage 2

Melinda, *Tree of Life*
Melinda's artwork reflects her journey living with HIV.

Gary L Mao, PhD
Statistics & Decision Sciences
Janssen Research & Development, LLC

May 22, 2023

lmao@its.jnj.com

Janssen | PHARMACEUTICAL COMPANIES OF Johnson & Johnson

Disclosures

- STREAM Stage 2 study was sponsored by Vital Strategies and designed and managed by Medical Research Council (MRC) Clinical Trials Unit at UCL
- USAID, Janssen Research & Development, LLC, and MRC are funding partners of the study
- Information shared in this presentation is based on the primary manuscript of the study*

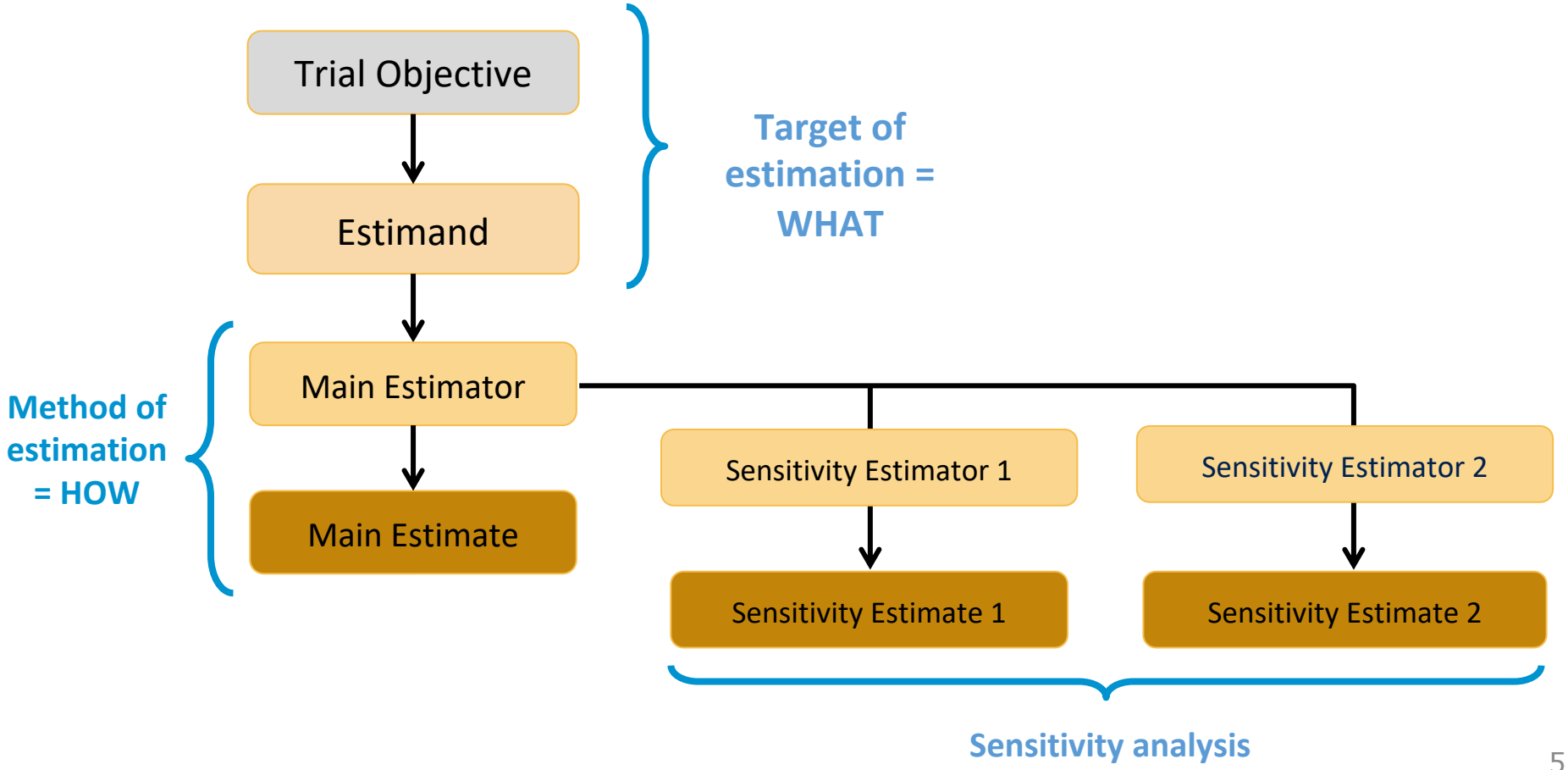
* Goodall RL, Meredith SK, Nunn AJ, et al. Evaluation of two shorter standardised regimens for the treatment of rifampin-resistant tuberculosis (STREAM stage 2): an open-label, multicentre, randomised, non-inferiority trial. Lancet 2022; 400: 1858–68

Outline

- **Practical points on the implementation of the estimand framework – a pharma perspective**
  - Good practices when defining the attributes of an estimand
  - Strategies for intercurrent events (ICEs)
  - Steps from Estimand to Estimator
- **A MDR-TB case study: STREAM stage 2**
- **Key takeaways**

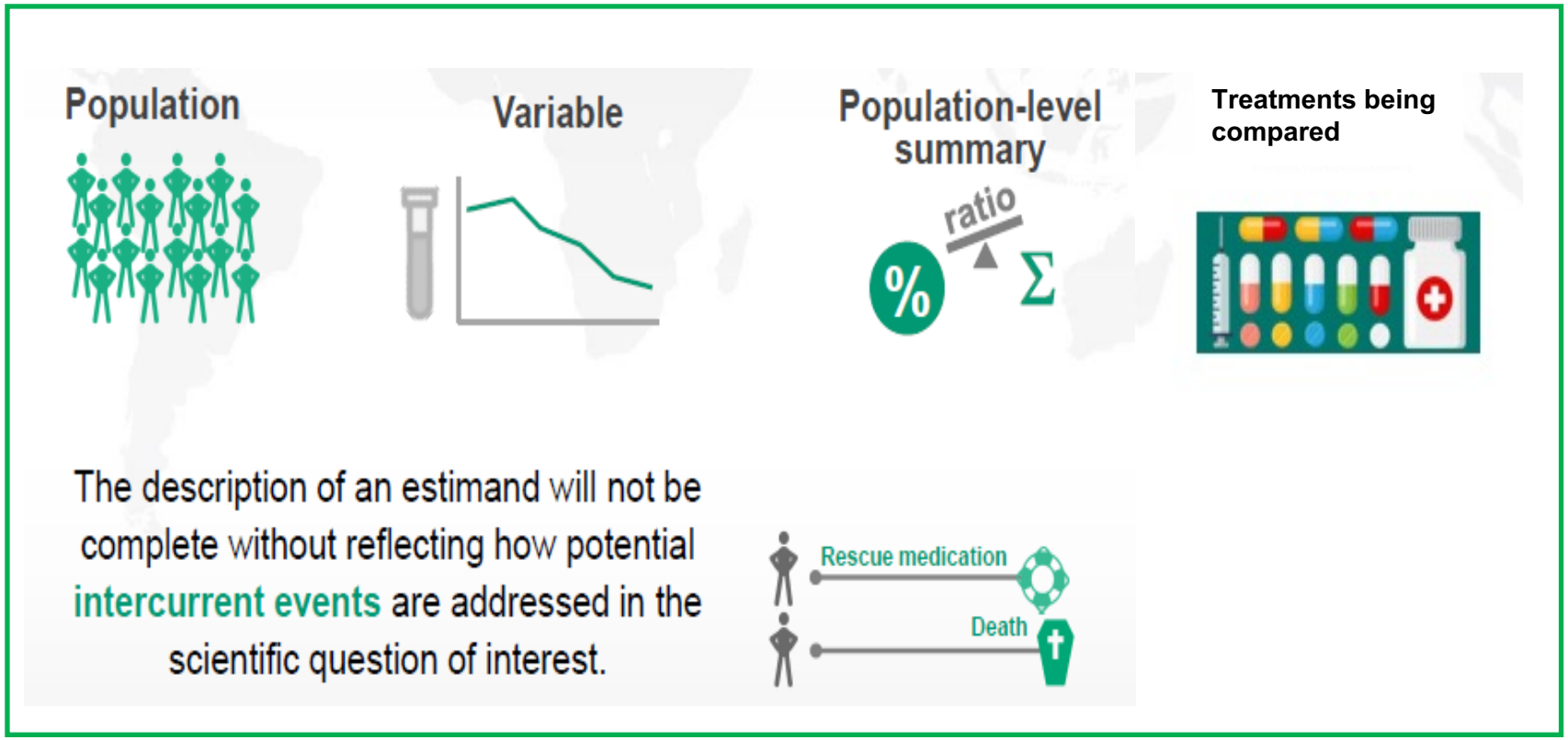
Practical points on the implementation of the estimand framework – a pharma perspective

Aligning target of estimation, method of estimation, and sensitivity analysis, for a given trial objective

