IMPAACT Network Annual Meeting TB Scientific Committee, June 18, 2018

Innovative Statistical Methods for Phase III Tuberculosis Trials in Children

Soyeon Kim, ScD Statistical and Data Analysis Center

Frontier Science Foundation

International Maternal Pediatric Adolescent AIDS Clinical Trials Network

MPAACT

Outline

- Setting
- Placebo and Active-Controlled Superiority Trials
- Non-Inferiority Trials
- Risk-Benefit Approach
- Adaptive Trials
- Entry Points, Stratification, Subgroups
- Outcomes

Setting

Current Treatment

- For RR/MDR-TB WHO recommends 9-12 month Bangladesh regimen (2016) which is an improvement over the 18-24 month standardized regimen (2011).
- The "Bangladesh" regimen remains long, has a high pill burden, and requires use of daily injections during the 4-6 month intensive phase.
- The injectables are associated with irreversible hearing loss >20% of children which has drastic implications for neurodevelopment and longterm functioning.

TB Drug Pipeline

- Newly licensed or repurposed drugs are promising candidates use in MDR-TB treatment.
- Multiple ways to combine drugs from various classes into potentially efficacious all-oral, shortened duration regimens that are less toxic.

Possible Benefits and Risks of All-Oral, Shortened Regimen

Benefits

- Shorter
- Potentially fewer specific-AEs, e.g., less permanent hearing loss and therefore better neurodevelopmental and long term outcomes (eg educational)
- Potentially better tolerated and more acceptable

Risks

- Potential loss of efficacy compared to current regimens
- Potential increase in other AEs, including cardiotoxicity

Problem with Extrapolation from Adults

Children <u>tend</u> to be more:

- Paucibacillary
- Smear- and culturenegative
- Non-cavitary
- Good outcomes compared to adults

Adults <u>tend</u> to be more:

- High bacillary load
- Smear- and culturepositive
- Cavitary with necrotic lesions
- Poorer outcomes compared to children

→Studying regimens in adults may result in the dismissal of regimens that could benefit children.

Placebo and Active-Controlled Superiority Trials

Ideal Trial



Active-Controlled Trial



Sample Size for Hypothetical Active-Controlled Trial



Since the difference in proportion with favorable outcomes between arms, θ , is likely smaller for an active-control vs placebo, then the sample size may be large.



Non-Inferiority Trials

Non-Inferiority Trials

- Increasingly popular method to evaluate treatments that might not improve efficacy but have better safety, tolerability, or other benefits that would justify a decrease in efficacy.
- Generally used when the Experimental arm has slightly worse efficacy than the Active-Control arm.
- Can be used when the Experimental arm has better efficacy than the Active-Control arm.

Superiority vs Non-inferiority Objective

Superiority Efficacy:

If the experimental regimen is expected to be better than the control then a trial can be done to show superiority.



Superiority vs. Non-inferiority Objective

Non-inferiority :

- Control has been shown to be efficacious in high quality studies
- Experimental arm is expected to be the slightly worse than the control, however it is more tolerable (e.g., injectable sparing or palatable), has fewer adverse events, or of a shorter duration
- $-\Delta$ is the non-inferiority margin



Adapted from Committee for Proprietary Medicinal Products (CPMP). Points to consider on switching between superiority and non-inferiority Br J Clin Pharmacol, 52, 223-228

Non-inferiority Margin Justification 1

• Statistical Reasoning



A confidence interval for θ̂ can be used to exclude Δ, e.g.
95% CI corresponding to a 2.5% significance level test.

17

Adapted from D'Agostino et al. Non-inferiority trials: design concepts and issues – the encounters of academic consultants in statistics. Statist. Med. 2003; 22:169–186.

Non-inferiority Margin Justification 2

Clinical Judgement based on anticipated benefits:

- Shorter duration of treatment
- Lower pill burden
- All-oral (vs injection-containing) regimen
- Fewer Adverse Events
- Better tolerability
- Outcome measure (components included in composite and timing)
- "Patients' preferences ... based on anticipated benefits and risks" (Mauri et al. N Engl J Med 2017;377:1357-67)

Non-Inferiority Margin Justification 3

- The hard part is translating the anticipated benefits of the experimental arm into a noninferiority margin Δ.
- The NI margin will be unique to anticipated benefits (reduction in risks) and also to the study population, treatment regimen, primary outcome, etc.
- D'Agostino et al. (2003) "we have never seen a case where [clinical judgement] has actually been employed."

