Advances in Clinical Trial Design for Development of New Treatments for Tuberculosis
Advances in Clinical Trial Design for Development of New Treatments for Tuberculosis

About this PLOS Collection

With recent opportunities to test combinations of new and repurposed drugs for the treatment of TB, new questions arise for optimal clinical trial design. To address these, in March 2018 WHO organized a consultation on “Advances in Clinical Trial Design for New TB Treatments” to identify and outline the optimal characteristics of clinical trial designs to inform policy guidance for new TB regimens. This PLOS Medicine Special Collection assembles a series of articles on current reflections and describes essential new steps in clinical research that will pave the way for the development of tomorrow’s optimal treatment for all forms of TB.

To read the Special Collection online, visit https://collections.plos.org/novel-tb-treatments

PLOS Special Collection Partners

French National Research Institute for Sustainable Development | World Health Organization
# Table of Contents

**Introduction**

**Advances in clinical trial design for development of new TB treatments: A call for innovation**
Christian Lienhardt, Payam Nahid
*PLOS Medicine*
Published: March 22, 2019 | DOI: 10.1371/journal.pmed.1002769 ................................................................. 1

**Definition of new regimens: early clinical development**

**Advancing the development of new tuberculosis treatment regimens: The essential role of translational and clinical pharmacology and microbiology**
Kelly E. Dooley, Debra Hanna, Vidya Mave, Kathleen Eisenach, Radojka M. Savic
*PLOS Medicine*
Published: July 5, 2019 | DOI: 10.1371/journal.pmed.1002842 ................................................................. 6

**Accelerating the transition of new tuberculosis drug combinations from Phase II to Phase III trials: New technologies and innovative designs**
Geraint Davies, Martin Boeree, Dave Hermann, Michael Hoelscher
*PLOS Medicine*
Published: July 9, 2019 | DOI: 10.1371/journal.pmed.1002851 ................................................................. 20

**New trial designs and how they may facilitate regimen development**

**Keeping phase III tuberculosis trials relevant: Adapting to a rapidly changing landscape**
*PLOS Medicine*
Published: March 22, 2019 | DOI: 10.1371/journal.pmed.1002767 ................................................................. 30

**Designing noninferiority tuberculosis treatment trials: Identifying practical advantages for drug regimens with acceptable effectiveness**
Piero L. Oliaro, Michel Vaillant
*PLOS Medicine*
Published: July 12, 2019 | DOI: 10.1371/journal.pmed.1002850 ................................................................. 45
Considering subjects: trial adherence and special populations

The importance of adherence in tuberculosis treatment clinical trials and its relevance in explanatory and pragmatic trials
Andrew Vernon, Katherine Fielding, Rada Savic, Lori Dodd, Payam Nahid
PLOS Medicine
Published: December 10, 2019 | 10.1371/journal.pmed.1002884 ......................................................... 59

Inclusion of key populations in clinical trials of new antituberculosis treatments: Current barriers and recommendations for pregnant and lactating women, children, and HIV-infected persons
Amita Gupta, Michael D. Hughes, Anthony J. Garcia-Prats, Katherine McIntire, Anneke C. Hesseling
PLOS Medicine
Published: August 15, 2019 | 10.1371/journal.pmed.1002882 ............................................................... 69

From trials to guidance: bringing evidence to best programme and practice
Development of new TB regimens: Harmonizing trial design, product registration requirements, and public health guidance
Christian Lienhardt, Andrew A. Vernon, Marco Cavaleri, Sumati Nambiar, Payam Nahid
PLOS Medicine
Published: September 6, 2019 | 10.1371/journal.pmed.1002915 ............................................................... 95

Advances in clinical trial design for development of new TB treatments—Translating international tuberculosis treatment guidelines into national strategic plans: Experiences from Belarus, South Africa, and Vietnam
Grania Brigden, Nguyen Viet Nhung, Alena Skrahina, Norbert Ndjeka, Dennis Falzon, Matteo Zignol
PLOS Medicine
Published: October 18, 2019 | 10.1371/journal.pmed.1002896 ............................................................... 108

Conclusion
Advances in clinical trial design: Weaving tomorrow’s TB treatments
Christian Lienhardt, Andrew Nunn, Richard Chaisson, Andrew A. Vernon, Matteo Zignol, Payam Nahid, Eric Delaporte, Tereza Kasaeva
PLOS Medicine
Published: February 27, 2020 | 10.1371/journal.pmed.1003059 ............................................................... 120
Advances in clinical trial design for development of new TB treatments: A call for innovation

Christian Lienhardt, Payam Nahid

1 Unité Mixte Internationale TransVIHMI (UMI 233 IRD–U1175 INSERM, Université de Montpellier), Institut de Recherche pour le Développement (IRD), Montpellier, France, 2 Division of Pulmonary and Critical Care Medicine, University of California San Francisco, San Francisco, California, United States of America

* christian.lienhardt@ird.fr

After decades of stagnation, research in tuberculosis (TB) therapeutics is experiencing a renaissance, with an increasing number of new and repurposed compounds undergoing evaluation as part of novel treatment regimens. This is much welcome progress, since current regimens are not ideal due to the long duration of treatment required, toxicities, drug–drug interactions, and high costs—particularly for treatment of the various forms of drug-resistant TB (DR-TB).

The development of new TB drugs is, however, complex, lengthy, and costly [1], and the pathway to proven new TB treatment regimens is fraught with numerous obstacles and uncertainties [2]. In this PLOS Medicine Collection, “Advances in Clinical Trial Design for Development of New Tuberculosis Treatments,” we highlight key obstacles and identify potential solutions that will help avoid misadventures and in turn maximize the likelihood of success in identifying new drugs and regimens through a rejuvenated global interest in TB therapeutics.

With the emergence of several new chemical entities expected to transition into clinical testing in the next 5 years, the possibility of ultrashort (i.e., requiring treatment for weeks rather than months) regimens for active TB is no longer fanciful. Investigators in the field have learned much from recent TB clinical studies, and we anticipate that well-designed and conducted clinical trials evaluating the next generation of drugs and regimens will, with some good fortune, lead to identification of the ultrashort, safe, and effective regimens so desperately needed.

Treatment of TB relies on a synergistic combination of drugs (traditionally categorized as bactericidal or sterilizing) administered for sufficient time to achieve definitive nonrelapsing cure and to prevent selection of drug-resistant mutants [3]. The treatment of drug-susceptible TB (DS-TB) is well codified, with a standard combination of 4 drugs given for a duration of 6 months [4]. This regimen is the result of a series of clinical studies conducted in several countries, which demonstrated the efficacy of short-course regimens of 6–8 months’ duration in patients with pulmonary disease [5]. These trials played a key role in the establishment of short-course chemotherapy worldwide, allowing treatment of DS-TB to be based on the best available evidence [6]. Since then, clinical trials and programmatic experience have shown that the standard 6-month isoniazid/rifampicin-based regimen, when adhered to, performs consistently well in a wide variety of settings and can serve as a reliable control regimen against which investigational regimens can be compared [4]. The situation is, however, more complicated for DR-TB. In the absence of controlled trials comparing different regimens to a recognized “gold standard” treatment, the current recommendations for therapy rely on early-phase culture-conversion results, observational studies, and a few late-phase clinical trials [7]. The number and type of drugs required to treat patients with DR-TB has long been a matter of
debate and controversy despite agreement on basic principles such as the minimum number of drugs to use and minimum duration of treatment. As a result, the efficacy of recommended DR-TB treatment regimens has been shown to vary widely in clinical studies and programs [8,9].

The need for solid evidence from randomized controlled trials has led the TB research community to adopt a design widely used in HIV research for the development of new antiretrovirals, in which patients are randomized to receive either a new drug or placebo in addition to a defined “optimized background regimen,” usually the best available standard of care [10]. This approach has been used in the development of bedaquiline [11] and delamanid [12], the first two new drugs approved for TB treatment since the late 1980s. While this research design assesses the added value, if any, of a given investigational drug, the approach leaves unresolved the question of the optimal drug combination in which to include the new agent [13]. As a result, additional clinical trials are then needed to identify the best options for treatment using new drugs in variable combinations, resulting in additional years of delay in producing the best evidence for global policy-making decisions. In parallel, practical recommendations are needed for the use of any newly approved drugs, along with guidance for countries and programs as to which combinations are safe, tolerable, and efficacious, an endeavor that requires systematic reviews and meta-analyses of observational cohort studies and programmatic data, which carry significant limitations. This approach is not sustainable, practical, or efficient and raises the need for a shift to a more efficient and seamless development process that allows the testing of novel treatment regimens, including one or more promising new or repurposed medicines, early in the clinical development pathway. Some stakeholders, such as the TB Alliance, therefore proposed a “unified approach to TB regimen development” addressing the joint development of new drugs and regimens for both DS-TB and DR-TB [14]. Also, the International Union against Tuberculosis and Lung Diseases opted to investigate the safety and efficacy of a set combination regimen of 9–12-months’ duration for the treatment of DR-TB in parallel through a randomized controlled trial [15] and observational studies undertaken within programmatic research conditions [16]. However, the availability of results from these various studies at different points in time and questions arising from the challenge of interpreting and integrating data from various methodologies were found to limit the adequacy of these complementary approaches for development of therapies [17].

Duly concerned with the need to base its normative treatment recommendation on the best available evidence [18], and to produce guidelines that would be readily usable in daily practice in all settings, WHO opted to establish minimal and optimal benchmarks for TB regimen development using industry-accepted target product profile (TPP) principles [19]. These TPPs for new anti-TB regimens, referred to as “target regimen profiles” (TRPs), describe the minimum and optimal attributes and characteristics of future TB regimens to guide the development process [20]. A population-level modeling analysis evaluating the potential impact of various regimen characteristics on the TB epidemic highlighted the paramount importance of regimen efficacy to exert the largest impact on reduction of TB cases and deaths, both for DS-TB and DR-TB [21]. Other characteristics such as shorter duration of, or increased adherence to, treatment were shown to have important effects by enabling more people with TB to receive appropriate and timely therapy. Most importantly, this model highlighted the difficulty of improving all potential characteristics simultaneously in a single regimen, leading developers to consider weighing in inevitable trade-offs (e.g., higher cure rates may be difficult to achieve simultaneously with shorter treatment duration, and simpler or better-tolerated regimens may be less robust to emergence of drug resistance) that are duly addressed in the TRPs.

Given the recommended regimen characteristics, the implementation of TRPs stimulated the question of which clinical trial designs and features should be optimally used for the
development of new anti-TB regimens. Major challenges exist along the current lengthy development pathway [1], including the lack of direct indicators of treatment response, the lack of reliable surrogate markers of treatment outcomes, and the lack of predictive quantitative relationships between Phase II and Phase III outcomes [22]. To accelerate and streamline the development of new TB regimens, the therapeutics research community needs to establish clear and rationally justified approaches for the choice of drug combinations, trial design, selection of endpoints, and analysis [23,24], taking into account new developments in individual drugs’ pharmacokinetic and pharmacodynamic characteristics, microbiological aspects, use of biomarkers, standardization of approaches and data collection, as well as drug effects in key patient populations.

From the regimen developer’s perspective, it is apparent that a new treatment regimen must bring a value proposition, beyond efficacy or safety targets. Products with broader applications (e.g., for eligible populations) gain in terms of delivery and scalability/distribution or cost and can bring substantial impact and value that define the developmental pathway. Sponsors and donors should evaluate the needs of the market and develop programs based on those needs. In conjunction, decisions about progress from Phase II to Phase III studies continue to involve significant uncertainty, and these limitations need to be considered when designing Phase III trials. Also, the issue of the control groups most appropriate for a given trial situation needs careful consideration. It thus appears that each development program needs to determine the most appropriate approach to trial design, depending on the situation and the questions to be addressed.

From both programmatic and patient perspectives, the recent pooled individual patient-level analysis of three treatment-shortening trials examining the efficacy and safety of 4-month combination regimens, including third-generation fluoroquinolones for the treatment of DS-TB [25–27], provided critically important insights relevant to TB treatment in the field and to therapeutics research [28]. Whereas these trials independently failed to show noninferiority of the 4-month experimental regimens tested, as compared to the 6-month control regimen, 80% of patients were cured. The pooled analysis of these trials found that patients with minimal disease, defined as low bacterial burden or absence of lung cavities, would be eligible for 4-month treatments [28]. Conversely, patients with high baseline smear, cavitation on chest X-ray, HIV coinfection, and low body mass index defined hard-to-treat phenotypes that would need more than the standard 6-month treatment duration to achieve the highest possible cure rates. In addition, even minimal nonadherence (i.e., missing 1 in 10 doses) to the current standard regimen was found to be a significant risk factor for unfavorable outcome, independent of treatment duration. These findings provide a strong evidence-based framework for investigating different approaches to achieving better patient-oriented treatment—such as the stratified medicine approach—and emphasize the importance of maximizing adherence in clinical trials and in real-world conditions.

These issues illustrate the need for obtaining maximally informative and reliable data from controlled trials, as these are paramount for the development of policy for wide public health use and for guideline development. To address these coherently, in March 14–16, 2018, WHO organized a technical consultation on “Advances in Clinical Trial Design for New TB Treatments” to identify and outline, through expert consensus, the optimal characteristics of clinical trial designs to inform policy guidance for the development of new TB regimens. Building on the lessons learned from the rich history of TB clinical trials, the WHO technical consultation [29] reviewed the various research designs and tools currently used in the conduct of clinical trials for development of new TB treatments and made a series of proposals to advance these further, seeking to move from evolutionary change informed by history to a bolder approach to innovation geared to the future. These are the aims of this PLOS Medicine Collection, which
we are launching on World TB Day 2019, beginning with the accompanying paper from Patrick Phillips and colleagues [30] on the changing landscape of clinical trial design for development of TB therapeutics. Further articles will be added to the Collection in due course, and the Collection will be available in its entirety alongside this paper once all the articles have been published.

References


Advancing the development of new tuberculosis treatment regimens: The essential role of translational and clinical pharmacology and microbiology

Kelly E. Dooley*, Debra Hanna, Vidya Mave, Kathleen Eisenach, Radojka M. Savic

1 Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, United States of America, 2 Bill & Melinda Gates Foundation, Seattle, Washington, United States of America, 3 Byamjee Jeejeebhoy Government Medical College–Johns Hopkins University Clinical Research Site, Pune, India, 4 TB or NOT TB Consulting, LLC, Little Rock, Arkansas, United States of America, 5 University of California San Francisco, San Francisco, California, United States of America

* kdooley1@jhmi.edu

Summary points

- Translational and clinical pharmacology are the state-of-the-art tools used by drug developers to efficiently move compounds and regimens through all drug development phases. Tuberculosis drug and regimen development, though, has traditionally underutilized these modern, model-based drug development approaches, despite the urgent need to understand major pharmacological aspects not only of the new candidates but also of existing drugs.

- Translational platforms that include drug combinations are critical and should encompass data from multiple preclinical drug development tools (in vitro and in vivo models) to select the best regimens to be moved forward into clinical development.

- Quantitative pharmacokinetic (PK)–pharmacodynamic (PD) approaches should be incorporated into all phases of drug development and be used for selection of optimal dose and schedule, assessment of drug–drug interactions, and dose determination in key populations including pregnant women, children, and people living with HIV. Quantitative pharmacology models should further be utilized for clinical trial design using clinical trial simulations.

- Microbiology determinants such as precisely assessed minimum inhibitory concentrations (MICs) as well as quantitative longitudinal cultures integrated with PK-PD assessment will substantially inform and enhance all phases of drug development.

- Commitment of all stakeholders, data sharing, and resource investment are required for development and utilization of these tools, which are necessary for successful TB regimen development.
Introduction

Application of clinical pharmacology best practices is essential to the efficient and rational development of drugs. In general, knowledge gained about exposure–response relationships in preclinical models aids drug and dose selection in human studies, and biomarkers and pharmacokinetic (PK) data one collects in early to middle drug development can be used to predict the dose and treatment response of promising therapeutics in definitive phase 3 trials. The essentiality of sound clinical pharmacology in tuberculosis (TB) drug and regimen development is heightened by unique challenges in assessing drugs for this disease—aspects of the organism’s biology, the variability in lung pathology, uncertainties about how to link treatment outcomes seen in preclinical models with those seen in humans (which thwarts preclinical–clinical translational work), the lack of predictive early clinical biomarkers, and the high variability in treatment response across patients and populations (Fig 1). In TB disease, *Mycobacterium tuberculosis* (*M. tb*) bacilli are detected in necrotic granulomas, large cavities with liquefied contents, and intracellularly within macrophages. We believe that drugs must access each of these compartments to achieve cure in patients [1]. We also believe that TB drugs and regimens must kill bacilli in different metabolic states, from actively multiplying to semidormant [2,3]. Both in vitro and in vivo preclinical models are leveraged to assess the clinical utility of new TB drugs and drug combinations. These models vary both in their ability to assess efficacy relative to the shifting metabolic states of *M. tb* infection and in their ability to recapitulate human disease. Still, two models are proving to be highly informative. The mouse model of infection has been invaluable in selecting rank-ordered drug combinations, whereas the

![Fig 1. Schema of preclinical and clinical pharmacology studies important for TB drug and regimen development. By phase of development, in green are the questions to be addressed, in blue are the tools to use to answer the questions, and in red are the outputs. ADME, absorption, distribution, metabolism, excretion; DDI, drug–drug interaction; D3, disease; MIC, minimum inhibitory concentration; PBPK, physiologically based PK; PD, pharmacodynamic; Ph2A, phase 2A; Ph2B/C, phase 2B and C; Ph3, phase 3; PK, pharmacokinetic; TB, tuberculosis; y.o., year-olds.](https://doi.org/10.1371/journal.pmed.1002842.g001)

---

**Provenance:** Not commissioned; externally peer reviewed.
now-validated in vitro pharmacodynamic (PD) system (IVPDS, or “hollow fiber model” for TB) has significantly improved our understanding of the PK drivers of treatment response in various growth and physiologic states [4,5]. In the IVPDS, an elaborate system of dialysis-like tubing allows the investigator to reproduce human-like concentration–time curves and see how different PK profiles affect killing of bacilli that are living in the system. Dose-fractionation studies, for example, can be carried out, and one can determine whether a drug’s activity is time dependent or, rather, concentration dependent. Or one can test a drug’s activity when the organism is nutrient starved, in log-phase growth, or intracellular. Whereas these models are informative, there remain gaps in our ability to bridge preclinical and clinical data using modern translational quantitative modeling [6]. There are also gaps in our ability to link surrogate end points in early-phase clinical trials (namely, longitudinally collected sputum cultures) and clinically relevant end points of treatment failure, relapse, and death in later-phase trials [7,8]. The identification of accurate tools that identify those patients who are unlikely to achieve cure with shortened regimens (specifically, patients with a disease phenotype that is “hard to treat”) would have immense value to both clinical trialists and TB clinicians [9]. The TB clinical pharmacology field has the opportunity to apply state-of-the-art quantitative pharmacology tools to bridge preclinical and clinical data more effectively and to enhance learning across the continuum of clinical development [9–13]. In this paper, based on discussions occurring at a WHO workshop held in March 2018, we describe our views on best practices for incorporating translational, PK-PD, and microbiologic assessments into drug development [14].

The importance of understanding PK-PD relationships by phase of regimen development

Key uncertainties and questions regarding the use of clinical and translational pharmacology, biomarkers, and microbiology in the evaluation of novel TB treatments are listed in Table 1. Herein, we review these, focusing in on the implications relevant to each developmental phase.

Table 1. Key uncertainties and questions about the use of clinical and translational pharmacology, biomarkers, and microbiology to advance TB treatments that were addressed at the WHO-sponsored workshop, advances in clinical trial design for development of new TB treatments. (Adapted from [15]).

<table>
<thead>
<tr>
<th>Topic Area</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical Pharmacology</td>
<td>What is the importance of understanding PK-PD relationships by phase of regimen development?</td>
</tr>
<tr>
<td>Pharmacometrics</td>
<td>How does quantitative modeling and simulation integrate PK and microbiology-based PD measures (e.g., MIC, bacterial burden as predictive covariates of treatment response) to inform drug development decision-making, especially in later stages of regimen evaluation?</td>
</tr>
<tr>
<td>Preclinical/Translational Pharmacology</td>
<td>Can dynamic experiment-level in vitro assessments (i.e., HFS-TB) be integrated with patient-level bacteriological data to improve quantitative clinical PK-PD predictions and streamline model development?</td>
</tr>
<tr>
<td>Biomarkers</td>
<td>What would be the most efficient framework for bacteriologically based biomarker identification and characterization in clinical trials to enable integration in modeling and simulation-based analyses?</td>
</tr>
<tr>
<td>Bacteriology</td>
<td>Should quantitative PK-PD models describing relevant bacteriologically based covariates be used to guide dose finding and dose optimization in key populations during early development?</td>
</tr>
<tr>
<td>Drug Development</td>
<td>How do we make use of PK-PD across clinical development phases to identify pharmacology-guided drug regimens?</td>
</tr>
</tbody>
</table>

Abbreviations: HFS-TB, hollow-fiber in vitro pharmacodynamic system for assessing TB drugs; MIC, minimum inhibitory concentration; PD, pharmacodynamic; PK, pharmacokinetic; TB, tuberculosis

https://doi.org/10.1371/journal.pmed.1002842.t001
Preclinical drug development

Investigation of PK-PD properties of a candidate drug in the preclinical space is critical for advancing new TB drugs and for building effective combination regimens with a clear rationale for contribution of each new agent. The combination of multiple new chemical entities requires an enhanced understanding of PK-PD across the development paradigm and refined understanding of penetration and mechanism of action within the granuloma. Several experimental tools are utilized at the preclinical stage for new TB drug assessment, each providing unique results that can be used synergistically for decision-making [16]. The European Medicines Agency (EMA)-qualified in vitro hollow-fiber system generates PK-PD data that can be used to refine in vivo animal experiments [16,17]. These in vitro data are integrated with data from multiple in vivo models, which include acute, chronic, and relapsing BalbC (older, well-validated standard mouse model) and Kramnik (newer model with more human-like pathology) mouse models of infection as well as marmoset and rabbit models of disease [18,19]. All of these models provide valuable information on the drug’s potential for microbiologic activity and sterilization (killing of semidormant bacilli) but generate only limited PK-PD data [4,5]. Collectively, these data contribute to understanding spatial distribution of candidate drugs at the site of action in the lung, PK-PD relationships for a single drug, contribution of an individual drug to the entire regimen, synergies of drugs within combinations, and potential for shortening treatment duration. Systems pharmacology models or PK-PD–driven translational platforms are essential in order to adequately assess the true potential of investigational regimens. In the absence of PK-PD–driven translational data, regimens move into late-phase trials with important uncertainties (about the real likelihood that a regimen will produce cure at rates equal to or better than that of the standard-of-care regimen) that traditional microbiologic surrogate markers like sputum culture conversion cannot adequately address [6]. Further, current lack of informative, translational biomarkers that are portable across TB clinical drug development stages for regimen and dose optimization puts further weight on preclinical analyses for de-risking regimen development [20].

The best way to identify new quantitative and translational tools for discovery and optimization of new TB treatment regimens is by investing in and enhancing data collaboration and translational modeling activities. This would support development of a universal preclinical–clinical mechanistic PK-PD system for TB drug combinations with high translational and predictive features to answer questions such as the following: What is the human equivalent dose/schedule of a candidate drug used in a combination regimen that will maximize its contribution to reducing treatment duration? What is the likelihood of achieving treatment durations of 1–3 months with a putative treatment-shortening regimen? And is it possible to shorten treatment duration in all disease phenotypes and all patients? A translational, data-driven mechanistic tool would be able to predict comparative efficacy and intended treatment-shortening potential of new candidate regimens based on preclinical data and optimized translational simulations. The major features of such a translational tool would ideally include: (1) quantification of bacterial growth dynamics in the absence of treatment, (2) quantification of the immune system response in the absence and presence of treatment and as a function of bacterial load and infection time, (3) quantification of the contribution of each drug (concentration–response relationship) to the observed total efficacy of drug combinations, (4) quantification of the interplay between disease pathology and drug response including description of tissue penetration, (5) a fully estimated set of model parameters with variability and uncertainty, and (6) appropriate scaling functions to human PK and PD to allow for accurate translational simulations. Because these components span numerous approaches to drug evaluation, from in vitro studies to clinical trials, data often need to be obtained from multiple
sources; once assembled through collaborations, they can permit accurate translational simulations to help address the critical questions and enable decision-making by stage–gate regimen developers (developers that have divided development into stages with go/no-go decisions at the end of each stage). Integration of these principles will allow for rational selection of the best regimens to be moved forward into clinical development, selection of rational dose ranges to be studied in clinical phases, quantitative predictions of clinical trial outcomes, and informed choice of clinical trial designs.

**Clinical PK and PD**

Prior to initiating a phase 3 registration trial of a new regimen, it is also important to understand the PK of experimental drug(s), exposure–response relationships, PK–toxicity relationships, risk and magnitude of drug interactions, drug safety, sources of variability (in PK, safety, treatment response) in the population, and PK in key populations (Fig 1). Phase I trials provide basic PK and safety information. It is important from a practical standpoint to assess food effect early, as this may impact administration requirements and may complicate approaches to coadministration with companion drugs, as some are taken on an empty stomach (e.g., rifampicin), whereas others are absorbed better with food (e.g., rifapentine, delamanid, bedaquiline) [21–24]. It is also necessary to determine whether weight-based dosing will be required. The requirement for weight banding adds complexities to field implementation as well as reduces the opportunity to coformulate companion drugs into fixed-dose combinations. Additionally, caution is noted, as systematic underdosing of lower-weight individuals can occur when weight banding is used without a reliable clinical PK-PD evidence base [25].

Over the course of phase 1 and 2 testing, assessment of drug PK in geographically and ethnically diverse populations is also invaluable, as variability of drug exposures across populations has been noted [22]. Sparse PK sampling can be employed after identifying optimal sampling times, and population PK modeling is then used to identify factors associated with variability in drug exposures (e.g., sex, race, HIV coinfection, malnutrition). For example, in individuals of black race, bedaquiline exposures are 50% lower than in persons of other racial backgrounds; rifampicin concentrations are very low in children who are malnourished or who have HIV infection; and isoniazid clearance is dependent on N-acetyltransferase 2 acetylator status [22,26,27]. Drug–drug interaction studies should be pursued in middle drug development and not left for late phases of development, particularly for interactions between TB and HIV drugs. The need for drug–drug interaction studies can be assessed based on knowledge of the putative TB drug(s) and standard-of-care HIV drugs’ metabolic pathways and their proclivity for inducing or inhibiting metabolizing enzymes or transporters. If interaction studies of HIV and TB drugs are not conducted early, the impact of the new TB regimens on anti-retroviral therapies (and vice versa) and the resultant effects on viral load suppression and on achieving durable cure from TB will not be understood; as a consequence, the inclusion of HIV patients into late-phase trials will be hindered, limiting the assessment of the safety, tolerability, and efficacy of the regimen in this key population. Moreover, drug–drug interaction studies will still be needed, and substudies will need to be designed, adding significant complexity and delays when they are embedded into late-phase, confirmatory clinical trials [28,29].

In early phase 2A trials, in which a drug or drug combination is administered for 7–14 days to small cohorts of patients \( (n = 15–20) \), a range of doses and schedules is tested for early bactericidal activity (EBA). Data on safety, tolerability, and longitudinal quantitative sputum bacillary loads are collected, and semi-intensive PK sampling is performed to characterize individual drug exposures and PK-PD relationships, which can narrow the doses to be tested in subsequent trials. In phase 2B trials, microbiologic responses to treatment are assessed.
through serial sputum cultures up to 8–16 weeks of treatment. In phase 2C trials, the experi-
mental regimens are administered for their intended duration (e.g., 3 or 4 months), and
patients are followed to collect information on longer-term clinical outcomes (failure, relapse,
death) [30]. Such phase 2B and phase 2C studies are typically multinational and can produce
rich PK and microbiologic data from geographically diverse settings. We propose that sparse
PK sampling be embedded in all phase 2B/C trials. If feasible, sparse PK sampling obtained on
more than one occasion—for example, early in treatment (intensive phase) and then later in
treatment (continuation phase)—would allow the quantification of longitudinal drug expo-
sures that, in turn, characterize exposure–response relationships necessary for selecting the
accurate dose(s) to evaluate in phase 3 trials [13,31]. Additionally, population PK-PD model-
ing can define the relative contributions of factors that lead to delayed culture conversion,
assessing within the model the full suite of potential features, from low drug exposures to clini-
cal factors such as disease severity or patient characteristics. Patient and disease severity char-
acteristics are important to incorporate into models, as the hardest-to-treat phenotypes of
disease disproportionately drive the unfavorable outcomes in contemporary phase 3 trials
[10,11,32]. To date, clinical PK-PD analyses have been unable to adequately inform decision-
making on selecting an optimal regimen duration. A PK-PD tool that predicts the optimal
treatment duration based on data from preclinical studies, phase 2 trials, and both successful
and unsuccessful phase 3 trials would be extremely valuable.

In middle development (at the phase 2 stage), PK–toxicity studies are also needed to define
the therapeutic margin and ensure that dose(s) used in phase 3 are likely to be safe and well tol-
erated. PK–safety relationships influence both dose and schedule (duration, dosing frequency),
with some drugs displaying toxicity associated with cumulative exposure (e.g., linezolid, eth-
ambutol) and others causing more adverse effects when given on an intermittent schedule
(e.g., rifamycins administered thrice or once weekly) [33–35]. Overlapping toxicities can also
be explored with PK data in hand to help discern relationships. For example, prolongation of
the QT segment on the electrocardiogram, a cardiac toxicity that can lead to torsades de
pointes, can be related to concentrations of the parent drug or metabolite and is of increased
concern when QT-prolonging drugs are administered concurrently [36].

Phase 3 trials provide the first opportunity to assess drug efficacy by comprehensively col-
lecting data on drug exposures, adherence, microbiological response over time, safety, and
long-term clinical outcomes; furthermore, this often is the only setting in which reduced treat-
ment durations are tested. Phase 3 studies also offer larger numbers of patients from key popu-
lations (e.g., people living with HIV and children). Because of these features, we recommend
that phase 3 trials include sparse PK sampling whenever feasible and that samples be collected
on all patients. PK-PD assessments can be performed on a subset of study participants to iden-
tify the reasons for poor treatment outcomes. If the trial was successful, these samples would
allow analyses that inform future use and scale-up of the regimens; if the trial was not success-
ful, these data help ascertain the reason(s) why and are critical for determining next steps.

Microbiology and quantitative pharmacology

Although microbiology (e.g., in vitro minimum inhibitory concentration [MIC]) is widely
accepted as an important determinant of response to treatment, integrated PK and microbiol-
ogy-based PD measures built into late-stage clinical trials to confirm relationships are rarely
undertaken. MICs are assessed in preclinical drug development, and the choice of dose and
schedule is driven by the desire, for example, to maintain plasma drug levels above MIC for a
defined duration. This approach has shortcomings, as PK-PD indices are often derived based
on plasma PK, which is often suboptimal compared with the site-of-action PK; traditional
assessment of MIC usually lacks precision [37], and by definition, MIC values indicate inhibition of bacterial growth rather than bacterial killing, which is key for cure. Bacteria with higher MIC are harder to eradicate, and patients with high-MIC bacteria might need more aggressive regimens or longer treatment to achieve cure. Similarly, pretreatment bacterial burden in sputum is highly associated with treatment response. The recent TB Reanalysis of Fluoroquinolone Clinical Trials (TB ReFLECT) meta-analysis revealed that patients with low bacterial burden at baseline could be effectively treated with a shortened (4 month)-duration experimental regimen [9]. Similarly, time-to-culture conversion on standard treatment appears to be shorter in patients with low baseline bacillary load. However, a number of questions remain unanswered—Is there a correlation between baseline bacterial burden and MIC? Does MIC change over time with treatment? Under which circumstances can higher bacterial load or MIC be overcome with higher doses, strong companion drugs, or longer treatment? Should we index PK parameters (Cmax, area under the concentration–time curve [AUC]) to MIC or to a data-informed factor of the MIC?

To address these questions, collection of microbiology data is key in all stages of clinical trials, especially late-stage trials followed up with adequate analysis. Collecting sputum specimens for MIC and bacterial load determinations in clinical drug development so that their value (in subsequent studies and in clinical practice) can be determined is important. *M. tb* isolates should be available for MIC determination from baseline and last positive cultures, and sputum specimens should be assessed over time for changes in bacterial load. Knowledge of strain lineage (e.g., Haarlem, Latin American/Mediterranean, W/Beijing) may also be helpful, as there may be strain heterogeneity with regard to virulence and drug response. A standardized method for providing robust and accurate MIC determinations, such as the 14-drug microtiter plate (ThermoFisher) employed by the Comprehensive Resistance Prediction for Tuberculosis: An International Consortium (CRyPTIC) [38], should be used. Techniques for measuring bacterial burden including time to positivity in mycobacterial growth indicator tube (MGIT) culture, cycle threshold in GeneXpert MTB/RIF assay, or potential novel biomarkers (e.g., quantification of sputum lipoarabinomannan [LAM] levels) should be routinely included in clinical trials to enable investigation of predictive bacterial burden biomarkers [39,40]. Assays such as GeneXpert cycle threshold have the advantage of producing results in real time, though one disadvantage is that DNA from both live and dead bacilli can be detected. Lastly, it is important to align new drugs with new diagnostics. Specifically, detection and characterization of resistance is a key component of TB drug development, and whole-genome sequencing can identify mutations that are associated with decreased susceptibility of *M. tb* strains to new drugs.

**Key populations: Optimal design to extend treatment advances to all**

Young children and pregnant women with TB may be at particularly high risk of adverse outcomes resulting from inadequate TB treatment [41,42]. There are limited data to inform use of TB drugs in pregnant women because they are routinely excluded from clinical trials, and there is no requirement to study them from any regulatory authority. This may change with the report of the Task Force on Research Specific to Pregnant Women (PRGLAC) released in September 2018 [43]. In Europe, a pediatric investigation plan (PIP) is required for drug registration, but there is no requirement for data from the pediatric patient population. In the United States, because TB is considered an orphan disease, the Pediatric Research Equity Act (PREA) does not apply to drugs developed for TB, relieving companies of the requirement to study TB drugs in children for registration. As it may be difficult to recruit children and
pregnant women with TB in any given location, design of clinical trials in these populations must be optimized for efficiency and yield of safety and PK data that will be needed to support dosing recommendations. Full efficacy trials are generally not required [44].