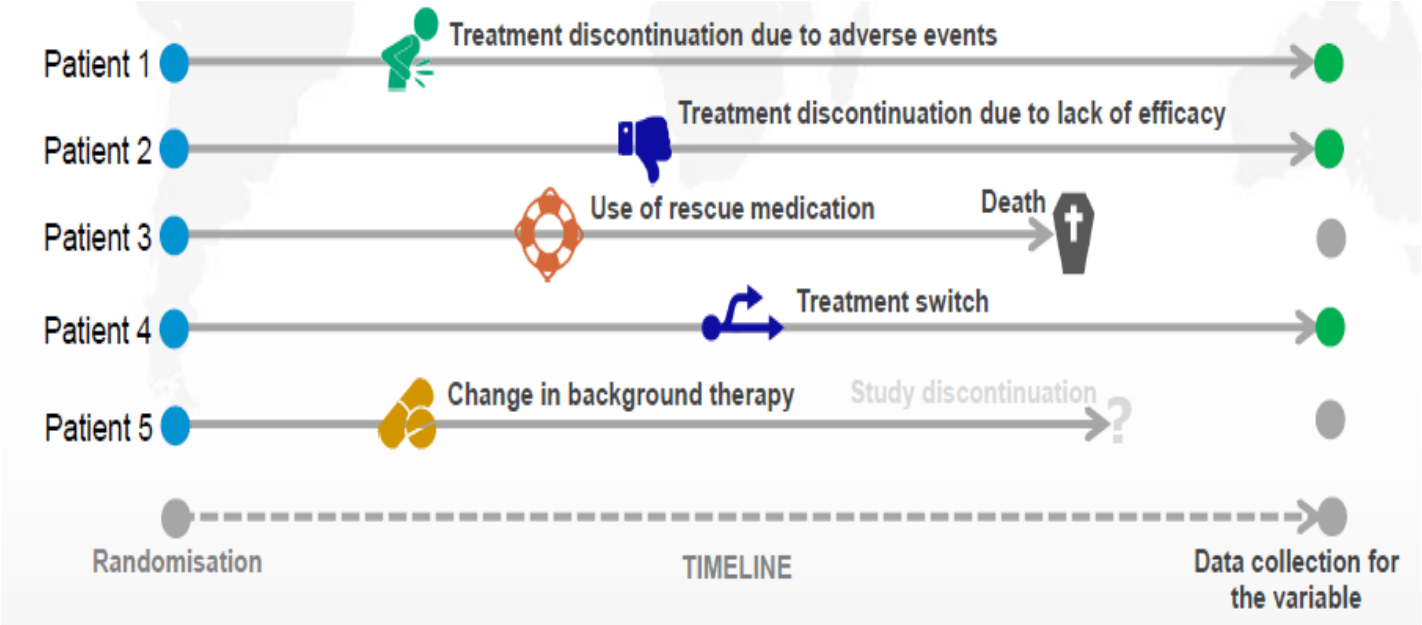


Description of an Estimand

- **An estimand defines in detail what needs to be estimated to address a specific scientific question of interest.**
  - a description of the estimand includes **5 attributes**



Intercurrent Events (ICE) Examples



- Each identified ICE requires a particular strategy.
- The chosen strategies need not be the same for all ICEs.

ICE Strategies

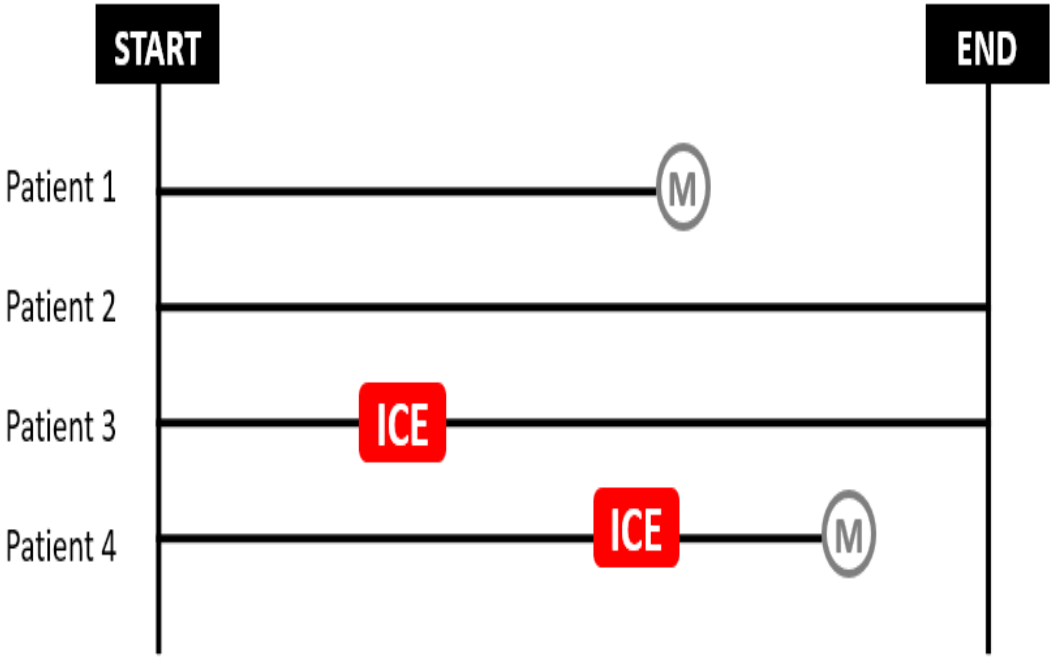
STRATEGY	DESCRIPTION: How is the ICE addressed?
Treatment Policy	In general, as part of the **treatment**/by disregarding its occurrence – NA when values of the variable do not exist after the ICE (e.g. after death) – more generally acceptable to support regulatory decision making
Hypothetical	As **model-based predictions** under the assigned treatment – consider reasonable scenarios
Principal Stratum	As part of the **population** – target population is a subset than the general patient population – can have implications on the label if considered for the primary estimand
Composite	As part of the **variable** – the ICE itself is meaningful and is thus reflected in the value that will be assigned to the variable (e.g., a non-responder status) More rarely, as part of the **population-level summary**: e.g. difference of trimmed means, where observations of patients with ICEs are trimmed
While-on-Treatment	As part of the **variable** – interest is on the response to treatment prior to occurrence of the ICE – variable is defined such that it is independent or incorporates the patient specific treatment duration (e.g., slope/rate, AUC); summary measure cannot be defined for a fixed time period

ICEs and Missing Data

Missing Data:
Data that would be meaningful for the analysis of a given estimand but were not collected. They should be distinguished from data that do not exist or data that are not considered meaningful because of an intercurrent event.

- Missing data:
  - After study withdrawal
  - After trial termination
  - Due to missed visits or measurements
- Data that do not exist:
  - After death
- Data that are not considered meaningful due to an ICE (data not used in analysis):
  - After an ICE addressed with a hypothetical strategy

ICE and Missing Data



Example: M = Study withdrawal

Patient 1 **has missing data**, but **NO ICEs**.

Patient 2 has **NO missing data** and **NO ICEs**.

Patient 3 **has an ICE** and **NO missing data**. Whether or not the post-ICE data is relevant depends on the ICE strategy.

Patient 4 **has an ICE** and depending on the ICE strategy the patient **may or may not have missing data**.

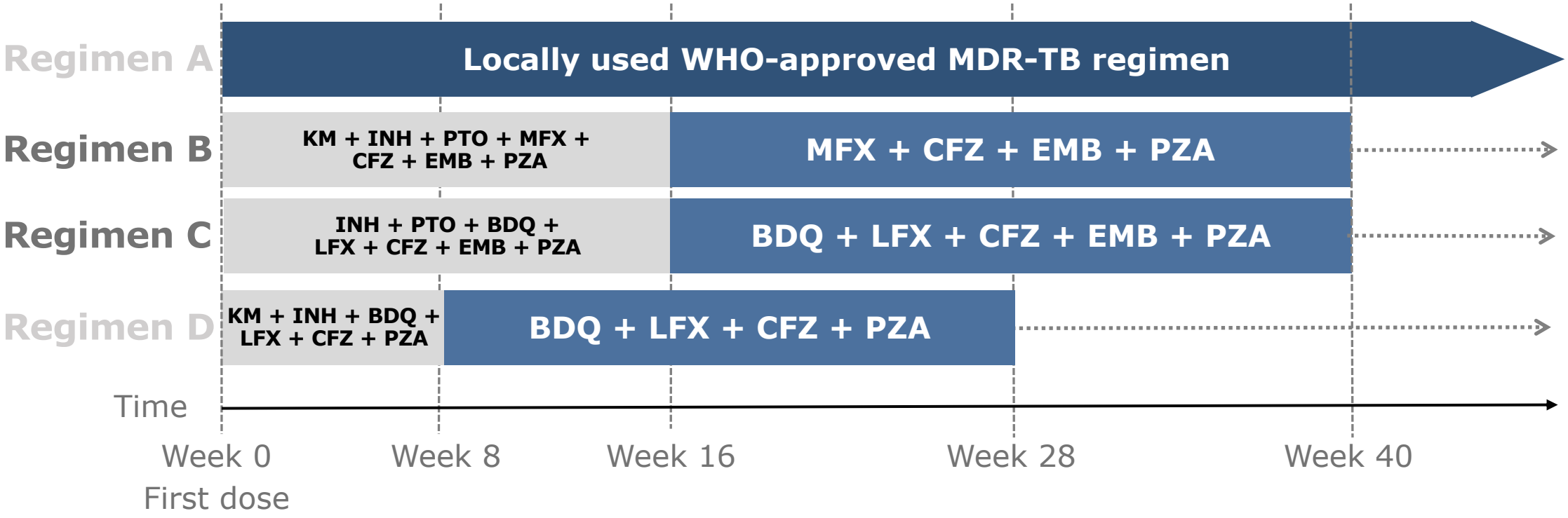
Steps from Estimand to Estimator

- Define the question of interest linked to the trial objective, for which the estimand is defined
- Define all attributes of an estimand, considering all intercurrent events relevant to the trial
- Consider the trial design and any key implementation elements needed to address that estimand
- For estimator specifications:
  - Define data to be included vs data not used or missing
  - Specify assumptions, including model assumptions and the assumptions for data that are not used or missing
  - Define the Main estimator
  - Define Sensitivity estimator(s) – describe what assumptions of the main estimator are changed

Implementation of the estimand framework
MRD-TB case study: STREAM stage 2

STREAM stage 2 Study Design

Evaluation of shorter standardized regimens of anti-tuberculosis drugs for patients with MDR-TB



Goodall RL, Meredith SK, Nunn AJ, et al. Evaluation of two shorter standardised regimens for the treatment of rifampin-resistant tuberculosis (STREAM stage 2): an open-label, multicentre, randomised, non-inferiority trial. Lancet 2022; 400: 1858–68

From Research Question to Estimand Specification

Research question of interest: For patients with multidrug resistant tuberculosis, is treatment with the all oral short-course regimen non-inferior to the injectable-containing short-course control regimen, measured by difference in proportion of patients achieving a favorable outcome at week 76?

For patients with multidrug resistant tuberculosis	population
Is treatment with	
the all oral short-course regimen non-inferior to the injectable-containing short-course control regimen,	treatment
measured by	
difference in proportion of patients achieving a	population-level summary measure
favorable outcome at week 76	*variable*
?	