D'Agostino et al. Non-inferiority trials: design concepts and issues – the encounters of academic consultants in statistics. Statist. Med. 2003; 22:169–1986.

Hypothetical Non-Inferiority Trial



Hypothetical Non-Inferiority Trial



Control Arm in Non-Inferiority Trials

- Control arm must be an active control with demonstrated efficacy.
 - This may be challenging in some settings and may be based on historical data.
 - A less efficacious treatment may also stand in for placebo.

Beware of Biocreep

p=true probability favorable outcome

	p _{Rx2}	p _{Rx1}
р _{кх3}	p _{Rx2}	
р _{кх3}		
	P _{Rx3}	p _{Rx3}

In trial 4, the experimental arm could be shown to be non-inferior but is actually no different than placebo.

Components Included in Primary Efficacy Outcome - Relevance to the non-inferiority margin

Example Adult Trial Primary Outcome Definition

Primary outcomes have been a composite of discordant components of differing clinical importance.

RIFAQUIN (Jindani et al. N Engl J Med. 2014, 23; 371(17): 1599–1608)

- Culture+ most recent follow-up visit related to efficacy
- Restart or change of treatment for any reason other than making up for missed doses or becoming pregnant – can be related to AE, tolerability or efficacy
- Death before the end of scheduled treatment for reasons other than violence or trauma – can be related to AE or efficacy
- Death after the end of treatment with evidence that confirmed or suggested possible treatment failure or relapse of their tuberculosis – related to efficacy
- Failure to complete treatment without a negative culture result at the end of the scheduled follow-up period can be due to many different reasons

In addition to measures of efficacy, it includes LFU, change of treatment for any reason (including AE, tolerability) as unfavorable

Composite Primary Outcomes and the Non-Inferiority Margin

- The NI margin needs to take into account the actual components included in unfavorable outcome.
 - —If it includes domains where the Exptl Arm is supposed to be beneficial (eg Rx discontinuation for AEs), we are incorporating the benefits of the experimental regimen into the primary outcome. This may warrant a smaller NI margin than if we used an outcome that focuses exclusively on efficacy measures.
 - Mauri et al. recommend that trialists "avoid using composite end points that include discordant components."
 - WHO/CPTR meeting (March 2018) participants from regulatory agencies suggested that they would be receptive to larger NI margins for primary outcomes that focus on efficacy.

Reference: Mauri et al. *N Engl J Med* 2017;377:1357-67

SMaRT Kids Co-Primary Objectives

- 1. Non-inferior <u>efficacy</u> with carefully defined "favorable outcome" that focuses on efficacy.
- 2. Superior <u>safety and tolerability</u>.
- Both objectives need to be achieved.
- Sample size chosen to allow high power for both objectives.

Risk-Benefit Approach

Risk-Benefit Approach

- Montepiedra et al. propose a different paradigm for assessing interventions using a totality of outcomes approach for TB
 - Assessing regimens based on efficacy, AEs, QoL using an integrated outcome that accounts for the differing importance of heterogeneous events.

Hypothetical Results Which Regimen is Better?

Regimen A

Reg	imen	E

	Treatment success	No treatment success
No serious adverse event	85	95
Serious adverse event(s)	110	10

Treatment success rate: 65% Serious adverse event rate: 40%

	Treatment success	No treatment success
No serious adverse event	140	10
Serious adverse event(s)	80	70

Treatment success rate: 73% Serious adverse event rate: 50%

References: Montepiedra Journal of Clinical Tuberculosis and Other Mycobacterial Diseases 4 (2016) 9–13

Simple Ordinal Outcome

Rank	Outcome
1	Treatment success, no SAE
2	Treatment success, SAE
3	Treatment failure, no SAE
4	Treatment failure, SAE

A Wilcoxon rank sum test of the previous scenario favors Regimen B (p=0.018).