**Children**

Opening doses in different pediatric age cohorts are more likely to be accurate when based on models that incorporate adult PK data and information about developmental pharmacology and when evidence-based target PK ranges are defined explicitly, rather than relying on empiric dose selection (e.g., same mg/kg dose as adults) [45]. Given that drug disposition is most variable between the ages of 0 and 2 years (and changes especially rapidly in the first 3–6 months of life), we suggest including a larger number of children in the youngest cohort to ensure full knowledge of drug disposition in that rapidly developing age group; the sample size of adolescents can be relatively smaller because drug disposition is similar in teens and adults. Key features of a pediatric PK–safety study include model-informed initial dose selection; early interim analysis of PK results in each age cohort (with dose adjustment and model updating); use of optimal sampling theory, a data-driven approach that informs the selection of the most informative time points for PK sampling while minimizing the number of required samples; defining the timing and content of safety visits to reflect knowledge of each drug’s preclinical toxicology and adult toxicity information; clear and evidence-based selection of PK target ranges for parent drug and metabolite(s); and model-based analysis of data by a pharmacometrician.

**Pregnant women**

Studies in pregnant women should take into account pregnancy-related physiologic changes, including changes in renal clearance, drug metabolism, and protein binding [46,47]. Some metabolizing enzymes have higher activity during pregnancy (cytochrome [CYP] P450 2A6, 3A4, 2D6; uridine 5’-diphospho-glucuronosyltransferase [UGT] 1A4), whereas others have lower activity (CYP1A2, 2C19); the magnitude of difference in enzyme activities in pregnant versus nonpregnant women differs by trimester [48]. As selection of doses most likely to achieve (but not exceed or fall significantly short of) therapeutic targets is especially crucial in pregnant women with TB, model-based dose selection is best from both scientific and ethical standpoints. PK assessments should be performed in the second and third trimester and then postpartum so that timing of dose adjustments can be assessed. Depending on the duration for which a given drug is administered, each woman may serve as her own control, reducing variability. Pharmacometric modeling should be used in the analysis so that specific effects of pregnancy on drug absorption, distribution, and clearance can be estimated while considering other cofactors that may affect the drug’s disposition, and recommendations for dose adjustments can be made with maximal knowledge. With regard to safety, whereas a very strong safety signal may be detected in a study powered to detect PK changes, a much larger cohort of women is needed to characterize the full safety profile of a drug in pregnancy for the mother and fetus. Pharmacovigilance via pregnancy registries is one way to achieve this (e.g., http://www.apregistry.com/ for antiretrovirals).

**Site-of-disease PK: Relevance for drug development and optimization**

*M. tb* bacilli are present in multiple compartments in a patient with pulmonary TB but are most numerous in large cavitary lesions that contain liquefied, caseous material. To effect cure, it is currently believed that drugs must penetrate necrotic granulomas and cavitary
<table>
<thead>
<tr>
<th>Question</th>
<th>Consensus</th>
<th>Options</th>
<th>Research</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>What is the importance of understanding PK-PD relationships by phase of regimen development?</strong></td>
<td>PK studies should be included throughout drug/regimen development phases, in both early and late stages of development. PK samples should be collected in all treatment trials with clear documentation of dosing history.</td>
<td>Other PK studies should be performed in the spirit of modern drug development, including the following:</td>
<td>Optimal timing and frequency of PK sampling by type of trial (e.g., phase 2A, 2B, 2C) to yield the most information in the most efficient way.</td>
</tr>
<tr>
<td></td>
<td>A guidance that outlines information to be collected and parameters to be identified at each phase of drug development is needed. This guidance should be organized by sections of minimum information and optimal information. This could be undertaken by a group of individuals with expertise in PK-PD research, such as the WHO Task Force on the PK-PD of TB medicines.</td>
<td>Drug–drug interaction studies, especially with companion TB drugs or antiretrovirals.</td>
<td>Translational modeling and quantitative pharmacology to link preclinical, early-mid clinical (with microbiology outcomes), and definitive trial (with clinical outcomes) results. Role of clinical trial simulation with phase 2 data to inform phase 3 design.</td>
</tr>
<tr>
<td></td>
<td>Importance of PK in phase 2 trials to allow understanding of dose–exposure–response relationships for dose selection in definitive trials.</td>
<td>Evaluation of PK–toxicity relationships for key toxicity concerns (e.g., QTc).</td>
<td>Validation and refinement of translational tools and modeling activities (mouse model, HFS, systems pharmacology model) through data sharing.</td>
</tr>
<tr>
<td></td>
<td>Critical importance of PK–safety assessment in phase 2/3 to inform the need for dose/schedule adjustments. Particularly important for narrow therapeutic index drugs.</td>
<td>Sparse PK collection in phase 3 to strengthen population PK modeling and to explore exposure differences in relevant subgroups including poor responders.</td>
<td>Biomarker (host, microbiology) explorations to find better ways to identify best regimens to carry forward from middle drug development.</td>
</tr>
<tr>
<td></td>
<td>Population PK modeling to understand sources of variability (e.g., sex, race, age, HIV status) in drug exposures and response.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Phase 2B/C studies with arms testing different doses and duration and collection of treatment outcomes will be most informative for identifying regimens most likely to be successful for treatment shortening.</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>How does quantitative modeling and simulation integrate PK and microbiology-based PD measures (e.g., MIC, bacterial burden as predictive covariates of treatment response) to inform drug development decision-making, especially in later stages of regimen evaluation?</strong></td>
<td>Importance of gaining a better understanding of the relevance and value of MIC measurements as well as baseline quantitative bacterial burden in assessments of exposure–response relationships.</td>
<td>M. tuberculosis isolates should be stored, including at a minimum the baseline isolate and that of the last positive culture.</td>
<td>Key research questions to answer by quantitative pharmacology by time of registration:</td>
</tr>
<tr>
<td></td>
<td>Collection of specimens for MIC (genotypic, phenotypic, whole-genome sequencing, etc.) in clinical drug development will allow for value assessment. Isolates should be collected at baseline and during midterm and late-stage development.</td>
<td>Bacterial burden should be quantified longitudinally via collection of serial sputum samples.</td>
<td>PK-PD underpinnings to support dose recommendations, including in hard-to-treat patients and key populations.</td>
</tr>
<tr>
<td></td>
<td>Specific guidance from WHO PK-PD Task Force to provide details on standardized approaches for collection of isolates (which isolates, how to collect, how to store, when to collect, what type of assay would be needed)</td>
<td></td>
<td>PK–toxicity relationships.</td>
</tr>
</tbody>
</table>

(Continued)
lesions to inhibit or kill the viable bacilli that are not expelled by coughing. We should ensure that the drugs achieve adequate bactericidal concentrations in the lesions where bacilli are present. Preclinical and clinical research focused on drug quantification in these matrices may help inform regimen selection for treatment-shortening trials, including drugs, doses, duration, and companion drugs. This is an area in which translational PK-PD research may be particularly valuable. In rabbit models of pulmonary TB that have human-like pathology, it has been observed that some drugs have excellent penetration into lesions, as assessed spatially by matrix assisted laser desorption/ionization (MALDI) mass spectrometry or quantitatively by laser capture dissection and laser capture microdissection liquid chromatography mass spectrometry (LCM-LC/MS), whereas others display poor lesion penetration [49–51]. Of note, all four current first-line TB drugs reach therapeutic concentrations in TB lung lesions [52].

Patients with highly drug resistant TB who must undergo lung resection for cure have participated in research aimed at measuring drug concentrations in lung compartments following observed dosing [1]; data from such investigations help bridge preclinical and clinical studies and provide evidence concerning the contribution of drug PK to acquisition of drug resistance in lung microenvironments [53]. With rabbit lesion penetration data for a novel drug as well as data on human plasma PK and treatment outcomes, translational models can be built that shed light on the differential response to TB treatment that results from differences in lung pathology [19]. These strategies may help reduce the risk of late-phase failure for drugs with promising preclinical and early clinical microbiologic efficacy by identifying early those compounds with poor penetration into critical lung lesions. Translational models may also be important in developing therapies for other manifestations of TB disease, like central nervous system or extrapulmonary TB.

### Table 2. (Continued)

<table>
<thead>
<tr>
<th>Question</th>
<th>Consensus</th>
<th>Options</th>
<th>Research</th>
</tr>
</thead>
<tbody>
<tr>
<td>Can dynamic experiment-level in vitro assessments (e.g., HFS) be integrated with patient-level microbiology data to improve quantitative clinical PK-PD predictions and streamline model development?</td>
<td>Development and validation of novel biomarkers should be integrated in all PK-PD activities to allow for rapid assessment of the biomarkers and properties of future potential surrogates for bacterial load.</td>
<td>Investment in development of translational tools and modeling activities (mouse model, HFS, systems pharmacology model) that can inform regimen composition.</td>
<td>Drug–drug interactions with companion TB and HIV drugs.</td>
</tr>
<tr>
<td>What would be the most efficient framework for microbiology-based biomarker identification and characterization in clinical trials to enable integration in modeling and simulation-based analyses?</td>
<td></td>
<td>Culture-free (and sputum-free) systems as alternatives to existing culture-based systems are urgently needed.</td>
<td>Evaluation of value of MIC (static drug concentration in relevant medium) versus dynamic susceptibility information in drug and regimen assessment.</td>
</tr>
<tr>
<td>Should quantitative PK-PD models describing relevant microbiology-based covariates be used to guide dose finding and dose optimization in key populations during early development?</td>
<td>Design of studies in key populations should be supported by clinical pharmacology principles (dosing regimen) and aided by model-based design.</td>
<td></td>
<td>Investment in development of translational tools and modeling activities (mouse model, HFS, systems pharmacology model) that can inform regimen composition.</td>
</tr>
</tbody>
</table>

Abbreviations: HFS, hollow fiber system; MIC, minimum inhibitory concentration; PD, pharmacodynamic; PK, pharmacokinetic; QTc, corrected QT interval on electrocardiogram; TB, tuberculosis

https://doi.org/10.1371/journal.pmed.1002842.t002
Conclusions

Modern drug development tools using quantitative and translational pharmacology and microbiology are proving to be invaluable when applied to TB drug and regimen development programs. We can further improve on these tools by constructing predictive, fully translational models that fully integrate data and knowledge from diverse models and sources including in vitro susceptibility data, drug(s) mechanism-of-action characteristics, hollow-fiber model PK-PD data, cure results from multidrug studies in different animal models, phase 2 longitudinal microbiologic data, and information on PK-PD, adherence, and well-defined clinical outcomes from carefully conducted phase 3 trials (Table 2). Comorbidities, sites of disease, characteristics of the infecting strain, and host immune status are also highly relevant; information on these elements can further enhance model performance. For the clinical phases of development, studies of drug interactions with relevant ART agents should be conducted early to allow the inclusion of HIV-infected patients in definitive trials. Similarly, children and pregnant women with TB should also be included in well-designed safety and PK studies.

Acknowledgments

Thanks to WHO for organizing the conference in Montreux and especially to Christian Lienhardt and Payam Nahid for their leadership in this effort.

References


COLLECTION REVIEW

Accelerating the transition of new tuberculosis drug combinations from Phase II to Phase III trials: New technologies and innovative designs

Geraint Davies1*, Martin Boeree2, Dave Hermann3, Michael Hoelscher4,5

1 University of Liverpool, Liverpool, United Kingdom, 2 Radboud University, Nijmegen, the Netherlands, 3 The Bill and Melinda Gates Foundation, Seattle, Washington, United States of America, 4 Division of Infectious Diseases and Tropical Medicine, Klinikum of the University of Munich, Munich, Germany, 5 German Center for Infection Research (DZIF), partner site, Munich, Germany

* gdavies@liv.ac.uk

Summary points

- Preclinical models of tuberculosis have significant limitations in selecting composition and duration of regimens for tuberculosis.
- Innovative early-phase clinical trial methodologies and technologies have the potential to reduce risk and accelerate drug development in tuberculosis.
- Phase IIA monotherapy studies are optional for proof of concept but may be useful for dose-finding in conjunction with pharmacokinetic–pharmacodynamic methods.
- Innovative Phase IIB designs are increasingly common in tuberculosis drug development, utilising multiarm selection designs, sometimes in an adaptive format.
- Novel biomarkers including liquid culture and nucleic acid amplification appear capable of replacing conventional solid culture in early-phase development.
- Phase IIC and ultrashort noninferiority designs attempt to mitigate the problem of estimating duration of treatment regimens from Phase II results alone.

Background

Tuberculosis (TB) remains the single biggest killer among infectious diseases, and treatment is prolonged, complex, and vulnerable to the development of resistance. Innovation in TB therapy is desperately needed, but the transition from first-in-patient studies to pivotal clinical trials in the treatment of TB is plagued by uncertainty for several reasons. In preclinical development, a limited understanding of mycobacterial physiology in vivo and an inability to closely mimic human pathology in animal models limit confidence in the translational predictions that may be used to plan early-phase trials. Furthermore, these trials face problems in accurately and rapidly measuring response to treatment in accessible clinical samples because of a continuing reliance on mycobacterial culture of sputum and constraints posed by
emergence of resistance during prolonged monotherapy. Irreversibility of the response precludes the use of crossover designs, and the need for combination treatment regimens containing three or more drugs limits the scope of dose-finding studies in clinical development. Anti-TB drugs have widely differing and sometimes poorly understood mechanisms of action capable of producing qualitatively different patterns of response, as well as variable physicochemical characteristics, distribution, and toxicity profiles. It is important, therefore, to recognise that a strategic approach to development pathways for different drugs may need to reflect these differences while identifying and preserving the core elements of a successful critical path to registration. At the same time, it is generally accepted that drug developers in TB need to focus more on codevelopment of regimens rather than drawing only on the lessons learned from development of individual drugs, which poses additional challenges in evaluation of both safety and efficacy [1]. Policymakers are increasingly seeking to integrate innovative clinical trial approaches into their decision-making to facilitate early and effective deployment of the best regimens [2], as emphasised by this collection. In the TB context, the key information to be gained in Phase II development includes obtaining initial proof of concept, finding the optimal dose for individual agents, selecting the best combinations of drugs to be further tested in Phase III trials, and predicting the likely necessary duration of a future regimen. Regulators have recently demonstrated openness to innovative approaches in these areas and flexibility around when and how key information on efficacy and safety of individual agents may be obtained during development. A number of new directions promise to improve the means by which these goals are currently achieved in TB trials, which we present in this paper (Fig 1).

Phase IIA monotherapy and combination studies

The traditional approach to Phase IIA (‘first-in-patient’) studies in TB since the 1980s has been short-term trials of treatment based on reduction of colony-forming units of *M. tuberculosis* in repeated sputum samples over the first 14 days of treatment, termed ‘early bactericidal activity’ (EBA) studies [3]. Such studies, using small sample sizes (10–15 per arm), have been used for initial dose-finding in humans, studying pharmacodynamic interactions between drugs, and more recently for selecting combinations, often based on indications from various mouse models [4]. They are economical relative to other phases of development and have the potential to demonstrate the contribution of individual drugs in humans prior to commencing study of combinations. This proof of concept has been particularly important in TB, given that pharmacodynamics cannot be investigated in healthy volunteer studies. However, a number of drugs that are known to influence long-term outcomes of treatment have little or no impact on quantitative bacteriology in the early phase of therapy [5,6], suggesting that positive results in a Phase IIA trial may be dispensable for some important agents, though it remains uncertain whether these agents can currently be clearly identified in advance during preclinical development. In addition, although a small Phase IIA trial may contribute to initial evaluation of safety in patients under conditions of monotherapy, their small sample size contributes only modestly to expanding the safety database from Phase I studies.

For these reasons, the value of Phase IIA monotherapy studies has been questioned, since positive results are neither sufficient nor even necessary for progression in development. They do represent, however, the first and last time that evidence on the contribution of individual drugs can be obtained during a development programme, and for this reason, a number of modifications to improve this approach have been suggested. Since it is clear that Phase IIA studies are prone to misinterpretation due to differing patterns of pharmacodynamics and relatively high interindividual variability, modern studies are usually conducted for a full 14 days [7], which is widely considered to be the ethical limit for monotherapy, beyond which the risk
of generating resistance is too great. This period appears long enough to capture the full pharmacodynamic behaviour of drugs in the early phase of treatment. A major innovation in recent years has been the systematic application of pharmacokinetic–pharmacodynamic (PK–PD) methods to the analysis of Phase IIA studies, which have helped to discriminate different patterns of response, account for important baseline prognostic variables, and improve the power of comparisons while also clarifying dose–response relationships using pharmacokinetic data.

Another innovation is the replacement of traditional bacteriology performed on selective solid media by other methods, such as liquid culture in the mycobacterial growth indicator tube (MGIT) system, which appears to have similar variability to solid culture and may remain positive for longer, providing a more plausible link to later-phase studies [8]. However, the relationship between colony counts on solid media and time to positivity in liquid culture changes over time, and the two measurements may not be completely interchangeable in longer studies [9]. Most recently, molecular assays promoted as candidates to replace culture-based techniques have begun to be evaluated. The DNA-based Xpert MTB RIF assay appears to lack the dynamic range required to be useful in early-phase studies, though it has been used...
An important recent trend has been for Phase IIA studies to go beyond 14-day proof-of-concept and dose-finding studies for individual drugs, and increasingly, focus has shifted to evaluating combinations of drugs. This is important because a recent meta-analysis of Phase II studies in TB noted that there was almost no overlap between the regimens studied in Phase IIA and Phase IIB [17], suggesting that decisions on drug selection and combination have historically been based largely on considerations other than performance in EBA studies, usually combination studies in the mouse model. However, uncertainty remains about how predictive such studies may be in humans, and the ability to confirm and select among at least a subset of promising regimens in clinical studies would clearly be desirable. Combination Phase IIA studies may follow separate initial monotherapy studies [18], or a monotherapy run-in of 14 days may be followed by a 14-day study of combinations (14+14 design) [19]. Such an approach representing a fusion of dose-finding and selection of combinations appears to be time-efficient and would represent a more significant expansion of the safety database in terms of numbers and duration. A similar 7+7 design has been successfully employed in a formal maximum tolerated dose study for dose-finding of rifampicin [20]. However, although the duration of exposure of patients to novel or higher doses of existing agents must clearly be determined by preexisting preclinical toxicology data, given that ethical concerns about resistance are minimised by the use of combinations, restricting Phase IIA combination studies to a period of 14 days appears arbitrary, restricting information about longer-term response. It also remains unclear whether some of these studies are large enough to formally discriminate among regimens reliably.

Phase IIB studies

Traditionally, Phase IIB studies in TB have relied on the endpoint of culture conversion at 8 weeks, based on its simplicity and the abundance of historical data showing that it is a useful, yet imperfect, surrogate endpoint for long-term treatment response [21,22]. However, reliance on this binary endpoint typically mandated relatively large sample sizes for such trials, for statistical reasons, and posed problems in using them for dose-finding and selection of combinations [23,24]. It also resulted in an inability to directly compare quantitative bacteriological results obtained in Phase IIA studies with the 8-week endpoint, meaning that there was almost no direct translational linkage between the two.

Over the last decade, however, new approaches to Phase IIB studies based on longitudinal statistical modelling of quantitative bacteriology, time to positivity in MGIT, or time-to-culture conversion data have been increasingly adopted by investigators (Fig 2) [25–27]. These studies have continued to be performed over the first 8 weeks of treatment but have employed more intensive sampling of sputum at earlier time points. As experience with effect sizes using these novel analyses has grown, such approaches have facilitated greater flexibility and economy in trial designs with reduced sample sizes (40 per arm) [28], although as yet they lack the support from long-term studies that is enjoyed by 8-week culture conversion. However,
because these outcomes are measured on a continuous rather than binary scale, they offer longer-term advantages in terms of validation over the 8-week endpoint, which is approaching 100% culture conversion in the comparator arm as regimens improve in efficacy. The confidence attached to the 8-week endpoint is not easily transferred to other single time points with which there is less experience. Worthy of special emphasis is the importance of laboratory quality. For any of these dichotomous and quantitative bacteriologic endpoints to be maximally informative, significant investment in and oversight of quality of microbiologic assays and laboratory procedures are essential.

The advent of these more quantitative endpoints has led to reconsideration of the purpose and design of Phase IIB trials and their possible use to achieve some of the objectives of Phase IIA trials of combination regimens. A multiarm Phase IIB trial method has been successfully used to select among members of the fluoroquinolone class substituted into the first-line combination regimen for progression to Phase III [29]. The results of these pivotal trials initially appeared disappointing and highlighted difficulties in accurately predicting duration of regimens from effect sizes in these novel IIB designs. However, recent pooled reanalyses are suggestive that the treatment-shortening potential of these regimens may in fact be confined to a majority subgroup of patients [30]. Similar Phase IIB designs have subsequently been used to select regimens containing the novel agents pretomanid and bedaquiline for Phase III trials, though the latter have not yet been completed [31]. More recently, three other similar Phase

---

**Fig 2. Schematic illustration of alternative bacteriological approaches to measurement of elimination of organisms in respiratory specimens over time in clinical trials of TB.** After conversion of cultures to negative in the first weeks of treatment, subsequent stable cure is defined only by the absence of relapse (return of positive cultures). CFU/MGIT, modelling of colony-forming units or mycobacterial growth indicator tube data; TB, tuberculosis.

https://doi.org/10.1371/journal.pmed.1002851.g002
IIB trials were used for dose-finding among the rifamycins, rifampicin, and rifapentine [32–34]. All these trials were explicitly designed and analysed using a PK-PD modelling approach, and two of them successfully demonstrated relationships of bacteriological response with dose and/or plasma concentrations, which have also been used to plan ongoing Phase III trials.

The demonstration of the feasibility of Phase IIB designs for dose-finding and selection of combinations has naturally led to exploration of adaptive designs that may offer the opportunity to test a broader range of both combinations and dose levels without increasing the number of patients enrolled prohibitively. An early attempt to apply this approach was unsuccessful and pointed to limitations on opportunities to adapt when using an 8-week primary endpoint [35]. However, a multiarm multistage (MAMS) design, similar to those developed in oncology, has been subsequently successfully implemented, eliminating two of its five arms, which contained the investigational agent SQ-109, prior to full enrollment [36,37]. This study demonstrated that meaningful adaptation in TB trials is possible but also drew attention to some of the challenges, particularly delays in receiving culture information, which must be balanced against a relatively slow rate of recruitment and the challenges of medication management of many diverse treatment regimens, which would usually preclude complete masking of treatment for purposes of safety assessment.

Predicting long-term outcome and duration

In order for regimens to be reliably selected in Phase II, investigators need to have reasonable confidence that the intermediate bacteriological endpoints on which they currently rely can be trusted to correctly predict treatment effects on definitive long-term outcomes, such as treatment failure and relapse. Several analyses have addressed this question, largely focusing on the 8-week culture conversion endpoint. In an early meta-regression analysis of the historical British Medical Research Council trials, the 8-week endpoint was found to be a reasonable predictor of long-term outcome for rifampicin-based regimens [20]. An extension of this analysis confirmed these results and developed a prediction model for the duration of a regimen required to produce acceptable rates of relapse, which appeared to perform well when applied to new datasets involving classes of drugs not included in the training set [38]. Finally, an analysis comprising all historical regimens included in TB trials confirmed the usefulness of this meta-regression approach while showing that the relevant relationships may be different for regimens that do or do not contain rifampicin [39]. These data, while supporting the utility of a meta-regression approach, suggest that 8-week culture conversion is a useful but imperfect surrogate endpoint for long-term treatment response and may be subject to drug class effects. Although this is also likely to be true for other intermediate bacteriological endpoints, data to support a similar analysis based on time-to-event or bacillary elimination rates are as yet too sparse to replicate this approach.

Alternatives to this indirect approach involve collecting varying degrees of follow-up data after Phase IIB trials have been completed. The simplest way to do this is to follow all the patients enrolled in a Phase IIB trial to the end of their complete regimen and for a defined period post-treatment, usually 12 months. This design, termed STEP Phase IIC by its proponents, permits estimation of a Bayesian prediction interval for the likely results of a future Phase III trial, with the advantage that the prediction is less dependent on intermediate outcomes than in the meta-regression approach [40]. However, slightly larger sample sizes (80 per arm) than those typically used for a Phase IIB selection design are desirable, and the intermediate results are still used as a threshold to prevent participants being exposed to very poorly performing regimens. The Phase IIC design is thus related to that used in the Phase III TRUNCATE-TB trial, which is evaluating ultrashort 2- to 3-month regimens in a similar way but with
fully powered noninferiority comparisons for which a much larger sample size is required [41]. A third approach, which has not yet been implemented in TB, is a fully seamless Phase II/III design in which adaptive evaluation of regimens in the Phase II stage is followed by enrichment of the successful arms with additional participants in the Phase III stage to achieve appropriate power for comparisons on the long-term outcomes in the selected arms [42].

Preclinical information could also be a useful source of data for initial predictions of duration of regimens, based either on bacterial elimination rates or relapse experiments. An approach based on translational PK-PD modelling of mouse data has been used to predict the results of a number of clinical trials with reasonable success [43]. Similar approaches combining preclinical and clinical data using Bayesian approaches have the potential to guide early-phase clinical development decisions in real time.

**Safety considerations**

Similarly to other complex therapeutic areas, effective therapy for TB relies on combining at least three and as many as seven drugs. Although the safety profile of many existing or repurposed drugs is relatively well characterised, the concept of universal regimens combining multiple novel agents raises some issues of interpretation of toxicity signals, which may only be clearly resolved when there are relatively extensive safety data generated during Phase I and II monotherapy studies. However, although such studies can address short-term toxicities, longer-term issues can only be addressed by good laboratory practice (GLP) standard toxicity studies in nonhuman animal models conducted at the appropriate time in the development programme. Regulatory guidance recognises this problem and specifies that, provided that complete preclinical development is also carried out for each component of a combination, an animal combination toxicity study equivalent to the duration of planned clinical trials, up to a maximum of 90 days, in a single relevant species would be sufficient to support marketing [44]. In some cases this requirement may be waived if no overlapping toxicities are observed in the preclinical programmes for each component, but further studies may be necessary if unexpected toxicities not observed with any of the components occur in the combination study.

**Conclusions**

Methods for transitioning of TB drugs and regimens through Phase II to Phase III have evolved rapidly in the last decade. Innovation in clinical trials methodology and evaluation of biomarkers have increased the confidence with which multiarm selection and adaptive designs have been adopted by the TB trials community. Although clinical monotherapy studies do not appear mandatory from a regulatory point of view, they may assist developers when significant uncertainty in preclinical development requires strong proof of concept in humans. The recently increased emphasis on development of combinations, however, may benefit from an increased variety of possible trial designs based on longitudinal bacteriological responses over various durations of therapy and making use of adaptation within sustainable clinical trial platforms. Acknowledging the imperfect nature of current intermediate outcomes, investigators are also seeking to increase confidence in Phase III planning by obtaining limited long-term follow-up in extended Phase IIC designs or in bypassing Phase IIB altogether. Translational modelling approaches may complement these data by explicitly bringing preclinical data to bear on clinical development decisions. These innovations promise to reduce risk and accelerate early-phase clinical development in TB, increasing the confidence that regimens selected for Phase III trials contain the right drugs at the right doses and maximising the possibility of successfully achieving reductions in treatment duration for patients.
References


Keeping phase III tuberculosis trials relevant:
Adapting to a rapidly changing landscape


1 School of Medicine, UCSF Center for TB, University of California San Francisco, San Francisco, California, United States of America,
2 Harvard Medical School, Boston, Massachusetts, United States of America,
3 School of Public Health, University of Minnesota, Minneapolis, Minnesota, United States of America,
4 Unité Mixte Internationale TransVIHMI (UMI 233 IRD, U1175 INSERM, Université de Montpellier), Institut de Recherche pour le Développement Montpellier, Montpellier, France,
5 Medical Research Council Clinical Trials Unit at University College London, London, United Kingdom

*patrick.phillips@ucsf.edu

Summary points

• The landscape of tuberculosis (TB) treatment has evolved considerably over the last 10 years, necessitating careful consideration of various trial design aspects to ensure that TB phase III trials are still impactful at trial completion, often more than 4–5 years after initial design.

• The choice of control is guided by the specific trial objectives, weighing the relative merits of internal validity and external generalizability alongside randomization in making the correct inference. A particular challenge occurs when international or national guidelines change during the trial.

• Improved execution and relevance of noninferiority trials for TB require greater emphasis on study quality, especially maximizing treatment adherence and minimizing missing outcome data; preferred use of intention-to-treat rather than per-protocol analyses; more careful justification of the margin of noninferiority; and consideration of recent innovations such as a Bayesian approach to noninferiority.

• Many adaptive trial designs are well suited to optimization of TB treatment. A thorough understanding of type I error rates and biases in treatment effect estimates is critical for regulatory approval and consideration in establishing World Health Organization (WHO) guidelines.

• Treatment stratification is an area of limited experience for TB trials, and trialists must learn from well-established methodology in other disease areas.

• Explanatory trials are important for evaluating the efficacy of an intervention under close to ideal conditions. However, no single trial can address all relevant questions about a given therapeutic intervention at one time, and pragmatic trials will be essential for public health and policy decision-making purposes.

• TB treatment trials today should favor bold and creative approaches that can produce high-quality evidence for effective, patient-centered care made accessible to all 10 million new TB patients, including the half-million with drug-resistant TB (DR-TB), each year.
Introduction

One of the first multicenter randomized trials was the British Medical Research Council (MRC) streptomycin trial [1]. From the first meeting of the special committee to “plan trials of streptomycin in tuberculosis” in September, 1946, the primary trial results from 107 participants followed for 12 months were published in the British Medical Journal two years later in October, 1948 [2]. Although treatment with a single drug was subsequently shown to be inadequate because of the generation of drug resistance [3], the results of the trial changed clinical practice [1].

It has, however, become difficult to conduct phase III clinical trials in the 21st century in any disease area in such a short time frame. Trials often require more patients to show benefit, and their initiation is often protracted because of the need for independent ethical review(s), approval by national regulatory bodies, and the training and compliance with Good Clinical Practice (GCP) that is necessary to ensure that the trial is designed and conducted to the highest standards. All of these changes have been aimed, rightly, at protecting participants and ensuring reliable results, but they have also limited the ability to conduct clinical trials to respond quickly to important public health questions, especially in the context of a rapidly evolving disease and treatment landscape. The interval from start of enrollment to first public presentation of results of recent phase III tuberculosis (TB) trials ranges from 4.6–8.4 years [4–7]. This does not include time for design, planning, and ethical and regulatory approvals prior to start of recruitment, which commonly takes at least a year, and is consistent with a systematic review of time to publication of results across other disease areas [8].

The landscape of TB treatment has evolved considerably over the last 10 years—particularly in the management of drug-resistant TB (DR-TB). Changes include the earlier diagnosis of DR-TB with widespread implementation of newer tests such as GeneXpert [9,10] and Line Probe Assays [11], a better understanding of the pharmacology and bactericidal activity of the various drugs used [12], and the introduction of new drugs (bedaquiline with accelerated approval by the United States Food and Drug Administration [FDA] in December, 2012 and delamanid with conditional approval by the European Medicines Agency [EMA] in November, 2013), as well as observational studies and clinical trials investigating various combinations of current, new, and repurposed drugs in an attempt to shorten DR-TB therapies [13–17]. These developments are reflected in evolving World Health Organization (WHO) guidance for DR-TB (see Table 1). Furthermore, knowledge about the epidemic itself continues to evolve with a recognition of the growing importance of the transmission of DR-TB [18,19] and increasing levels of second-line drug resistance [20]. Thus, in 2017, among the 10.0 million people developing TB disease, 558,000 (5.6%) developed a form that was resistant to at least rifampicin, the most effective first-line drug, and 230,000 died of it. The severity of national epidemics varies widely among countries. Estimated prevalence of rifampicin-resistant TB (RR-TB) among new TB cases ranges from 1.3% in Kenya to 38.0% in Belarus among the 30 high-TB–burden countries [20].

Given the unavoidably protracted duration of phase III TB trials in the 21st century, the status of TB as a global health priority (the first ever United Nations [UN] General Assembly high-level meeting on TB was held in September, 2018), and the increasing trial costs relative to a huge shortfall in research and development funding [21], those who conduct clinical trials are obligated to design them in such a way that they will have a direct impact on policy and practice of TB treatment at the projected time of trial completion. In this paper, as part of a PLOS Medicine Collection on Advances in Clinical Trial Design for Development of New TB Treatments [22], we discuss how phase III TB trials could be designed with such “future-proofing” in mind.
Choice of control

It is usual for the comparator in a clinical trial to be the standard of care treatment so that the results can be interpreted in relation to current practice [23]. Furthermore, the principle of clinical equipoise provides an ethical obligation to ensure patients on the control arm receive the best available standard of care [24]. In drug-sensitive TB (DS-TB), a 6-month regimen of rifampicin and isoniazid, supplemented by pyrazinamide and ethambutol in the first 2 months, is the recognized standard of care [25]; all recent phase III trials have therefore used...
this regimen as the internal control. For DR-TB, WHO guidelines provide a recipe for constructing an effective regimen based on the combination of drugs from various classes, leading to variability in terms of regimen composition across patients and trial sites. The most recent guidelines go further and recommend both long and short regimens [26]. For these reasons, and in the absence of an established standard, trials have selected various approaches to the choice of control (see Table 2 for a discussion of advantages and limitations). For example, the design that adds a single new drug (or placebo) to a background regimen has shown its limitations because it provides no information on how to use the drug in a regimen, which is essential for programmatic implementation; effect of drug can be masked if background regimen is highly effective.