Primary Outcome Measure – STREAM stage 2

Patient-level measure/variable: favorable status at 76 weeks, defined as a negative culture for *M tuberculosis* at week 76 and on the preceding visit, with no intervening positive culture or previous unfavorable outcome.

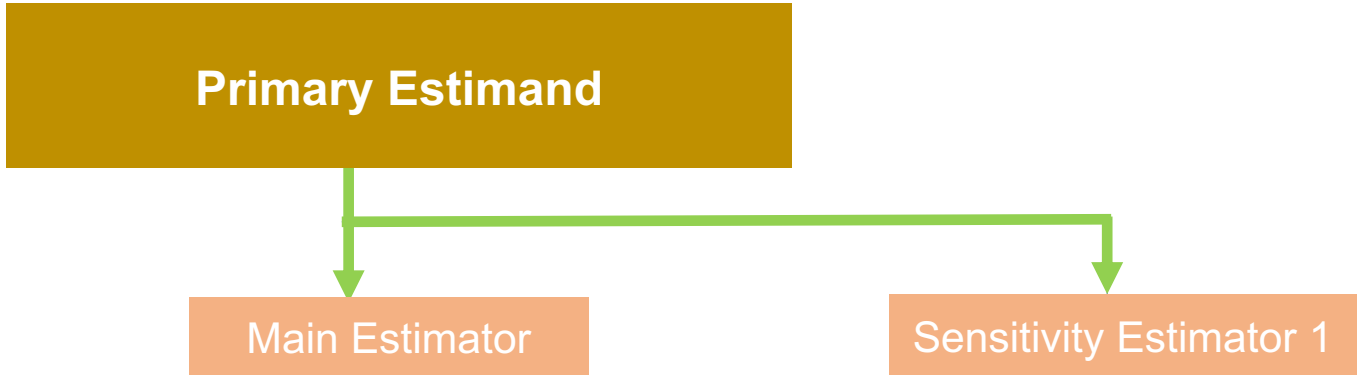
Unfavorable outcomes include:

- Treatment discontinuation (ICE)
- Initiation with bedaquiline, kanamycin, linezolid, or two or more other drugs if they were not part of the assigned regimen (ICE)
- Treatment extension beyond the permitted duration (ICE)
- Death from any cause (ICE)
- Reinfection (ICE)
- A positive culture from one of the two most recent specimens

Primary Estimand: Intercurrent Events

Intercurrent event	Strategy addressing ICE
Treatment discontinuation	**Composite strategy:** this ICE is defined as an unfavorable outcome • Rationale: event likely represents poor treatment outcome • Limitation: certain D/Cs might not imply poor treatment outcome
Major treatment modifications (treatment switch, treatment extension, treatment addition)	**Composite strategy:** this ICE is defined as an unfavorable outcome • Rationale: events likely represent inadequate/lack of efficacy • Limitation: certain modifications might have ignorable impact on treatment outcome
Death from any cause	**Composite strategy:** this ICE is defined as an unfavorable outcome • Rationale: event might be a result of disease progression • Limitation: not all deaths are treatment-related (accident, trauma)
Reinfection	**Composite strategy:** this ICE is defined as an unfavorable outcome • Rationale: event may represent poor treatment outcome • Limitation: reinfection could be with a different strain

From Estimand to Estimators



Data used: mITT analysis set

Assumptions:
- treatment outcome for HIV+ and HIV- patients are different
- treatment outcome for patients randomized under different protocol versions are different

Missing data handling:
- Missing culture data at week 76 due to COVID-19 or other reasons was imputed to unfavorable

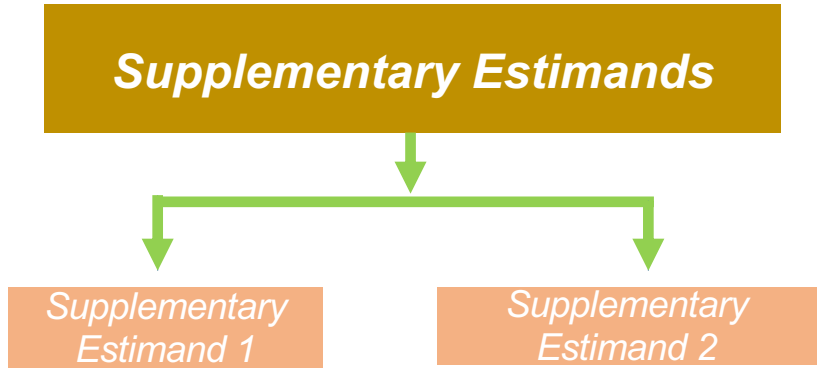
Estimator:
- Difference in proportion of patients with favorable outcome at week 76 estimated using Cochran-Mantek-Haenszel weight, adjusting for HIV status and randomization protocol

Data used: mITT analysis set

(Alternative) Assumptions:
- treatment outcome for HIV+ and HIV- patients are not different
- treatment outcome for patients randomized under different protocol versions are different

Missing data handling: same as main estimator

Estimator:
- Difference in proportion of patients with favorable outcome at week 76 estimated using Cochran-Mantek-Haenszel weight, adjusting for randomization protocol



(Alternative) Population: patients with confirmed diagnosis of MDR-TB completing protocol-adherent course of treatment (Per-protocol analysis set)

Main Estimator

Data used: Per-protocol analysis set

Missing data handling: same as main estimator

Estimator: same as main estimator

(Alternative) ICE strategy:
- Treatment policy strategy for ICE of reinfection due to a different strain

Main Estimator

Data used: mITT analysis set

Missing data handling: same as main estimator

Estimator: same as main estimator

Key Takeaways

Implementation challenges:

- Multiple objectives & scientific questions of interest
  - The primary estimand <-> primary questions of interest
  - Supplemental estimands <-> additional questions of interest
  - Key estimands?

- Choice of ICE strategies plays major role in defining treatment outcome

- HA endorsement

Acknowledgements

Janssen Estimands and Missing Data Working Group

Elena Polverejan	Cristina Sotto		
An Vandebosch	Wouter Willems	Ilse Van Dromme	Sjouke Vanderberghe
Kim Hung Lo	Serene Jiang	Hsiaowei Deng	Kelle Van Lancker
Lixia Pei	Rui Qin	Liwei Wang	Hege Michiels
Yun Zhang	Nancy Chen		

HARVARD T.H. CHAN
SCHOOL OF PUBLIC HEALTH

Estimands for clinical endpoints in Tuberculosis randomized controlled trials: an application in a completed trial

Isabelle R. Weir, PhD
Center for Biostatistics in AIDS Research
Harvard T.H. Chan School of Public Health

SCT session “The What and the How: Practical challenges in aligning estimands and estimation using tuberculosis as a case study”

joint work with Suzanne Dufault and Patrick Phillips

May 22, 2023

iweir@sdac.harvard.edu

Disclosures

Nothing to disclose.

Primary outcome in TB treatment trials

Systematic review of 31 TB treatment trials[1]

- Broad consensus in the components of the composite outcome
- Heterogeneity in specific definition and application of statistical methods

Points with consensus

- Outcome determined after at least 1 year follow-up
- Components:
  - Events: deaths, treatment issues, bacteriologic findings, recurrence
  - Timing: *during* treatment versus *after* the end of treatment

[1]Hills et al. Trials 2021.

Limitations of the ‘standard approach’

Outcome definitions are not standardized

- Challenging to combine data, interpret results in the context of historic trials

Outdated

- Simple, unadjusted “per-protocol” analyses do not take into account causal framework
- At odds with regulatory guidance

Mixing of safety and efficacy events

- Inclusion of treatment changes due to AEs during treatment conflates safety/tolerability with efficacy

Barrier to identifying highly efficacious regimens

- Including losses to follow-up and non-TB death in composite outcome increases variability in treatment effect estimates and requires a larger sample size.
- Impedes progress in prediction modeling

ICH E9(R1) → opportunity for broad standardization

TB Estimand Proposal

Modernization of Endpoints and Estimands of Late-phase Tuberculosis Therapeutic Trials

Objective:

- To improve the specification and standardization of endpoint and estimand definitions across protocols, analysis plans, and reports from last phase pulmonary TB therapeutic trials.

Key components:

- Education on the ICH E9(R1) framework in the context of TB treatment trials
- Listing of 35 intercurrent events (ICE) reasonable to expect in most trials
- 4 estimands defined for various trial stakeholders
  - Differentiated by the strategies applied to handle ICEs

OSF

SCAN ME

Estimand Components

Intercurrent Events

Treatment Events

4 Estimands (the 'what')

TB Specific Estimand

- Focuses on the effect of the treatment exclusively on TB disease outcomes.
- To disaggregate efficacy events from non-TB related adverse events and other events.
- Hypothetical

Composite Estimand

- Cautious, assuming a worst-case scenario.
- To provide a cautious estimate of the treatment effect assuming many ICE indicate absence of cure, following a strict 'ITT' approach.
- Hypothetical, treatment policy, composite absence of cure

Assessable Estimand

- A middle group between the TB-specific and composite estimands, providing a link to previous trials
- To provide a treatment effect estimate more closely aligned with previous trial analyses.
- Hypothetical, treatment policy, composite absence of cure

Per Protocol Estimand

- Targets the treatment effect among the sub-group of TB patients that take an adequate treatment course.
- To provide the treatment effect in the group of participants who comply with key components of the protocol including treatment adherence.
- Hypothetical, principal stratum

Estimation (the ‘how’)

Strategies to handle ICEs **&** method for estimating the population summary measure

Composite strategy

- Occurrence of ICE is mapped to absence of cure (or in some cases, durable cure)

Treatment Policy strategy

- Occurrence of ICE is ignored when estimating the treatment effect

Hypothetical strategy

- Multiple Imputation (MI)
- Inverse probability of censoring weighting (IPCW)

Principal stratum strategy

- Bayesian statistical model (prior distribution incorporates model assumptions such as probability of ICE occurrence across levels of treatment)

Estimation (the ‘how’)

Strategies to handle ICEs **&** **method for estimating the population summary measure**

Population summary measure = Difference in risk of unfavorable clinical outcome

TB Specific, Composite, Assessable Estimands

- Cochrane Mantel Haenszel
- Kaplan-Meier

Per Protocol Estimand

- Posterior risk difference (using Bayesian framework)

Illustrative example: REMoxTB Trial

Phase 3, non-inferiority trial comparing two 4-month regimens against 6-month SOC

Primary outcome measure: proportion of participants who experienced a composite unfavorable outcome defined by bacteriologically or clinically defined failure or relapse within 18 months after randomization

N=1931

Non-inferiority for these regimens was not shown (margin 6%)

Originally analyzed with legacy mITT and PP principles

Four-month Moxifloxacin-based regimens for drug-sensitive Tuberculosis. *Gillesepie et al. NEJM 2014.*

Re-analysis methods

We reanalyzed the individual participant level REMoxTB trial data according to each of the 4 estimands and in the original mITT population.