The rankings can be more nuanced (e.g., ototoxicity vs subclinical lab toxicity can be treated differently) and can incorporate tolerability and other patient reported outcomes.

References: Montepiedra Journal of Clinical Tuberculosis and Other Mycobacterial Diseases 4 (2016) 9–13

Developing Rankings

- The rankings are inherently subjective.
- Consensus process to inform the development of rule-based ranking scheme based on primary outcome indicators with secondary outcome indicators to further differentiate outcomes or a Delphi process.
- The panel of stakeholders can include clinicians, public health authorities, and patients/ caregivers.

Risk-Benefit Analysis

- Standard methods of analysis for ordinal data, e.g., Wilcoxon rank sum test or Cochran Mantel-Haenszel chi-square test, can be used to detect differences in the ranked outcomes.
- Analyses based on separate efficacy and safety outcomes will still be useful and are always be important.

Challenges of Risk-Benefit Approach

- Requires a lot more work up front with collaboration with stakeholders.
- We need to apply these methods to existing datasets to understand its behavior before using as primary outcomes in actual trials.

→ IMPAACT trials are using this approach as secondary and other outcomes (P1078, SMaRT Kids)

Adaptive Multi-Arm Studies



Why?

• Multiple ways to combine drugs from various classes into potentially efficacious regimens.

Two-Arm Study Paradigm

Two-arm trials performed sequentially:



time

- Long duration to evaluate all experimental regimens.
- Because of non-contemporaneous enrollment of all arms, changing SOC, relative comparisons of all treatment arms are not straightforward.

Multi-Arm Study Designs

 Instead of separate two-arm trials performed sequentially, a single trial with experimental arms being compared to a shared control group.



Multi-Arm Multi-Stage Design (MAMS)

Example of a 5-arm phase 2 trial with a 3-stage MAMS design

Start of	First Interim	Second Interim	End of
Recruitment	Analysis	Analysis	Recruitment
1	:	1	:
Control Reg	gimen		
i			i
Novel regin	nen 1	Stop	
Novel regin	nen 2		
i		i	i
Novel regin	nen 3	Stop	ļ
		i	i
Novel regin	nen 4 Stop		
i Stag	je i St	agez _i S	stage 3 j

MAMS Features

- The maximum sample size is fixed but the actual sample size depends on the data.
- At each interim analysis, poorly performing arms can be discontinued.
- Criteria with high stagewise power and low type I error can be used at each interim analysis to prevent the discontinuation of effective arms but allow for the discontinuation of clearly inferior arms.
- Patients are more likely to receiving promising regimens.
- More than one efficacious regimen may be identified.
- At the end of the trial, the remaining arms should be balanced with respect to patient characteristics and the analysis is straightforward and intuitive.
- Variations: Seamless Phase II/III trials (Bratton *et al. BMC* Medical Research Methodology 2013, 13:139)

Challenges of MAMS Design

- To be drop poorly performing arms at each stage and reduce the total study size, outcomes that are quickly available need to be used for decision making.
 - Outcomes that are delayed (like 27 month culture) and inefficiencies in study processes or waiting for cultures to grow reduces the efficiency of adaptive designs because participants may be enrolled in dropped arms. At the extreme with a fast recruitment rate, the study might be fully enrolled before any study adaptations are made.
 - If an intermediate outcome (biomarker) is used, it must be sensitive and specific otherwise effective treatments may be dropped.
- In the context of pediatric MDR-TB this type of error may be problematic.

endTB Adaptive Randomization Treatment Shortening Non-Inferiority Trial for Rif-r FQ-s TB



http://www.resisttb.org/wp-content/uploads/2016/06/endTB_trial_june2016_Mitnick.pdf

Adaptive Randomization Features

Benefits

- Adapt randomization probabilities so that later stages randomize more participants to arms that perform better at interim analyses. Results in more data on the better arms and better overall treatment response for participants.
- Costing is easier because the sample size is fixed.

Challenges

- In open-label studies when randomization probabilities deviate from equal randomization, it may be difficult to prevent investigators from learning which arms may be doing better.
- May be logistically challenging.
- Analysis needs to take into account the randomization probabilities which was time- varying. Analysis is not straightforward.