### Table 2. Controls used in DR-TB trials.

<table>
<thead>
<tr>
<th>Choice of control</th>
<th>Examples</th>
<th>Strengths</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Placebo, added to optimized background regimen</td>
<td>Delamanid phase II and III trials [28,29], Bedaquiline phase II trial [30,31], Opti-Q [32]</td>
<td>Permits blinding of healthcare providers and participants, yielding unbiased estimates of the efficacy and safety of the individual drug. This design was used to inform regulatory approval of new drugs.</td>
<td>Yields little or no information on how to use the drug in a regimen, which is essential for programmatic implementation; effect of drug can be masked if background regimen is highly effective.</td>
</tr>
<tr>
<td>External control (historical or concurrent)</td>
<td>NiX-TB (NCT02333799), ZeNIX-TB (NCT03086486)</td>
<td>Smaller sample size and operational efficiencies due to absence of randomization and use of only one regimen. Considered the only option if there is no accepted standard of care. The justification for use of a historic control can only be used in the first successful trial in that patient population; subsequent trials could use the previous intervention as internal control.</td>
<td>Highly dependent on choice of external control, differences between patient populations and secular trends (with a historical control) affect interpretation of results. Challenging to quantify how much &quot;supportive care&quot; in the trial affected outcomes relative to control outside trial [33].</td>
</tr>
<tr>
<td>Randomized comparison in DS-TB, parallel uncontrolled DR-TB cohort with same regimen</td>
<td>STAND (NCT02342886), SimpliciTB (NCT03338621)</td>
<td>Randomized comparison in DS-TB provides strong evidence for safety of regimen in TB patients and efficacy in DS-TB. The parallel DR-TB cohort informs whether results differ between the two TB patient populations.</td>
<td>Only appropriate for regimens that are targeted for both DS- and DR-TB. Extrapolation from DS-TB comparison to DR-TB population requires assumptions.</td>
</tr>
<tr>
<td>Local standard of care (varying by site)</td>
<td>STREAM Stage 1 [34], endTB [35]</td>
<td>Better external validity because of randomization to genuine standard of care, operational efficiencies because sites use local standard for control arm participants.</td>
<td>Control regimen may differ by site and over time. This will increase variability in results and may need to be accounted for by increasing sample size.</td>
</tr>
<tr>
<td>Prescriptive regimen</td>
<td>NEXT (NCT02454205), STREAM Stage 2 [36]</td>
<td>Better internal validity because of clear randomized comparison of two regimens.</td>
<td>Limited external validity since choice of control regimen may not reflect standard in many places. This would change if a standardized regimen is widely adopted; there are currently variations in how the short regimen is used (for example, choice of fluoroquinolone and bedaquiline in South Africa).</td>
</tr>
</tbody>
</table>

**Abbreviations:** DR-TB, drug-resistant TB; DS-TB, drug-sensitive TB; TB, tuberculosis.

https://doi.org/10.1371/journal.pmed.1002767.t002

---

**Box 1. A case study: The STREAM trial [34,36]**

The STREAM trial was initiated to evaluate a novel 9- to 11-month regimen for the treatment of multidrug-resistant TB (MDR-TB) based on results from an observational cohort study in Bangladesh [13,16]. The primary objective of Stage 1 of this multicenter randomized trial was to determine whether a slightly modified version of this regimen
(with high-dose moxifloxacin replacing gatifloxacin) was safe and at least as effective as the recommended standard of care. The first trial participant was enrolled in July, 2012, with the first results due to be published in 2019. During the trial period, the landscape changed substantially (as described in main text), and the trial had to adapt in a number of ways.

**Incorporating new drugs in additional trial arms**

The availability of new drugs without data on how to use them in combination regimens prompted the investigators to consider transitioning from a two-arm study (STREAM Stage 1) to a four-arm study (STREAM Stage 2), with the aim that any new arm(s) added to the trial should be shorter or simpler to take and intended to be less toxic. Through wide consultation, an injectable-sparing regimen to avoid the hearing loss associated with aminoglycoside use was selected. There was also preference for a regimen that excluded prothionamide and ethionamide because these drugs cause nausea and vomiting that compromise the tolerability of any MDR-TB regimen. The decision was therefore made to add four arms, a 9-month completely oral regimen in which kanamycin was replaced by bedaquiline and a shorter 6-month regimen in which bedaquiline replaced prothionamide and kanamycin duration was reduced to 8 weeks. The primary objective of Stage 2 was to evaluate whether the bedaquiline-containing regimens were safe with efficacy not inferior to that of the 9- to 11-month control regimen.

The first patient in Stage 2 was randomized in April, 2016. In order to ensure timely completion of the main comparison of the fully oral regimen with the 9- to 11-month injectable-containing regimen and the increasing desirability of an injection-free regimen, it was decided to terminate enrollment to the 6-month injectable-containing regimen in 2018.

**Choice of control**

The locally used 20- to 24-month regimen consistent with 2011 WHO guidelines [39] was selected as the control arm in Stage 1. Although results from Stage 1 were not available at the time that Stage 2 was initiated, the STREAM investigators took the unconventional step of including two control regimens: (i) the 9- to 11-month regimen studied as the intervention in Stage 1 and (ii) the locally used WHO-recommended regimen that had been the control in Stage 1. The second control was considered as a “reserve internal control.” Although it was not included in the primary objective and only 1 in 7 participants were to be allocated to this arm, it was to be used as a comparator in secondary analyses to permit interpretation of the results of the trial as compared to 2011 WHO guidelines.

In May, 2016, the revised MDR-TB treatment guidelines from WHO recommended a short regimen very similar to that being evaluated in the STREAM trial (see **Table 1**) for patients who met specific inclusion criteria. Subsequently, countries adopting these revised guidelines were no longer able to enroll patients to the second control arm. The protocol was therefore amended to exclude enrollment to the “reserve internal control” in these countries, thereby unfortunately reducing the value of comparisons to that regimen because of fewer participants.

Continued evolution in WHO guidance [26], and its implications for the control arm, is under consideration by the investigators.
An added complication arises when international or national guidelines change during the course of a trial, as exemplified in the STREAM trial [34,36] (see Box 1). If significant new developments arise during the course of a trial that may impact a participant’s willingness to continue, investigators have a responsibility to inform patients; this is part of the Federal Code of Regulations in the US. It may not be feasible or ethical to continue the trial without modification under such circumstances; conversely, if the evidence base for change is weak [37] and randomization among treatment arms is still possible, no change in the trial design may be warranted [38]. The investigators, usually with advice of an independent group such as the data and safety monitoring board (DSMB) or a community advisory group, should evaluate the new information and make a judgment about whether the trial protocol should be modified and how, if at all, participants should be informed. Policy makers and guideline developers can help with this by including explicit wording that further research is still needed when making recommendations that are based on low certainty in the evidence.

Noninferiority, analysis populations, and “estimands”

Noninferiority trials are designed to evaluate whether a reduction in efficacy with the intervention as compared to the control does not exceed a prespecified threshold. The prespecified difference is denoted as the noninferiority margin (see Fig 1 for an illustration of the results of a noninferiority trial). The choice of the margin in trials of TB treatment regimens continues to be a major discussion issue. In order to have confidence that the new treatment is better than no treatment, it is accepted that the margin should be no larger, and considerably smaller, than the estimate of benefit of the chosen control as compared to no treatment. This effect is, however, estimated from historical data [41], and the statistical uncertainty of the estimate should be taken into account. For example, a somewhat conservative estimate of the success rate of standard therapy in DS-TB is around 85% [42], which, compared to the estimated 30% survival from untreated TB [43], gives an estimate of an absolute treatment effect of 55%. A declaration of noninferiority with a margin of 10% would therefore give confidence that more than 80% of this effect of the control is preserved, and 90% would be preserved with a margin of 5%.

Consideration of the expected benefits of the intervention shapes the final choice of margin. In TB, regimens that are shorter confer a direct benefit to patients and health systems. They are expected to result in better treatment adherence in addition to reduced patient and health...
system costs, although none of these can be easily measured in a clinical trial that is not close to usual practice. Anticipated reduction in toxicity is another factor that may influence the choice of margin; whether this is a consequence of the new treatment or not cannot be known until the trial has been completed. Collection of good safety data is critical to properly weigh the risks and benefits of an intervention. Combining efficacy and safety in a composite outcome or a formal risk–benefit scoring system [44] is useful to summarize this balance in a single measure. Such measures can, however, obscure differences between outcomes of varying severity. Papers summarizing the primary results of trials should, therefore, report safety and efficacy outcomes separately for others to make a judgment on the risk–benefit balance.

There has been a trend towards larger noninferiority margins in a number of recent protocols; this permits a reduction in sample size, resulting in a less expensive study and earlier completion, but leads to greater uncertainty as to the true efficacy of the regimen. Widening the margin increases the possibility that a substandard regimen could be accepted as a new gold standard, thereby increasing the risk of “biocreep,” whereby after several generations of noninferiority trials, considerably less effective regimens would become the standard of care simply because of the cumulative reduction in efficacy [45,46].

The intention-to-treat (ITT) analysis population includes all randomized participants in the groups to which they were allocated, irrespective of treatment received, loss to follow-up, or any protocol violations. In contrast, the per-protocol (PP) analysis population includes participants who achieve an adequate measure of compliance with the treatment and with the trial protocol [47]. The analysis population to be used in noninferiority trials has been the subject of recent debate since neither the ITT nor PP populations are free from bias, and reliance on either can increase the chance of falsely declaring noninferiority. In contrast to superiority trials, in which ITT is preferred because it provides “a secure foundation for statistical tests” [47], an ITT analysis can be biased towards noninferiority because of poor trial conduct diluting the treatment effect, whereas a PP analysis can also be biased in either direction when
postrandomization exclusions from the analysis may be directly or indirectly related to treatment allocation.

While a PP population has been recommended for noninferiority trials in the past [45,47], its importance has been re-evaluated. FDA guidance no longer recommends PP (or as-treated) analysis [41], even though the 2010 draft guidance accommodated one. There are limitations in PP analyses, and proposed improvements include correcting for noncompliance and dependent censoring using inverse probability weighting [48]. Current guidance suggests, instead, to focus on ensuring trial quality to reduce the bias in the ITT analysis; consideration is also given to multiple imputation as a way to counter bias due to attrition [41]. The 2017 addendum ("estimands and sensitivity analysis in clinical trials" [49]) to the 1998 International Council for Harmonization of Technical Requirements for Pharmaceuticals for Human Use (ICH) E9 document ("statistical principles for clinical trials" [47]) goes beyond just specification of an analysis population by recommending the use of an estimand as a framework for aligning the target and method of estimation of a treatment effect with the objectives of the clinical trial. The estimand “defines in detail what needs to be estimated to address a specific scientific question of interest” [49] and includes four attributes: the population, the endpoint, the specification of how to account for non-endpoint intercurrent events, and the population-level summary. Practically, defining the primary estimand(s) of interest in the protocol before the trial starts promotes clarity and coherence in how a treatment effect is estimated and how it links back to the trial objective.

A further recent innovation has been the application of Bayesian methodology to the interpretation of noninferiority trials. Presentation of the results as a simple binary statement as to whether or not noninferiority has been achieved is of limited value because it gives no indication as to how close in efficacy the intervention is likely to be to the control and places undue emphasis on the often arbitrary noninferiority margin. A much more informative approach is to use a Bayesian analysis to provide the probability that the difference is less than some given percentage, say 5% [50].

### Role of adaptive trial designs

An adaptive clinical trial permits changes to various trial design features after trial initiation in response to accruing data [51]. Although potential changes must be prespecified in the protocol so as not to undermine trial validity and integrity, adaptive trial designs are nevertheless useful to account for uncertainty when a trial starts or for anticipating potential landscape changes that may occur during the course of the trial. Most common are (i) the inclusion of interim analyses that permit early stopping for overwhelming efficacy or lack of benefit when evidence is sufficiently compelling with a smaller sample size than anticipated, and (ii) sample size re-estimation during recruitment using a preplanned algorithm to ensure that the final size will be adequate to answer the research question (particularly relevant when there is uncertainty in the efficacy of the control arm).

When there are many potential combination regimens that might be considered for evaluation, one might consider designs that select among multiple regimens either by stopping recruitment to poorly performing arms after fixed-interval interim analyses (an example being the multiarm multistage [MAMS] design [52,53]) or by adjusting randomization probabilities to enroll more patients in more promising arms (Bayesian adaptive randomization [35]). When the toxicity of a regimen is unknown, one might consider designs in which the eligibility criteria are widened during the trial as more safety data accrue. This can, for instance, be performed by starting to recruit patients with extensively drug-resistant TB (XDR-TB) because few treatment options are available, then expanding to multidrug-resistant TB (MDR-TB) and
DS-TB if safety thresholds are reached. Adaptations can, however, introduce bias in the estimate of treatment effect or inflate the probability of a false positive result (type I error rate). For example, in a two-stage multiarm trial in which only the intervention with the highest efficacy in the first stage is taken forward to the second stage, the uncorrected estimate of efficacy for this intervention at the end of the trial will be markedly biased and higher than the true efficacy [54]. Thorough understanding of these aspects is critical for regulatory approval and consideration in establishing WHO guidelines.

**Strategy trials incorporating treatment stratification**

Treatment stratification, the process of splitting a patient population into a small number of groups who receive different treatments for the same disease based on a predictive biomarker, is being widely studied in other disease areas [55–57]. In TB, it has long been recognized that disease prognosis is affected by certain baseline factors such as pretreatment extent of cavitation and viable counts of TB bacteria [58]. Wallace Fox sowed the seeds of stratified medicine in 1981 [59] by noting that good prognostic factors could be used to tailor treatment duration. The first trial incorporating treatment stratification had an enrichment design (one in which eligibility criteria are restricted to or enriched for a particular subgroup of participants) that evaluated a 4-month regimen with no new drugs in patients with noncavitary disease and culture negativity at 2 months. Before recruitment finished, the trial was stopped by the safety monitoring committee because of an apparent increased risk for relapse in the 4-month arm [7]. However, the completion of several large multicenter randomized trials in DS-TB showed that a 4-month fluoroquinolone-based regimen may well be adequate for patients with noncavitary disease [60]. Subsequent analyses describing an algorithm to more precisely identify subgroups of patients with lower or higher risk of failure and relapse [61] have provided important evidence to support the evaluation of treatment stratification in TB trials. These data are only from rifampicin-containing regimens for DS-TB to date. Nevertheless, the principles are likely also relevant for DR-TB, for which reducing duration for patients who do not need it is even more important, given the high levels of toxicity of drugs and the longer duration of treatment [62].

Several trials are under development to evaluate new treatment strategies to assess different durations, drug combinations, or drug dosages according to patient risk factors [63]. These predictive biomarker validation trials are designed to “confirm” a stratification algorithm in a randomized comparison against the standard-of-care strategy of a fixed duration regimen for all patients [64]. They are distinct from more exploratory trials designed to “learn” or develop and optimize the stratification algorithm [55]. Such trials tend to be smaller or have highly adaptive designs and are also important to incorporate newer biomarkers into the stratification algorithms, often to be evaluated in a subsequent larger confirmative trial. With appropriate stratification, it is expected that it may be possible to target treatment strategies with superior efficacy to standard of care, thereby avoiding many of the pitfalls of noninferiority trials.

**The need for more pragmatic trials**

In general, trials can be classified as explanatory (with the objective of evaluating the benefit an intervention produces under ideal conditions, i.e., efficacy) or pragmatic (with the objective of evaluating the benefit the treatment produces in routine clinical practice, i.e., effectiveness) [65,66], although this is more of a continuum than a dichotomy [67]. Trials that are more explanatory are needed to understand the efficacy and safety of a new drug under conditions as ideal as possible. However, the context in which an explanatory trial is conducted can be so far removed from routine practice that the results cannot readily be assumed to be transferable
to clinical care. This is particularly the case when there are considerable changes in the landscape, as has been seen in DR-TB. The current acceptance of bedaquiline as a safe and efficacious drug in the treatment of MDR-TB is due less to the pivotal phase II background regimen study [40], which initially led to WHO guidelines recommending bedaquiline only under certain conditions [68], than to the extensive nonrandomized data gathered outside of a trial setting [15], mostly under program conditions. These programmatic data were influential in bedaquiline becoming one of the three priority medicines in the revised 2018 WHO guidelines for DR-TB [26], albeit based on low-quality evidence [69]. This reflects the way in which, in the absence of pragmatic trials, WHO guidelines have been based almost exclusively on observational data that yield conditional recommendations based on low-quality evidence.

Pragmatic randomized trials with broader eligibility criteria, greater geographical spread, use of programmatically relevant primary endpoints, use of best available standard of care as control (see Table 2), and delivery and adherence strategies that are closer to “real-life” conditions greatly increase generalizability of the results and lead to faster and more evidence-based changes to policy and practice. Such pragmatic trials have been necessary in evaluating effective treatment strategies for HIV using previously licensed drugs (the START trial [70], for example), and they will also be needed in TB. Pragmatic trials can also be embedded within implementation programs to evaluate population-level effects of an intervention, an example being the XTEND study, which was designed to evaluate the effect of the GeneXpert MTB/RIF during implementation in South Africa [71]. Clearly, no single trial can address all relevant questions about a given therapeutic intervention at one time, and pragmatic trials will be invaluable for public health and policy decision-making purposes.

Conclusions

Just over 10 years ago, calls to action were published for innovations in drug development, capacity building for TB trials, and execution of clinical trials of treatment for DR-TB [72–74]. Since November, 2007, 538 TB clinical trials have been posted on clinicaltrials.gov; 27 (5%) of these have been for DR-TB. Between 1997 and 2007, these numbers were 127 and 4 (3%), respectively. Although the objectives and quality of these trials vary hugely, these raw numbers suggest that some progress has been made in clinical trial conduct.

The present review comes at a time when new drugs, new diagnostics, and new methods make possible real transformation in TB treatment. Today, there are 8 and 6 new compounds known to be in phase I and phase II clinical development, respectively (https://www.newtbdrugs.org/pipeline/clinical), with many more in preclinical development. The advances in clinical trial methodology that have been mentioned above alongside the promise of a variety of host-directed therapies [75] contrast starkly with the relative stagnation in treatment of DS- and DR-TB since the 1990s. The delivery of new regimens to patients demands nimbleness in an endeavor that is long and cumbersome. Trials must be designed and implemented in such a way that their relevance persists through completion. Careful choices of trial design, comparator, sample size, biomarker stratification, estimands, analysis population(s), and non-inferiority margin are critical from the outset. Changes in some of these characteristics after trial initiation—through predefined adaptation and protocol amendments—must also be entertained. Continued weighing of implications for time, cost, interpretation, and impact on practice is essential; whether the trial is primarily explanatory or pragmatic is a decision based on the balance among these competing priorities for any given trial. Transparency around assumptions and factors influencing decision making is critical to interpretation by guidance developers, practitioners, and patients. Consultation with external experts, including community advisory boards, can facilitate this transparency.
In conclusion, we strongly believe that TB treatment trials today should favor innovative approaches that are able to produce high-quality evidence for high-quality, patient-centered care that can be made accessible to all 10 million new TB patients, including the half-million with DR-TB, each year.

References


COLLECTION REVIEW

Designing noninferiority tuberculosis treatment trials: Identifying practical advantages for drug regimens with acceptable effectiveness

Piero L. OlliaroID1*, Michel VaillantIIID2*

1 Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom, 2 Competence Center in Methodology and Statistics, Department of Population Health, Luxembourg Institute of Health, Strassen, Luxembourg

☯ These authors contributed equally to this work.
* piero.olliaro@ndm.ox.ac.uk

Summary points

- The noninferiority design is being adopted in tuberculosis treatment trials to identify regimens that may have practical advantages over current standard therapy (e.g., being shorter, easier to adhere) and thus are more efficient in real-life settings, even while accepting that they might be less effective to a certain degree.

- This margin of acceptance is called the noninferiority margin, or delta. How narrow or wide the margin should be, and how this translates into acceptable losses and desired gains, is a matter of debate.

- Noninferiority trials are trials of ‘trade-offs’, in which one has to decide what one can lose in terms of pure efficacy against what one expects to make up in terms of effectiveness, tolerability, deployability, affordability, or else when replacing an existing intervention with a new one.

- This paper is about the principles behind identifying a ’meaningful noninferiority margin’—that is, a margin that is meaningful from a statistical, ethical, clinical, and health standpoint.

- Pragmatic approaches to expressing treatment effects using the number needed to treat (NNT), the reciprocal of the absolute risk reduction, with NNT for one patient to benefit (NNTB) and NNT for one patient to be harmed (NNTH) are useful to understand the implications of outcome definition and find a way to quantify gains and losses.

- Applying the noninferiority design to pragmatic (effectiveness) trials in addition to efficacy/safety trials would help quantify the trade-offs in real life.
Introduction
Identifying effective regimens for tuberculosis (TB) is challenging; trials are long between treatment and follow-up and require large sample sizes, so they take a long time to complete and are expensive. Often times, they are also inconclusive. Lienhardt and Nahid [1] and Phillips and colleagues [2] call for innovation in trial design that would allow for identifying effective regimens more quickly and efficiently.

Hardly present in the medical literature before the year 2000, the noninferiority design has gained in popularity across disciplines and medical interventions in the past 2 decades. A recent paper [3] and the ensuing debate it generated [4–7] illustrate some of the controversies regarding this approach. An extension of the Consolidated Standards of Reporting Trials (CONSORT) statement covers the reporting of noninferiority trials [8], and there is regulatory guidance on the design of noninferiority trials [9,10].

The noninferiority design is generally chosen when it is felt that a new medicine or intervention conveys benefits over the existing approved standard of care (such as better tolerability, real-life effectiveness, accessibility, or affordability), which would be enough to justify a ‘trade-off’ [4] between these advantages and an ‘acceptable’ loss of efficacy. A change in practice would be warranted if the new intervention is as effective or better (superiority would be preferred) but not if it is worse than the standard of care by a predefined noninferiority margin (also known as delta) [8]. The challenge with this design is 2-fold: (1) to identify an appropriate noninferiority margin so as to avoid retaining a harmful treatment because it has wrongly been judged noninferior [11], but also inappropriately discarding a treatment that brings a true benefit for the patient [12], and (2) to quantify how gains may offset losses.

Central to the design of these trials is therefore establishing a noninferiority margin, which should ‘preserve a minimum clinically acceptable proportion of the effect of the active treatment compared with placebo. This margin cannot be greater than the smallest effect size for the active treatment that would be expected in a placebo-controlled trial’ [13]. However, the delta should be ‘meaningful’ not just in statistical terms but also for patients and health systems on clinical, ethical [14,15], and public health grounds.

With the noninferiority design, the null hypothesis is that treatments are different, a type I error is to wrongly accept an inferior intervention, and a type II error is to reject a noninferior intervention [8]. The statistical procedure to test noninferiority is typically a one-sided test with a 97.5% level of significance or, preferably, a two-sided test with a 95% level of significance [3,8]. When the treatment outcome is binary (e.g., success or failure), regimens are compared by calculating either a relative risk (RR), an odds ratio (OR), or an absolute risk reduction (ARR, also known as risk difference) and then calculating the (crude or adjusted) difference in failure (or success) rates between test and control treatment and the confidence interval (CI) around it. In order for a new treatment to be deemed noninferior to the comparator standard treatment, the lower bound of the CI (in the case of the risk difference between failure rates between control and test treatment) must be within the noninferiority margin (see Fig 1).

Although opinions have shifted over the years, it is now generally agreed that conclusions should take into account the result of the analyses of both the modified intent-to-treat (mITT) and the per-protocol (PP) population and that conclusions are more robust when the results of the analyses of both sets are consistent [8,16]. The sample size of a noninferiority trial will depend on how narrow or wide the noninferiority margin is, the level of confidence, and the power chosen.

In this paper, we consider the implications of the noninferiority design for TB treatment trials, identify specific issues, and propose practical options. In particular, we focus on the choice of the noninferiority margin and clinically relevant end points; how these can be taken into account to weigh losses versus gains; and how to link statistical, clinical, ethics, patients' and...
public health imperatives in a way that these studies can be designed and interpreted with a view to informing policy decisions and ultimately improving health outcomes.

**Noninferiority design in TB treatment trials**

The noninferiority design has been adopted in explanatory treatment trials of active TB for newly diagnosed (expectedly drug-sensitive) TB (DSTB) (five trials completed and reported
and one systematic review [22]) and three for drug-resistant TB [DRTB] [23–25]. In these trials, the noninferiority margin ranged from 4% to 12% and is wider for multidrug-resistant TB (MDRTB) than DSTB (see Table 1).

The standard treatment for newly diagnosed DSTB is a 6 month regimen made of a 2 month, four-drug intensive phase with daily isoniazid (H), rifampicin (R), pyrazinamide (Z), and ethambutol (E), followed by a 4 month, two-drug phase with H and R (4HRZE/2HR) [26]. This regimen is generally very effective if adhered to but is usually less so in routine practice, in which compliance is lower than in trial conditions; its performance varies even across clinical studies, also depending on trial methodology [27], including the type of culture used (solid versus liquid media), the population analysed (PP versus mITT) and the efficacy end points adopted—the latter being particularly relevant here, and it will be further discussed in this paper.

The current standard WHO-recommended ‘conventional’ regimen for MDRTB requires 18–20 months [28] with an (up to) 8 month intensive phase with four or more second-line drugs followed by a 12 month (or more) continuation phase with three or more second-line drugs. A shorter regimen of 9–12 months may be used in patients with R-resistant TB or MDRTB who were not previously treated with second-line drugs and in whom resistance to fluoroquinolones and second-line injectable agents was excluded or is considered highly unlikely [29]. Patient retention with such long, cumbersome, and potentially toxic regimens is a major challenge [30].

Of the trials listed in Table 1, so far, noninferiority has been demonstrated in DSTB in the following cases: a 6 month fluoroquinolone-substitution regimen delivered intermittently in the continuation phase including rifapentin versus standard 6 month daily regimen [18] (non-inferiority margin 6%); fixed-dose versus loose (separately formulated drugs) combination [23, 24] (noninferiority margin 4%); and a 4 month fluoroquinolone-substitution regimen (with either gatifloxacin or moxifloxacin) versus a standard 6 month regimen in a meta-

### Table 1. Overview of TB noninferiority treatment trials.

<table>
<thead>
<tr>
<th>Indication</th>
<th>Study [Reference]</th>
<th>Regimen</th>
<th>Comparator</th>
<th>Delta</th>
<th>Outcome</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>DSTB</td>
<td>Jindani, 2004 [19]</td>
<td>8HRZE</td>
<td>6HRZE</td>
<td>5%</td>
<td>inferior</td>
<td></td>
</tr>
<tr>
<td>DSTB</td>
<td>Jindani, 2004 [19]</td>
<td>8HRZE (weekly*)</td>
<td>6HRZE</td>
<td>5%</td>
<td>inferior</td>
<td></td>
</tr>
<tr>
<td>DSTB</td>
<td>Gillespie, 2014 [17]</td>
<td>4HRZM</td>
<td>6HRZE</td>
<td>6%</td>
<td>inferior</td>
<td></td>
</tr>
<tr>
<td>DSTB</td>
<td>Gillespie, 2014</td>
<td>4RZEM</td>
<td>6HRZE</td>
<td>6%</td>
<td>inferior</td>
<td></td>
</tr>
<tr>
<td>DSTB</td>
<td>Jindani, 2014 [18]</td>
<td>4HRZM</td>
<td>6HRZE</td>
<td>6%</td>
<td>inferior</td>
<td></td>
</tr>
<tr>
<td>DSTB</td>
<td>Jindani, 2014</td>
<td>6HRZM</td>
<td>6HRZE</td>
<td>6%</td>
<td>noninferior</td>
<td></td>
</tr>
<tr>
<td>DSTB</td>
<td>Merle, 2014 [21]</td>
<td>4HRZG</td>
<td>6HRZE</td>
<td>6%</td>
<td>inferior</td>
<td></td>
</tr>
<tr>
<td>DSTB_noncavitary</td>
<td>Johnson, 2009 [20]</td>
<td>4HRZE</td>
<td>6HRZE</td>
<td>5%</td>
<td>inferior</td>
<td></td>
</tr>
<tr>
<td>DSTB_noncavitary</td>
<td>Alipanah, 2016 [22]</td>
<td>4HRZM/E</td>
<td>6HRZE</td>
<td>6%</td>
<td>noninferior</td>
<td>meta-analysis</td>
</tr>
<tr>
<td>DSTB</td>
<td>Lienhardt, 2011 [24]</td>
<td>6HRZE_fixed</td>
<td>6HRZE_loose</td>
<td>4%</td>
<td>noninferior</td>
<td></td>
</tr>
<tr>
<td>DSTB</td>
<td>Aseffa, 2016 [23]</td>
<td>6HRZE fixed</td>
<td>6HRZE_loose</td>
<td>4%</td>
<td>noninferior</td>
<td></td>
</tr>
<tr>
<td>DSTB</td>
<td>TBTC Study 31</td>
<td>4RHE; 4RptZHE</td>
<td>6HRZE</td>
<td>6.6%</td>
<td>enrolling</td>
<td></td>
</tr>
<tr>
<td>DSTB + DRTB</td>
<td>STAND</td>
<td>PaMZ</td>
<td>6HRZE</td>
<td>12%</td>
<td>active, nonrecruiting</td>
<td>NCT02342886</td>
</tr>
<tr>
<td>MDRTB</td>
<td>STREAM [25]</td>
<td>40–48 weeks</td>
<td>18–24 months</td>
<td>10%</td>
<td>noninferior</td>
<td></td>
</tr>
<tr>
<td>MDRTB</td>
<td>endTB</td>
<td>5 arms</td>
<td>SOC</td>
<td>12%</td>
<td>enrolling</td>
<td>NCT02754765</td>
</tr>
<tr>
<td>MDRTB</td>
<td>PRACTECAL</td>
<td>2 arms</td>
<td>SOC</td>
<td>12%</td>
<td>enrolling</td>
<td>NCT02589782</td>
</tr>
</tbody>
</table>

* In the continuation phase.

Abbreviations: DRTB, drug-resistant TB; DSTB, drug-sensitive TB; E, ethambutol; G, gatifloxacin; H, isoniazid; M, moxifloxacin; MDRTB, multidrug-resistant TB; Pa, pretomanid; R, rifampicin; Rpt, rifapentine; SOC, standard of care; STAND, Shortening Treatments by Advancing Novel Drugs; TB, tuberculosis; TBTC, Tuberculosis Trials Consortium; Z, pyrazinamide.

https://doi.org/10.1371/journal.pmed.1002850.t001
analysis of noncavitary TB (noninferiority margin 6%) [22]. Of the studies in DRTB, a 9–11 month regimen proved noninferior to the standard 20 month regimen [25].

Nunn and colleagues [31] illustrate a procedure to adopt when estimating the margin of noninferiority and the issues related to using the PP or mITT populations and dealing with missing data. Critical in this calculation is the choice of the trial end point, which in Nunn and colleagues is the relapse rates, assuming an insignificant number of primary on-treatment failures. Nunn and colleagues expect the relapse with current standard regimen in trial conditions to be 5%, which is broadly consistent with the findings of a trial by Jindani and colleagues [19] and a Cochrane systematic review [32] in which the relapse rate of reference regimens given for 4.5–12 months was 3.2 (95% CI 2.5%–4%). Using early TB trial data, Nunn and colleagues concluded that, when shortening the treatment from 6 to 4 months (a one-third reduction in duration), the expected difference in relapse rate would be 9%–10% with the current standard regimen for DSTB.

Noninferiority trials of DSTB so far have used noninferiority margins ranging from 4% to 6.6% (mostly 6%) (Table 1) and, instead of relapse rates, a composite end point (‘unfavourable outcome’, including primary failure during treatment, relapse during follow-up, and death) [17–19, 21]. In these trials, the overall rate of unfavourable outcomes at an 18 month follow-up with the standard regimen ranged from about 13% [21] to 20% [20].

A 4-percentage-point shift from a 6% to a 10% margin (see Fig 2) would mean that one of the fluoroquinolone-substitution regimens would have been deemed noninferior had the larger delta been adopted [21].

‘Meaningful’ noninferiority margin

This example further illustrates how critical it is to establish a noninferiority margin that is ‘meaningful’ statistically, clinically, and programmatically and is ethically acceptable. But how can the results of a trial be made to speak to clinicians and policy makers?

Fig 2. Unfavourable treatment outcome with fluoroquinolone-substitution regimens versus standard regimen for DSTB expressed as ARR with 95% CIs interpreted against a 6% (predetermined) and a 10% (post hoc) noninferiority margin. 18m, 18 month follow-up from treatment start; 24m, 24 month follow-up after treatment end; ARR, absolute risk reduction; CI, confidence interval; DSTB, drug-sensitive tuberculosis; mITT, modified intent-to-treat set; PP, per-protocol set.

https://doi.org/10.1371/journal.pmed.1002850.g002
We propose to translate the ARR into number needed to treat (NNT). Although objections have been raised to the use of NNTs [33, 34], and despite some statistical limitations, the NNT conveys a message that is easier for clinicians and policy makers to understand when it comes to quantifying the trade-offs between two interventions [35, 36]. We also support the use of the terms NNT for one patient to benefit (NNTB) and NNT for one patient to be harmed (NNTH) with the test regimen when compared with the control regimen, as proposed by Altman [37].

The NNT is easy to calculate: it is the reciprocal of the ARR (NNT = 1 / ARR); similarly, the CI is calculated by inverting and exchanging the upper limit (UL) and lower limit (LL) of the CI for ARR [1 / UL (ARR), 1 / LL (ARR)]. However, complications arise when there is no difference between treatments because, when the ARR is zero, the NNT is infinite, and the CI of the NNT will comprise infinity, thus violating the continuity between the CI limits. The classical Wald’s CIs suffers from a series of limitations (see, for instance, Newcombe [38]), and alternatives have been proposed, such as Cook and Sacket’s [39]—which we use for our calculations in this paper—Schultzer and Mancini’s [40], and Wilson scores [41]. We use here the NNT scale proposed by Altman [37] (Fig 3, Table 2).

Example 1: fluoroquinolone-substitution trials in DSTB. In the OFLOTUB trial [21], if we take the 18 month follow-up unfavourable outcome end point with the 4 month gatifloxacin-containing regimen versus standard treatment, the ARR (95% CI) is \(-6.4\% (−10.2\% to −2.4\%)\) on the PP population analysis and \(-7.4\% (−10.7\% to −4.2\%)\) on the mITT population—thus, this regimen is not noninferior and is inferior, respectively, to the standard regimen, as all confidence limits sit on one side of the no-difference line. This translates into an NNTH ranging from 41 to 9 between the two analysis populations, which means that a one-third reduction in treatment duration will cause one more patient to fail (compared with the standard 6 month regimen) between every 41 (best case) and 9 (worse case) patients treated.