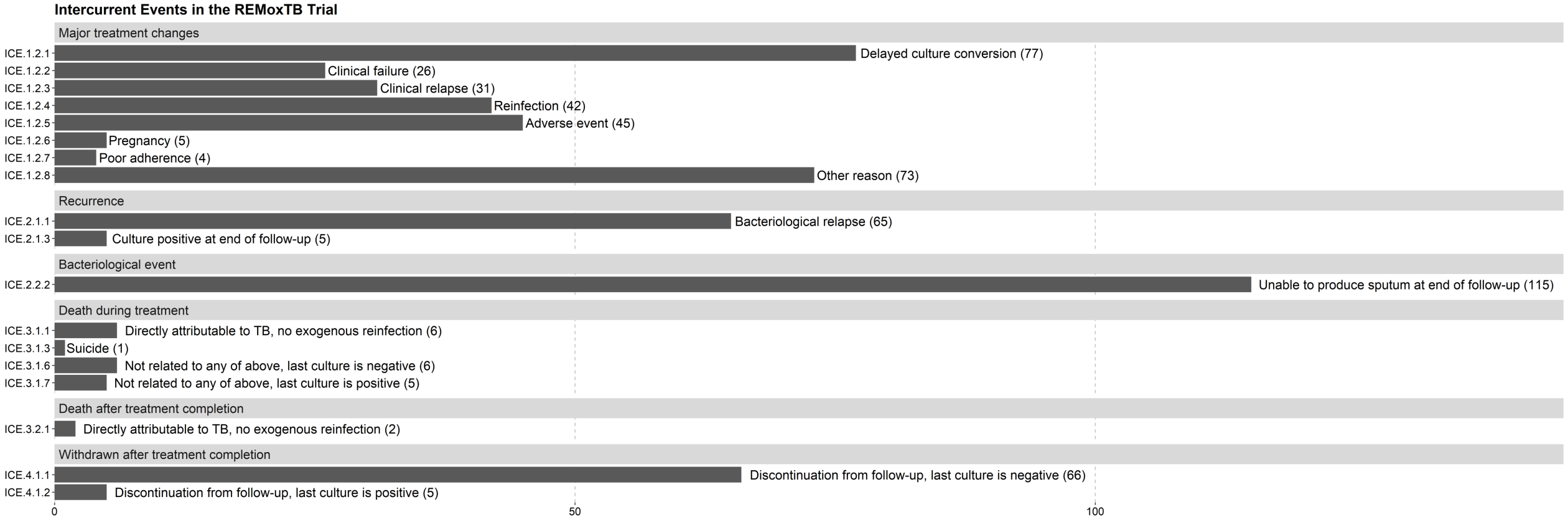
1. Identification of intercurrent events
2. Estimation of difference in risk of unfavorable clinical outcome at 18 months after randomization
   - Comparing each experimental arm against SOC
   - Same 97.5% CI and non-inferiority margin of 6%
   - 4 Estimands
   - Re-analysis of the mITT population from original trial analysis

Reanalysis methods

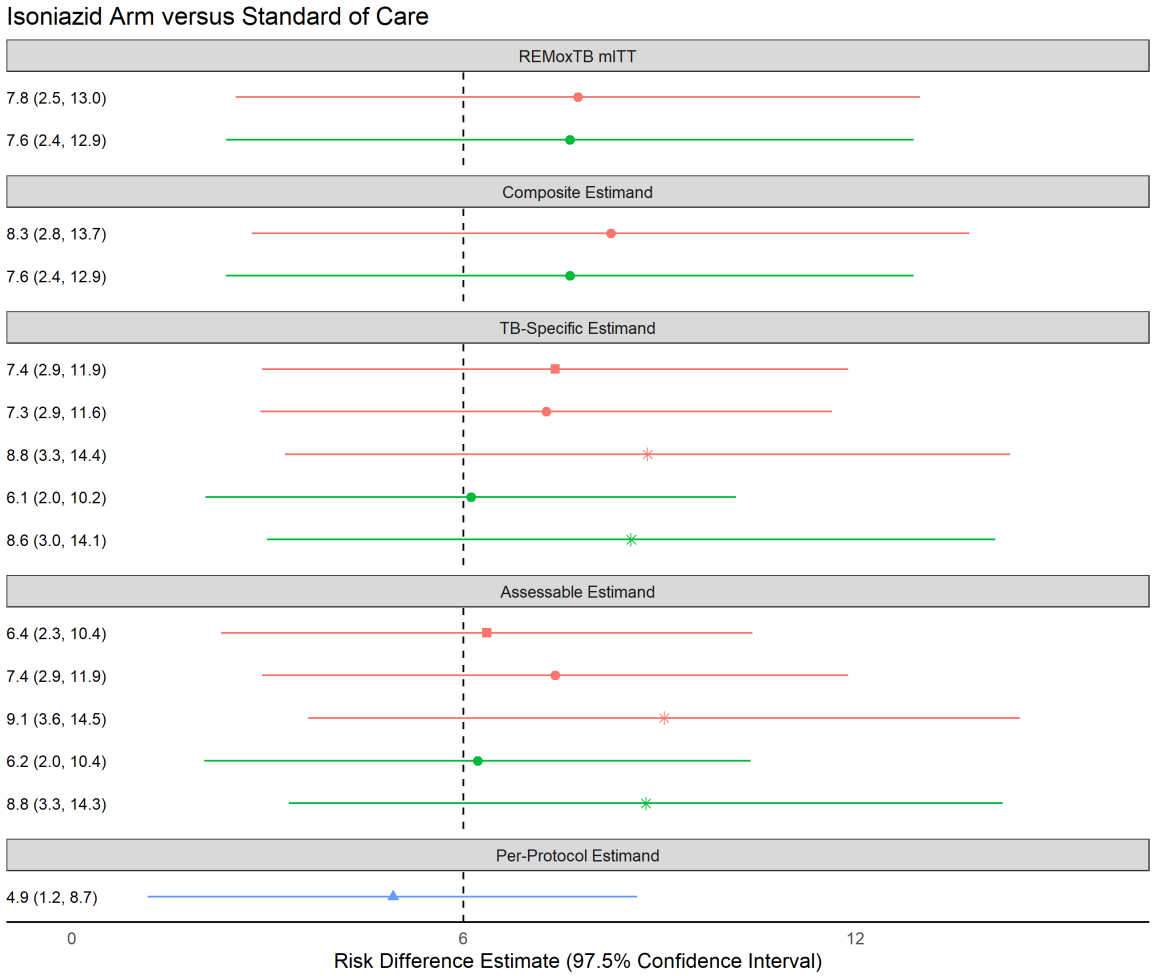
Estimand(s)	Cochrane Mantel Haenszel	Kaplan-Meier	Bayesian Framework
• TB-Specific • Assessable	• Naïve approach assuming hypothetical ICEs are durable cure • MI	• Naïve approach censoring participants with hypothetical ICEs at the time of the ICE occurrence • MI • IPCW	
• Composite • REMoxTB mITT	• No advanced methods	• No advanced methods	
• Per-Protocol			• Individual membership within principal strata is based on the unobservable distribution of ICEs given the observed and counterfactual intervention assignments. • The estimand is not fully identifiable, but inference is obtained by placing Bayesian priors reflective of the necessary assumptions on the probability model.