Entry Points, Stratification, Subgroups

Entry Points – The Challenges

- Unlike adult trials, microbiological confirmation and DST are available in ~40% of children.
- Even when results are ultimately available, they may be <u>delayed until after study entry</u>.
- It may not be in best interest of the child to wait for culture results to randomize and initiate Rx.

Why Not Exclude Clinically Diagnosed TB Cases?

- Because clinically diagnosed TB cases are more likely to have minimal disease and be paucibacillary, they might be the subgroup most likely to benefit from shorter, all-oral treatment.
- They are a large portion of the population of children with TB and they also urgently need evidence-based efficacious treatment.

Pragmatic Solutions – Entry Points

- Enroll both confirmed and clinically-diagnosed TB cases.
- Use adult contact TB diagnostic results, coupled with welldefined signs/symptoms and radiology, to clinicallydiagnose TB and provide DST profile for the child.
- Primary randomized comparison groups should be based on information or samples collected <u>prior to</u> randomization. No post-randomization samples will be used in defining the analysis population.
- Exclusions to the analysis populations are consistently applied to all randomized Rx arms.

 \rightarrow Preserves the benefits of randomization

Stratification and Subgroups

- Stratification of randomization can only be done on data that is available at the time of randomization.
 - Confirmed vs clinically-diagnosed TB may not be a feasible stratification factor.

Subgroup Analysis and Covariate Adjustment

- Subgroups can be defined based on data or samples obtained <u>prior to</u> randomization even if results were not available for stratification of the randomization.
 - It does not preclude stratified analysis to explore robustness of results or identify heterogeneity.
 - Can be used for covariate adjustment to improve efficiency of comparisons.

Primary Outcomes

Implications of Entry Points for Pediatric Outcome Definitions

- Inclusion of clinically-diagnosed TB poses challenges for defining outcomes, because it precludes the use of culture conversion in the outcome definition.
- Rigorously defined clinical definitions of probable cure will be needed.
- Review by Independent Outcomes Committee.

Duration for Evaluating Primary Outcomes for a Treatment Shortening Trial



Thoughts on Follow-up for Differing Trial Durations

Option 1: Same post-Rx F/U

• allows similar time for relapse.

Option 2: Same overall duration allows same time to capture combined cure (failure), relapse, and reinfection.

- May miss more <u>relapse</u> in the control arm
- May capture more <u>reinfection</u> during the longer post-Rx F/U period for Experimental Arm.
 - explanatory perspective might not be desirable.
 - pragmatic perspective this may be part of the benefit of the longer regimen.
- The post-Rx F/U duration can be chosen to balance capturing the relapses without capturing reinfection.

Blinding

- When regimens differ drastically in composition, duration, or modes of administration in combination, blinding would require the use multiple placebos and increase complexity of all regimens.
- The use of a placebo-injection may negatively affect adherence which is known to decrease regimen efficacy.
- Placebos to lengthen shorter regimens may not be acceptable to patients/caregivers and IRBs.
- Blinding investigators who do not provide patient care, independent review committees evaluating CXR or outcomes, etc still desirable.

Summary

 Newly licensed and repurposed drugs can be combined into all-oral, shorter duration regimens and can be evaluated in high quality trials that preserve type I error and have high power and minimize bias.



Anneke Hesseling, Tony Garcia-Prats, Simon Schaaf, James Seddon, Pauline Howell, Leavitt Morrison, Betsy Smith, Patrick Jean-Philippe and other members of the IMPAACT 2020 (SMaRT Kids) Study Team.

Grace Montepiedra, Sachiko Miyahara, Ritesh Ramchandani and others from the Risk-Benefit working group.

Overall support for the International Maternal Pediatric Adolescent AIDS Clinical Trials Network (IMPAACT) was provided by the National Institute of Allergy and Infectious Diseases (NIAID) with co-funding from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) and the National Institute of Mental Health (NIMH), all components of the National Institutes of Health (NIH), under Award Numbers UM1AI068632 (IMPAACT LOC), UM1AI068616 (IMPAACT SDMC) and UM1AI106716 (IMPAACT LC), and by NICHD contract number HHSN2752018000011. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