By contrast, with a 6 month moxifloxacin regimen with rifapentin given intermittently in the continuation phase (the RIFAQUIN trial [18]), the confidence limits stretch across the no-difference line, and the NNT includes infinity (e.g., NNTB 25 to infinity to NNTH 42 in the mITT analysis). This means that with this regimen, an NNTB better than 25 is unlikely (i.e., that in order to obtain one more success over standard treatment, one would need to treat at least 25 participants). At the same time, an NNTH worse than 42 is also unlikely (i.e., that for one more patient to be harmed, at least 42 will have to be treated).

Example 2: STREAM trial in MDRTB [23]. This trial compared a shorter (9–11 month) regimen to the ‘classical’ 20 month regimen with a 10% noninferiority margin. The failure rates in the test and control arms in the mITT population (n = 253 and 130, respectively) were 21.2% versus 20.2%, with an unadjusted ARR for failure between control and test treatment of \(-1\% (95\% CI 7.5\% to −9.5\%)\). This translates into NNT \(-100\) (NNTB 13 to infinity to NNTH 11), which means that an NNTB better than 13 and an NNTH worse than 11 are unlikely. Similar conclusions are derived from the PP set: ARR (95% CI) \(0.7\% (10.5\% to −9.1\%)\) for NNT (95% CI) \(143\) (NNTB 7 to infinity to NNTH 11).

**Composite versus individual study end points**

Using a composite end point is practical (as it summarises findings into a single message), but we must be aware of two potential issues.

One is that, as mentioned earlier, changing from ‘relapse’ to ‘unfavourable outcome’ (generally including primary failure, relapse, and death) inflates the failure rate and has effects on the power and sample size calculation of the study. For instance, a change from 5% to approximately 10%–15% failure rate (depending on the population analysed) means that the required sample size...
could double or triple; for a noninferiority margin set at 6%, the sample size would increase by 1.8%–2.6%, 1.9%–2.8%, 2.1%–3.1%, and 2.3%–3.5% for risk differences from 1% to 4%, respectively. An example of implications for sample size calculation is presented in Fig 4.

Table 2. NNTs with Cook and Sacket’s 95% CIs based on 18 month unfavourable outcome with fluoroquinolone-substituted regimens versus standard TB regimen in the PP and mITT populations.

<table>
<thead>
<tr>
<th>Clinical trial</th>
<th>PP</th>
<th>mITT</th>
</tr>
</thead>
<tbody>
<tr>
<td>OFLOTUB</td>
<td>NNTH = 16 (41–10)</td>
<td>NNTH = 13 (24–9)</td>
</tr>
<tr>
<td>REMOX (H)</td>
<td>NNTH = 15 (NNTH 5 to 39)</td>
<td>NNTH = 17 (NNTH 6 to 47)</td>
</tr>
<tr>
<td>REMOX (E)</td>
<td>NNTH = 9 (14–6)</td>
<td>NNTH = 11 (16–8)</td>
</tr>
<tr>
<td>RIFAQUIN, 4M</td>
<td>NNTH = 8 (15–5)</td>
<td>NNTH = 9 (19–6)</td>
</tr>
<tr>
<td>RIFAQUIN, 6M</td>
<td>NNNTB = 59 (NNNTB 17 to 40)</td>
<td>NNNTB = 120 (NNNTB 25 to 42)</td>
</tr>
</tbody>
</table>

Abbreviations: CI, confidence interval; E, ethambutol; H, isoniazid; mITT, modified intent-to-treat; NI, noninferior; NNTB, number needed to treat for one patient to benefit; NNTH, number needed to treat for one patient to be harmed; PP, per-protocol; TB, tuberculosis.

https://doi.org/10.1371/journal.pmed.1002850.t002
The other complication with a composite end point is that there might be discordant results within its individual components. An illustrative example can be found in the OFLOTUB trial (Fig 5) [21].

The primary efficacy end point was a composite unfavourable outcome including on-treatment events (failure, death, other adverse event, drop-out, and withdrawal) plus posttreatment recurrence by month 24 on the mITT population. This gives an unadjusted ARR (95% CI) of −3.8% (0.4% to −8%). However, when one looks at these two components separately, the short regimen is significantly better than the standard treatment for on-treatment outcomes (ARR [95% CI] 3.6% [0.7%–6.6%]) but significantly worse for relapse (−7.5% [−4.2% to −10.7%]); at the same time, it has also significantly fewer posttreatment losses to follow-up (5.9% [2.4%–9.5%]).

It is therefore prudent to dissect composite end points in order to verify that the individual components do not have conflicting implications for regimen effectiveness [3]. Analysing the granularity of the results is important also for practical reasons. We must know what we have to watch out for when it comes to decide what type of losses we are prepared to accept in terms of efficacy. Type and timing of failure is of paramount importance. Patient retention is a challenge, especially after treatment is completed; long-term posttreatment follow-up is required in TB to make sure the patient does not relapse. For instance, in Jindani and colleagues [19], in the 6 month standard therapy group, there were three times as many patients lost because they did not report to a posttreatment follow-up visit as those not reporting while on treatment for DSTB (12% versus 4%). Primary (on-treatment) failures are easier to detect, especially in clinical trials and in routine practice when treatment is supervised; posttreatment relapses may be
more challenging, as patients are generally less compliant with follow-up visits, at least for as long as they are not unwell.

Trading gains for losses

Now the question is, Using the previously mentioned examples, how would stakeholders (national TB programme managers, caregivers, patients) weigh losses and gains?

In the case of a one-third reduction of treatment for DSTB given under directly observed treatments (DOTs), what would be the practical gains for the health system (e.g., more time for patient visits, reduced costs, increased efficiency) versus having to deal with one more failure every 10 rather than 40 cases? Would the advantages of a shortened treatment and faster resolution, along with the smaller reduction in wages for the patients, outweigh the disadvantages of excess relapses? Can a national TB programme gear up for actively and systematically following up with patients in order to identify and deal with relapses promptly?

Another example for DSTB: How would health providers and patients value a regimen that is given for the same total duration but weekly (instead of daily) in the 4 month continuation
phase [18] when this regimen is estimated to be producing a benefit every approximately 20
treatments or one more failure every approximately 40?

Similarly, for MDRTB, when a patient is now on treatment for 1.5–2 years, how would they
value a regimen that is half as long and might either produce a benefit every 7–13 patients
treated or one more failure every 11? The gains here may, however, be offset by the need for
drug sensitivity testing and by the toxicity of injectable aminoglycosides used in these shorter
regimens, at least until evidence is gained on replacing them with safer drugs [28], or by a
higher risk of selecting for drug-resistant bacteria.

Where would one draw the line? Information is required on a number of variables which,
together, can help quantify gains and losses and thus inform both study design and treatment
policy decisions. These cover a range of outcomes—not just efficacy and safety but also
patient’s preferences and satisfaction, quality of life, healthcare provider’s preferences and per-
formance, emergence of drug resistance, etc., which are rarely collected in clinical trials. Fig 6
(derived from the Cochrane systematic review of fixed-dose versus loose combination treat-
ment [43]) offers an example of some of the criteria which could be used to compare gains and
losses (limited information was collected on patient’s satisfaction, so this could not be plotted).

Together with key stakeholders, we need to identify the critical questions and score the
answers in order to inform both study design (and identify a ‘meaningful noninferiority
margin’ in case of a noninferiority trial) and policy and practice (how to take advantage of the benefits and how to handle the potential problems).

Traditionally, we tend to derive this information from explanatory (‘phase 3’) trials. However, they suffer from an inherent limitation in this regard, in that they typically seek to measure the efficacy (and safety) gains introduced by a new intervention by minimising confounders and standardising eligibility criteria—the very elements we are interested in to decide on the trade-offs. Instead, in order better to quantify gains and losses of a new treatment and its associated effects, it would be useful to apply the noninferiority design also to pragmatic (effectiveness) trials, analysed on the (m)ITT (as well as PP) population.

Conclusions

Though it has been wrongly used to justify ‘me-too’ medicines, the noninferiority design is having a growing place in diseases like TB, which require treatments that are long and cumbersome for patients and health systems alike, and where attributes like adherence, user-friendliness, and tolerability are critical to real-life effectiveness. Although the noninferiority design may be applied to treatment trials in both DSTB and DRTB against the current standard of care, this does not take away the responsibility for finding both more effective and easier-to-comply regimens, especially for DRTB.

This design responds to the need for a planned trade-off between what we think we can afford to lose in terms of efficacy against what we expect to gain in terms of safety, effectiveness, ease of use, costs, etc. It is generally applied when a net gain in efficacy cannot realistically be shown within the conditions of a typical trial, though a superiority test can be applied if noninferiority is demonstrated. However, more work is required to develop end points for TB treatment trials, which will identify regimens that better serve the needs of patients as well as country TB programmes and health providers. Also, using the noninferiority design in pragmatic trials would provide useful information.

The noninferiority margin is a central element in study design and interpretation. Identifying and weighing the appropriate parameters for gains and losses is crucial towards defining a ‘meaningful’ noninferiority margin.

Acknowledgments

PLO was also a staff member of the UNICEF/UNDP/World Bank/WHO Special Programme on Research and Training in Tropical Diseases (TDR), World Health Organization, Geneva, Switzerland, at the time of the writing of this paper. MV is a staff member of LIH.

The opinions expressed are those of the authors and do not necessarily represent the views and opinions of their employers.

References


The importance of adherence in tuberculosis treatment clinical trials and its relevance in explanatory and pragmatic trials

Andrew Vernon†1*, Katherine Fielding2,3, Rada Savic4, Lori Dodd5, Payam Nahid6

1 Clinical Research Branch, Division of TB Elimination, NCHHSTP, US Centers for Disease Control & Prevention, Atlanta, Georgia, United States of America, 2 TB Centre, London School of Hygiene & Tropical Medicine, London, United Kingdom, 3 Faculty of Health Sciences, University of Witwatersrand, Johannesburg, South Africa, 4 Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, California, United States of America, 5 National Institute for Allergy and Infectious Disease, National Institutes of Health, Washington DC, United States of America, 6 Division of Pulmonary and Critical Care Medicine, University of California, San Francisco at San Francisco General Hospital, San Francisco, California, United States of America

* anv3@cdc.gov

Summary Points

- Adherence to prescribed treatment remains a critical component of clinical trials in tuberculosis (TB) treatment. Recent evidence indicates that adherence strongly influences the outcome of therapy; attention to its quantification and measures to assure its implementation should increase.

- In the context of a World Health Organization (WHO) Technical Consultation on “Advances in Clinical Trial Design for Development of New TB Treatments,” we reviewed the challenges related to adherence confronting the trials community.

- We discuss the importance of adherence to therapy in TB clinical trials, consider several definitions and measures of adherence, comment on the standard provided by directly observed therapy (DOT), and briefly review evolving electronic methods for the assessment of adherence.

- Adherence affects both the outcome of therapy and the risk of acquired drug resistance. Assessment of adherence should consider not only overall adherence but also the timing and intensity of nonadherence.

- Appropriate methods for pooling and analyzing electronic data on adherence are needed.

- Better methods are needed for linking information on adherence to individual pharmacokinetics and pharmacodynamics and to individual patient outcomes.

Introduction

Medication adherence remains the most underrated and understudied factor affecting the outcome of tuberculosis (TB) therapy. Its importance has been appreciated since the time of the
Initial South India trial conducted by the Tuberculosis Research Center and the British Medical Research Council (MRC), comparing in-patient and domiciliary treatment [1]. Twenty-five years later, Fox wrote “It is paradoxical to insist on the importance of 100% success with primary chemotherapy and to use self-administered chemotherapy as a means of achieving it” [2]. In their 1999 encyclopedic review of the MRC TB trials, Fox, Mitchison, and Ellard reported that a common feature of those trials was “the effort made [including hospitalization for the full treatment] to ensure that the patients actually took the prescribed regimen throughout the trial period” [3]. These examples illustrate the importance of adherence to treatment for the validity of a clinical trial and for the success of individual and programmatic care. Despite the clear and obvious need to ensure optimal treatment adherence, “full supervision,” in the form of directly observed therapy (DOT) as currently delivered, has not consistently been associated with improved outcomes. Thus, significant challenges persist in measuring and maximizing adherence with antituberculosis therapy; recent data and analyses provide evidence that the absence of full adherence in TB trials has important implications for TB regimen development and for the durability of new regimens. In March 2018, the World Health Organization (WHO) held a Technical Consultation on Advances in Clinical Trial Design for Development of New TB Regimens, which is the topic of the Collection of which this paper is part [4]. In this context, we reviewed the importance of treatment adherence, the implications of a drug or regimen’s “forgiveness for missed doses,” and emerging novel approaches to measuring and maximizing adherence in clinical trials and in patient care.

**Importance of adherence**

Adherence affects patient outcomes and is thus an important factor to consider when evaluating regimens in clinical trials. Differing adherence across treatment arms could potentially lead to misleading conclusions about treatment arm performance. For example, consider (as a hypothetical example) a study with poor adherence in the control arm but perfect adherence in the experimental arm. If the goal of a study is to measure the efficacy of a new regimen, the relatively poor adherence in the control arm will give an overly optimistic estimate of the improvement in outcomes with the experimental treatment. However, if the goal of the study is to evaluate effectiveness (i.e., performance under real-word conditions), the relative difference in adherence may accurately reflect the real-word difference in the 2 regimens. One complication is that the level of adherence may vary widely across different populations and cultural or economic settings, raising concerns about whether estimates of effectiveness are broadly generalizable. The relation of adherence to regimen effectiveness (the usual target outcome of “pragmatic” trials) in trial versus program settings was noted nearly 50 years ago and continues to challenge the generalizability of trial findings [5].

Adherence may have a substantial impact on the interpretation of clinical trial findings. Adherence is often an active choice by each patient on how to comply with the assigned therapy. Adjustments in analysis based on observed adherence may alter the balance introduced by randomization. Restricting analyses only to those with high adherence focuses on a subset of the population that may have fundamentally different risk than those who are not adherent. A classic example of this circumstance is provided by the Coronary Drug Project trial assessing a lipid-lowering drug in men with recent myocardial infarction: participants with good adherence had low and equivalent mortality in the test and placebo arms, whereas poor adherers did better in the test arm [6]. Still, understanding trial outcomes among participants who take drugs as prescribed (i.e., a “per protocol [PP] analysis”) has some appeal, even though such an analysis is not protected by randomization. In recent TB treatment trials, adequate adherence was defined by a threshold of 76%–80% of intended doses taken, to identify the PP population.
This is consistent with analytic practice in the reporting of the MRC trials (which defined an “excessive interruption” with exclusion from the relapse analyses if less than approximately 77% of intended doses were received) [9] and with recent practice in prominent United States TB control programs (e.g., New York City) [10].

A recent meta-analysis of three Phase III trials of fluoroquinolone-based 4-month TB treatment regimens found that nonadherence was the single most potent factor associated with unfavorable treatment outcome. The adjusted hazard ratios (aHRs) were 5.7 (95% CI 3.3–9.9) for test arm participants who missed 10% or more of prescribed doses and 1.4 (95% CI 1.0–1.9) for test arm participants who had less than 10% nonadherence, compared with participants who completed treatment without any missed doses; the aHRs were similar in the control arm participants (Table 1) [11]. The same trend was seen in PP analysis, which excluded participants who failed to complete at least 75%–80% of intended doses.

Such a potent influence of nonadherence serves to emphasize the often-noted importance of the quality of performance in noninferiority trials; it further suggests that PP analyses might examine more than 1 threshold for nonadherence (e.g., 80% and 95%) to help in more robustly assessing efficacy. A stronger analytic approach might evaluate the effect on trial outcomes of baseline pre-randomization variables associated with poor adherence [12]; by definition, baseline variables should be approximately balanced in large randomized trials, thereby not introducing bias in the assessment of outcomes.

### Definitions and measures of adherence

Adherence refers to the completeness with which participants or patients follow medical instructions. Because adherence can vary so greatly among different individuals, it can have an important influence on treatment outcomes. Adherence more broadly may also involve changes required by the protocol (e.g., in response to elevated liver function tests) that are not active choices by the participant. In their recent review on this topic, Blaschke and colleagues observe that adherence is a major source of variability affecting the outcome of TB therapy [13]. Adherence, in turn, is affected by diverse individual and social factors [14]. Other sources of variability include the formulation of the test medications, the prescribed dosing, and the pharmacokinetics and pharmacodynamics of each agent employed, as well as key features of the infecting *Mycobacterium tuberculosis* strains (for example, the minimal inhibitory concentrations of each drug employed), and inherent characteristics of the host patient (including genetic determinants of drug metabolism, immunologic competence, and the architecture of TB lesions). The latter sources of variability are already determined at the onset of therapy and are therefore likely to be balanced between treatment arms by the process of randomization. In contrast, adherence is subject to ongoing variability during treatment, which complicates its effects. Although genetic factors affecting drug exposure should be comparable at randomization, their impact may vary by the drugs used in each arm. The recent availability of electronic methods for monitoring adherence has made it possible to measure adherence quite precisely;

---

### Table 1. Pooled mITT analysis of 3 TB treatment-shortening trials showing impact of adherence on unfavorable outcome.

<table>
<thead>
<tr>
<th>Prescribed doses</th>
<th>Test arms (4 months, with FQ)</th>
<th>Control arms (6 months, no FQ)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Unfavorable</td>
<td>Total</td>
</tr>
<tr>
<td>Received 100% of prescribed doses</td>
<td>238 (18%)</td>
<td>1,348</td>
</tr>
<tr>
<td>Received 90%–99% of prescribed doses</td>
<td>64 (22%)</td>
<td>288</td>
</tr>
<tr>
<td>Received &lt;90% of prescribed doses</td>
<td>15 (47%)</td>
<td>32</td>
</tr>
</tbody>
</table>

Abbreviations: FQ, fluoroquinolones; mITT, modified intent-to-treat

https://doi.org/10.1371/journal.pmed.1002884.t001
these novel methods have become the gold standard for compiling dosing histories. At least 3 aspects of adherence are specifically relevant to antituberculosis therapy [15]:

1. the total quantity of nonadherence (i.e., what proportion of doses are missed, in relation to the total number of doses in the intended treatment regimen?);

2. the timing of nonadherence (i.e., does it occur at the outset of therapy, throughout therapy, or primarily at the end of the intended course of therapy?); and

3. the intensity and patterns of nonadherence (i.e., are many consecutive doses missed, or are missed doses distributed relatively evenly throughout the course of therapy?).

The third aspect in particular can exert an important influence upon drug pharmacokinetics and thus may predispose to either loss of efficacy or emergence of drug resistance. Consecutive lapses in dosing can lead to lower-than-usual peak drug concentrations and lower total drug exposures, whereas extra doses can result in risk of toxicity due to higher-than-usual peak concentrations and total exposures (the review by Blaschke and colleagues includes a figure that nicely illustrates these risks [13]). The term “forgiveness” of a regimen is intended to reflect the impact of variable lapses in dosing. Although “forgiveness” was originally defined as “the post dose duration of therapeutically effective drug action, minus the recommended interval between doses” [13], the shift from action to no action is likely to be gradual and to vary among patients.

In the circumstance of treatment for TB, several examples come readily to mind:

1. The work of Imperial and colleagues demonstrated the association of overall nonadherence with the treatment outcome of short-duration fluoroquinolone-based regimens [11];

2. The timing of nonadherence is likely critical, because nonadherence in the presence of high bacillary loads typically seen in the intensive phase is likely to have greater impact than the same degree of nonadherence later during the continuation phase, when bacillary loads are generally several logs lower; this is particularly an issue in the presence of immunosuppression, because bacillary multiplication will resume more rapidly when such patients become nonadherent;

3. Similarly, a gap of several doses (i.e., intensity) would likely have greater impact in the presence of high bacillary loads, such as during the early intensive phase. Recent guidelines have advised against the use of highly intermittent regimens for this reason, with substantial supporting evidence [16].

There is thus an urgent need for improved measures and more sophisticated means of analyzing such patterns and types of nonadherence in relation to treatment outcomes. In Phase I and Phase II studies, optimal adherence is imperative to make decisions on regimens to move forward to late-stage development. In Phase III studies, the objectives would drive the decision on adherence implementation and measurement. In both scenarios, there is a need to measure and report adherence appropriately, to understand better the performance of the tested interventions.

**Methods for assessment of adherence**

Currently available methods for assessing adherence are limited, but considerable work is underway to develop better approaches (Table 2). Among methods that have been used in past investigations are (1) clinic-based DOT, in which ingestion of each medication dose is observed in clinic by a health worker, thereby allowing exact counting of each study dose given or missed; (2) home-based DOT by a health worker, a community worker, or a family
Some methods offer direct demonstration of adherence, whereas others provide only indirect readouts. Newer methods are currently under investigation, including the quantification of drug levels in hair [17] and the measurement of changes in skin color associated with specific medications (e.g., rifamycins or clofazimine); these both indicate cumulative adherence rather than dynamic patterns.

There were several early trials of interventions to improve adherence (with the goal of achieving better outcomes), but the knowledge base overall remains sparse, and recent systematic reviews underline the need for further investigation and a substantially enlarged evidence base [18]. Electronic methods for measuring or estimating adherence are increasingly available, and some TB programs and countries are moving forward rapidly with such digital tools.

<p>| Table 2. Strengths and weaknesses of methods for encouraging and/or assessing adherence. |
|---------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|</p>
<table>
<thead>
<tr>
<th><strong>Methods</strong></th>
<th><strong>Description</strong></th>
<th><strong>Strengths</strong></th>
<th><strong>Weaknesses</strong></th>
<th><strong>Notes</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Home-based or work-based DOT by HCW</td>
<td>Ingestion of each medication dose is observed and recorded by an HCW at home/work.</td>
<td>Monitors adherence in real time; convenience for patient.</td>
<td>Cost; patient confidentiality.</td>
<td></td>
</tr>
<tr>
<td>In clinic DOT by HCW</td>
<td>Ingestion of each medication dose is observed and recorded by an HCW at the clinic.</td>
<td>Monitors adherence in real time; lower cost than home or work-based DOT by HCW.</td>
<td>Inconvenience to patient; cost to health system.</td>
<td></td>
</tr>
<tr>
<td>Family member DOT</td>
<td>Ingestion of each medication dose is observed and recorded by a designated family member.</td>
<td>Convenience for patient; lower cost versus DOT by health worker.</td>
<td>Confidence in the reports from family members; data on real-time adherence not available unless transmitted to HCW on a daily basis.</td>
<td></td>
</tr>
<tr>
<td>Live video DOT [27,28]</td>
<td>Ingestion of each medication dose is videoed by patient and observed by an HCW in real time.</td>
<td>Monitors adherence in real time; convenience for patient.</td>
<td>Cost (HCW review of live video; smartphone); patient and HCW acceptability.</td>
<td></td>
</tr>
<tr>
<td>Recorded video DOT</td>
<td>Ingestion of each medication dose is videoed by patient and sent to HCW to be viewed later.</td>
<td>Convenience for patient.</td>
<td>Cost (HCW review; smartphone); patient and HCW acceptability; depending on when videos are viewed, may not monitor adherence in real time; privacy concerns.</td>
<td></td>
</tr>
<tr>
<td>Direct monitoring:</td>
<td>Blood testing</td>
<td>Direct measurement of dose ingested.</td>
<td>Feasibility/logistics; cost; depends on timing of blood sample relative to time of ingestion; limited time window.</td>
<td></td>
</tr>
<tr>
<td>1. Blood testing</td>
<td>Urine testing</td>
<td>Direct measurement of dose ingested.</td>
<td>Feasibility/logistics; cost; sensitivity may vary depending on acetylator status; limited time window.</td>
<td></td>
</tr>
<tr>
<td>2. Urine testing</td>
<td>Blood sample taken to measure plasma levels of TB medications.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Swallowed pill sensors [29]</td>
<td>Urine testing for drug metabolites (e.g., isoniazid).</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indirect monitoring, device facilitated</td>
<td>Medications placed in pill box/bottle. Opening/closing of box/bottle, a proxy for dose taken, is documented in real time via SIM card.</td>
<td>Monitors adherence in real time (if pill box/bottle cap transmits); low cost (relative to HCW DOT).</td>
<td>Pill box opening/bottle cap removal may not reflect an ingestion of dose; nonopening may not reflect noningestion of dose if medications are not stored in box/bottle.</td>
<td></td>
</tr>
<tr>
<td>Indirect monitoring, patient facilitated SMS text messages</td>
<td>Patient sends SMS message to HCW when a dose has been ingested.</td>
<td>Monitors adherence in real time; low cost.</td>
<td>Patient needs to be familiar with text messaging; text message sent may not reflect an ingestion of dose; nonreceipt of SMS may not reflect noningestion of dose.</td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: DOT, directly observed therapy; HCW, healthcare worker; SIM, subscriber identity module; SMS, short message service; TB, tuberculosis

https://doi.org/10.1371/journal.pmed.1002884.t002

a HCW observation could be replaced by face recognition and motion-detection software.
b Non–real time use of pillbox also possible where data on pill box opening/closing are downloaded at regular intervals, at a pharmacy refill, for example.
c Costs are important for all modalities; these often vary by setting or country and vary for newer technologies.

https://doi.org/10.1371/journal.pmed.1002884.1002
generating large quantities of data. Appropriate methods for pooling and analyzing these data are needed, as are methods for linking information on adherence to individual pharmacokinetics and pharmacodynamics and to individual outcomes [19]. A recently published systematic review of newer digital technologies, including short message service (SMS), video-observed therapy (VOT), and medication monitors (MMs), for TB treatment adherence identified few comparative studies for inclusion and concluded that the evidence on the effect of digital technologies to improve TB care remained limited [20]. For the studies included in the review, no statistically significant effect on treatment completion was identified when SMS was added to standard care or when VOT was used as an alternative to in-person DOT. It was noted that MMs increased the probability of cure (risk ratio 2.3, 95% CI 1.6–3.4) in one observational study [21] and in one trial, significantly reduced missed treatment doses relative to standard care (adjusted means ratio 0.58, 95% CI 0.42–0.79) [22].

Overall, the systematic review concluded that more studies of better quality are needed for the evaluation of technologies applicable to measuring and maximizing adherence. There are also few studies that have assessed the accuracy of digital adherence technologies in measuring ingestion of medication doses. A study in China assessed the MM box (box opening between 6 and 24 hours before urine sample taken) against detecting rifampicin in urine and found a sensitivity of 99% and specificity of 95% [23]. In India, SMS responses (from the cellphone-based monitoring system known as “99DOTS”) over a 48-hour period, indicating dose taken, were compared with isoniazid detection in urine, and a sensitivity and specificity of 68% and 62% were observed, respectively [24]. A recent randomized trial in the United Kingdom noted considerable success in use of VOT to assure dosing, compared with traditional DOT [25]. Adequate study designs for evaluating accuracy of adherence monitoring devices are critical to provide realistic tests of performance. Bias may be introduced, for example, if patient knowledge that a urine sample will be collected inflates adherence around the scheduled time. Unannounced collections may mitigate this. Timing of collections throughout the full treatment period may also be important, for example, if adherence drops later during treatment. The recent technical consultation report on “Advances in Clinical Trial Designs for Development of New TB Treatments” also strongly endorsed the need for further investigations in this domain and noted that trials offer an excellent platform for substudies in these areas [26].

Evidence for the benefit of traditional DOT has not been entirely consistent, and its role remains controversial [18,31,32]. Some investigators favor relatively strict application of in-person DOT, whereas others feel this is excessive and does not contribute to achievement of objectives in properly randomized and implemented trials. Some investigators favor implementation of non-family-member in-person DOT, whereas others feel it is more reasonable to allow local determination of what types of adherence support would be most useful. Better means to measure adherence and its association with outcomes would contribute usefully to this discussion [15,16]. Some of the digital health approaches being assessed in pragmatic trials may be combined with differentiated care; in this approach, for example, those identified as poor adherers through the digital health measures are assisted further with more traditional approaches to maximizing adherence, with actual observation by health workers in the most extreme cases [25,33].

Likewise, there is no consensus on a single criterion for “clinically important” nonadherence. Assessment of the degree of nonadherence that should be deemed “clinically important” depends on multiple factors specific to each trial setting, including the component drugs of the regimen, the dosing schedule, the pharmacokinetics of the individual drugs, and other risk factors and comorbidities that could influence the risk of treatment failure or relapse. Embedded in this discussion is consideration of the concept of “forgiveness” of a regimen (i.e., as noted previously, a reference to the types and levels of nonadherence that would not substantively
alter the likelihood of treatment effectiveness of a regimen). Although it could be considered
that this aspect should be reflected in the regimen’s efficacy and requires no other adjustment,
it can be argued that this aspect should be considered in the determination of the noninferior-
ity margin [34]. Further, some note that in the rational design and composition of new TB reg-
imens, the “forgiveness” of a regimen for missing doses should be considered with
significantly greater deliberation than is currently common, particularly given that adherence
in practice will never be perfect.

**Adherence and acquired drug resistance**

Recently released WHO target regimen profiles (TRPs) for TB identify the barrier to emer-
gence of resistance as an important characteristic to address in the development of new drugs
and regimens [35]. The association of nonadherence with acquisition of drug resistance has
been well reviewed [36,37,38], but the mechanisms underlying the association remain largely
speculative [39]. In the WHO TRPs, it is suggested, based on expert opinion, that each compo-
nent of the regimen should permit no greater mutation rate (in unselected bacterial popula-
tions) than 1/10^7 mutations/bacterium/generation and that new resistance to one or more
drugs in the regimen should emerge in fewer than 2% of treatment courses when taken as pre-
scribed and when there is no preexisting resistance to the drugs in the regimen. This minimal
target is based on an acquired resistance rate of 0%–2% when 5 effective drugs are used in the
WHO-recommended multiply drug resistant (MDR) regimen [40]. The reality of reduced
adherence in the field, as compared with clinical trial settings, and the potential impact such
real-world usage of a regimen may have on risks for emergence of resistance need further
study, representing another outcome of interest in how much “forgiveness” a putative new reg-
imen may carry for missed doses.

**Summary**

In conclusion, medication adherence remains a critical, yet understudied, factor influencing
outcomes of TB therapy. Its importance has been recognized since the advent of effective anti-
tuberculosis therapy, and the vital role adherence plays in the conduct of TB clinical trials has
been further highlighted in contemporary clinical trials [11]. The growing importance of non-
inferiority trial designs and the challenge of interpreting PP analyses have focused more atten-
tion on the issue of precisely measuring adherence and adherence patterns. Adherence is
important in both superiority and noninferiority trials and in both intent-to-treat (ITT) and
per-protocol (PP) analyses; both should be performed, and both should assess the impact of
variation in adherence. Only recently has our ability to measure adherence improved. Novel
(in particular, electronic) methods for assessing and encouraging adherence hold promise, and
efforts to develop a robust evidence base to support them are growing. Our understanding of
the impact of nonadherence on key outcomes (treatment success, emergence of resistance) in
TB treatment trials is relatively modest but is also receiving increased attention from investiga-
tors. Continued development of more convenient, more reliable, and less costly means to
achieve high levels of adherence will serve both trials and programs well.

**Acknowledgments**

The authors are grateful to James Neaton, Abdel Babiker, and Patrick Phillips for helpful
observations and suggestions; to Christian Lienhardt for acquisition of support for the work-
shop and for enthusiasm, helpful critique, and creative energy; and to Carla Winston and Amy
Sandul for careful review and constructive commentary on the draft manuscript.
The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the US Centers for Disease Control and Prevention or the authors’ affiliated institutions.

References


Inclusion of key populations in clinical trials of new antituberculosis treatments: Current barriers and recommendations for pregnant and lactating women, children, and HIV-infected persons

Amita Gupta1*, Michael D. Hughes2, Anthony J. Garcia-Prats3, Katherine McIntire1, Anneke C. Hesseling3

1 Department of Medicine, Division of Infectious Diseases, Johns Hopkins University School of Medicine, Baltimore, Maryland, United States of America, 2 Center for Biostatistics in AIDS Research, Harvard T. H. Chan School of Public Health, Boston, Massachusetts, United States of America, 3 Desmond Tutu TB Centre, Department of Paediatrics and Child Health, Faculty of Medicine and Health Sciences, Stellenbosch University, Tygerberg, South Africa

* agupta25@jhmi.edu

Summary points

- Pregnant women, children < 15 years old, and HIV-infected persons contribute approximately 20% of the global tuberculosis (TB) burden, with an estimated 216,000, 1,000,000, and 1,040,000 cases each year, respectively, yet these populations are currently largely excluded from TB clinical trials, leading to suboptimal treatment and poor access to new therapeutics.

- Special considerations in these populations include specific TB disease spectrum and severity, lower sensitivity of commonly used TB diagnostic tests, potential differential drug dosing and treatment responses, drug–drug interactions, and challenges in acquiring high-quality data through clinical trials.

- To counter the automatic exclusion of pregnant and lactating women that currently pervades the TB trial landscape, early discussions among trialists, pharmaceutical companies, maternal–child clinical experts, ethicists, and regulatory bodies are needed to address risks, benefits, and compelling rationale for inclusion. Reconsenting women when pregnancy occurs on a trial to allow continuation of study drug by informed choice is a practical and valuable approach to expand the currently limited evidence base.

- Children tend to have less severe, often paucibacillary TB disease and may respond better to treatment than adults. Consequently, trials of shorter, less intense TB treatment regimens in children are needed; pharmacokinetic and safety studies should be initiated earlier and involve age groups in parallel rather than in an age-de-escalation approach. More rapid development of child-friendly drug formulations is needed.