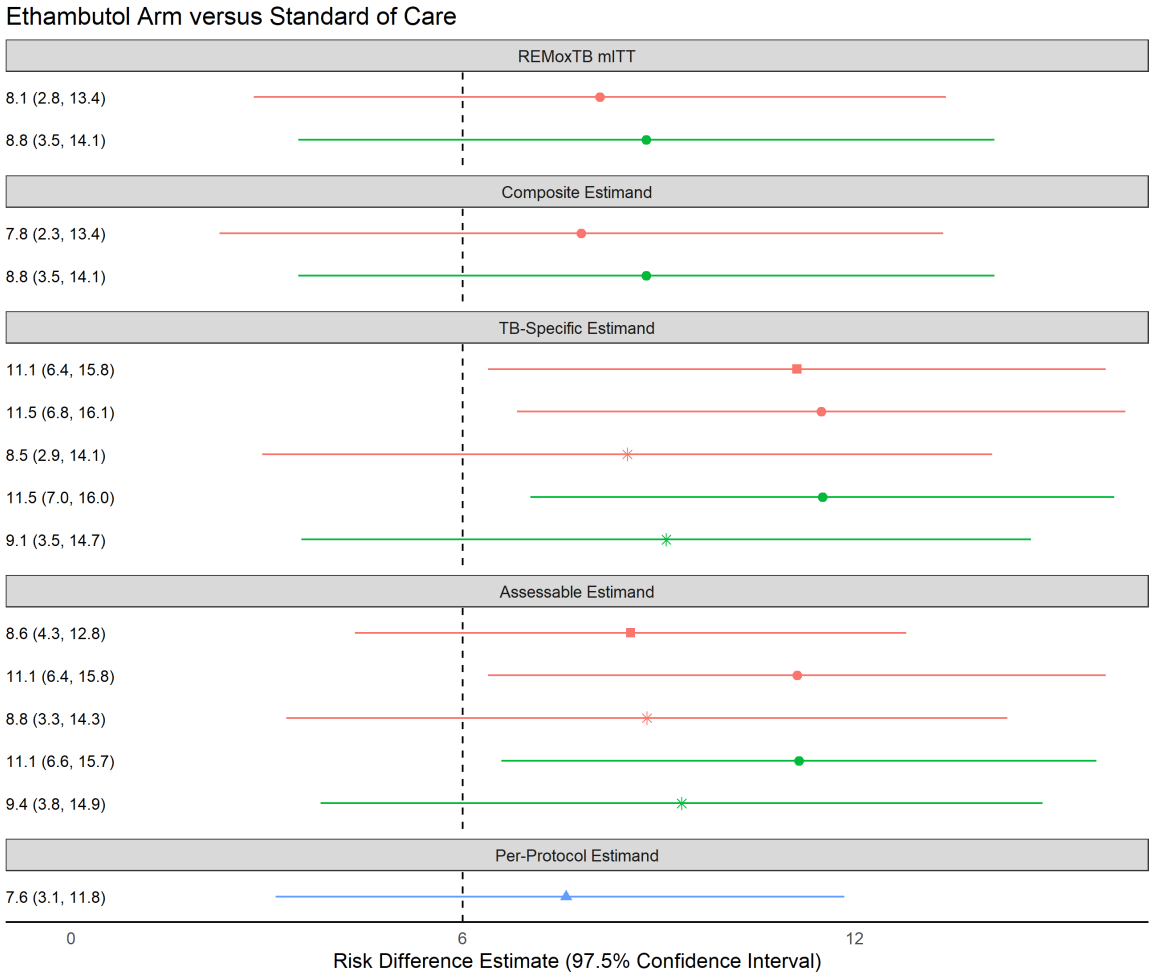
Results - ICEs

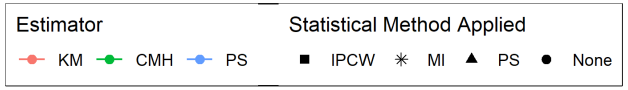
- 68% had ‘durable cure’ at 18 months
- 32% experienced an ICE



Results – estimation







Conclusions

1. We have proposed 4 estimands for use in late phase TB treatment RCTs and have demonstrated their use with the historic REMoxTB trial. We found consistent conclusions with each estimand and method of estimation.
    - Applying more complex statistical analysis methods did not lead to sizable differences in the treatment effect estimates.
    - No single estimand gives a more true or less biased treatment effect estimate.
    - Our analysis is limited by the retrospective application of the estimand framework.
2. Our proposed estimand framework aligns with ICH E9(R1) and gives trialists a thorough starting point for estimand specification when designing future TB treatment randomized controlled trials.
3. Our TB estimand proposal is a living document to be updated with implementation in future trials

University of California
San Francisco

Center for
Tuberculosis

University of California
San Francisco

Beyond estimand specification

Considerations for estimand-aligned estimation

Suzanne M. Dufault, PhD

2023-05-22

ICH E9 (R1) addendum provides a framework for increasing consistency and transparency in clinical trials.

Estimand

a precise description of the treatment effect reflecting the clinical question posed by the trial objective (i.e. "what is to be estimated")

5 Attributes

1. Treatment
2. Population
3. Handling of ICEs
4. Outcome ☹☺
5. Population-level summary θ

Once this is defined… how will you perform estimation?

Goal: estimand-aligned estimation

Source: ICH E9(R1) Training Materials

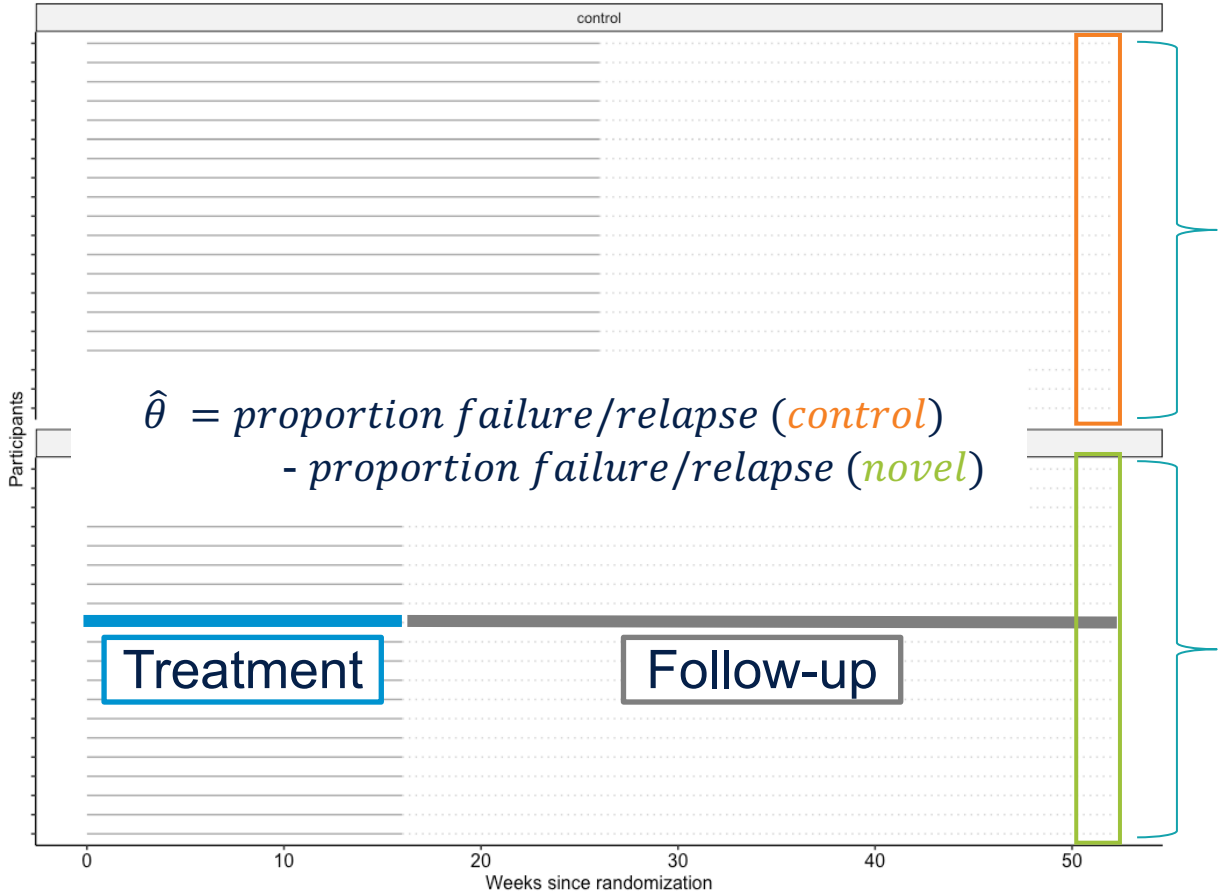
Goal: Provide practical guidance for estimand-aligned estimation.

We used a simulation study to identify:

1. Appropriate estimators for complex estimands.
   - We learned:
     - Real-world practicalities
     - Roadblocks
     - Computational constraints
2. How different ICE settings affect estimation.
   - We investigated:
     - Which estimands are easy to estimate, but perhaps prone to misleading results?
     - Which types of ICEs introduce the most estimand deviation?

Are we asking the right questions and are we estimating efficiently?

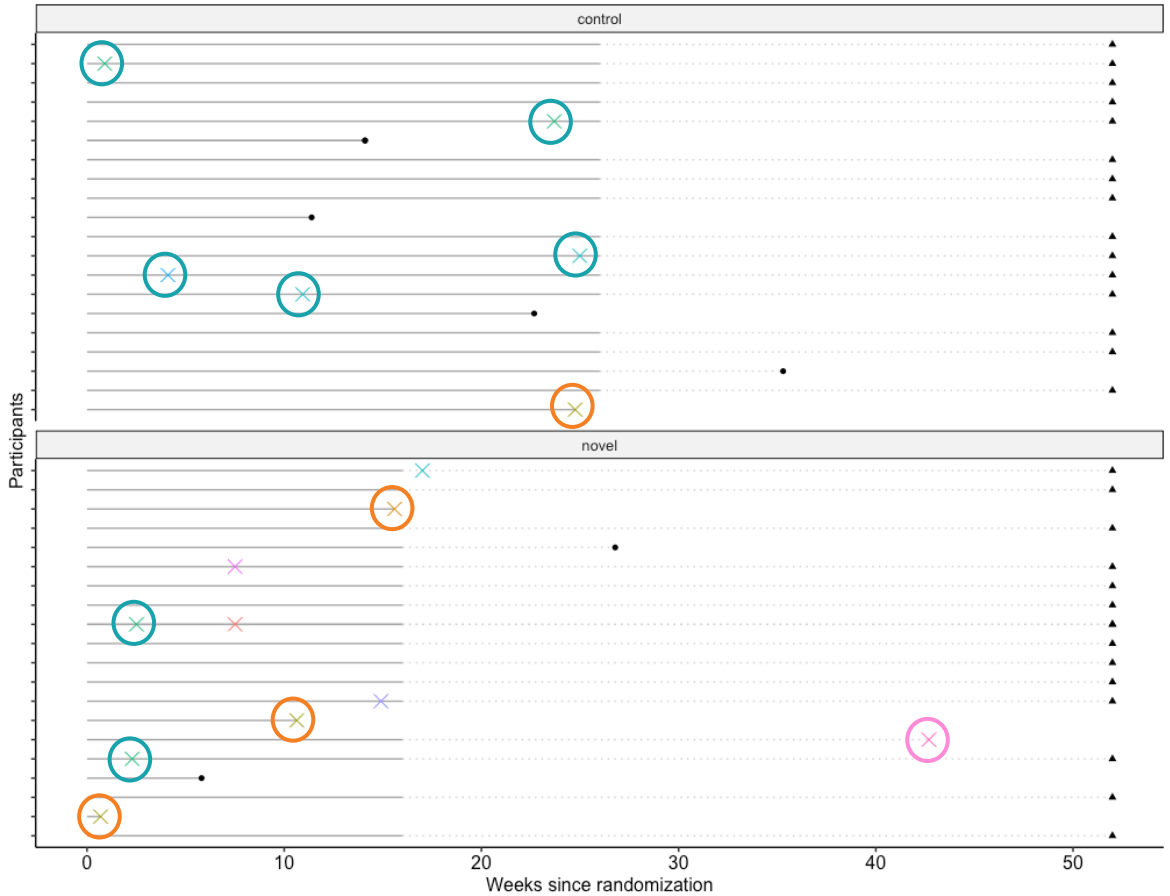
Description of simulation study



Participants randomized to control regimen

Participants randomized to novel regimen

UCSF



Major treatment changes

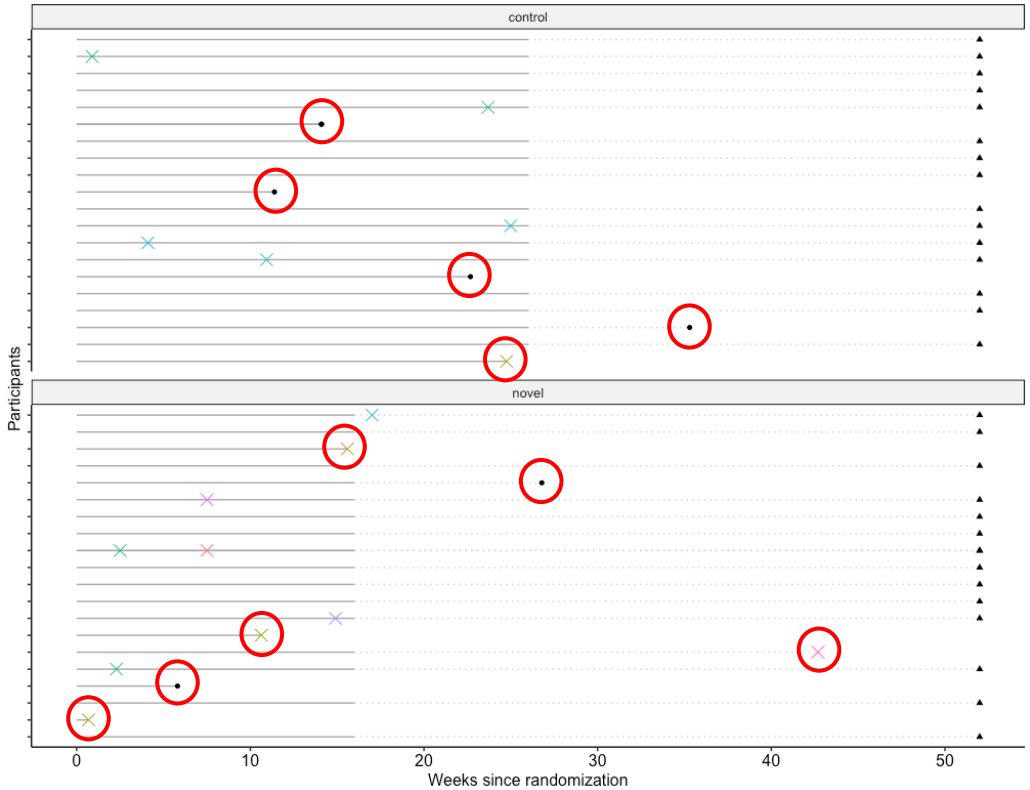
Bacteriological events (e.g., reinfection without change in treatment)

Death during treatment

Loss to follow-up

Recent work found 35 distinct ICEs in Phase III TB trials, each with different implications for estimation.

UCSF



4 failures/relapses
+ 1 unobserved outcome
= 5 unfavorable outcomes

Composite Estimand

$$\hat{\theta} = \frac{5}{20} - \frac{6}{20} = -0.05$$

2 failures/relapses
+ 4 unobserved outcomes
= 6 unfavorable outcomes

UCSF

4 failures/relapses
+ 1 hypothetical failure
= 5 unfavorable outcomes

Assessable Estimand

$$\hat{\theta} = \frac{5}{20} - \frac{4}{20} = 0.05$$

2 failures/relapses
+ 2 hypothetical failures
= 4 unfavorable outcomes

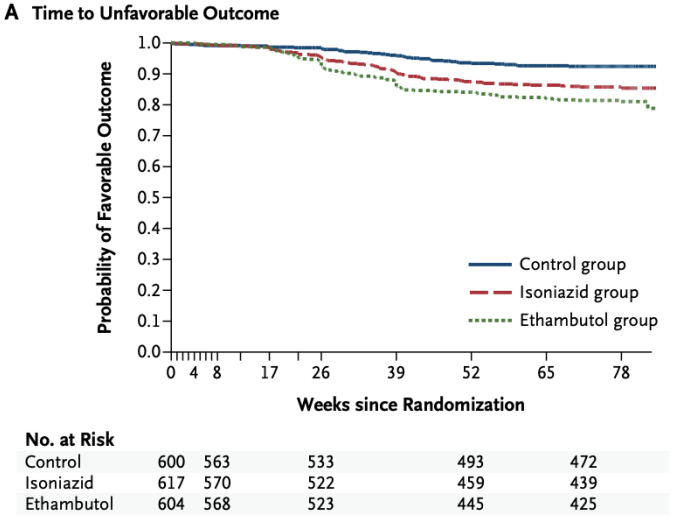
Estimand	ICE Strategies	Estimation
Composite	Ignoring the following ICEs (treatment policy): (#7) Minor treatment changes (#9) Discontinuation from treatment (#10) Bacteriological events: reinfection without change in treatment (#11) Other bacteriological events: isolated positive culture Incorporating ICE into the outcome (composite strategy): (#1) Unfavorable outcome (failure/relapse) (#2) Major treatment changes: reinfection (#3) Major treatment changes: adverse event (#4) Major treatment changes: poor adherence (#5) Death during/after treatment: suicide, related to treatment or other (last culture negative) (#6) Withdrawal or loss to follow-up after treatment completion (last culture is negative) (#8) Major treatment changes: other reason including pregnancy (#12) Other bacteriological events: unable to produce sputum at end of follow-up (#13) Death during treatment: accident or trauma	1. Generate a new composite outcome. 2. "Ignore" treatment policy ICEs. (If available, use outcome as recorded. If not available, drop or censor these individuals). 3. Perform Kaplan-Meier estimation.

Estimand	ICE Strategies	Estimation
Assessable	Ignoring the following ICEs (treatment policy): (#7) Minor treatment changes (#9) Discontinuation from treatment (#10) Bacteriological events: reinfection without change in treatment (#11) Other bacteriological events: isolated positive culture Use IPW or MI for the following ICEs (hypothetical): (#2) Major treatment changes: reinfection (#3) Major treatment changes: adverse event (#6) Withdrawal or loss to follow-up after treatment completion (last culture is negative) (#8) Major treatment changes: other reason including pregnancy (#13) Death during treatment: accident or trauma (#14) Death after treatment completion: suicide, accident or trauma, other (last culture negative) Incorporating ICE into the outcome (composite strategy): (#1) Unfavorable outcome (failure/relapse) (#4) Major treatment changes: poor adherence (#5) Death during/after treatment: suicide, related to treatment or other (last culture negative) (#12) Other bacteriological events: unable to produce sputum at end of follow-up	1. Generate new composite outcome. 2. “Ignore” treatment policy ICEs. (If available, use outcome as recorded. If not available, drop **or censor** these individuals). 3. Generate individual-level propensities for ICE occurrence. 4. Perform IPCW *weighted* Kaplan-Meier estimation.

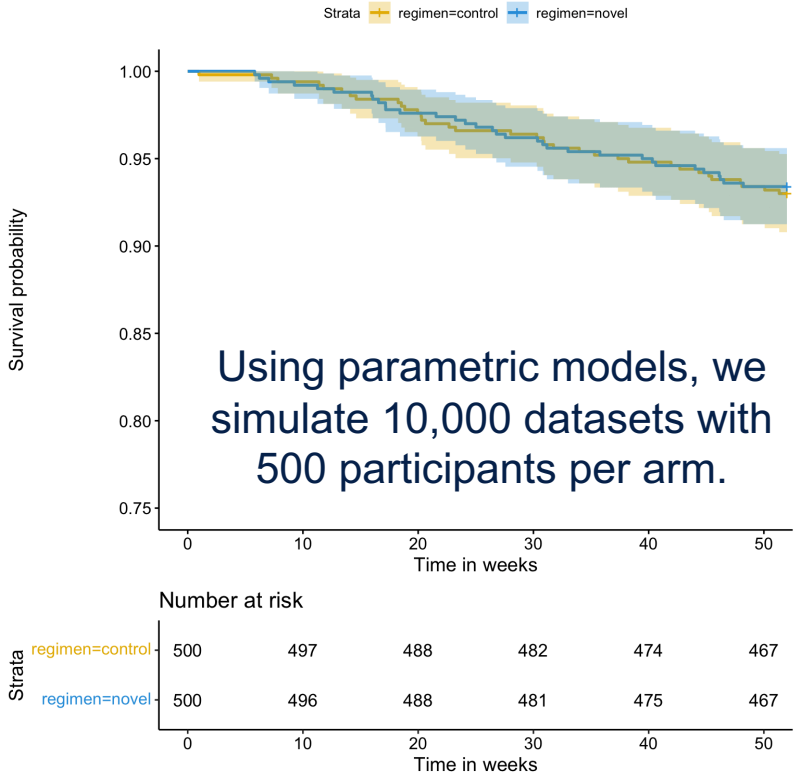
Simulating the primary outcome

Time-to-event: Cure versus unfavorable outcome (e.g., failure/relapse)

Source: Gillespie, Stephen H., et al. "Four-month moxifloxacin-based regimens for drug-sensitive tuberculosis." *New England Journal of Medicine* 371.17 (2014): 1577-1587.