- All HIV-infected populations, including those with advanced disease, who are likely to be the intended population of the TB therapy, should be involved in Phase Ib and/or
Introduction

Globally, 10 million cases of active tuberculosis (TB) disease and 1.6 million TB-related deaths occurred in 2017 [1]. Pregnant and postpartum women, children < 15 years old, and HIV-infected persons account for 20% of the global TB burden, with an estimated 216,000, 1,000,000, and 1,040,000 cases each year, respectively [1,2]. Special considerations in these populations include TB disease spectrum and severity, lower diagnostic sensitivity, possible differential treatment responses, drug dosing and interactions, and challenges in acquiring high-quality data through clinical trials [3–5]. Without clear consideration of actual risks and benefits of trial participation, pregnant women have been uniformly excluded from TB therapeutic trials, especially for multidrug-resistant (MDR) TB [6,7], based on fears of harming the fetus and legal liability [8]. Children have better treatment outcomes than adults for most forms of TB, but they present different pharmacologic responses to drugs and typically require higher mg/kg doses, especially if very young [9–11]. HIV-infected persons experience complicated drug–drug interactions (DDIs) and worse TB treatment outcomes than HIV-uninfected persons and have 2–3 times greater likelihood of TB-related mortality [12]. In March 2018, the World Health Organization (WHO) held a technical consultation focused on advancing clinical trial design for more successful development of new TB treatments [13], including enrollment of key populations that may be currently underrepresented in clinical trials. Although many such populations exist, including migrants, prisoners, homeless people, and healthcare workers, the technical consultation discussions were concentrated on three populations and were framed around five questions (Box 1). This review is part of a Collection, “Advances in Clinical Trial Design for the Development of New TB Treatments: A Call for Innovation,” and highlights key aspects, barriers, and potential solutions to conducting TB therapeutic clinical trials in pregnant and lactating women, children, and HIV-infected persons [14].

Box 1. Five questions addressed during discussions about key populations in clinical trials of TB therapeutics [13]

1. Aside from the use of well-designed trials based on solid preclinical data conducted under the protections outlined in existing regulations, what are the biggest barriers to including key populations in clinical trials? What approaches or measures might stimulate greater inclusion of key populations in trials, including greater community engagement and awareness?
2. What would make the inclusion of key populations easier for researchers?
3. What special considerations need to be taken into account to include key populations into trials? Can they be included as an additional arm of study? A part of a larger patient group?
4. At what phase is it most appropriate to include key populations?
5. Areas where key populations are included should be prioritized based on burden. What are these priority areas, and what are the requirements for each population?
Why is it important to include key populations in clinical trials?

After unanticipated harm occurred from in utero exposure to thalidomide and diethylstilbestrol in the 1960s and 1970s, the United States Food and Drug Administration (FDA) enacted policies to protect women research participants of reproductive age from teratogenic exposure [15]. An unintended consequence has been the uniform exclusion of pregnant women from Phase III trials of TB therapies, even for MDR and extremely drug-resistant (XDR) TB [7,8]. Exclusion has been based on concerns of legal liability as well as new or increased frequency/severity of adverse events and potential unpredictability of such events in pregnancy or the postpartum period. Ethical complexities and insufficient market interests for developing pediatric formulations and concerns of potential DDIs among antiretrovirals and TB therapies are among the factors preventing adequate trial data from being collected from child and HIV–TB-coinfected populations, particularly those with advanced immunosuppression.

Although concerns of potential harm from TB therapeutics are understandable, a scientific and ethical foundation exists for including pregnant and lactating women and other key populations in trials of TB medicines for prevention and treatment [16,17]—namely, the need for effective treatment and evidence-based answers to enable patients to make fully informed choices for themselves (and the developing fetus) based on risks and benefits of specific therapies. However, these data are rarely available [8,16–20]. Pregnant and lactating women, children, and HIV-infected persons each have unique features. Thus, assumptions made from therapeutic TB trials excluding these populations are not always applicable, and data cannot be reliably extrapolated from other populations. Without high-quality data from targeted studies, many unanswered questions remain concerning optimal TB regimens, optimal dosing of new/existing TB drugs, and their safety.

Although the landmark zidovudine trial paved the way for rigorous study of HIV antiretrovirals in pregnancy [21], this has yet to translate to the TB arena. TB treatment in pregnancy and lactation is mostly based on case reports and small case series [6,7,22]. As a result, medications, including those for TB, are often prescribed in pregnancy without the knowledge required to achieve appropriate doses for optimal therapeutic effect [23,24], and WHO and Centers for Disease Control and Prevention (CDC) recommend conflicting treatment guidelines for drug-susceptible TB (i.e., 6-month regimen, including pyrazinamide versus 9-month regimen, excluding pyrazinamide, respectively) [25,26]. Overall, uncertainty persists concerning optimal drug selection, safety, and timing of TB treatment initiation and whether safety signals differ by trimester.

In pediatrics, off-label drug use is a common practice and is largely based on adult studies without rigorously conducted pharmacokinetics (PK), dose-finding, or formulation studies in children [27]. Children, however, are not small adults. The age-related risk of progressing to disease after TB infection and excess risk of disseminated forms of TB in children mandate the study of new therapies in this group. Additionally, it is critical to include young, small children in trials given that the effects of age and weight on PK are most pronounced and challenging to predict in this subgroup. Notably, the 2011 revised WHO dosing guidelines for first-line TB drugs in children <12 years old were based on studies suggesting that young children require higher mg/kg doses [28]. However, the evidence supporting these dosing recommendations was limited and especially lacking in studies using high-quality drug formulations. With a wide spectrum of disease, children with paucibacillary intrathoracic TB may in fact require lower total drug exposures (lower dose and/or shorter regimen), whereas children with more severe pulmonary TB or disseminated disease (e.g., TB meningitis) may require higher doses than adults.

Regardless of age, HIV-infected persons are at highest risk of developing TB and have a high TB-related mortality. In this population, differential responses to TB treatment and
preventive regimens and overlapping toxicities between HIV therapies and TB therapies are such that safety, toxicity, and DDIs cannot be predicted by modeling alone. In particular, adults and children with advanced HIV disease have more complex and unknown responses, toxicities, and DDIs than HIV-infected persons with higher CD4 T-cell counts. This subgroup is important to include in TB trials, as they may benefit from new TB therapies, but this needs to be ascertained carefully and is best done in a clinical trial setting.

Clearly, gathering evidence under rigorous scientific conditions is among the most compelling reasons for inclusion of key populations in TB drug research [16,17,23,29,30], especially because safety signals can be more readily interpreted in a clinical study setting. Controlled trials are also essential to assess specific TB treatment–associated outcomes and adverse effects. However, there are also issues of justice and access to the benefits of research participation. Inclusion in clinical trials is likely the only way for pregnant/lactating women, children, adolescents, and HIV-infected persons to access or accelerate access to new regimens and medications.

Overview of trial design considerations for key populations

Pregnant and lactating women

Overview of TB in pregnant and lactating women. In most countries, TB incidence peaks in women of reproductive age, irrespective of HIV [22]. Pregnancy is not routinely included in national/international TB registries, but worldwide, at least 216,000 TB cases are reported to occur in pregnancy annually [2]. Immune changes in pregnancy may alter the risk of disease, TB presentation, and diagnosis [4,31,32]. Complications of TB developing during pregnancy and lactation are well known and can include maternal death, preeclampsia, vaginal bleeding, and maternal death as well as prematurity, low birth weight, and fetal or infant death, particularly if TB is inadequately treated [22,33,34]. Notably, many TB drugs are categorized by the US FDA as former category C (Table 1), and many have undetermined placenta crossing, fetal, or lactation compatibility [6] (Table 1). In addition, drug absorption, distribution, metabolism, and elimination may be modified in pregnancy and lactation [35,36], and increased clearance of some drugs requires dose modification, particularly in the third trimester [37]. Lastly, there is often a significant time gap between licensure of medicines and pregnancy-specific data being obtained. HIV antiretrovirals, which have more data in pregnancy, still had a median gap of 6 years from licensure to access [38].

TB trial design considerations and recommendations for pregnant and lactating women. In 2018, the US FDA and the US Federal Task Force on Research Specific to Pregnant Women and Lactating Women (PREGLAC) issued separate documents to accelerate inclusion of pregnant and lactating women in clinical trials. The FDA draft guidance [23] outlines prerequisites for “reasonable” and “ethically justifiable” inclusion of pregnant women in premarking studies (i.e., “adequate” preclinical data plus the potential to provide unique clinical benefit to the woman or fetus) and postmarketing studies (i.e., “adequate” nonclinical data plus established safety in nonpregnant women and no alternate means to extrapolate efficacy and/or assess safety). Generally, Phase I and II trials should be conducted in nonpregnant women of reproductive age, and inclusion of pregnant women should be considered in Phase III or IV trials based on clear risks and benefits assessment. Critical trial components include PK data with minimum requirements (i.e., gestational age at enrollment, gestational timing/duration of drug exposure, and pregnancy outcomes [adverse maternal, fetal, and neonatal events]), obstetrical care meeting recognized standards for pregnant women on trial, and follow-up safety data among infants of mothers with investigational drug exposure. The FDA also provides guidance regarding evaluation of systemic drug exposure to fetus/newborn,
Table 1. FDA/WHO pregnancy classification and select maternal–fetal and reproductive toxicity characteristics of drugs used to treat TB.

<table>
<thead>
<tr>
<th>Drug Name</th>
<th>FDA Grouping</th>
<th>WHO Grouping</th>
<th>Crosses Placenta <em>(Cord: Maternal Ratio)</em></th>
<th>Fetal Toxicity</th>
<th>Breastfeeding Compatible</th>
<th>Teratogenic in Reproductive Toxicity Studies</th>
<th>Additional Concerns in Pregnancy and Postpartum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Isoniazid</td>
<td>C</td>
<td>1</td>
<td>Y</td>
<td>CNS defects</td>
<td>Yes (&lt;5%)</td>
<td>No</td>
<td>Possible increased hepatotoxicity</td>
</tr>
<tr>
<td>Rifampin</td>
<td>C</td>
<td>1</td>
<td>Y</td>
<td>Hemorrhage</td>
<td>Yes (minimal passage, approximately 0.05% to &lt;5%)</td>
<td>Yes</td>
<td>Possible postpartum hemorrhage; interacts with NNRTIs, PIs, decreases efficacy of hormonal contraceptives</td>
</tr>
<tr>
<td>Ethambutol</td>
<td>C</td>
<td>1/C</td>
<td>Yes</td>
<td>Jaundice</td>
<td>UD (minimal passage, &lt;5%)</td>
<td>Yes (low incidence)</td>
<td>–</td>
</tr>
<tr>
<td>Pyrazinamide</td>
<td>C</td>
<td>1/C</td>
<td>Unknown</td>
<td>Jaundice</td>
<td>UD (excreted in breast milk)</td>
<td>UD</td>
<td>Differential recommendation between US CDC and WHO for use in TB treatment in pregnancy</td>
</tr>
<tr>
<td>Rifabutin</td>
<td>B</td>
<td>–</td>
<td>UD</td>
<td>–</td>
<td>UD</td>
<td>No</td>
<td>Possible postpartum hemorrhage; interacts with NNRTIs, PIs, decreases efficacy of hormonal contraceptives</td>
</tr>
<tr>
<td>Rifapentine</td>
<td>C</td>
<td>–</td>
<td>UD</td>
<td>–</td>
<td>UD</td>
<td>Yes</td>
<td>Possible postpartum hemorrhage; interacts with NNRTIs, PIs, decreases efficacy of hormonal contraceptives</td>
</tr>
<tr>
<td>Aminoglycosides</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Capreomycin</td>
<td>C</td>
<td>Not A–C</td>
<td>Yes</td>
<td>–</td>
<td>UD</td>
<td>Yes</td>
<td>–</td>
</tr>
<tr>
<td>Streptomycin</td>
<td>D</td>
<td>C</td>
<td>Yes</td>
<td>Ototoxicity, thrush, diarrhea</td>
<td>Yes (minimal passage)</td>
<td>No</td>
<td>–</td>
</tr>
<tr>
<td>Kanamycin</td>
<td>D</td>
<td>Not A–C</td>
<td>Yes</td>
<td>Ototoxicity</td>
<td>Yes (minimal passage)</td>
<td>No</td>
<td>–</td>
</tr>
<tr>
<td>Amikacin</td>
<td>D</td>
<td>C</td>
<td>Yes</td>
<td>Ototoxicity</td>
<td>UD</td>
<td>UD</td>
<td>–</td>
</tr>
<tr>
<td>Levofoxacin</td>
<td>C</td>
<td>A</td>
<td>Yes</td>
<td>Possible bone</td>
<td>Yes</td>
<td>No</td>
<td>–</td>
</tr>
<tr>
<td>Moxifloxacin</td>
<td>C</td>
<td>A</td>
<td>Yes</td>
<td>Possible bone</td>
<td>UD</td>
<td>No</td>
<td>–</td>
</tr>
<tr>
<td>Gatifloxacin</td>
<td>C</td>
<td>Not A–C</td>
<td>UD</td>
<td>Possible bone</td>
<td>UD</td>
<td>No</td>
<td>–</td>
</tr>
<tr>
<td>Ethionamide/ prothionamide</td>
<td>C</td>
<td>C</td>
<td>UD</td>
<td>Developmental anomalies</td>
<td>UD</td>
<td>Yes</td>
<td>Developmental abnormalities in human case series</td>
</tr>
<tr>
<td>P-aminosalicylic acid</td>
<td>C</td>
<td>C</td>
<td>UD</td>
<td>Diarrhea</td>
<td>No</td>
<td>No</td>
<td>–</td>
</tr>
<tr>
<td>Cycloserine</td>
<td>C</td>
<td>B</td>
<td>UD</td>
<td>–</td>
<td>Yes</td>
<td>UD</td>
<td>Congenital sideroblastic anemia</td>
</tr>
<tr>
<td>Terizidone</td>
<td>–</td>
<td>B</td>
<td>UD</td>
<td>–</td>
<td>Yes</td>
<td>UD</td>
<td>–</td>
</tr>
<tr>
<td>Thioacetazone</td>
<td>–</td>
<td>Not A–C</td>
<td>UD</td>
<td>–</td>
<td>UD</td>
<td>UD</td>
<td>–</td>
</tr>
<tr>
<td>Clofazimine</td>
<td>C</td>
<td>B</td>
<td>UD</td>
<td>Reversible skin pigmentation</td>
<td>UD</td>
<td>No</td>
<td>–</td>
</tr>
<tr>
<td>Clarithromycin</td>
<td>C</td>
<td>Not A–C</td>
<td>Yes (0.15)</td>
<td>–</td>
<td>UD</td>
<td>No</td>
<td>–</td>
</tr>
<tr>
<td>Amoxicillin-clavulanic acid</td>
<td>B</td>
<td>Not A–C</td>
<td>Yes (0.56)</td>
<td>Necrotizing enterocolitis, transaminitis</td>
<td>UD</td>
<td>No</td>
<td>–</td>
</tr>
<tr>
<td>Linezolid</td>
<td>C</td>
<td>A</td>
<td>UD</td>
<td>–</td>
<td>UD</td>
<td>No</td>
<td>Case report of reduced PK in pregnancy</td>
</tr>
<tr>
<td>Imipenem/meropenem</td>
<td>C</td>
<td>C</td>
<td>UD</td>
<td>–</td>
<td>UD</td>
<td>No</td>
<td>–</td>
</tr>
<tr>
<td>High-dose isoniazid</td>
<td>C</td>
<td>Not A–C</td>
<td>Yes (0.73)</td>
<td>CNS defects</td>
<td>UD</td>
<td>No</td>
<td>Possible hepatotoxicity</td>
</tr>
<tr>
<td>Bedaquiline</td>
<td>B</td>
<td>A</td>
<td>UD</td>
<td>–</td>
<td>UD</td>
<td>No</td>
<td>Drug accumulation in tissues</td>
</tr>
<tr>
<td>Delamanid</td>
<td>Not approved</td>
<td>C</td>
<td>UD</td>
<td>–</td>
<td>UD</td>
<td>Yes</td>
<td>Embryofetal toxicity at maternally toxic doses in rabbits; breast milk concentration 4× higher than blood in rats</td>
</tr>
</tbody>
</table>

(Continued)
women who become pregnant on study, obtaining adequate nonclinical reproductive and developmental toxicology data, identifying trial populations standing to benefit most while minimizing risk, gestational timing of investigational drug exposure relative to fetal development, and appropriate control populations. In its report, PREGLAC highlighted 15 recommendations to encourage research on therapies during pregnancy and lactation, the majority of these being of particular relevance to TB therapeutics [18].

An international group of experts has also issued recommendations with particular reference to TB treatment trials: pregnant and lactating women should be eligible for Phase III MDR TB trials unless a compelling reason for exclusion exists, drug companies should be encouraged to complete reproductive toxicity studies of TB drugs before beginning Phase III studies, trials of shortened treatment regimens for latent TB infection (LTBI) should be designed to improve completion rates and reduce risk of progression in pregnancy and lactation, targeted PK studies should be nested in all TB studies when evidence is lacking, and a TB pregnancy registry should be established to accumulate data on maternal–infant outcomes [6]. These were discussed at the March 2018 WHO technical consultation discussions, and the following propositions were made.

**Trial designs for active TB disease in pregnant and lactating women.** Inclusion in Phase III trials is likely the only way to access more optimal regimens/newer agents and generally the only way to obtain safety, PK, and outcome data in this population, as postmarketing studies are not prioritized for funding or by regulatory bodies. In this respect, because MDR TB has significant morbidity and mortality and because many MDR TB drugs are associated

<table>
<thead>
<tr>
<th>Drug Name</th>
<th>FDA Grouping</th>
<th>Crosses Placenta (Cord: Maternal Ratio)</th>
<th>Fetal Toxicity</th>
<th>Breastfeeding Compatible</th>
<th>Teratogenic in Reproductive Toxicity Studies</th>
<th>Additional Concerns in Pregnancy and Postpartum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretomanid</td>
<td>Not approved</td>
<td>– UD</td>
<td>– UD</td>
<td>UD</td>
<td>UD</td>
<td></td>
</tr>
<tr>
<td>Sutezolid</td>
<td>Not approved</td>
<td>– UD</td>
<td>– UD</td>
<td>UD</td>
<td>UD</td>
<td></td>
</tr>
</tbody>
</table>

Table adapted from [6].

* The former FDA categories were defined as follows: category A: adequate and well-controlled studies have failed to demonstrate a risk to the fetus in the first trimester of pregnancy (and there is no evidence of risk in later trimesters); category B: animal reproduction studies have failed to demonstrate a risk to the fetus, and there are no adequate and well-controlled studies in pregnant women; category C: animal reproduction studies have shown an adverse effect on the fetus, and there are no adequate and well-controlled studies in humans, but potential benefits may warrant use of the drug in pregnant women despite potential risks; category D: there is positive evidence of human fetal risk based on adverse reaction data from investigational or marketing experience or studies in humans, but potential benefits may warrant use of the drug in pregnant women despite potential risks; category X: studies in animals or humans have demonstrated fetal abnormalities, and/or there is positive evidence of human fetal risk based on adverse reaction data from investigational or marketing experience, and the risks involved in use of the drug in pregnant women clearly outweigh potential benefits. The US FDA now uses narrative summaries to communicate what information is known and not known for individual drugs. However, the former risk categorization is still felt to be useful and has been used in this table. https://www.fda.gov/about-fda/economic-impact-analyses-fda-regulations/summary-content-and-format-labeling-human-prescription-drug-and-biological-products-requirements.

Additional information about each drug can be found at https://www.accessdata.fda.gov/scripts/cder/daf/index.cfm.

*Information on breast milk transfer of TB drugs is collated on LactMed, the National Library of Medicine searchable database of drugs to which breastfeeding mothers may be exposed. https://toxnet.nlm.nih.gov/newtoxnet/lactmed.htm.*

*Approved by European Medicine Association and other non-FDA agencies outside the US.

Abbreviations: CDC, Centers for Disease Control and Prevention; CNS, central nervous system; FDA, Food and Drug Administration; NNRTI, nonnucleoside reverse transcriptase inhibitor; PI, protease inhibitor; PK, pharmacokinetics; UD, undetermined; WHO, World Health Organization
with substantial intolerance and adverse effects, it is reasonable to consider inclusion of pregnant and lactating women in Phase III MDR TB treatment trials when there is no teratogenicity signal from reproductive toxicity. However, to our knowledge, no Phase III trial of MDR TB treatment has included pregnant women to date. To counter the automatic exclusion of pregnant women that currently pervades the TB trial landscape, early discussion among trialists, pharmaceutical companies, maternal–child experts, ethicists, and regulatory bodies are needed to address risks, benefits, and compelling rationale for inclusion [7].

Another important approach is to capture pregnancy outcomes among women who become pregnant while participating in a therapeutic trial. Current practice is to discontinue study drugs at the time pregnancy is identified and define the participant as “unassessable.” Instead, newly pregnant participants should be reconsented, offering the option to continue the study drug unless teratogenicity is known or suspected. All current information concerning the drug/regimen during pregnancy should be reviewed and communicated, including any shifts in risk–benefit balance, and carefully described to the patient. Examples of such secondary consent forms have been developed and are already used in some clinical trials [4]. Furthermore, support and mandates to standardize systematic data collection and reporting to a global pregnancy TB treatment registry is urgently needed. Similar to the HIV antiretroviral therapy (ART) registry, data from pregnancy, delivery, and infancy until age 6 months should be mandated [39, 40]. Whether from trials or registries, collecting PK and outcome data among pregnant women will be invaluable and can be pooled for analysis once sufficient data have accumulated. Novel physiologically based PK and pharmacodynamics (PD) modeling can also be applied to estimate drug dosing in pregnancy, but prediction of safety and toxicity profiles still requires trial data [41].

The postmarketing opportunistic PK model illustrated by International Maternal Pediatric Adolescent AIDS Clinical Trials Network (IMPAACT) P1026s [42] is another approach to advance the evidence base (Table 2). This protocol is enrolling pregnant and lactating women

Table 2. Ongoing and planned clinical trials in pregnant and lactating women (as of December 2018).

<table>
<thead>
<tr>
<th>Study/Trial Number</th>
<th>Funding/ Sponsor</th>
<th>Phase TB Type</th>
<th>Purpose</th>
<th>Design</th>
<th>Regimen</th>
<th>Study Population</th>
<th>Location</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>IMPAACT P2001/ NCT02651259</td>
<td>NIH NIAID, NICHD</td>
<td>I/II LTBI</td>
<td>PK, tolerability, and safety of 3HP for LTBI</td>
<td>Open-label, non-randomized trial</td>
<td>12 once-weekly doses of P and H (3HP)</td>
<td>Pregnant (≥14 weeks GA)/lactating women (18 years+), HIV + (any CD4, compatible ARV)/HIV–, with LTBI or known recent pulmonary TB exposure</td>
<td>Haiti, Kenya, Malawi, Thailand, Zimbabwe</td>
<td>Fully accrued/ results expected early 2020</td>
</tr>
<tr>
<td>IMPAACT P1078/ NCT011964038</td>
<td>NIH NIAID, NICHD</td>
<td>IV LTBI</td>
<td>Safety of antepartum versus postpartum-initiated IPT for TB prevention in HIV + pregnant women in high-TB-burden settings</td>
<td>Randomized, double-blind, placebo-controlled trial</td>
<td>Immediate H (entry through week 28), then placebo through week 40 postpartum versus placebo (entry through week 12 postpartum), then H through week 40 postpartum</td>
<td>Pregnant (≥14 weeks GA)/lactating women (13 years+), HIV + (any CD4, any ARV) without active TB</td>
<td>Botswana, Haiti, India, South Africa, Tanzania, Thailand, Uganda, Zimbabwe</td>
<td>Completed/ primary results presented CROI 2018 [49]</td>
</tr>
</tbody>
</table>
## Table 2. (Continued)

<table>
<thead>
<tr>
<th>Study/Trial Number</th>
<th>Funding/ Sponsor</th>
<th>Phase</th>
<th>TB Type</th>
<th>Purpose</th>
<th>Design</th>
<th>Regimen</th>
<th>Study Population</th>
<th>Location</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>IMPAACT CS 5021</td>
<td>NIH NIAID, NICHD</td>
<td>IV</td>
<td>LTBI</td>
<td>Safety, tolerability, optimal timing, and PK of 1HP versus 3HP in pregnant and postpartum women</td>
<td>Open-label, randomized, 4-arm factorial design trial</td>
<td>1HP versus 3HP in HIV-infected pregnant and postpartum women</td>
<td>Recently exposed or LTBI+, HIV+ (any CD4, compatible ARV) pregnant (≥24 weeks GA) women; subset of HIV– for PK and safety under consideration</td>
<td>Multisite international</td>
<td>Planned</td>
</tr>
<tr>
<td>DS TB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tshepiso</td>
<td>NIH NICHD</td>
<td>IV</td>
<td>DS</td>
<td>PK of first-line TB drugs</td>
<td>Open-label, nonrandomized trial</td>
<td>First-line TB drugs with and without ARVs</td>
<td>HIV+ (any CD4, any ARV)/HIV – pregnant and postpartum/lactating women</td>
<td>South Africa</td>
<td>Completed. Some results published [41,44,45]</td>
</tr>
<tr>
<td>PK of first-line TB drugs in pregnancy</td>
<td>NIH NICHD</td>
<td>IV</td>
<td>DS</td>
<td>PK of first-line TB drugs</td>
<td>Open-label, nonrandomized trial</td>
<td>First-line TB drugs with and without ARVs</td>
<td>HIV+ (any CD4, any ARV)/HIV – pregnant and postpartum/lactating women</td>
<td>India</td>
<td>Ongoing</td>
</tr>
<tr>
<td>DR TB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VirTUAL/ NCT03923231</td>
<td>EDCTP</td>
<td>DS/ DR</td>
<td>PK/PD modeling to predict doses for pregnant women, lactating women, and children</td>
<td>PK studies and modeled data</td>
<td>First- and second-line TB drugs with and without ARVs</td>
<td>HIV+/HIV – pregnant (≥20 weeks GA) and lactating women on first-line TB treatment or second-line MDR TB treatment. NCT03923231 assessing atazanavir/ritonavir with rifampin, specifically</td>
<td>South Africa, Uganda, United Kingdom, and Italy</td>
<td>Ongoing</td>
<td></td>
</tr>
<tr>
<td>ACTG 5300B</td>
<td>NIH NIAID, NICHD</td>
<td>III</td>
<td>DR</td>
<td>Efficacy and safety of De versus IPT for MDR TB prevention in high-risk household contacts (HIV+, non-HIV immunosuppression, LTBI, and children &lt;5 years)</td>
<td>Open-label, randomized trial</td>
<td>De x26 weeks versus H x26 weeks</td>
<td>Children and adult household contacts of MDR TB case. HIV+ (any CD4, any ARV)/HIV–, possible opportunistic substudy of PK among women who become pregnant during study drug intake</td>
<td>27 sites on 3 continents</td>
<td>Accrual expected to start mid-2019. Pregnancy study under consideration</td>
</tr>
<tr>
<td>BDQ in pregnancy</td>
<td>South Africa MRC</td>
<td>IV</td>
<td>DR</td>
<td>PK of BDQ in pregnancy</td>
<td>Open-label, nonrandomized trial</td>
<td>BDQ in optimized regimen</td>
<td>HIV+ (any CD4, compatible ARV)/HIV – pregnant and postpartum women on MDR TB treatment</td>
<td>South Africa</td>
<td>Ongoing</td>
</tr>
</tbody>
</table>

(Continued)
to assess the safety and PK of first- and second-line TB drugs routinely used in clinical practice as regimens evolve [43]. Assessments are made by pregnancy trimester, at delivery, and postpartum, with careful monitoring/ascertainment of maternal, fetal, and infant outcomes. PK of multiple TB drugs are captured in maternal plasma by pregnancy stage and from cord blood, breast milk, and infant samples along with relevant maternal–fetal–infant safety and clinical outcomes. This model also allows for study of DDIs between TB drugs and both antiretrovirals and postpartum contraceptives [44,45].

**Trial designs for TB preventive therapy in pregnant and lactating women.** Despite the large burden of LTBI and risk of progression to active TB, pregnant women have been systematically excluded from the >12 Phase III and postmarketing clinical studies of TB preventive therapy [6,46]. Data from nonpregnant individuals and small observational studies have informed the guidance for isoniazid preventive therapy (IPT) in pregnancy [47,48]. The first randomized placebo-controlled trial to assess safety and optimal timing of IPT in HIV-infected pregnant women in high-TB-burden settings (IMPAACT P1078) was recently completed (Table 2) [49]. The relative risks and benefits of immediate antepartum versus deferred postpartum IPT initiation was assessed and included careful monthly monitoring of maternal, fetal, pregnancy, and infant outcomes. No differences in maternal safety outcomes, maternal–infant TB, or infant safety outcomes were found between arms, but an increase in composite adverse pregnancy outcomes was observed in the immediate IPT arm. Shorter-course, efficacious TB preventive therapy regimens have been studied in nonpregnant adults [50,51].

### Table 2. (Continued)

<table>
<thead>
<tr>
<th>Study/Trial Number</th>
<th>Funding/ Sponsor</th>
<th>Phase</th>
<th>TB Type</th>
<th>Purpose</th>
<th>Design</th>
<th>Regimen</th>
<th>Study Population</th>
<th>Location</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>IMPAACT P1026s/ NCT0042289</td>
<td>NIH NIAID, NICHD</td>
<td>IV</td>
<td>DS/ DR</td>
<td>PK of ARVs and first- and second-line TB drugs (including BDQ and De) in pregnant women and their infants and ARVs in postpartum before/after initiation of hormonal contraceptives</td>
<td>Open-label, nonrandomized trial</td>
<td>ARVs without TB drugs; ARVs with TB drugs; no ARVs with TB drugs; +/- ARVs with second-line TB drugs; ARVs with postpartum hormonal contraceptives</td>
<td>HIV+ (any CD4, compatible ARV)/HIV – pregnant (≥20 weeks GA) and postpartum/ lactating women on first-line TB treatment or second-line MDR TB treatment</td>
<td>US and international sites (TB mostly from South Africa)</td>
<td>Accrual expected mid-2019/ results expected 2025</td>
</tr>
<tr>
<td>IMPAACT 2026</td>
<td>NIH NIAID, NICHD</td>
<td>IV</td>
<td>DS/ DR</td>
<td>PK of first- and second-line TB drugs in pregnant women with and without HIV</td>
<td>Open-label, nonrandomized trial</td>
<td>ARVs, contraception, and TB-related drugs during and after pregnancy</td>
<td>HIV+/HIV+ (any CD4, compatible ARV), pregnant (≥20 weeks GA) and postpartum/ lactating women on first-line TB treatment or second-line MDR TB treatment</td>
<td>TBD</td>
<td>Concept sheet in development</td>
</tr>
</tbody>
</table>

IMPAACT trial protocols can be found at [https://impaactnetwork.org/studies/index.asp](https://impaactnetwork.org/studies/index.asp); NCT is the [https://clinicaltrials.gov/](https://clinicaltrials.gov/) identification number; trials including HIV-infected (HIV+) are demarcated using bolded “HIV+” in the Study Population column.

Abbreviations: 1HP, 1 month of daily H and P; 3HP, 3 months of weekly H and P; ACTG, AIDS Clinical Trials Group; ARV, antiretroviral; BDQ, bedaquiline; CROI, Conference on Retroviruses and Opportunistic Infections; De, delaminid; DS, drug-sensitive; DR, drug-resistant; EDCTP, European & Developing Countries Clinical Trials Partnership; GA, gestational age; H, isoniazid; HIV, human immunodeficiency virus; IMPAACT, International Maternal Pediatric Adolescent AIDS Clinical Trials Network; IPT, isoniazid preventive therapy; LTBI, latent TB infection; MDR, multidrug-resistant; MRC, Medical Research Council; NIH, National Institutes of Health; NIAID, National Institute of Allergy and Infectious Diseases; NICHD, Eunice Kennedy Shriver National Institute of Child Health and Human Development; P, rifapentine; PD, pharmacodynamics; PK, pharmacokinetics; TB, tuberculosis; TBD, to be determined

[https://doi.org/10.1371/journal.pmed.1002882.t002](https://doi.org/10.1371/journal.pmed.1002882.t002)
greater advocacy and effort on behalf of groups focused on high-quality data for pregnant
women, postmarketing trials assessing shorter LTBI regimens are also now underway or in
development for pregnant women (Table 2). These include IMPAACT P2001 (PK and safety
of 3 months of weekly isoniazid and rifapentine [3HP]) and IMPAACT Concept 5021 (safety,
tolerability, optimal timing, and PK of 3HP versus 1 month of daily isoniazid and rifapentine
[1HP]).

The IMPAACT network serves as an excellent example of how a group focused on therapeutics in pregnant women can make major strides to close the evidence gap (Table 2). Establishing a global TB registry and inclusion of pregnant women into relevant Phase III TB trials should be the next step. TB therapeutic protocols under development should be reviewed by experts in the care of TB in pregnant women, maternal–fetal medicine specialists, regulatory authorities, and bioethicists who can further comment on the risks and benefits of including pregnant women during the trial planning stage.

Children

Overview of TB in children. Globally, approximately 10% of TB cases occur among chil-
dren (0–14 years) annually. Of the estimated 1,000,000 cases in 2017, only 360,000 were noti-
fied to WHO, yet children < 5 years old are particularly vulnerable, accounting for >50% of
child TB cases and approximately 80% of child TB-related deaths [1]. In contrast to the situa-
tion in adults, children display a wide spectrum of TB disease phenotypes ranging from nonse-
vere, often paucibacillary pulmonary/intrathoracic TB (usually uncomplicated lymph node
disease) to severe disseminated TB and TB meningitis, a major cause of TB-related morbidity
and mortality in children [52]. Paucibacillary intrathoracic TB (minimal or nonsevere TB) is
more prevalent overall, and TB treatment outcomes are generally good for drug-sensitive (DS)
and drug-resistant (DR) TB (provided treatment is initiated early), even when considerably
lower doses of antituberculosis drugs were used for DS TB [53]. However, risk of progression
from infection to active TB disease varies substantially by age and with HIV infection. PK also
varies because of effects related to child age and size. Young children, particularly <2 years
old, are at much higher risk of developing TB and severe disease forms [54] and typically
require higher mg/kg doses of most TB drugs to reach adult therapeutic targets. Finally, TB
diagnosis and treatment response monitoring rely on clinical, more subjective measures in at
least 60% of children, as young children cannot spontaneously produce sputum for examina-
tion, and paucibacillary disease (sputum smear negative) is diagnosed by culture, the current
diagnostic gold standard, in only 30%–40% of cases [55].