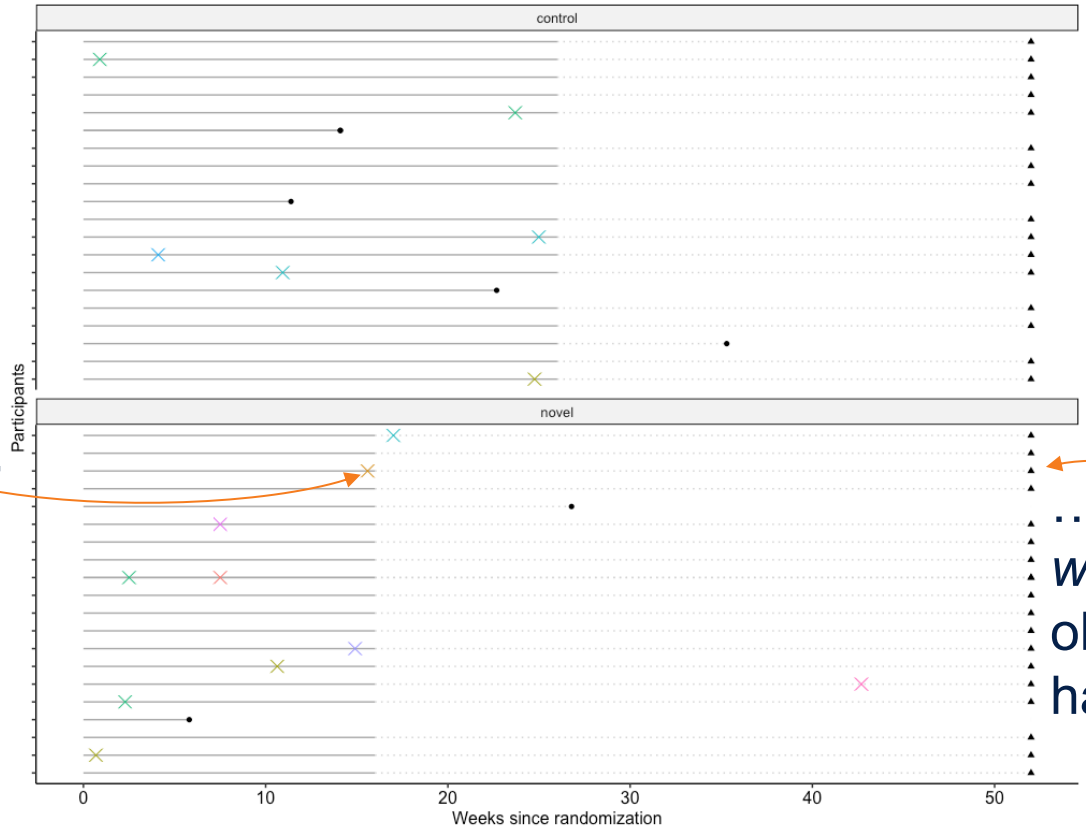


No. at Risk					
Control	600	563	533	493	472
Isoniazid	617	570	522	459	439
Ethambutol	604	568	523	445	425



Example: Simulated data

Simulate ICEs…



… and what *would* have been observed if ICE hadn't occurred

Results

Goal: Provide practical guidance for estimand-aligned estimation.

We used a simulation study to identify:

1. Appropriate estimators for complex estimands.
   - We learned:
     - Real-world practicalities
     - Roadblocks
     - Computational constraints
2. How different ICE settings affect estimation.
   - We investigated:
     - Which estimands are easy to estimate, but perhaps prone to misleading results?
     - Which types of ICEs introduce the most estimand deviation?

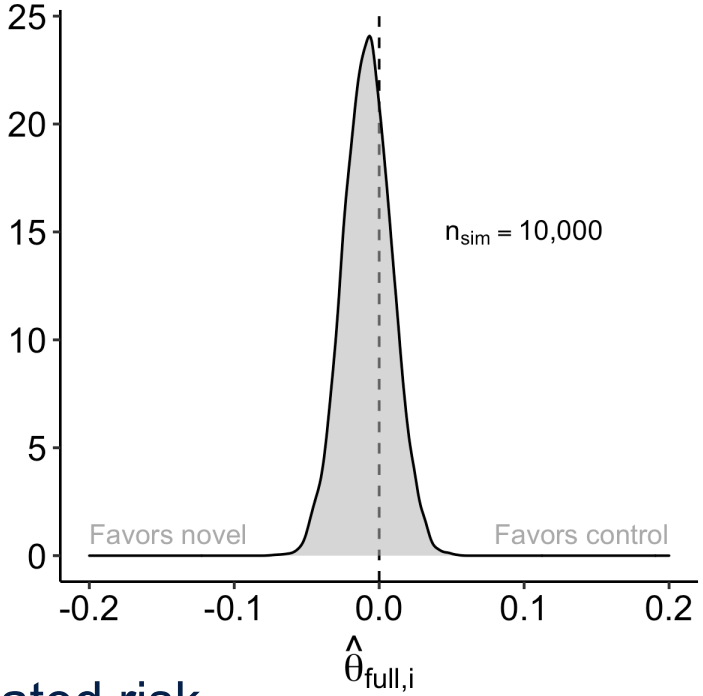
Are we asking the right questions and are we estimating efficiently?

Full Data Results

Checking for errors in the simulation data.

Conditions

- $n_{novel} = n_{control} = 500$
- $S_{novel}(52) = S_{control}(52) = 0.93$



What is the simulation distribution of estimated risk difference, $\hat{\theta}_{full} = \hat{S}_c(52) - \hat{S}_n(52)$, if *no* ICEs had occurred?

If you have experience with the simsurv package in R and complex user-written hazard functions, I would love to pick your brain!

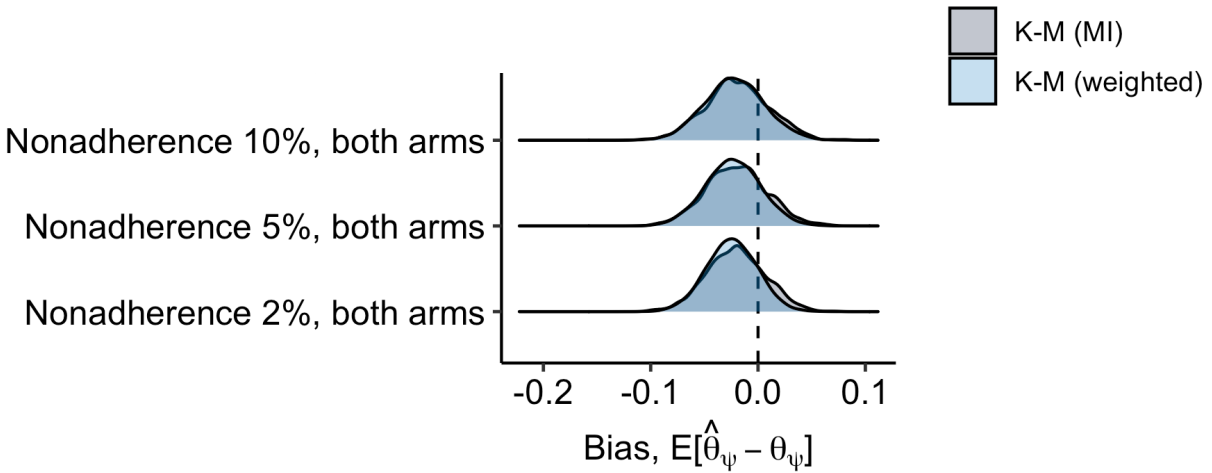
UCSF

Estimator Performance

Are we estimating efficiently?

When we've chosen an estimand, which methods are preferred for robust results?

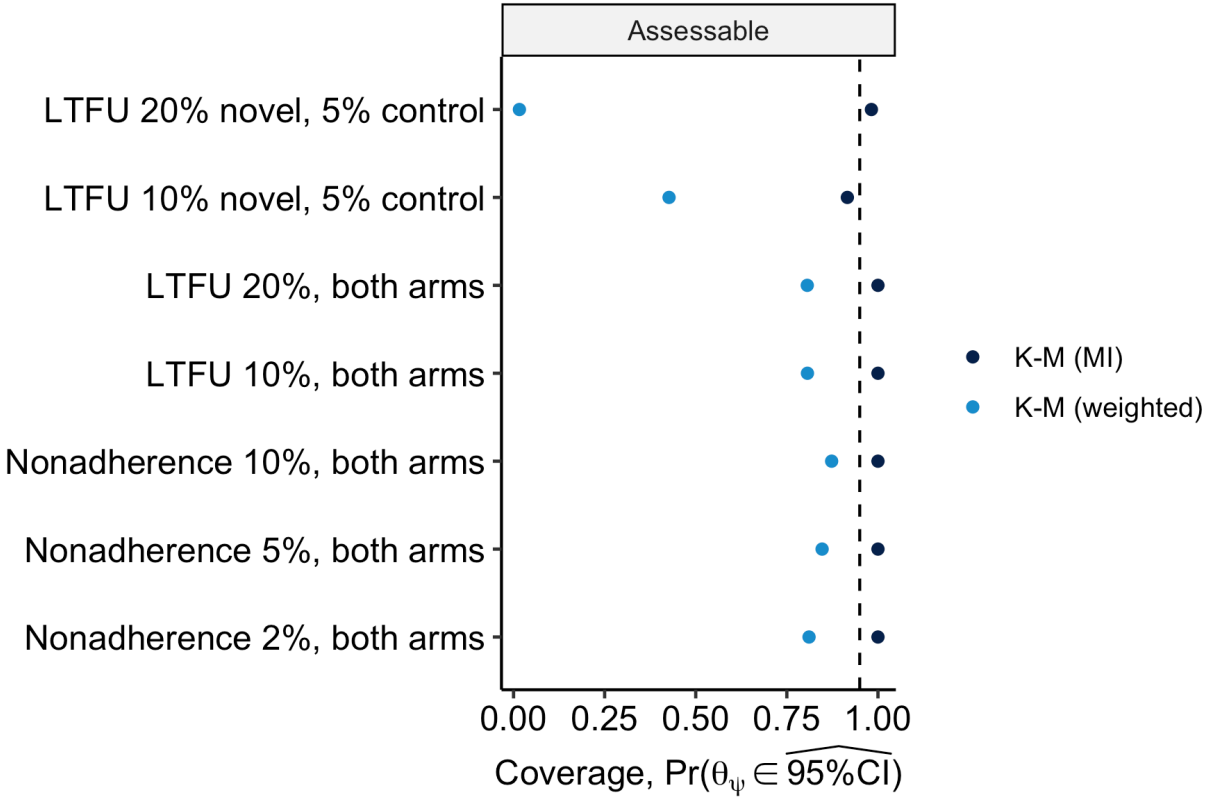
Assessable



UCSF

Are we estimating efficiently?