TB trial design considerations and recommendations for children. With concerted
effort and advocacy along with academic and government funding and recognition from regu-
larly agencies, the pediatric TB trial landscape has substantially improved, as evidenced by
the number of ongoing and planned studies of treatment for the diverse forms of TB in chil-
dren (Table 3). The ways in which pediatric and adult TB differ inform the type of pediatric
TB drug trials needed and their key design considerations. If children are to be included in
adult trials, different inclusion and exclusion criteria may be needed, and definitions used to
determine study endpoints (e.g., unfavorable outcome) require careful consideration because
of differing clinical features and diagnostic challenges of TB in children compared with adults.
Diagnosis, treatment response monitoring, and characterization of treatment outcome in chil-
dren often depend on clinical measures that are relatively imprecise compared with the diag-
nostic standard used in adults. Limited availability of pediatric-friendly formulations also
poses a barrier to enrollment of younger children. Large Phase III clinical trials may not be fea-
sible or always needed for children, yet timely PK and safety data in children, especially in
Table 3. Ongoing and planned TB clinical trials in children (as of December 2018).

<table>
<thead>
<tr>
<th>Study/Trial Number</th>
<th>Funding/ Sponsor</th>
<th>Phase</th>
<th>TB Type</th>
<th>Purpose</th>
<th>Design</th>
<th>Regimen</th>
<th>Study Population</th>
<th>Location</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>LTBI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TBTC Study 35/ NCT03730181</td>
<td>TBTC, CDC</td>
<td>I/II</td>
<td>LTBI</td>
<td>Optimal dose, PK, and safety of 3HP for LTBI in HIV +/− children</td>
<td>Open-label PK and safety trial of P and H coformulation</td>
<td>P in fixed dose combination + H + P single formulation</td>
<td>Infants and children (0–12 years old), HIV +/−, modified age de-escalation, population PK modeling</td>
<td>South Africa</td>
<td>Accrual expected to start 2019</td>
</tr>
<tr>
<td>IMPAACT CS 5019</td>
<td>NIH NIAID, NICHD</td>
<td>I/II</td>
<td>LTBI</td>
<td>PK, safety, and tolerability of 1HP in HIV-infected and uninfected children with exposure to DS TB</td>
<td>Multicenter, open-label dose-finding and safety study</td>
<td>1HP with integrate inhibitors in HIV-infected children</td>
<td>Infants, children, and adolescents &lt;12 years old, HIV+/−</td>
<td>Multisite international</td>
<td>Planned</td>
</tr>
<tr>
<td>iTIPS/NCT02613169</td>
<td>Thrasher Research Fund</td>
<td>II</td>
<td>LTBI</td>
<td>Efficacy of INH to prevent MTB in HIV-exposed uninfected infants</td>
<td>Randomized control trial</td>
<td>Daily H ×12 months versus no H</td>
<td>Infants (6 weeks), HIV-exposed</td>
<td>Kenya</td>
<td>Fully accrued</td>
</tr>
<tr>
<td>P4v9 Trial/ NCT00170209</td>
<td>Canadian Institutes of Health Research</td>
<td>III</td>
<td>LTBI</td>
<td>Efficacy, safety, and tolerability of R and H for LTBI</td>
<td>Multicenter, open-label, randomized positive-controlled trial</td>
<td>R ×4 months versus H ×9 months</td>
<td>Children and adolescents (&lt;18 years), children with LTBI at high risk of TB</td>
<td>Canada, Australia, Benin, Ghana, Indonesia</td>
<td>Fully accrued</td>
</tr>
<tr>
<td>TB-CHAMP/ ISRCTN92634082</td>
<td>Joint Global Health Trials Scheme, South African MRC</td>
<td>III</td>
<td>LTBI</td>
<td>Efficacy of Le for MDR TB prevention in HIV +/− child household contacts</td>
<td>Multicenter, cluster randomized, double-blind, placebo-controlled, superiority trial</td>
<td>Daily Le ×6 months versus placebo</td>
<td>Infants and children (0 to &lt;5 years old), HIV+/−/HIV−, household randomization, IGRA+/−</td>
<td>South Africa</td>
<td>Enrolling</td>
</tr>
<tr>
<td>V-QUIN/ ACTRN12616000215426</td>
<td>Australian National Health and MRC</td>
<td>III</td>
<td>LTBI</td>
<td>Efficacy of Le for MDR TB prevention in adult and adolescent household contacts</td>
<td>Multicenter, randomized, double-blind placebo-controlled, superiority trial</td>
<td>Daily Le ×6 months versus placebo</td>
<td>Adolescents and adults, HIV+/−, household randomization, TST+</td>
<td>Vietnam</td>
<td>Enrolling</td>
</tr>
<tr>
<td>A5300B I2003B/ NCT03568383 (PHOENIx)</td>
<td>NIH NIAID, NICHD</td>
<td>III</td>
<td>LTBI</td>
<td>Efficacy and safety of De versus standard-dose H for MDR TB prevention in high-risk household contacts</td>
<td>Multicenter, open-label, randomized superiority trial</td>
<td>Daily De ×26 weeks versus daily H + vitamin B6 ×26 weeks</td>
<td>Adults, adolescents, children, infants, HIV+/−, household randomization</td>
<td>Botswana, Brazil, Peru, India, Philippines, Haiti South Africa, Thailand, Kenya</td>
<td>Planned to open 2019</td>
</tr>
<tr>
<td>DS TB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAtiC/NICHD069175 NCT01637358</td>
<td>NIH NICHD</td>
<td>I</td>
<td>DS</td>
<td>PK of first-line TB drugs using 2010 WHO guidelines across pediatric populations</td>
<td>Intensive PK sampling of HRZE</td>
<td>ATT no ART, ATT + LPV/r-based ART; no ART on LPV/r-based ART, ATT + NVP-based ART</td>
<td>Children and infants (0–12 years), HIV+/HIV−, malnutrition, drug–drug interactions, population PK modeling</td>
<td>South Africa, Malawi</td>
<td>Fully accrued</td>
</tr>
<tr>
<td>OptiRif Kids</td>
<td>TB Alliance, Unitaid</td>
<td>I</td>
<td>DS</td>
<td>PK, safety, and dose optimization of R for TB treatment in children and infants</td>
<td>Intensive PK sampling</td>
<td>High-dose R</td>
<td>Infants and children (0–12 years old), HIV−</td>
<td>South Africa</td>
<td>Fully accrued</td>
</tr>
</tbody>
</table>

(Continued)
<table>
<thead>
<tr>
<th>Study/Trial Number</th>
<th>Funding/ Sponsor</th>
<th>Phase</th>
<th>TB Type</th>
<th>Purpose</th>
<th>Design</th>
<th>Regimen</th>
<th>Study Population</th>
<th>Location</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>IMPAACT P1101/ NCT01751568</td>
<td>NIH NIAID, NICHD</td>
<td>I/II</td>
<td>DS</td>
<td>Safety and tolerability of raltegravir with R-containing TB regimen in infants and children</td>
<td>Open-label, dose-finding, safety, tolerance, and PK study of raltegravir</td>
<td>Chewable raltegravir tablets + 2NRTIs + R-containing TB regimen</td>
<td>HIV+/TB-coinfected children (≥4 weeks to &lt;12 years old), received ≥1 week and ≤20 weeks of R-based TB therapy prior to ARV initiation</td>
<td>South Africa</td>
<td>Enrolling</td>
</tr>
<tr>
<td>TBM-KIDS/ NCT02958709</td>
<td>NIH NICHD</td>
<td>II</td>
<td>DS</td>
<td>Efficacy, PK, and safety of high-dose R +−/−/− Le for TB meningitis in children</td>
<td>Open-label, randomized trial</td>
<td>High-dose R + EHZ x2 months/10HR versus high-dose R + LeHZ x2 months/10HR versus standard of care (2REHZ/10HR)</td>
<td>Children and infants (6 months–12 years), HIV+, intensive PK, population PK modeling</td>
<td>India, Malawi</td>
<td>Enrolling</td>
</tr>
<tr>
<td>SHINE study/ ISRCTN63579542</td>
<td>Joint Global Health Trials Scheme</td>
<td>III</td>
<td>DS</td>
<td>Efficacy and safety of shortened first-line TB regimen using 2010 WHO-recommended doses for minimal TB in children</td>
<td>Open-label, randomized, noninferiority trial</td>
<td>2HRZ/E/2HR versus 2HRZ/E/4HR</td>
<td>Children, adolescents, and infants (0–16 years old), HIV+/HIV−, nested PK studies, drug–drug interactions</td>
<td>India, Uganda, Zambia, South Africa</td>
<td>Fully accrued</td>
</tr>
<tr>
<td>SURE-TBM/ ISRCTN40829906</td>
<td>MRC, DFID, NIH, Wellcome Trust</td>
<td>III</td>
<td>DS</td>
<td>Efficacy and safety of high-dose R, Le, and H with Z for shortening TB meningitis treatment to 6 months</td>
<td>Open-label, randomized, noninferiority trial</td>
<td>Higher dose (6LLeHZ) versus WHO standard of care regimen (2HRZE/10HR)</td>
<td>Infants, children, and adolescents (28 days–15 years old), HIV+/−</td>
<td>Vietnam, India, Uganda, Zambia, Zimbabwe</td>
<td>Planned</td>
</tr>
<tr>
<td>Rifabutin PK trial</td>
<td>ICMR, NACO</td>
<td>IV</td>
<td>DS</td>
<td>PK and safety of rifabutin</td>
<td>PK and safety</td>
<td>Rifabutin</td>
<td>Adults, children, HIV−</td>
<td>India</td>
<td>Planned</td>
</tr>
<tr>
<td>DR TB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MDR-PK 1</td>
<td>NIH NICHD</td>
<td>I/II</td>
<td>DR</td>
<td>PK, safety of second-line drugs for MDR TB, particularly Mo, Le, and Li</td>
<td>Semi-intensive PK sampling, model-based analysis</td>
<td>Ethionamide, Le, ofloxacin, Mo, high-dose H, PZA, terizidone, PAS</td>
<td>Children, infants, adolescents (&lt;18 years), HIV+/−, drug–drug interactions</td>
<td>South Africa</td>
<td>Fully accrued</td>
</tr>
</tbody>
</table>

(Continued)
### Table 3. (Continued)

<table>
<thead>
<tr>
<th>Study/Trial Number</th>
<th>Funding/ Sponsor</th>
<th>Phase</th>
<th>TB Type</th>
<th>Purpose</th>
<th>Design</th>
<th>Regimen</th>
<th>Study Population</th>
<th>Location</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>MDR-PK2</td>
<td>NIH NICHD</td>
<td>I/II</td>
<td>DR</td>
<td>PK, safety of second-line drugs for MDR TB, particularly Mo, Le, and Li</td>
<td>Semi-intensive PK sampling, model-based analysis</td>
<td>Li, Mo, Le, clofazimine, BDQ</td>
<td>Children, infants, adolescents (&lt;18 years old), HIV +/-, drug–drug interactions</td>
<td>South Africa</td>
<td>Fully accrued</td>
</tr>
<tr>
<td>IMPAACT P1108 NCT02906007</td>
<td>NIH NIAID, NICHD</td>
<td>I/II</td>
<td>DR</td>
<td>PK, safety, and tolerability of BDQ for MDR TB</td>
<td>Open-label, single-arm, dose-finding and safety study</td>
<td>BDQ × 24 weeks + routine background MDR therapy</td>
<td>Children, infants, adolescents (0–18 years old), HIV +/-, population PK modeling, modified age de-escalation</td>
<td>South Africa, India, Haiti</td>
<td>Enrolling</td>
</tr>
<tr>
<td>232 and 233 NCT01856634 NCT01859923</td>
<td>Otsuka</td>
<td>I/II</td>
<td>DR</td>
<td>PK, safety, tolerability, and efficacy of De + MDR TB therapy in HIV−</td>
<td>Multiple doses of De × 6 months + OBR</td>
<td>Open-label, single-arm dose-finding trial</td>
<td>Children, infants, adolescents (0–17 years old), HIV−, population PK modeling, age de-escalation</td>
<td>Philippines, South Africa</td>
<td>Fully accrued</td>
</tr>
<tr>
<td>IMPAACT 2005 NCT03141060</td>
<td>NIH NIAID, NICHD</td>
<td>I/II</td>
<td>DR</td>
<td>PK, safety, tolerability of De + OBR for MDR TB in HIV +/- children</td>
<td>Multisite, open-label, single-arm dose-finding trial</td>
<td>De × 6 months + OBR</td>
<td>Children, infants, adolescents (&lt;18 years old), HIV−, population PK modeling</td>
<td>Botswana, India, South Africa, Tanzania</td>
<td>Enrolling</td>
</tr>
<tr>
<td>Janssen C211 NCT02354014</td>
<td>Janssen</td>
<td>II</td>
<td>DR</td>
<td>PK, safety, tolerability of BDQ + OBR for MDR TB</td>
<td>Multicenter, open-label, single-arm, dose-finding and safety trial</td>
<td>BDQ × 24 weeks + OBR</td>
<td>Children, infants, adolescents (0–18 years old) HIV−, age de-escalation</td>
<td>Russian Federation, South Africa, Philippines</td>
<td>Enrolling</td>
</tr>
<tr>
<td>IMPAACT 2020 “Smart Kids”</td>
<td>NIH NIAID, NICHD</td>
<td>II</td>
<td>DR</td>
<td>Safety, efficacy of oral 6-month regimens for RR/ MDR/pre-XDR/ XDR TB</td>
<td>Multicenter, open-label, randomized trial</td>
<td>Oral 6-month regimen BDQ, De, Li, Le (clofazimine for FQ resistant)</td>
<td>Infants, children, adolescents (0–15 years old), HIV+/-</td>
<td>Multisite</td>
<td>Planned</td>
</tr>
<tr>
<td>IMPAACT P1106 NCT02383849</td>
<td>NIH NIAID, NICHD</td>
<td>IV</td>
<td>DS/ DR</td>
<td>PK and safety of R and H with NVP or LPV/r in low-birth-weight infants</td>
<td>Open-label, nonrandomized PK study of ARVs and TB medicines</td>
<td>NVP versus NVP + H versus NVP + H + R versus H alone or H + R versus LPV/r + 2NRTIs +/- H versus LPV/r + 2NRTIs + R +/- H</td>
<td>Infants (7–14 days old), HIV+/-, low birth weight, premature</td>
<td>South Africa</td>
<td>Enrolling</td>
</tr>
</tbody>
</table>

IMPAACT trial protocols can be found at [https://impaactnetwork.org/research-areas/tuberculosis.htm](https://impaactnetwork.org/research-areas/tuberculosis.htm); NCT is the [https://clinicaltrials.gov](https://clinicaltrials.gov) identification number; trials including HIV-infected (HIV+) persons are demarcated using bolded ”HIV+” in the Study Population column.

Abbreviations: 1HP, 1 month of daily isoniazid and rifapentine; 3HP, 3 months of weekly isoniazid and rifapentine. AECID, Agencia Española de Cooperación Internacional para el Desarrollo (Spanish Agency for International Development Corporation); ART, antiretroviral therapy; ARV, antiretroviral; AFD, French Development Agency; ATT, antituberculosis therapy; BDQ, bedaquiline; CDC, Centers for Disease Control and Prevention; De, delamanid; DFID, British Department for International Development; DS, drug-sensitive; DR, drug-resistant; E, ethambutol; FQ, fluoroquinolone; H, isoniazid; HIV, human immunodeficiency virus; ICMR, Indian Council of Medical Research; IGRA, interferon gamma release assay; IMPAACT, International Maternal Pediatric Adolescent AIDS Clinical Trials Network; INH, isoniazid; Le, levofloxacin; Li, linezolid; LPV, lopinavir; LTBI, latent TB infection; Mo, moxifloxacin; MDR, multidrug-resistant; MRC, Medical Research Council; MSF, Médecins Sans Frontières; NACO, National AIDS Control Organization; NIAID, National Institute of Allergy and Infectious Diseases; NICHD, Eunice Kennedy Shriver National Institute of Child Health and Human Development; NIH, National Institutes of Health; NIHR, National Institute for Health Research; NRTI, nucleoside reverse transcriptase inhibitor; NVP, nevirapine; OBR, optimized background regimen; P, rifapentine; PAS, P-aminosalicylic acid; PK, pharmacokinetics; PZA, pyrazinamide; R, rifampin; RR, rifampicin-resistant; SDC, Swiss Agency for Development and Cooperation; TB, tuberculosis; TBTC, Tuberculosis Trials Consortium; TST+, tuberculin skin test positive; WHO, World Health Organization; XDR, extremely drug-resistant; Z, pyrazinamide.

[https://doi.org/10.1371/journal.pmed.1002882.t003](https://doi.org/10.1371/journal.pmed.1002882.t003)
young and HIV-infected children, is critical to inform policy guidance on new therapies deemed to be safe and efficacious in adolescent and adult populations. Modified study designs should be explored to accelerate implementation of PK and safety studies in children while ensuring the validity of the trial results and the safety of all child participants. Unlike younger children, adolescents (typically ≥ 10 years old) have TB disease characteristics similar to adults, including frequent cavitating disease. Adolescents should therefore be routinely considered for inclusion in adult Phase IIb and III trials. However, similar to pregnant and lactating women, legal requirements for child participation in clinical trials are often barriers (perceived or real) and vary by country. When feasible and justified through appropriate consultation, the inclusion of children should be carefully considered and supported early during protocol development. Summaries of considerations for the types of trials needed for children, including practical and ethical considerations regarding inclusion of children in TB trials, can be found elsewhere [5, 56]. Highlights and considerations discussed at the WHO technical consultation are discussed below based on updated information.

**Trial designs for active TB disease in children.** Considering scenarios in which disease progression and/or response to an intervention are expected to differ among adults and children, the classical approach is to conduct PK studies in children to establish appropriate dosing followed by safety and efficacy trials. For example, because children often develop less severe, paucibacillary TB, it is plausible that children would respond equally well (i.e., treatment would have at least equal efficacy) to shorter, less intense, and less complex regimens than adults while potentially improving their tolerability, safety, acceptability, and adherence. Identifying such regimens would require an efficacy study in children, as regimens that could be effective in children may be rejected in adult trials. Based on these assumptions, the currently ongoing Shorter Treatment for Minimal TB in Children (SHINE) trial (ISRCTN63579542) investigates whether a shorter 4-month regimen can be used for children with less severe disease than the standard 6-month adult regimen (Table 3). Other examples include the treatment of LTBI (discussed below) and TB meningitis. TBM Kids (NCT 02958709) is the only currently open trial to assess the treatment of TB meningitis, which especially affects very young children.

In contrast, when considering scenarios in which children and adults are expected to have similar disease progression, response to an intervention, and exposure response, then it is logical to conduct PK studies to achieve drug exposures similar to adults, followed by safety trials at the proper dose. For individual TB medications, it is reasonable to assume that the response in children would be at least as good as in adults. Therefore, repeating formal efficacy studies for individual TB drugs in children is unnecessary. Instead, the focus should be on trials to establish PK, dose, and safety in children. Many of the trials shown in Table 3 are such studies, including the pediatric trials of the recently approved drugs bedaquiline and delamanid. Another example is the Opti-Rif Kids trial (South African trial identifier 27-0117-5411), which aims to characterize rifampin doses among children 0 to <12 years old that approximate exposures observed in adults receiving higher rifampin doses (≥35 mg/kg) in adult trials [57]. Both age and weight have an impact on PK in children and must be considered in the design of pediatric PK studies of TB drugs. It is especially critical to include young, small children given that the effects of age and weight are most pronounced in this subgroup. Traditionally, age-de-escalation studies have been a major feature of pediatric PK-focused Phase I/II trials whereby children have been studied in series, rather than in parallel, starting with older children and progressing to younger children. This approach, however, should be avoided if possible: it is costly and time consuming; older children may have limited ability to inform dosing and safety in the youngest children, for whom there is the most uncertainty; and regulatory agencies do not strictly require age de-escalation [5]. HIV infection and malnutrition are additional,
important covariates to consider when designing pediatric trials, and these children should be included in TB therapeutic trials.

If the exposure response to an intervention is expected to differ among children and adults, then PK/PD should be conducted to establish the exposure response in children followed by safety studies. If a PD marker is unavailable to assess pharmacologic response, as is typically the case in bacteriologically unconfirmed TB (i.e., clinically diagnosed TB), then PK studies should be followed by safety and efficacy studies [56]. The traditional assumption that exposure response is similar among children and adults for all types of TB is being questioned. For example, most children with pulmonary TB are sputum smear and culture negative and therefore have different bacillary burden compared with adults with cavitating disease. Given that childhood TB may differ in disease type and severity compared with adult TB, target concentrations for treatment of many forms of childhood TB may differ from those in adults. This provides additional justification for efficacy trials in children in some instances. For example, there are no data from trials investigating regimens to prevent MDR TB in either adult or child household contacts. TB-CHAMP (ISRCTN92634082) is a Phase III cluster-randomized placebo-controlled study that is specifically powered to evaluate the efficacy of 6 months of levofloxacin versus placebo for the prevention of TB in young child household contacts (age < 5 years) of MDR TB cases. Although not powered for efficacy in children, the PHOE-Nlx trial (A5300/12003) plans to study adult, HIV-infected, and child contacts of MDR TB using delamanid versus isoniazid and is a good example of how key populations can be studied within a single Phase III efficacy trial (Table 2).

Lastly, child-friendly formulations are important to ensure accurate, acceptable, and palatable doses in young children. The development and implementation of bioequivalence studies of pediatric formulations is lengthy and should start much earlier during the drug development process. A potential temporary solution is to better understand how manipulating the adult formulation affects the PK to inform pediatric use. The TASK-002 study successfully assessed the relative bioavailability of 100-mg bedaquiline tablets suspended in water versus when administered in healthy adult volunteers to inform its use in children [58]. This does not eliminate the need for making pediatric formulations available but does improve access to much-needed medications during the timeframe following trial completion and drug registration until routine medication availability.

HIV-infected persons

Overview of TB in HIV-infected persons. Worldwide, an estimated 1,040,000 TB cases and 300,000 TB deaths occurred among HIV-infected persons in 2017–86% of reported HIV-associated TB deaths occurred in sub-Saharan Africa [1]. TB is 20–30 times more likely in the context of HIV and remains the leading cause of death in this population. Adults and children with advanced HIV disease (low CD4 count) are especially vulnerable. This subgroup has a particularly high mortality rate [59] and is more likely to have disseminated TB disease and more rapid disease progression. Despite this, a 2011 review revealed that many TB trials exclude HIV-infected persons with CD4 counts < 200–350 cells/mm³ [60]; our review of recent [61–64], currently enrolling, and registered (clinicaltrials.gov) randomized TB trials suggests recent expansion of inclusion criteria, but HIV-infected persons with very low CD4 counts (<50–100 cells/mm³) remain frequently excluded (Table 4). Overall, clinical management of dual TB–HIV disease is complex [12,65]. As in children, smear-negative TB disease is common in the context of HIV, which poses challenges for TB diagnosis and treatment monitoring. In addition, polypharmacy arising from treatment of HIV, TB, and new/existing comorbidities may increase adverse events and impact adherence and tolerability. Drug
Table 4. Ongoing and planned TB clinical trials in HIV-infected persons 12 years and older (as of December 2018).

<table>
<thead>
<tr>
<th>Study/Trial Number</th>
<th>Funding/ Sponsor</th>
<th>Phase</th>
<th>TB Type</th>
<th>Purpose</th>
<th>Design</th>
<th>Regimen</th>
<th>Study Population</th>
<th>Location</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>LTBI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WHIP3TB/ NCT02980016</td>
<td>USAID</td>
<td>III</td>
<td>LTBI</td>
<td>Efficacy of 3HP given once or annually to reduce TB</td>
<td>Open-label, randomized trial</td>
<td>Part A: 6H versus 3HP; part B: 3HP once versus annually</td>
<td>Adults, adolescents, children (2+ years old), HIV+ on ART 3+ months or not eligible for ART, any CD4</td>
<td>South Africa, Mozambique, Ethiopia</td>
<td>Enrolling</td>
</tr>
<tr>
<td>TBTC37</td>
<td>CDC</td>
<td>III</td>
<td>LTBI</td>
<td>Efficacy and safety of 6 weeks of HP daily</td>
<td>Open-label, randomized trial</td>
<td>6 weeks daily HP versus 3HP versus 4HR daily versus 4R daily</td>
<td>Adults and adolescents (12 + years old), HIV +/-, on compatible ART, any CD4</td>
<td>US, TBTC international sites TBD</td>
<td>In development</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DS TB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NCT03563599</td>
<td>Qurient</td>
<td>IIa</td>
<td>DS</td>
<td>Assess early bactericidal activity of Telaceboc</td>
<td>Open-label, randomized trial</td>
<td>Multiple doses of Telaceboc (Q203) versus Rifafour e-275 (RHZE)</td>
<td>Adult (18–65 years old), new treatment-naïve smear-positive DS TB, no HIV exclusion criteria stated</td>
<td>South Africa</td>
<td>Enrollment complete March 19, 2019</td>
</tr>
<tr>
<td>ReDEFINE/ NCT02169882</td>
<td>Universitas Padjadjaran, USAID</td>
<td>IIb</td>
<td>DS</td>
<td>Dose finding for R to treat TB meningitis</td>
<td>Double-blind randomized trial</td>
<td>Standard dose versus R$<em>{900}$ or R$</em>{1350}$ + HEZ &gt;6 months</td>
<td>Adults (15+ years old), no pregnancy/breastfeeding, on ATT &lt;3 days with clinical suspicion of TBM, no HIV exclusion criteria stated</td>
<td>Indonesia</td>
<td>In data analysis</td>
</tr>
<tr>
<td>APT/ NCT02256696</td>
<td>FDA</td>
<td>IIb</td>
<td>DS</td>
<td>Mycobacterial activity of Pa824</td>
<td>Open-label, randomized trial</td>
<td>Pa824$_{(200)}$ ×12 weeks added to HRZ</td>
<td>Adults (18+ years old); HIV−/HIV + CD4 ≥350 cells/mm$^3$ and not on ART</td>
<td>South Africa</td>
<td>Enrolling</td>
</tr>
<tr>
<td>ACTG5362 CLOFAST</td>
<td>NIH NIAID</td>
<td>IIc</td>
<td>DS</td>
<td>Dose finding for C to treat DS TB</td>
<td>Double-blind randomized trial</td>
<td>(4C$<em>{10}$ versus 4C$</em>{100}$ versus 4placebo) + 4HP$_{120}ZE/2placebo$ versus 2placebo versus 2HP</td>
<td>Adults (18+ years old), no pregnancy/breastfeeding, HIV + CD4 ≥100 cells/mm$^3$, compatible ARV or about to start</td>
<td>ACTG sites TBD</td>
<td>In development</td>
</tr>
<tr>
<td>NCT02836483</td>
<td>LegoChem Biosciences</td>
<td>II</td>
<td>DS</td>
<td>Early bactericidal activity, safety, and PK of oral delpazolid</td>
<td>Open-label, randomized trial</td>
<td>Multiple doses of delpazolid versus Li</td>
<td>Korean adults (19–70 years old) with smear-positive pulmonary TB. No HIV exclusion criteria stated</td>
<td>Korea</td>
<td>Enrolling</td>
</tr>
<tr>
<td>TBTC Study 31 ACTG 5349/ NCT02410772</td>
<td>AIDS Clinical Trials Group, CDC</td>
<td>III</td>
<td>DS</td>
<td>Efficacy of 2 shortened rifapentine-containing regimens for pulmonary TB</td>
<td>Open-label, randomized, controlled clinical trial</td>
<td>Standard 6-month regimen versus 4-month regimen substituting P for R versus 4-month regimen substituting P for R and M for E</td>
<td>Children and adults (12 years+), AFB or GeneXpert-positive, documented HIV status, if HIV + CD4 &gt; 100 cells/mm$^3$</td>
<td>USA, Brazil, China, Haiti, India, Kenya, Malawi, Peru, South Africa, Thailand, Uganda, Vietnam, Zimbabwe</td>
<td>Enrollment completed. Follow-up ongoing.</td>
</tr>
</tbody>
</table>

DR TB

(Continued)
### Table 4. (Continued)

<table>
<thead>
<tr>
<th>Study/Trial Number</th>
<th>Funding/ Sponsor</th>
<th>Phase</th>
<th>TB Type</th>
<th>Purpose</th>
<th>Design</th>
<th>Regimen</th>
<th>Study Population</th>
<th>Location</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACTG 5356</td>
<td>NIH NIAID</td>
<td>IIa</td>
<td>DR</td>
<td>Dose-finding for Li in all oral regimens for MDR TB</td>
<td>Open-label, randomized trial</td>
<td>Li (600 qd/1,200 qd) + Bdq200 +De200 + Le (if FQ S) or C (if FQ R)</td>
<td>Adults and adolescents (&gt;12 years old); if HIV + CD4 ≥ 50 cells/mm³</td>
<td>ACTG sites TBD</td>
<td>In development</td>
</tr>
<tr>
<td>TBTC Study 32, OPTI-Q</td>
<td>CDC, NIH NIAID</td>
<td>II</td>
<td>DR</td>
<td>Efficacy, safety, and tolerability of using Le in regimen for MDR TB</td>
<td>Blinded, randomized PK/PD trial</td>
<td>4 doses of Le + OBR</td>
<td>Adults (18+ years old), smear-positive/culture-positive MDR TB; HIV+ included must have viral load and CD4 count within 3 months</td>
<td>Peru, South Africa</td>
<td>Enrolling</td>
</tr>
<tr>
<td>ACTG 5343/ NCT02583048</td>
<td>NIH NIAID</td>
<td>II</td>
<td>DR</td>
<td>Safety, tolerability, and PK of BDQ and De (alone and in combination) + OBR for RR/MDR TB</td>
<td>Open-label, randomized trial</td>
<td>6 months of BDQ + OBR versus De + OBR versus BDQ + De + OBR, dolutegravir + 2 NRTIs for HIV+ only</td>
<td>Adults (18+ years old), documented RR/MDR pulmonary TB, documented HIV status, if HIV+: CD4 &gt; 100 cells/mm³ and one fully active NRTI available if on ART &gt;6 months and viral load &gt; 500 copies/mL</td>
<td>Peru, South Africa</td>
<td>Active, not recruiting</td>
</tr>
<tr>
<td>MDR END/ NCT02619994</td>
<td>University Seoul, Korea</td>
<td>II</td>
<td>DR</td>
<td>Safety, efficacy of shortened injection-free regimen for MDR TB</td>
<td>Open-label, randomized controlled clinical trial</td>
<td>De + Le + Li + Z x 9 or 12 months versus 24 OBR</td>
<td>Adults 19+ years old, no FQ resistance, no HIV exclusion criteria stated</td>
<td>Korea</td>
<td>Enrolling</td>
</tr>
<tr>
<td>SODOCU</td>
<td>EDCTP</td>
<td>II</td>
<td>DR</td>
<td>Dose-finding study of U</td>
<td>Open-label dose-finding trial</td>
<td>3U (0 mg qd + 600 mg qd versus 1,200 mg qd versus 600 mg bid versus 800 mg bid) + 3BdqDeM</td>
<td>Adults</td>
<td>TBD</td>
<td>In development</td>
</tr>
<tr>
<td>SimpliciTB/ NCT03338621</td>
<td>Global Alliance for TB Drug Development</td>
<td>II/III</td>
<td>DS/ DR</td>
<td>Efficacy, safety, and tolerability of a new, shorter oral regimen for DS/DR TB</td>
<td>Open-label, partially randomized trial</td>
<td>DS TB: BDQPaMoZ ×4 months versus HRZE/HR ×6 months; DR TB: BdqPaMoZ ×6 months</td>
<td>Adults (18+ years old), new smear-positive DS/DR TB; HIV + criteria: CD &gt;100 cells/mm³, Karnofsky score &gt;60%, no IV antifungal in past 90 days, and WHO clinical stage &lt;4 disease</td>
<td>10 countries in Africa, Asia, Europe, and South America</td>
<td>Enrolling</td>
</tr>
<tr>
<td>TB PRACTICAL/ NCT02589782</td>
<td>MSF, Global Alliance for TB Drug Development, WHO, THINK</td>
<td>II/III</td>
<td>DR</td>
<td>Safety (Phase II) and efficacy (Phase III) of short regimens containing B and Pa for MDR/XDR TB</td>
<td>Open-label, randomized trial</td>
<td>6 months of BdqPaLiMo, BdqPaLiC, or BdqPaLi versus local WHO SOC MDR/XDR TB regimen</td>
<td>Adults (18+ years old), with microbiologically confirmed TB resistant to at least R; HIV+ included regardless of status</td>
<td>Belarus, South Africa, Uzbekistan</td>
<td>Enrolling</td>
</tr>
</tbody>
</table>

(Continued)
Table 4. (Continued)

<table>
<thead>
<tr>
<th>Study/Trial Number</th>
<th>Funding/ Sponsor</th>
<th>Phase</th>
<th>TB Type</th>
<th>Purpose Design</th>
<th>Regimen</th>
<th>Study Population</th>
<th>Location</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>NExT-5001/ NCT02454205</td>
<td>University of Cape Town</td>
<td>II/III</td>
<td>DR</td>
<td>Efficacy, safety, tolerability of shortened, injection-free regimen for MDR TB</td>
<td>Open-label, randomized controlled clinical trial</td>
<td>LiBdqLeZ + E or high-dose H ×6–9 months versus conventional empiric injection-based 21–24 month regimen</td>
<td>Adults (18+ years old), new culture or GeneXpert-positive MDR TB; if HIV + CD4 &gt; 50 cells/mm³</td>
<td>South Africa</td>
</tr>
<tr>
<td>ACTG 5273 FIRST</td>
<td>NIH NIAID</td>
<td>III</td>
<td>DR</td>
<td>Efficacy of new regimens for H mono-resistant TB</td>
<td>Open-label, randomized clinical trial</td>
<td>6H 15mg/kg RZE versus 2RZELe/2RLe</td>
<td>Adults, adolescents, and children; any CD4, any ARV</td>
<td>ACTG sites TBD</td>
</tr>
<tr>
<td>STREAM/ NCT02409290</td>
<td>IUATLD</td>
<td>III</td>
<td>DR</td>
<td>Efficacy of different regimens for MDR TB</td>
<td>Open-label, randomized clinical trial</td>
<td>Local 2011 WHO MDR TB regimen versus CEMZ ×40 weeks + H, kanamycin, prothionamide × first 16 weeks versus 40 weeks oral regimen BdqCELeZ + H and prothionamide × first 16 weeks versus 28 weeks BdqCLeZ + H and kanamycin × first 8 weeks</td>
<td>Adult (15+ years old), AFB or GeneXpert positive, resistant to rifampicin and isoniazid, if HIV+: willing to start ART and CD4 &gt; 50 cells/mm³</td>
<td>Ethiopia, Georgia, India, Republic of Moldova, Mongolia South Africa, Uganda</td>
</tr>
<tr>
<td>Nix-TB (B-Pa-L)/ NCT02333799</td>
<td>Global Alliance for TB Drug Development</td>
<td>III</td>
<td>DR</td>
<td>Safety, efficacy, tolerability, and PK of BDQ + Li ×6 months for MDR/XDR TB</td>
<td>Open-label trial</td>
<td>6–9 months of BdqPaLi</td>
<td>Children and adults (14+ years old), XDR TB or nonresponsive MDR TB, culture-positive, documented HIV status; if HIV + CD4 &gt; 50 cells/mm³</td>
<td>South Africa</td>
</tr>
<tr>
<td>ZeNix NC-007/ NCT03086486</td>
<td>Global Alliance for TB Drug Development</td>
<td>III</td>
<td>DR</td>
<td>Safety and efficacy of various doses and treatment duration of Li + Pa + BDQ for MDR, pre-XDR, and XDR TB</td>
<td>Open-label, partially blinded, randomized clinical trial; even allocation across arms by HIV status and TB type</td>
<td>Li(1,200) ×26 weeks + Pa + BDQ versus Li(1,200) ×9 weeks + Pa + BDQ versus L(600) ×26 weeks + Pa + BDQ versus Li(600) ×9 weeks + Pa + BDQ</td>
<td>Children and adults (14+ years old), documented HIV status, culture or molecular test positive and documented resistance; if HIV + CD4 &gt; 100 cells/mm³</td>
<td>Georgia, Republic of Moldova, Russian Federation, South Africa</td>
</tr>
<tr>
<td>endTB/ NCT02754765</td>
<td>UNITAID</td>
<td>III</td>
<td>DR</td>
<td>Evaluating newly approved oral, shortened regimens for MDR TB (FQ sensitive)</td>
<td>Open-label, randomized controlled noninferiority clinical trial</td>
<td>BdqLiMoZ ×39 weeks BdqLiCleZ ×39 weeks BdqDeLiLeZ ×39 weeks DeLiCleZ ×39 weeks DeCmOZ ×39 weeks versus control (Z)</td>
<td>Children and adults (15+ years old) with documented pulmonary MTB resistant to R, no HIV exclusion criteria stated</td>
<td>Georgia, Kazakhstan, Kyrgyzstan, Lesotho, Peru, and South Africa</td>
</tr>
<tr>
<td>endTB-Q</td>
<td>UNITAID/ MSF</td>
<td>III</td>
<td>DR</td>
<td>Evaluating newly approved oral, shortened regimens for MDR TB (FQ sensitive)</td>
<td>Open-label, randomized controlled noninferiority clinical trial</td>
<td>6BdqDeLiC versus 10BdqDeLiC versus OBR</td>
<td>Children and adults (15+ years old) with documented pulmonary MTB resistant to R, no HIV exclusion criteria stated</td>
<td>India, Pakistan, Kazakhstan, Lesotho, Peru</td>
</tr>
</tbody>
</table>
metabolism, absorption, and toxicity profiles may be altered in HIV, making longer courses of
treatment and side effects, such as neuropathy, liver injury, and rash, more likely [66,67].
Immune reconstitution inflammatory syndrome (IRIS)/paradoxical worsening, specific cyto-
chrone interactions, poor nutritional status, and chronic inflammation further impact HIV-
infected populations. As in children and pregnant women, physiologically based PK modeling
can help inform TB drug dosing in the setting of HIV but cannot replace data generated from
trials. In recent years, high-quality evidence has dramatically evolved the use and timing of TB
treatment in relation to ART [68]—persons with advanced HIV who are diagnosed with TB
are currently recommended to start ART within 2 weeks [69,70]. However, potential DDIs
remain a major concern for TB treatment in HIV-infected persons, particularly between anti-
retroviral agents such as protease inhibitors and integrase inhibitors, and rifamycins, key TB steriliz-
ing agents [12,65]. DDIs and adverse effects cannot always be readily identified from
observations in HIV-uninfected populations. A healthy-volunteer study assessing a TB-pre-
ventive regimen (rifapentine and isoniazid) and interaction with dolutegravir (HIV antiretro-
viral) found significant toxicity and was terminated early, yet these effects were not observed
in a larger study of HIV-infected persons [71,72]. It is important that TB trials assess the full
spectrum of HIV/TB and be sufficiently powered to evaluate the impact of HIV [41,60].

**Trial design considerations and recommendations for TB disease and preventive thera-
pies in HIV-infected persons.** Inclusion of HIV–TB-coinfected populations in TB clinical
trials poses a number of challenges. To enhance their enrollment, TB trials should be con-
ducted, at least in part, in geographic locations where HIV and TB epidemics coincide and
interact. Partnering with public-funded trials networks specializing in recruitment of HIV-
infected persons can facilitate this. For example, the US CDC Tuberculosis Trials Consortium
(TBTC)/AIDS Clinical Trials Group (ACTG) partnership has enhanced enrollment of HIV-
infected people in the Phase III randomized trial of rifapentine-containing shortened treat-
ment for pulmonary TB (NCT02410772). Requesting culture-confirmed disease for trial

<table>
<thead>
<tr>
<th>Study/Trial Number</th>
<th>Funding/ Sponsor</th>
<th>Phase</th>
<th>TB Type</th>
<th>Purpose</th>
<th>Design</th>
<th>Regimen</th>
<th>Study Population</th>
<th>Location</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>BEAT TB</td>
<td>South Africa</td>
<td>III</td>
<td>DR</td>
<td>Safety and efficacy of short regimen for MDR TB</td>
<td>Strategy trial</td>
<td>6BdqDeLeLi400C (drop Li if FQ sensitive; drop Le if FQ resistant)</td>
<td>Adults</td>
<td>South Africa</td>
<td>In development</td>
</tr>
<tr>
<td>DRAMATIC/ NCT03828201</td>
<td>US Department of Defense</td>
<td>III</td>
<td>DR</td>
<td>Efficacy and tolerability of shortened injection-free regimen for MDR TB</td>
<td>Open-label, randomized controlled clinical trial</td>
<td>4BdqDeLe400Li1200C versus 8BdqDeLe400Li1200C versus 2011 WHO regimen</td>
<td>Adults and adolescents 12 + years old, HIV−/ HIV+ any CD4</td>
<td>TBD</td>
<td>In development</td>
</tr>
</tbody>
</table>

Data in this table obtained from clinicaltrials.gov and adapted from a table compiled by Michael J. Vjecha, MD, on behalf of TBTC Core Science Group.

 Abbreviations: 3HP, 3 months of weekly isoniazid and rifapentine; ACTG, AIDS Clinical Trials Group; AFB, acid-fast bacilli; ART, antiretroviral therapy; ARV, antiretroviral; ATT, antituberculosis therapy; Bdq, bedaquiline; C, clofazimine; CDC, Centers for Disease Control and Prevention; De, delamanid; E, ethambutol; EDCTP, European & Developing Countries Clinical Trials Partnership; FDA, Food and Drug Administration; FQ, fluoroquinolone; HIV, human immunodeficiency virus; H, isoniazid; IUATLD, International Union Against Tuberculosis and Lung Disease (The Union); Le, levofloxacin; Li, linezolid; LTBI, latent TB infection; MDR, multidrug-resistant; Mo, moxifloxacin; MSF, Médecins Sans Frontières; NIAID, National Institute of Allergy and Infectious Diseases; NIH, National Institutes of Health; NRTI, nucleoside reverse transcriptase inhibitor; OBR, optimized background regimen; P, rifapentine; Pa, pretomanid; PD, pharmacodynamics; PK, pharmacokinetics; R, rifampicin; RR, rifampicin-resistant; s, sensitive; SOC, standard of care; TB, tuberculosis; TBD, to be determined; TBTC, Tuberculosis Trials Consortium; THINK, TB&HIV Investigative Network; U, sutezolid; USAID, United States Agency for International Development; WHO, World Health Organization; XDR, extremely drug-resistant; Z, pyrazinamide

[https://doi.org/10.1371/journal.pmed.1002882.t004](https://doi.org/10.1371/journal.pmed.1002882.t004)
eligibility also limits enrollment of HIV–TB-coinfected persons. Sensitivity of sputum smear and culture are limited by low bacillary load of TB in the context of HIV [73]. As in young children, less stringent measures, such as clinical TB diagnosis, could be incorporated. To ensure balanced treatment assignments among various trial subgroups, randomization could be stratified by HIV status (i.e., HIV-infected versus -uninfected) or by specific eligibility criteria (i.e., culture-confirmed versus nonconfirmed). Incorporating clinical TB diagnosis as a secondary outcome measure (ideally reviewed by an expert committee blinded to treatment assignment) may also be important for interpreting results in the overall trial population and in key subgroups. Outcome rates could also be assessed by HIV infection/HIV disease status and/or ART use, as treatment outcomes in HIV–TB-coinfected patients may be highly dependent on the specifics of ART management. Consistent with HIV and TB treatment guidelines, ART should be required or expected to be initiated within 4–8 weeks of initiating TB treatment. It is important to understand whether mortality or other poor outcomes in HIV–TB-coinfected patients is related to HIV or TB. Thus, data analysis should be stratified by HIV infection/HIV disease status (i.e., HIV-uninfected, HIV-infected with high CD4 count, and HIV-infected with low CD4 count) to reduce concerns about any potential imbalances in subgroup numbers between randomized arms.

Carefully designed DDI studies are a major element of clinical research of TB therapeutics for treatment and prevention of TB in HIV-infected people, including HIV-infected adults and children [74]. DDIs may be bidirectional, and the potential impact of host genetics is difficult to predict from small PK studies alone. To facilitate enrollment of HIV-infected individuals, DDI studies should be conducted early in drug development and/or nested in major trials [41]. The Phase III randomized ACTG 5279 trial, "Short-Course Rifapentine/Isoniazid for Treatment of Latent TB in HIV-Infected Individuals” (NCT01404312)[51], is an example of a nested DDI study: the first 90 participants that were on efavirenz-based ART and randomized to the rifapentine arm entered into a semi-intensive PK study [75] and were evaluated for PK/PD and potential HIV virologic failure to confirm that efavirenz PK and ART outcomes remained adequate. As in this example, the risk to a TB trial may be lower if PK of an HIV drug is the concern, particularly for shorter periods of TB drug use. If the potential DDI involves one of the TB drugs and may affect the randomized comparison, then an alternative trial design might be used: HIV-infected individuals could be excluded from randomization to the TB intervention but entered into a parallel PK cohort to evaluate the DDI. Once the potential DDI has been resolved, including by testing different drug dosing, randomization of HIV-infected individuals might proceed expeditiously. Alternatively, an observational study could be conducted whereby HIV-infected people who are on a targeted HIV drug and start a TB drug of interest would undergo PK/PD evaluations. IMPAACT P1026s (NCT00042289) uses this design to evaluate routinely used dosing of ART and TB (DS and DR TB) drugs during pregnancy in HIV-infected and uninfected women. The key is to have an ongoing, approved protocol in place that allows for targeted drugs to be studied without needing to develop a new study for each potential DDI. Irrespective of the design used, the respective advantages and disadvantages of intensive versus sparse drug sampling should be considered to facilitate rapid enrollment and availability of information about potential DDIs.

Conclusions

TB therapeutic trials that exclude key populations are often not followed by trials in those populations. Pregnant and lactating women, children, and HIV-infected persons contribute a large proportion of the global TB burden and require optimized TB treatment and access to the latest therapeutic advances. Overall, adequate inclusion and appropriate study of these
populations remain problematic, particularly for pregnant and lactating women; some advances are being made for children, yet pediatric TB trials lag far behind adult trials despite the potential for better TB treatment outcomes among children, and further evaluation of DDIs is needed in HIV–TB-coinfected populations to ensure that HIV-infected persons, particularly those with more advanced HIV disease, more fully benefit from therapeutic advances. Importantly, despite the differences among these populations, several cross-cutting themes exist and can serve as a way forward toward inclusion of key populations in TB clinical trials (Box 2).

Box 2. Summary of recommendations and cross-cutting issues among key populations

1. Pregnant and lactating women, children, and HIV-infected persons have increased susceptibility to TB and variable responses during TB treatment, which cannot be predicted by modeling data alone. Inclusion into clinical trials—especially Phase IIb and beyond—is often the best way to generate population-specific data, as postmarketing studies are not prioritized and cause delay in obtaining needed information.

2. Ethics are not a reason to exclude people from clinical trials, but careful consideration of design and involvement of content experts, regulatory agency inputs, and community participation is critical to ensure appropriate trial design and implementation. Inclusion will continue to require careful risks and benefits assessments, weighing direct benefits alongside potential risks of adverse effects from interventions on a case-by-case basis. The uncertainty cost of uniform exclusion results in lack of guidance to inform use of these important TB therapies.

3. Design of trials requires careful attention to how safety, risks, and benefits are defined and measured. Novel approaches may be useful, such as desirability of outcome ranking (DOOR)/response adjusted for duration of antibiotic risk (RADAR), a methodology that integrates overall clinical outcome and patient-level risks and benefits and was specifically developed for clinical trials comparing strategies to optimize antibiotic use [76].

4. Rigorous qualitative research is useful to inform trial design and elicit patient, caregiver, and family preferences regarding trial participation and regimens.

Acknowledgments

Thanks to WHO for organizing the conference in Montreux and especially to Christian Lienhardt and Payam Nahid for their leadership in this effort.

This review is an independent work by all authors. The content of this paper is solely the responsibility of the authors and does not necessarily represent the official views of the funders.
References


COLLECTION REVIEW

Development of new TB regimens: Harmonizing trial design, product registration requirements, and public health guidance

Christian Lienhardt1*, Andrew A. Vernon2, Marco Cavaleri3, Sumati Nambiar4, Payam Nahid5

1 Unité Mixte Internationale TransVIHMI, UMI 233 IRD–U1175 INSERM—Université de Montpellier, Institut de Recherche pour le Développement (IRD), Montpellier, France, 2 Division of TB Elimination, National Center for HIV, Viral Hepatitis, STD and TB Prevention, Centers for Disease Control and Prevention, Atlanta, Georgia, United States of America, 3 Office of Anti-infective and Vaccines, Human Medicines Evaluation, European Medicines Agency, Amsterdam, the Netherlands, 4 Division of Anti-Infective Products, Center for Drug Evaluation and Research, US Food and Drug Administration, Silver Spring, Maryland, United States of America, 5 Division of Pulmonary and Critical Care Medicine, Center for Tuberculosis, University of California, San Francisco, San Francisco, California, United States of America

* christian.lienhardt@ird.fr

Summary points

• Regulatory approval of new tuberculosis (TB) drugs can be based on data from trial(s) using a surrogate endpoint of treatment efficacy under an accelerated or conditional procedure. In such circumstances, policy makers and TB programs can be hampered in their ability to make recommendations on the optimal use of the drug(s), and consequently, the uptake by national or international public health institutions of such recommendations can be limited.

• Based on the essential need to produce high-quality evidence for policy decisions, this paper reflects on specific methodological issues in clinical trial design that need to be addressed to improve compliance with clinical, regulatory, and public health requirements.

• Established mechanisms for communication between drug developers and regulators already exist; however, equal engagement with policy makers is also essential for the optimal selection of trial designs, endpoints, and markers of treatment outcome and for giving consideration to public health and program aspects.

• The next generation of TB trials should better reconcile the research agenda with the need for global policies on access to TB medicines. Policy decision-makers should establish formal mechanisms for iterative feedback on regimen-development pathways. In this paper, we provide examples of how the need for interactions between regulators, trialists, and policy decision-makers can be addressed.

Introduction

Under the paradigm of adding a new drug to a regimen or substituting single drugs in a regimen one at a time, it would take 15–20 years to develop an entirely new tuberculosis (TB) regimen comprising three to four new drugs [1]. As has been noted in the papers of this Special
supplies or funding for PK sub-studies. One company, Sanofi, has provided 6 unrestricted grants to the CDC Foundation over the years 2007–2015 totaling ~$2.8 million to facilitate or support TBTC work related to rifapentine. These funds have supported several PK sub-studies, supported 3 contract research staff, have funded travel to TBTC scientific meetings for invited speakers (all in coach class), and have supported expenses related to fulfillment of company requests for data and data formats as part of their efforts to use TBTC data to support regulatory filings. None of these funds have otherwise benefited members of his research group.

Abbreviations: AE, adverse event; BMRC, British Medical Research Council; CDC, Centers for Disease Control; CDISC, Clinical Data Interchange Standards Consortium; Ctz, clofazimine; CROI, Conference on Retroviruses and Opportunistic Infections; Del, delamanid; DR, drug-resistant; DS, drug-sensitive; E, ethambutol; EBA, early bactericidal activity; EMA, European Medicine Agency; FDA, Food and Drug Administration; GRADE, Grading of Recommendations Assessment, Development, and Evaluation; H, isoniazid; HIV, human immunodeficiency virus; IND, investigational new drug; ITT, intention-to-treat; IUATLD, International Union Against Tuberculosis and Lung Diseases; Lfx, levofloxacin; Lzd, linezolid; M, moxifloxacin; MDR, multidrug-resistant; MGIT, mycobacterial growth-in-tube; MIC, minimum inhibitory concentration; mITT, modified intent-to-treat; MSF, Médecins Sans Frontiers; NI, noninferiority; NIAID, National Institute of Allergy and Infectious Diseases; NUS, National University of Singapore; OBR, optimized background regimen; orig, originally; ped, pediatric; PP, per protocol; Pret, pretomanid; R, rifampin 10 mg/kg; R35, rifampin at 35 mg/kg; Rpt, rifapentine; RR, rifampin-resistant; TB, tuberculosis; TBTC, TB Trials Consortium; TTP, time to positivity; USAID, US Development Aid Agency; WHO, World Health Organization; Z, pyrazinamide.

Provenance: Not commissioned; part of a Collection; externally peer reviewed.

Collection on Advances in Clinical Trial Design for Development of New TB Treatments [2–4], the major challenges in the development of new TB treatments include the long developmental pathway to identify best regimens, the uncertainties around the correlation between the treatment effect and existing surrogate endpoints, and uncertainties around the predictive quantitative relationships between Phase II and Phase III trial outcomes. Beyond measures of efficacy, the development of shorter, simpler regimens combining new and existing drugs also requires detailed information on their respective safety and toxicity, their potential for drug–drug interactions, their propensity for development of drug resistance while on therapy, and their use in specific patient populations such as persons infected with human immunodeficiency virus (HIV), pregnant women, and children [5].

Over the last decade, a series of clinical trials have been carried out to assess the safety and efficacy of new or repurposed drugs for the treatment of TB [6]. Although in some of these trials the endpoints were selected to address regulatory requirements, such endpoints were not always optimal to draw inferences for policy-recommending institutions, such as the World Health Organization (WHO), that provide guidance on the optimal use of these drugs in combination treatment regimens [2]. Ideally, clinical trials should provide results that are as meaningful as possible for clinical, regulatory, and programmatic perspectives. In situations when the regulatory approvals are conditional, based on surrogacy or on preliminary limited clinical data sets, the question is posed as to what extent policy makers can suitably generate comprehensive recommendations on the optimal use of the drug(s) in combination regimens. What needs to be considered in the design of a clinical trial to have relevance across regulatory and programmatic requirements? The design and choice of specific endpoints in trials of new TB drugs and regimens have implications for the development of guidelines and their adoption by national or international public health institutions. Starting from the need to produce evidence of high quality, this paper reflects on study designs and endpoints that respond best to the combined clinical, regulatory, and public health requirements.

The regulatory needs

In principle, regulatory authorities overseeing drug development have the primary responsibility of ensuring that the quality, efficacy, and safety of marketed medicinal products are adequate, conforming to currently defined standards. A key role of the regulatory authorities is to determine whether there is a positive benefit–risk balance to support use of the drug for the proposed indication and patient population.

Regulators also continue to reevaluate the benefit–risk balance after approval through pharmacovigilance activities and postmarketing studies. New data that emerge in the postapproval phase are taken into consideration in reassessing the benefit–risk balance, and information is communicated in product labeling as appropriate. Regulators, however, are not expected to consider cost-effectiveness or to perform in-depth evaluations of comparative effectiveness in assessing benefit and risk or for defining treatment policies. This role lies, rather, within the scope of public health recommending bodies, and, even if at times there seems to be some overlap, it is important to recognize and understand the implications of this distinction.

Some regulatory agencies have mechanisms for accelerated reviews and early approval of new drugs that address unmet needs according to specified criteria—e.g., the conditional marketing authorization pathway in the European Union where the benefit–risk balance of the new drug is such that immediate availability justifies acceptance of less comprehensive data than normally required [7,8]. In the United States, the accelerated approval pathway allows for the approval of a product for a serious disease with an unmet need based on a surrogate or an intermediate clinical endpoint that is reasonably likely to predict clinical benefit [9]. The
accelerated approval pathway has been used primarily in conditions in which the disease course is long and an extended period of time would be required to measure the intended clinical benefit of a drug. The implication is that, while awaiting further data to be generated post-approval, there may be limited data to support policy recommendations at this stage.

Development of new TB drugs and regimens is a good example of a scenario in which regulators need to establish that a drug submitted for licensure is safe and effective for the proposed use, whereas recommending bodies need to define how to use the drug optimally within a regimen in a way that addresses the public health need. Often, demonstrating the safety and effectiveness of a drug is the first step. Although a single clinical study cannot answer all research questions at once, it is still worth exploring clinical study designs that maximize the chance of gathering evidence that is informative both for assessing the benefit–risk of individual drugs and for determining their optimal use in the context of TB regimens. In view of the shift in focus toward the development of new treatment regimens, the European Medicine Agency (EMA) has proactively issued updated guidance to developers to address such scenarios [10].

In July 2017, the US Food and Drug Administration (FDA) held a public workshop regarding scientific and clinical trial design considerations for development of new TB drug regimens [11]. Of note, the FDA and EMA work collaboratively to provide advice to pharmaceutical sponsors or investigators on various aspects of the clinical trial design and to ensure that, whenever feasible, the same development program addresses the regulatory requirements of these agencies (for instance, the FDA pre–investigational new drug (IND) consultative process allows facilitated early communications between the FDA and potential drug sponsors or investigators [12]).

The public health needs

Countries, technical agencies, donors, and other TB stakeholders, routinely seek guidance and advice from WHO on optimal disease management practices to be adopted based on the evidence available. Over the last decade, WHO has published a series of normative guidance documents for the diagnosis and treatment of all forms of TB, with a particular focus on the needs of low- and middle-income countries [13]. In 2007, WHO adopted a procedure to guarantee that guidelines are based on the best available evidence and meet the highest international standards. Using the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) framework, which relies on the use of systematic reviews and meta-analyses, the findings of these reviews are then considered in the context of implementation and feasibility issues of stakeholder countries [14,15]. The GRADE framework provides an explicit and transparent approach to assess the level of certainty in the evidence across relevant studies and outcomes and to translate that evidence to recommendations. This framework incorporates multiple processes to minimize bias and optimize usability and requires rigor, fairness, and transparency in all judgments and decision-making.

To formulate evidence-based recommendations, four key aspects are taken into account:

1. the respective magnitude of benefits and harm conferred by the intervention under evaluation;
2. the consideration of resource use, feasibility, acceptability, and equity;
3. the certainty (“quality”) of evidence; and
4. patients’ values and preferences. Based on this assessment, the proposed recommendation is qualified as “strong” or “conditional” (i.e., “weak”), reflecting the extent to which one can, across the range of patients for whom the recommendation is intended, be certain in the evidence that the desirable effects of the given intervention outweigh the undesirable effects. The assessment of each of the above aspects leads, understandably, to the consideration of a number of nuances when moving from clinical trial results to public health policy making. As a result, the final qualification of the
recommendation ultimately has implications for the way policy makers, clinicians, and patients interpret and adopt the guidance, as shown in Table 1.

Recent developments highlight how trial results that are used as the basis for regulatory approval may allow only conditional recommendations for policy making due to the use of surrogate endpoints and limited data on patient- and population-relevant outcomes. As an example, the accelerated approval of bedaquiline by the US FDA in December 2012, based on the surrogate endpoint of sputum culture conversion at 6 months, allowed the drug to be readily used in the treatment of multidrug-resistant (MDR)-TB under certain conditions in the field [16]. However, the data gathered from the pivotal Phase II trial appeared inadequate for policy decision-making because of the absence of information on the outcomes of interest (nonrelapsing cure); further, the selected design did not provide information on the optimal use of the drug in combination with others or whether the addition of the drug would allow any modification in treatment duration. Finally, there was an excess of deaths in the experimental arm, the significance of which was uncertain given the small sample sizes and lack of long-term follow-up. These limitations in the available evidence at the time of regulatory review led to the adoption of a conditional recommendation that had implications in terms of wider scale-up of the intervention. Thus, for bedaquiline, results of the pivotal Phase II trial, in addition to relevant safety data, were adequate for obtaining regulatory approval but appeared insufficient for wider policy recommendations [17], thus calling for postlicensure evidence generation. The yield of a large body of observational data obtained over a subsequent period, associated with large individual-patient data meta-analyses, allowed WHO to update its recommendations for MDR-TB treatment in December 2018 [18], with significant changes in the assessment of the quality of evidence. As a result, bedaquiline is now strongly recommended for use in the treatment of MDR-TB, based on moderate-quality evidence—showing the importance of collecting additional data to complement early trial results. It should be noted that, at the time, the standard of care for rifampicin-resistant (RR)-TB treatment had low efficacy and high toxicity and was based on observational evidence. Though these conditions are now changing, a similar situation may present itself again in the future. Therefore, the experience with bedaquiline raises the question of whether specific trial features and designs can be used to produce endpoints with value for both the regulator and the policy maker. It is with this objective in mind that the Task Force on New Drug Policy Development established by WHO in 2012 worked together with drug developers, regulators, scientists, and program managers to define the policy needs and produce relevant documents [19].

### Methodological issues: How to fit both regulatory and programmatic decision-making needs

Could outcome definitions in clinical trials be redesigned to satisfy both regulatory and programmatic decision-making needs? We argue that this is feasible, and WHO Technical

<table>
<thead>
<tr>
<th>Target population</th>
<th>Strong recommendation</th>
<th>Conditional/weak recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Policy makers</td>
<td>The recommendation can be adapted as a policy in most situations</td>
<td>There is a need for substantial debate and involvement of stakeholders</td>
</tr>
<tr>
<td>Patient</td>
<td>Most people in this situation would want the recommended course of action, and only a small proportion would not</td>
<td>The majority of people in this situation would want the recommended course of action, but many would not</td>
</tr>
<tr>
<td>Clinician</td>
<td>Most patients should receive the recommended course of action</td>
<td>Be more prepared to help patients to make a decision that is consistent with their own values/decision aids and shared decision-making</td>
</tr>
</tbody>
</table>

**Abbreviation:** GRADE, Grading of Recommendations Assessment, Development, and Evaluation

https://doi.org/10.1371/journal.pmed.1002915.t001
Consultation on Advances in Clinical Trials Design for TB Treatment Regimen proposed features and designs that could address this need in greater detail and that are described in relevant papers of this Collection [2, 20].

Regulatory agencies rightfully seek to use conservative approaches to endpoint evaluation, relying upon the protection from bias provided by randomization. For certain diseases, including MDR-TB, the expedited approval pathway can be used based on a surrogate or an intermediate clinical endpoint that is reasonably likely to predict a clinical benefit. These endpoints, however, are not fit-for-purpose for programmatic and policy needs. Whereas intensive efforts are underway to identify improved intermediate surrogate markers of treatment outcome with the ability to measure and describe accurately the effect an experimental regimen will likely have on achieving nonrelapsing cure [21, 22], no marker has yet been identified that fully serves the needs of TB investigators and regulators, let alone policy makers [23]. The desire for an equivalent to the viral load in HIV and viral hepatitis trials has been often voiced but not yet attained, and current efforts are directed toward identification of markers that might reliably predict efficacy. In addition, combination of bacterial (e.g., minimum inhibitory concentration [MIC]) and host (e.g., pharmacokinetic characteristics, adherence, and perhaps genetic or other features) factors would be of value in dose selection and for predicting outcome [24, 25]. Relevant surrogate markers providing highly reliable estimates of treatment outcome, once realized, could provide sufficient evidence for guideline development beyond market approval [4], but until then, the TB therapeutics field has to look to novel trial designs, long-term endpoint definitions, and other trial features as a means to generating data pertinent to policy decisions [3].

The “composite” clinical trial endpoint (comprising multiple events such as a combination of failure, relapse, and death) has been used as a mechanism to capture multiple serious outcomes of interest with a programmatic perspective, often allowing for smaller sample sizes. The use of composite endpoints, however, poses some problems, the most significant being that respective endpoints are of differing individual and public health value (i.e., death is always a worse outcome than any other). Further, there are often varying levels of certainty around different endpoints (for example, cause of death is often uncertain in trials performed in low-resource settings). The choice of the components of a composite endpoint should be made carefully: because the occurrence of any one of the individual components is considered to be an endpoint event, each of the components is of equal importance in the analysis of the composite [26]. For these reasons, when composite outcomes are used, it is essential that information on all their components be collected in such a way that they can be disaggregated and individually reported. As an illustration, endpoints of currently conducted Phase II and Phase III trials of TB drugs or regimens are shown in Table 2.

Noninferiority (NI) design has become the design of choice in most Phase II and Phase III trials of new TB drugs and regimens over the last decade, either because of the high efficacy of the control regimens (as in drug-susceptible TB) or because of the interest in shortening treatment (as in the case of DR-TB). NI trial designs, however, pose a number of methodological questions, particularly in terms of analysis [27]. In NI trial designs, different analysis populations are of interest—the effect in all randomized patients and the effect in those who can adhere to treatment, which have historically been estimated using the intention-to-treat (ITT) and the per protocol (PP) populations, respectively [28]. The ITT principle allows virtually all patients to contribute information to the primary trial analysis. In this approach, all randomized patients are included in the analysis of results, and favorable status is assigned only to those patients whose favorable outcome is documented; all others are deemed unfavorable or nonassessable (including those lost to follow-up, those whose therapy is altered, those who die or withdraw early, etc.). The PP population, conversely, is composed of those randomized and
Table 2. Recent and current Phase II and Phase III trials of new TB drugs or regimens, with their respective endpoints. (Trial names shown with a blue background involve DS TB; those with a gray background involve DR-TB).