When we've chosen an estimand, which methods are preferred for robust results?



UCSF

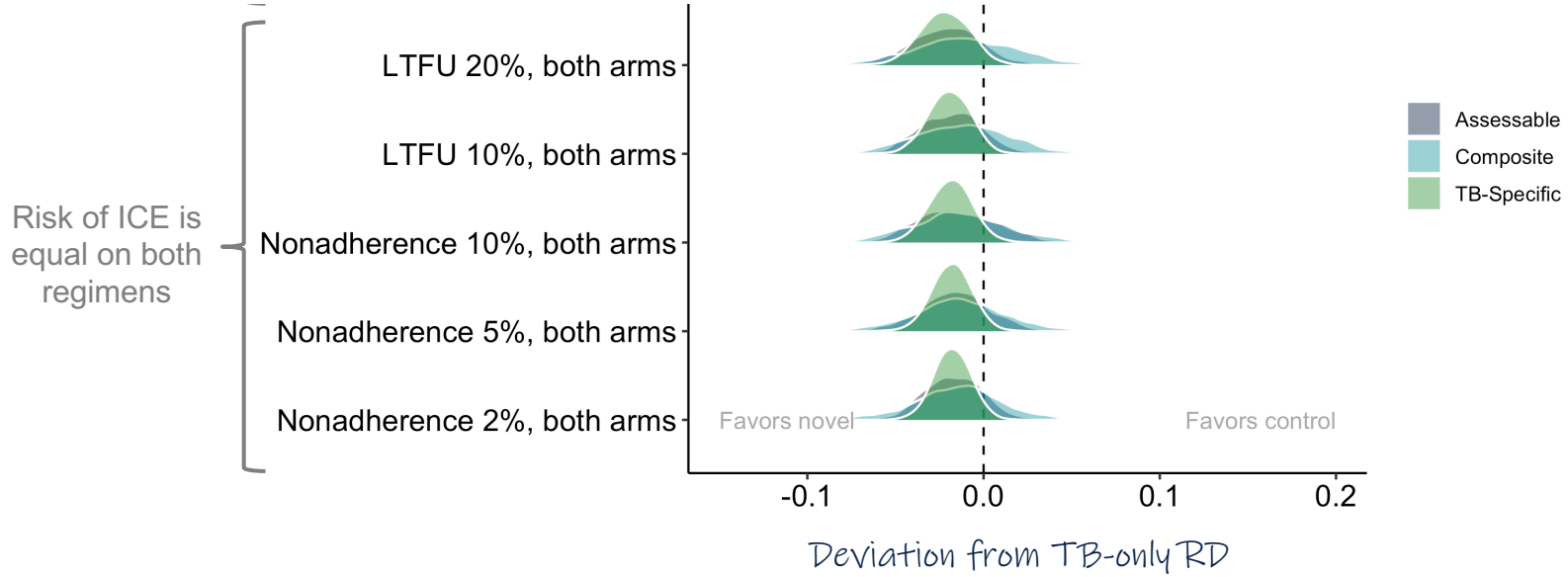
Estimand Selection

TB-Specific Estimand

What happens as ICEs occur unevenly?

TB-Specific Estimand

- Address the treatment effect, looking specifically at TB-specific efficacy endpoints.
- ICEs unrelated to TB disease or treatment are assumed not to have happened (hypothetical strategy).



Takeaways + Next Steps

Takeaways

Takeaway: Pragmatic guidance for estimand-aligned estimation is important and not as well established.

Let's ensure we are not getting the right answer for the wrong question.

Next:

- De-bug simsurv behavior. Please contact if you are interested in helping!

Forthcoming:

- Paper summarizing estimator performance and estimand comparison

suzanne.dufault@ucsf.edu

Thank you

I want to acknowledge the contribution of all individuals who have provided input to this work. A special thanks to Isabelle Weir (Harvard), Nancy Hills (UCSF), and Patrick Phillips (UCSF).

I gratefully acknowledge funding from the UCSF Center for TB and TB RAMP scholar program (NIH/NIAID R25AI147375).

Work performed with guidance from:

Patrick P. J. Phillips (UCSF)

Contact Info
suzanne.dufault@ucsf.edu
suzanne-dufault-phd.netlify.com

Please contact me with any questions, comments, or feedback.

UCSF

Center for Tuberculosis

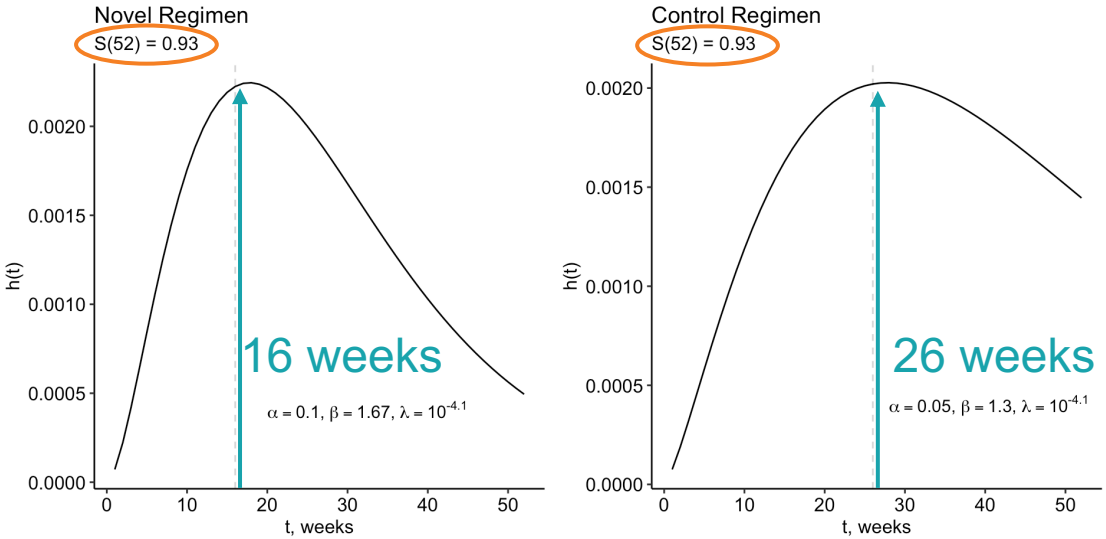
UCSF

University of California
San Francisco

tb.ucsf.edu

Simulating the primary outcome

Time-to-event: Cure versus unfavorable outcome (e.g., failure, relapse)



The primary outcome is simulated from a surge hazard function:

- Proportion unfavorable events: 7% by 52 weeks
- Peak hazard at end of treatment

Simulating the ICEs

Time-to-event variables that may prevent observation of primary endpoint

1. Reinfection or attributable death
2. Major treatment change: adverse event
3. Major treatment change: poor adherence
4. Death during treatment: suicide, related to study treatment, other (last culture negative)
5. Withdrawal or LTFU (last culture negative)
6. Minor treatment changes
7. Major treatment changes: pregnancy or other reason
8. Discontinuation from treatment: fail to complete adequate treatment
9. Bacteriological events: reinfection without change in treatment
10. Other bacteriological events: isolated positive culture
11. Other bacteriological events: unable to produce sputum at end of follow-up
12. Death during treatment: accident or trauma ← Could only occur during treatment
13. Death after treatment completion: suicide, accident or trauma, other (last culture negative) ← Could only occur after treatment

ICEs are drawn from an exponential hazard model (constant hazard), with risk intervals dependent on the ICE.

Estimand	ICE Strategies	Estimation
TB-specific	Ignoring the following ICEs (treatment policy): (#7) Minor treatment changes (#10) Bacteriological events: reinfection without change in treatment (#11) Other bacteriological events: isolated positive culture Use IPW or MI for the following ICEs (hypothetical): (#2) Major treatment changes: reinfection (#3) Major treatment changes: adverse event (#4) Major treatment changes: poor adherence (#5) Death during/after treatment: suicide, related to treatment or other (last culture negative) (#6) Withdrawal or loss to follow-up after treatment completion (last culture is negative) (#8) Major treatment changes: other reason including pregnancy (#9) Discontinuation from treatment (#13) Death during treatment: accident or trauma Incorporating ICE into the outcome (composite strategy): (#1) Unfavorable outcome (failure/relapse) (#12) Other bacteriological events: unable to produce sputum at end of follow-up	1. Generate new composite outcome. 2. "Ignore" treatment policy ICEs. (If available, use outcome as recorded. If not available, drop **or censor** these individuals). 3. Generate individual-level propensities for ICE occurrence. 4. Perform *weighted* Kaplan-Meier estimation.

Estimand	ICE Strategies	Estimation
Per protocol	Ignoring the following ICEs (treatment policy): (#7) Minor treatment changes (#10) Bacteriological events: reinfection without change in treatment (#11) Other bacteriological events: isolated positive culture Use IPW or MI for the following ICEs (hypothetical): (#5) Death during/after treatment: suicide, related to treatment or other (last culture negative) (#6) Withdrawal or loss to follow-up after treatment completion (last culture is negative) (#8) Major treatment changes: other reason including pregnancy (#13) Death during treatment: accident or trauma Incorporating ICE into the outcome (composite strategy): (#1) Unfavorable outcome (failure/relapse) (#2) Major treatment changes: reinfection (#3) Major treatment changes: adverse event (#12) Other bacteriological events: unable to produce sputum at end of follow-up Use to define population of interest (principal stratification) (#4) Major treatment changes: poor adherence (#9) Discontinuation from treatment	1. Generate new composite outcome. 2. “Ignore” treatment policy ICEs. (If available, use outcome as recorded. If not available, drop **or censor** these individuals). 3. Use multiple imputation to address the hypothetical ICEs. 4. Use Bayesian modeling to estimate the RD in the principal stratum.

Clinical Development

Phase I	Phase II	Phase III
Is it safe? At what dose?	Does it work?	How does it compare to standard of care?