<table>
<thead>
<tr>
<th>Phase II trials</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Trial name (registration no.)</strong></td>
</tr>
<tr>
<td>APT (NCT02256696)</td>
</tr>
<tr>
<td>HIGHRIF-1 Extension</td>
</tr>
<tr>
<td>Janssen C211 (NCT02354014)</td>
</tr>
<tr>
<td>NC-005 (NCT02193776)</td>
</tr>
<tr>
<td>OPTI-Q (NCT01918397)</td>
</tr>
<tr>
<td>Stage 2 STEP</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Phase II/III trials</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Trial name (registration no.)</strong></td>
</tr>
<tr>
<td>NC-008 SimpliciTB (DS) (NCT03338621)</td>
</tr>
<tr>
<td>NC-008 SimpliciTB (DR) (NCT03338621)</td>
</tr>
<tr>
<td>NExT (NCT02454205)</td>
</tr>
<tr>
<td>TB-PRACTECAL (NCT02589782)</td>
</tr>
</tbody>
</table>

(Continued)
<table>
<thead>
<tr>
<th>Trial name (registration no.)</th>
<th>Phase</th>
<th>Sample size</th>
<th>Study groups; +/- dates; locations; sponsor</th>
<th>Primary efficacy endpoint (per online registration)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRUNCATE-TB (NCT03474198)</td>
<td>2/3</td>
<td>900</td>
<td>2 months various new regimens versus standard 6 months; regimens including H + R35 + Z + E + Lzd, H + R35 + Z + E + Cfz, H + Z + Rpt + Lzd + Lfx, H + Z + E + Lzd + Bdq. Opened late 2017, results expected 2021; MAMS adaptive trial design. Thailand, Indonesia, Philippines, Singapore; BMRC, NUS</td>
<td>• Unsatisfactory clinical outcome at week 96 after randomization (active TB, TB treatment, or death)</td>
</tr>
<tr>
<td>MDR-END (NCT02619994)</td>
<td>2</td>
<td>238</td>
<td>9 or 12 months Del + Lfx (750 or 1,000 mg) + Lzd (600 mg daily for 2 months, 300 mg daily thereafter) + Z, versus local regimen Opened January 2016, results December 2019; Korea</td>
<td>• Treatment success 24 months after start of treatment (both &quot;cured&quot; and &quot;treatment completed&quot;)</td>
</tr>
</tbody>
</table>

### Phase III trials

<table>
<thead>
<tr>
<th>Trial name (registration no.)</th>
<th>Phase</th>
<th>Sample size</th>
<th>Study groups; +/- dates; locations; sponsor</th>
<th>Primary efficacy endpoint (per online registration)</th>
</tr>
</thead>
<tbody>
<tr>
<td>endTB (NCT02754765)</td>
<td>3</td>
<td>324</td>
<td>9 months Bdq + Lzd + M + Z daily, 9 months Bdq + Lzd + Cfz + Lfx + Z daily, 9 months Bdq + Lzd + Del + Lfx + Z, 9 months Del + Lzd + Cfz + Lfx + Z, or 9 months Del + Cfz + M + Z, versus local regimen Opened December 2016, results September 2020; Georgia, Kazakhstan, Kyrgyzstan, Lesotho, Peru; MSF, Partners in Health</td>
<td>• Proportion favorable at week 73 (not unfavorable, and culture negative at week 65–73, or earlier negative culture and no other evidence of unfavorable) • In addition, a companion phase 3 trial will be launched in drug-resistant TB patients, the &quot;end TB-Q&quot; trial (NCT03896685). This trial compares 6 months or 10 months of daily Bdq, Del, Lzd and Cfz versus WHO standard of care in DR patients with fluoroquinolone resistance.</td>
</tr>
<tr>
<td>Otsuka Trial 213 (NCT01424670)</td>
<td>3</td>
<td>511</td>
<td>2 months Del (100 mg twice daily) and 4 months Del (200 mg daily) plus OBR versus 6 months placebo plus OBR Opened September 2011, completed June 2016, preliminary findings presented at IUATLD October 2017, results published 2019; Otsuka</td>
<td>• Time to SCC, i.e., distribution of the time to SCC during the 6 months of study drug treatment</td>
</tr>
<tr>
<td>NC-006 STAND-DS (NCT02342886)</td>
<td>3</td>
<td>271 (orig 1,200)</td>
<td>4 months Pret (100 mg twice daily or 200 mg once daily) + M + Z daily, or 6 months Pret (100 mg twice daily) + M + Z daily, or 6 months Pret (200 mg once daily) + M + Z daily, versus standard 6-month therapy Opened February 2015, paused October 2016–May 2017; accrual not resumed; TB Alliance</td>
<td>• Incidence of combined bacteriologic failure or relapse, or clinical failure, at 12 months from start of therapy</td>
</tr>
<tr>
<td>NC-006 STAND-DR (NCT02342886)</td>
<td>3</td>
<td>13 (orig 300)</td>
<td>6 months Pret (200 mg) + M + Z daily, single-arm study Opened February 2015, paused October 2016–May 2017, accrual not resumed; TB Alliance</td>
<td>• Incidence of combined bacteriologic failure or relapse, or clinical failure, at 12 months from start of therapy</td>
</tr>
<tr>
<td>NiX-TB (NCT02333799)</td>
<td>3</td>
<td>109 (orig 300)</td>
<td>6 months Bdq (200 mg daily for 2 weeks and then 200 mg three times weekly) + Pret (200 mg daily) + Lzd (600 mg twice daily), single-arm study Opened March 2015, preliminary findings presented at CROI, 2017, accrual closed November 2017, with opening of NC-007 ZeNiX trial; TB Alliance</td>
<td>• Incidence of bacteriologic failure or relapse or clinical failure through follow-up until 6 months after the end of (6–9) months treatment</td>
</tr>
<tr>
<td>NC-007 ZeNiX (NCT03086486)</td>
<td>3</td>
<td>180</td>
<td>2 or 6 months Lzd (600 or 1,200 mg daily, double-blind) + Bdq (200 mg daily for 2 weeks, then 100 mg daily) + Pret (200mg daily) Opened November, 2017, results January, 2021; TB Alliance</td>
<td>• Incidence of bacteriologic failure or relapse or clinical failure through follow-up until 26 weeks after the end of treatment; culture conversion requires at least two consecutive culture negative/positive samples at least 7 days apart</td>
</tr>
<tr>
<td>RIFASHORT (NCT02581927)</td>
<td>3</td>
<td>800</td>
<td>2 months H + R (1,200 or 1,800 mg) + Z + E daily and 2 months H + R (1,200 or 1,800 mg) daily, versus standard 6-month therapy Opened February, 2017, results expected January, 2020; St George’s London, INTERTB</td>
<td>• Combined rate of failure and relapse 12 months after end of treatment in mITT • Grade 3–4 AEs</td>
</tr>
</tbody>
</table>

(Continued)
otherwise eligible participants who complete the trial without significant deviation from the intended trial behavior; in particular, such participants typically satisfy minimal requirements for adherence to the trial interventions. Analysis with each of these two populations should lead to similar conclusions for a robust interpretation [29]. The ICH E9 Guideline further specifies that “any differences between them can be the subject of explicit discussion and interpretation” [30]. This concern arises in part from the recognition that adherent participants differ in unknown ways from those who are not adherent, as they may have more favorable outcomes, no matter what their randomized therapy [31]. The analyses of these trials are most robust when there is a high level of adherence, as inadequate therapy in all trial arms may lead to equally poor performance across arms and nonadherers are imputed as treatment failures in the analysis of all randomized patients, risking creating a false conclusion of NI. Consequently, it is extremely important that trial protocols encourage a high level of adherence.

Finally, the generalizability of findings from preapproval clinical trials to the different populations and areas of interest to policy makers is also a significant concern. Some populations

Table 2. (Continued)

<table>
<thead>
<tr>
<th>Trial ID</th>
<th>Sample Size</th>
<th>Treatment Details</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>SHINE (ISRCTN63579542)</td>
<td>3</td>
<td>1,200 (ped minimal disease) 2 months H + R (600 mg) + Z + (in some) E daily, and Z, and (in some) E daily, and 2 months H + R (600 mg) daily versus standard 6-month therapy Opened third quarter of 2016, results 2020; treatment-shortening strategy trial for children with minimal TB; India, Uganda, South Africa, Zambia; BMRC</td>
<td>• Unfavorable outcome (failure, relapse, death) • Grade 3–4 AEs</td>
</tr>
<tr>
<td>STREAM Stage-1 (ISRCTN78372190)</td>
<td>3</td>
<td>424 4 months daily M + Cff + Z + E + high-dose H + kanamycin (daily for 3 months and then 3 times per week) + prothionamide, and 5 months of M + Cff + Z + E daily, versus local standard Opened 2012, closed to accrual June 2015, preliminary findings presented at IUATLD October 2017, results early 2019; IUATLD, MRC, USAID</td>
<td>• Proportion of patients with a favorable outcome 132 weeks after randomization having not previously had an unfavorable outcome or been retreated</td>
</tr>
<tr>
<td>STREAM Stage-2 (NCT02409290, ISRCTN18148631)</td>
<td>3</td>
<td>1,155 9 months M + Cff + E + Z daily, with initial 2 months of high-dose H + kanamycin + prothionamide daily, or 9 months Bdq + Cff + E + Lfz + Z daily, with initial 2 months high-dose H + prothionamide daily (all oral), or 6 months Bdq + Cff + Lfz + Z daily with initial 2 months high-dose H and kanamycin versus 20–24 month local regimen Opened April 2016, results expected April 2021; IUATLD, MRC, USAID, TB Alliance</td>
<td>• Proportion of patients with a favorable outcome at week 76 (noninferiority margin 10%)</td>
</tr>
<tr>
<td>TBTC 31/A5349 (NCT02410772)</td>
<td>3</td>
<td>2,500 2 months H + Rpt (1,200 mg) + Z + E daily, and 2 months H + Rpt (1,200 mg) daily, or 2 months H + Rpt (1,200 mg) + Z + M daily, and 2 months H + Rpt (1,200 mg) + M daily versus standard 6-month therapy Opened January 2016; results 2020; substudies include interactions of Rpt and efavirenz, intensive PK and pharmacodynamics of Rpt, and sputum biomarkers to predict outcomes; CDC TBTC, ACTG</td>
<td>• TB disease-free survival at 12 months after assignment • Proportion of participants with grade 3–5 AEs during treatment</td>
</tr>
</tbody>
</table>

Adapted from Tiberi and colleagues [6].

Abbreviations: AE, adverse event; Bdq, bedaquiline; BMRC, British Medical Research Council; Cff, clofazimine; CDC, Centers for Disease Control; CROI, Conference on Retroviruses and Opportunistic Infections; Del, delamanid; DS, drug-sensitive; DR, drug-resistant; E, ethambutol; EBA, early bactericidal activity; H, isoniazid; HIV, human immunodeficiency virus; IUATLD, International Union Against Tuberculosis and Lung Diseases; Lzd, linezolid; Lfz, levofloxacin; MGIT, mycobacterial growth in-tube; MSF, Médecins Sans Frontiers; M, moxifloxacin; mITT, modified intent-to-treat; NCT, identifying registration number on www.ClinicalTrials.gov; NIAID, National Institute of Allergy and Infectious Diseases; NUS, National University of Singapore; OBR, optimized background regimen; orig, originally; ped, pediatric; PK, pharmacokinetics; Pret, pretomanid; R, rifampin 10 mg/kg; R35, rifampin at 35 mg/kg; Rpt, rifapentine; SCC, sputum culture conversion; TB, tuberculosis; TBTC, TB Trials Consortium; TTP, time to positivity; USAID, US Development Aid Agency; Z, pyrazinamide.

https://doi.org/10.1371/journal.pmed.1002915.1002
may be underrepresented in clinical trials conducted for approvals (e.g., children, elderly people, pregnant women, persons with advanced comorbid illness), whereas others are excluded for reasons of feasibility (e.g., those living far away from a clinic or deemed unreliable for follow-up). Significant problems have arisen from the assumption of generalizability [32]. When a successful trial establishes the efficacy of a new agent or regimen, efforts are then needed to expand exploration of the regimens in broader populations, or through additional pragmatic trials, such as the endTB trial [33]. The need for such trials is unlikely to be addressed through any innovations in design, but the rationale for excluding special populations even from early and middle phases of development is currently being revisited in the TB therapeutics field [2, 3, 20].

The link between registration and public health recommendations: Implications for national TB programs and the way forward

For TB program managers and policy makers at the country level, the successful registration of a candidate drug is only one component of the decision-making process around adoption and use. Feasibility, acceptability, resource use, equity, and quality of life are also considered when formulating public health recommendations, and these rely on qualitative data that need to be collected in parallel to quantitative assessment of evidence.

WHO guidelines are key for the development of national policies for the care of TB patients. However, when reliable data are lacking, recommendations are predominantly based on low or very low certainty in the evidence, which creates challenges for the potential rapid adoption, successful implementation, and subsequent uptake of the new therapies—as has been the case with the treatment of DR-TB [34, 35]. Moreover, recommendations, even if based on low or very low certainty in the evidence, will often create the perception of a new “standard of care” that subsequently complicates the ability to fund and conduct pragmatic trials that would address the uncertainty left by the lack of data. Policy makers, donors, and ethical review bodies should be aware that significant uncertainty persists when recommendations based on very low or low certainty are adopted and that further research is essential to test the merits of the new standard of care proposed. Such additional research can generate postlicensure data that are important for the update of policies, as in the case of the recent WHO DR-TB treatment guidelines [18, 36] (Table 3).

Drug and regimen developers already have formal mechanisms of communication with regulators, but the engagement of policy recommendation institutions should be actively encouraged and pursued as early as possible at design stages. One example of the value of such communication relates to the definition of outcomes selected for trials. Discussions with regulatory authorities usually identify endpoints that address foundations of efficacy, safety, and tolerability in studies with shorter follow-up duration; however, these outcomes may not provide adequate information for guideline developers and policy makers to endorse a given drug for use in regimens. Integration of long-term outcomes into TB trials as much as is feasible, along with the standardization of outcomes with shorter follow-up duration; however, these outcomes may not provide adequate information for guideline developers and policy makers to endorse a given drug for use in regimens. Integration of long-term outcomes into TB trials as much as is feasible, along with the standardization of outcomes, should be a top priority for the TB therapeutics field, using, for example, the novel Phase IIC design, wherein follow-up is extended and the experimental regimens are used for their intended total duration [37].

Finally, standardized data collection and outcome definitions compatible with the Clinical Data Interchange Standards Consortium (CDISC) platforms are required by regulatory bodies. These have enhanced the ability to optimally use GRADE-based methodological approaches to evaluating the evidence, and should be similarly considered by policy makers. The application of such data standards to cohorts and the collection of national TB program data would be an invaluable step forward by allowing real-world data analyses that will greatly inform policy
decisions. Until then, TB clinical trialists and regimen developers are strongly encouraged to share individual patient–level data with policy makers to permit meta-analytic data synthesis approaches to be used in the GRADE methodology [38]. Data sharing in the domain of TB is a matter of global public good, and funders, donors, and implementers of trials should not only mandate such expectations for their clinical trials but also allocate funding to support the careful curation of data accessible to the public and to policy makers for future analyses.

Table 3. The interplay between trials and guidelines: Review of the proposals arising from WHO Technical Consultation on Advances in clinical trial design for development of new TB treatments (adapted from WHO [20]).

<table>
<thead>
<tr>
<th>Issue</th>
<th>Expert consensus</th>
<th>To be explored</th>
<th>Research gaps</th>
</tr>
</thead>
<tbody>
<tr>
<td>What clinical trial outcomes are required to inform regulatory and</td>
<td>A single clinical trial cannot address all relevant regulatory and policy/public</td>
<td>Consider postauthorization studies to answer some of the questions that cannot</td>
<td>Treatment success outcomes in recent trials of MDR-TB were much higher than</td>
</tr>
<tr>
<td>programmatic decision-making and need to be prioritized for</td>
<td>health questions. Explanatory trials, novel adaptive trials, pivotal trials for</td>
<td>be addressed in the registrational trial(s) to help bridge gaps in knowledge.</td>
<td>that reported in prior trials and across program settings. Further research is</td>
</tr>
<tr>
<td>prospective implementation in novel trial designs?</td>
<td>licensure need to be followed up with pragmatic trials to understand the optimal</td>
<td></td>
<td>needed to better understand the performance of the standard of care for</td>
</tr>
<tr>
<td></td>
<td>use of new drugs and regimens.</td>
<td></td>
<td>rifampicin-susceptible and rifampicin-resistant TB in various conditions and</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>settings.</td>
</tr>
<tr>
<td>How can current/novel clinical trial endpoints that are intended to</td>
<td></td>
<td>Operational research can help to translate clinical trial outcomes into WHO</td>
<td></td>
</tr>
<tr>
<td>support regulatory decisions be subsequently translated to support</td>
<td></td>
<td>guidance and add evidence for better programmatic implementation. Often,</td>
<td></td>
</tr>
<tr>
<td>programmatic implementation?</td>
<td></td>
<td>patients enrolled in trials are not reflective of the general population;</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>consider ways to make trial population more reflective of the population of</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>patients who will be receiving treatment in real life. Also consider pragmatic</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>studies for better evidence on programmatic implementation.</td>
<td></td>
</tr>
<tr>
<td>Should the assessment of clinical trial outcomes be updated for</td>
<td>Communication between drug/regimen developers, regulators, and recommendation</td>
<td>Approaches to collecting clinical outcomes data that can potentially address</td>
<td></td>
</tr>
<tr>
<td>harmonization across regulatory and programmatic objectives, and if</td>
<td>bodies is essential and should be encouraged and facilitated as early as</td>
<td>assessment of safety and efficacy of the product and answer questions that are</td>
<td></td>
</tr>
<tr>
<td>yes, how?</td>
<td>possible at design stages.</td>
<td>important from a programmatic perspective should address the following:</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• secondary/exploratory analyses are an option—but caution in overinterpreting</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>the data</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• sample size implications if multiple primary analyses considered</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• importance of prespecifying analyses; consistent definitions across different</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>trials are needed; limitations of using surrogate endpoints (e.g., 2-month</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>culture conversion) for development of guidelines.</td>
<td></td>
</tr>
<tr>
<td>How to ensure that trial data at the individual-patient level can be</td>
<td>Data should be collected using standard definitions, and use of data standards</td>
<td>As data quality improves, recommendations based on lower-quality data should be</td>
<td></td>
</tr>
<tr>
<td>pooled for enhanced meta-analysis when reviewing evidence for policy</td>
<td>for clinical trial is essential. Clinical trial data should be made available</td>
<td>reexamined. A relevant process to address this should be established.</td>
<td></td>
</tr>
<tr>
<td>making by WHO and other professional bodies</td>
<td>for sharing so as to conduct individual patient–level data analyses. Such</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>databases are used by WHO and other recommending bodies for policy development.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GRADE method should be well understood by all stakeholders</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: GRADE, Grading of Recommendations Assessment, Development, and Evaluation; MDR, multidrug-resistant; TB, tuberculosis; WHO, World Health Organization

https://doi.org/10.1371/journal.pmed.1002915.t003
Conclusion

Given the recent enthusiasm for pursuing novel trial designs in TB therapeutics [37, 39], more interactions will be needed between researchers responsible for designing the next generation of TB trials, regulators, and policy makers. This will allow better harmonization across the research pipeline and subsequent policies on access to TB medicines. Further, stakeholders, including donors and funders, need to acknowledge that both explanatory and pragmatic trials are needed to answer questions about efficacy and safety (explanatory) as well as expected effectiveness in programmatic conditions (pragmatic). In all cases, endpoints should be specific to the purposes. Late-phase clinical trial outputs that serve the objective of registration of a new TB drug or regimen can indeed meet the needs for development of public health guidelines, provided that data on long-term, patient-relevant, and population-relevant outcomes are being collected. Additionally, public health factors such as feasibility, acceptability, resource use, equity, and quality of life should be part of data collections, as these are necessary when formulating public health recommendations. The existing dialogue between drug developers and regulators should be expanded to policy makers under formal mechanisms of consultation, such as the one offered by WHO Task Forces [19]. More effective input from policy makers could greatly streamline and strengthen the value of TB clinical trial data in clinical settings. Such interactions with policy makers can be invaluable at the design stages and would result in better harmonization between the research pipeline and policies on access to TB medicines. The broad discussions that we propose would also ensure that secondary pooled analyses performed by WHO (or other policy-recommending bodies) are reliable and that the risk of conflicting interpretation and messaging provided by investigators and policy makers is reduced and usefully contribute to the generation of reliable and relevant data for further policy guidance on the treatment of all forms of TB [2].

Acknowledgments

The authors would like to thank Michael J. Vjecha, MD, for his critical review of Table 2. The views expressed in this article are the personal views of the authors and may not be understood or quoted as being made on behalf of, or reflecting the position of the regulatory agency/agencies or organisations with which the authors are employed/affiliated.

References


COLLECTION REVIEW

Advances in clinical trial design for development of new TB treatments—Translating international tuberculosis treatment guidelines into national strategic plans: Experiences from Belarus, South Africa, and Vietnam

Grania Brigden1*, Nguyen Viet Nhung2, Alena Skrahina3, Norbert Ndjeka4, Dennis Falzon5, Matteo Zignol5

1 Department of Tuberculosis, The International Union Against TB and Lung Disease, Geneva, Switzerland, 2 National Lung Hospital, Vietnam NTP, Vietnam, 3 Republican Scientific and Practical Centre for Pulmonology and TB, Minsk, Belarus, 4 Drug-Resistant TB, TB and HIV directorate, National Department of Health, Pretoria, South Africa, 5 Global Tuberculosis Programme, World Health Organization, Geneva, Switzerland

* grania.brigden@theunion.org

Summary points

• The World Health Organization (WHO) plays an important role in setting global norms and standards with a focus on public health and publishes international guidelines regularly to support Member States, particularly ministries of health, in the provision of the highest standard of healthcare.

• Over the last 5 years, multiple advances in diagnosis and treatment of tuberculosis (TB) have resulted in a number of new WHO guidelines for TB care, but these recent guidelines have not always been implemented in a timely fashion, raising issues in their adoption and scale-up at country level.

• We discuss the experiences of three countries with a high burden of multidrug-resistant TB (MDR-TB)—Belarus, South Africa, and Vietnam—in implementing recent WHO guidelines on bedaquiline, a drug recently registered and recommended for the treatment of MDR-TB and the standardised shorter treatment regimen (STR) for MDR-TB.

• The process of adopting and implementing new guidelines requires national TB programmes (NTPs) to interact with multiple agencies: both intergovernmental departments and external agencies such as regulators and donors. These processes are country specific, but there are some generalised challenges that NTPs in high-burden countries experienced when implementing recent WHO MDR-TB guidance.

• With multiple trials of new regimens for MDR-TB and new classes of drugs in the clinical treatment pipeline, the frequency of new guidelines for TB is expected to increase, and it is important to support NTPs to implement and scale-up these new developments in treatment.
Introduction

One of the key missions of national tuberculosis (TB) programmes (NTPs) is to issue policy and technical guidance for clinicians and healthcare workers involved in TB care at the country level. These national policies are generally developed based on international public health guidelines, such as those issued by the World Health Organization (WHO) [1, 2].

Updating national policies or technical guidelines in view of recent advances in TB diagnosis, care, and prevention has an important impact on TB patients, the health system, the community and is key to ensuring the best quality of care for people with TB.

WHO has a mandate to provide technical assistance to its Member States on different aspects of public health. The 13th General Programme of Work of WHO [3] outlines the organisation’s status as a science- and evidence-based agency setting global norms and standards, with a focus on public health. Translating research findings into policies may be a challenging task, given that the design of clinical studies may not always address the main public health priority directly, and recommended interventions require substantial adaptation to the particular programme conditions and settings [4].

In 2007, WHO established the Guideline Review Committee (GRC) to provide oversight to organisational efforts to ensure that policy guidance is up-to-date, trustworthy, feasible, and developed in a transparent way, in line with the highest international standards of care [5], and adheres to WHO principles for policy development [6]. The WHO-convened Guideline Development Group advises on the scope of the guidelines, assesses the quality of available evidence, and formulates recommendations using a systematic method termed Grading of Recommendations Assessment, Development, and Evaluation (GRADE) [7]. This approach requires experts who are formulating recommendations to base their judgements not only on trial evidence but also on other considerations, such as the balance of expected desirable and undesirable effects, equity, resource use, feasibility, and acceptability to the populations targeted by the guidance. These changes have contributed to an improvement in purpose, clarity, and the methodological quality of WHO guidelines in the last decade [7].

The pace of developments in new TB diagnostics, treatment, and patient support has increased substantially over the last decade, leading to the release of over 20 new or updated WHO guidelines on different aspects of TB care since 2010 [8]. This pace is expected to continue, and the PLOS Medicine Collection of which this paper is part [9] discusses the optimal characteristics of clinical trial designs to inform future policy guidance for new TB regimens.

Already in the last 5 years, NTPs have had to respond to a number of WHO policy updates on multidrug-resistant TB (MDR-TB) treatment as new medicines became available and results from studies on the use of novel drugs and the standardised shorter treatment regimen (STR) were communicated (e.g., bedaquiline, delamanid, and the 9–12-month-shorter MDR-TB regimen) [10–16]. Partly as a result of these rapid changes, a number of these new treatment policies have not been adopted or fully implemented by national programmes. A recent review [17] of national policies in 29 countries highlighted national policy gaps when compared to WHO policies. Thus, in the case of WHO’s recommended 9–12-month-shorter MDR-TB regimen, 45% of the countries had developed policies, but only 69% of those countries had implemented them. By the end of 2017, 62 countries, mostly in Africa and Asia, reported having used shorter MDR-TB regimens; between 2016 and 2017, the number of patients reported to have been started on the 9–12-month-shorter regimen globally increased from 2,400 to 10,000 [18]. With regard to the new drugs, bedaquiline and delamanid, 86% of countries had a policy on bedaquiline and 67% on delamanid, but the actual use of the new drugs reflected the implementation gap, with only 12,194 and 976 treatment courses procured globally for bedaquiline and delamanid, respectively, in 2017 [19].
There are multiple barriers to the adoption of international treatment guidelines, including factors relating to the acceptability and perceived feasibility of the recommendation, the individual opinion of clinicians, patient preferences, regulatory processes for new drugs, requirement for new resources, and the financial and political commitment from the Ministry of Health (MOH) [20].

The following case studies from the NTPs of three high-burden countries refer to national experiences in the introduction of new drugs and regimens for MDR-TB to illustrate how countries approached implementation of new policies for TB treatment. Belarus, South Africa, and Vietnam are all on WHO’s high-burden MDR-TB list but with different epidemic patterns (see Table 1). The case studies review the experiences of the countries in implementing the interim guidance for the use of bedaquiline in the treatment of MDR-TB, issued by WHO in 2013 [21], and the revised guidelines on treatment of MDR-TB issued in 2016 that recommend the use of the 9–12-month-shorter MDR-TB regimen under certain conditions [13].

Implementation of bedaquiline in Belarus

In 2017, there were an estimated 3,500 new TB cases in Belarus of which 2,500 had rifampicin resistance or MDR-TB [18]. In 2012, in anticipation of the approval of a new drug for TB, WHO released a handbook to advise countries on how to organise both spontaneous and active pharmacovigilance [22]. The national pharmacovigilance centre of the Belarus MOH, with its prior experience in active pharmacovigilance in the country for antiretrovirals [23], established strong links with the NTP to enhance pharmacovigilance among MDR-TB patients. The implementation of cohort event monitoring for MDR-TB treatment on regimens containing linezolid, and later bedaquiline, were labour-intensive activities for MOH staff, undertaken without additional resources [24] (Table 2).

In mid-2013, the national TB guidelines were updated in alignment with the new WHO policy on bedaquiline use (including translation into the Russian language) and staff training organised by the MOH under the guidance of the MDR-TB expert group (consilium). The MDR-TB consilium is a platform of multidisciplinary experts from Belarus with the aim to improve the quality of diagnosis and care and to reduce the time to initiation of effective MDR-TB treatment throughout the country. The NTP also benefited from reviews of its work

Table 1. Overview of the TB epidemic in Belarus, South Africa, and Vietnam [18].

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Belarus</th>
<th>South Africa</th>
<th>Vietnam</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total population (2017)</td>
<td>9.5 million</td>
<td>57 million</td>
<td>96 million</td>
</tr>
<tr>
<td>TB prevalence (all forms)</td>
<td>32.1 per 100,000</td>
<td>398.6 per 100,000</td>
<td>110.1 per 100,000</td>
</tr>
<tr>
<td>TB incidence (new and relapse cases)</td>
<td>29.3 per 100,000</td>
<td>386.3 per 100,000</td>
<td>107.0 per 100,000</td>
</tr>
<tr>
<td>HIV prevalence among TB</td>
<td>2.9 per 100,000</td>
<td>340 per 100,000</td>
<td>4.7 per 100,000</td>
</tr>
<tr>
<td>Incidence of MDR/RR-TB</td>
<td>26 per 100,000</td>
<td>25 per 100,000</td>
<td>7.4 per 100,000</td>
</tr>
<tr>
<td>Percent of new cases with MDR/RR-TB</td>
<td>38% (36–41)</td>
<td>3.4% (2.5–4.3)</td>
<td>4.1% (2.8–5.7)</td>
</tr>
<tr>
<td>Percent of retreated cases with MDR/RR-TB</td>
<td>67% (63–70)</td>
<td>7.1% (4.8–9.5)</td>
<td>17% (17–18)</td>
</tr>
<tr>
<td>TB treatment coverage</td>
<td>80%</td>
<td>68%</td>
<td>83%</td>
</tr>
<tr>
<td>MDR/RR-TB treatment success rate (2015)</td>
<td>64% (cohort size: 1,400)</td>
<td>55% (cohort size: 9,750)</td>
<td>74% (cohort size: 2,045)</td>
</tr>
<tr>
<td>XDR-TB treatment success rate (2015)</td>
<td>53% (cohort size: 508)</td>
<td>48% (cohort size: 427)</td>
<td>Not reported</td>
</tr>
</tbody>
</table>

Abbreviations: MDR/RR-TB, multidrug-resistant/rifampicin-resistant TB; TB, tuberculosis; XDR-TB, extensively drug-resistant TB
by WHO, Médecins sans Frontières, the Supranational Reference Laboratory, and other external experts. Measures were taken by the Council of Ministers to waive the requirements for drug registration for bedaquiline. The same mechanism was used subsequently to permit the use of other medicines, including clofazimine and delamanid. The support of the Global Fund was critical in securing resources to purchase bedaquiline and the companion medicines. By October 2018, 543 patients had started treatment with bedaquiline. In 2018, Belarus reported individual case-based data from programmatic cohorts of patients treated with bedaquiline-containing regimens to the pooled analysis for the latest update of WHO’s MDR-TB treatment guidelines [25, 26]. An important challenge faced by the MOH when implementing bedaquiline was for healthcare staff to adhere to proper criteria when selecting patients to be placed on regimens including this new agent. The MDR-TB expert consilium played an important role to ensure compliance. Another limitation was to have all the medicines needed for the regimen available at the time of start of treatment: this required coordination with all stakeholders (i.e., funders, logistics, facilities) to limit delays. The WHO-recommended 9–12 month STR MDR-TB regimen in Belarus is contraindicated in many because MDR-TB patients commonly have strains harbouring additional resistance to pyrazinamide and to key second-line drugs such as fluoroquinolones and injectable agents. This is why the focus has been on scaling up the use of bedaquiline, with other second-line drugs that have not been previously used in Belarus. Since late 2018, the NTP introduced under operational research conditions a shorter regimen of 9 months consisting of all group A and B medicines recommended in MDR-TB regimens.

In 2015, following WHO advice on active TB drug safety monitoring and management (aDSM) in patients treated with novel regimens and repurposed medicines [27], Belarus became an early adopter of aDSM as a standard of care and among the first countries to contribute records to WHO’s global aDSM database [28]. Using domestic and external funding, the Belarus MOH is updating the national electronic TB patient register to enhance future data management.

### Table 2. Key milestones in the successful introduction of new medicines for MDR-TB patients, Belarus.

<table>
<thead>
<tr>
<th>Actions to strengthen MDR-TB treatment</th>
<th>Actions to strengthen patient safety</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Aligning national TB guidelines to WHO recommendations</td>
<td>• 2012: Links cultivated between national TB programme and the NPV</td>
</tr>
<tr>
<td>• Training of clinical staff in the new policies and in case management</td>
<td>• 2012: NPV strengthens its methods for both spontaneous reporting and for active surveillance (using CEM)</td>
</tr>
<tr>
<td>• Piloting and subsequent scale-up of video-supported therapy as an adjunct to patient-centred care</td>
<td>• 2013: CEM for antiretroviral treatment started</td>
</tr>
<tr>
<td>• Strengthening of laboratory capacity to detect drug resistance using newer techniques and to perform increasing volumes of culture</td>
<td>• 2013: CEM for antiretroviral treatment extended to patients with HIV who had MDR-TB</td>
</tr>
<tr>
<td>• Changes in the drug procurement system, including ministerial waiver for the importation of new medicines</td>
<td>• 2014: CEM for linezolid-containing regimens started in MDR-TB</td>
</tr>
<tr>
<td>• Updated national electronic TB register to include information on adverse events and details on regimen</td>
<td>• 2016: aDSM introduced for all MDR-TB patients on treatment</td>
</tr>
<tr>
<td>• Funding proposal to the Global Fund to provide financial resources</td>
<td>• 2017: aDSM data reported to global database</td>
</tr>
<tr>
<td>• Technical support provided by WHO and by Médecins sans Frontières</td>
<td>-------------------------------------------------------------------------------------------------------</td>
</tr>
</tbody>
</table>

Abbreviations: aDSM, active TB drug safety monitoring and management; CEM, cohort event monitoring; MDR-TB, multidrug-resistant TB; NPV, national pharmacovigilance centre; TB, tuberculosis; WHO, World Health Organization

https://doi.org/10.1371/journal.pmed.1002896.t002
The articulated response from the MOH, including strengthening the surveillance and preventive and curative components of the NTP [29], has resulted in high case detection of TB, TB/HIV, and drug-resistant TB and treatment success in new and relapsed TB patients approaching 90% [30].

Introducing bedaquiline in South Africa

South Africa is a country with high TB, MDR-TB, and HIV burden. The country contributes approximately 10% of global MDR-TB cases diagnosed and reported, with treatment success similar to the global rate at 54% and mortality at just above 20% [18].

The use of bedaquiline in the country started in December 2012, when the South Africa Medicines Control Council (MCC) approved the drug as part of a clinical access programme [31]. The programme was implemented at five sites and was later scaled up to 12 sites in 2014 after early successful results were obtained [32]. Once bedaquiline received full registration with the MCC, the inclusion criteria were broadened, and from 2017, bedaquiline use was decentralised to the district level to facilitate scale-up (Fig 1). In June 2018, South Africa announced that bedaquiline would be available to all eligible patients with rifampicin resistance, replacing the injectable agents in both the recent WHO-recommended longer treatment regimens as well as variants of the STR [26]. The STR has been included in national policies since 2015 [33], but similar to Belarus, the eligibility criteria for the STR have meant that its use has been limited in a population with increasingly complex resistance patterns. However, since September 2018, the South African NTP recommended a modified injectable-free STR nationwide. This regimen has the addition of linezolid for 2 months, with bedaquiline replacing the injectable agent and given for 6 months and levofloxacin replacing moxifloxacin [34].

The primary challenge to adoption and implementation of bedaquiline use has been the full regulatory approval required from the MCC, as the initial approval was only for a compassionate-use programme. The process to reach full regulatory approval took 18 months. Once registered, there was hesitancy of clinicians on the use of a new drug for which programmatic data were initially extremely limited. Subsequently, data were collected from pilot sites and published. A National Clinical Advisory Committee was formed to support implementation of WHO guidance by helping physicians design effective treatment regimens and establishing provincial committees to discuss difficult clinical cases. The NTP discussed WHO guidelines with local researchers and academia to ensure the guidance was customised to the national context and translated into practice. An additional challenge to the scale-up was maintaining a secure supply of stocks, particularly as bedaquiline was not on national tender.

Improving diagnosis and treatment of MDR-TB in Vietnam

Vietnam is one of the 20 countries considered to have both a high TB and MDR-TB burden [18]. In 2016, Vietnam had 106,527 registered cases of TB, and it is estimated that 20% of cases are not detected [18]. To address this problem, the NTP developed the 2X strategy (Xray-
Xpert MTB/RIF) to enhance early TB and MDR-TB detection. This strategy, in line with WHO guidance on the use of Xpert MTB/RIF [35–36] and chest radiography [37], aims to screen for and confirm TB infection and disease, including rifampicin resistance status, at the start of treatment.

The scale-up of newer diagnostics was coupled with a patient triage strategy with bedaquiline and the STR part of the strategy. As clofazimine, a key drug in the shorter regimen and a companion drug to bedaquiline, was not registered in the country, the NTP had to apply for an investigation study to be approved by the institutional review board of the MOH so as to allow importation of the drugs needed. Bedaquiline was introduced under import waiver in December 2015 with the shorter treatment regimen introduced in April 2016, in three pilot provinces, and with the implementation for the STR expanded to an additional eight provinces after 18 months [38]. The expansion occurred after WHO’s recommendations on the short-course regimen in 2017 [14]. During this stepwise scale-up of the use of bedaquiline and the STR, the scale-up was interrupted because of a 7-month interruption pending MOH approval of the expansion. During this time, the STR enrolment declined from 32% to 11%; and bedaquiline use in those eligible declined from 92% to 40% (Fig 2). Following these pilots, the STR was included in the national guidance in 2018 and is now a major treatment option for MDR-TB countrywide.

The long-term plan in Vietnam is to continue to scale up the use of bedaquiline. Based on local cohort studies, laboratory capacity was available to identify susceptibility of almost all drugs before indication of the regimen for individual patients, and the Vietnam NTP decided to apply modified STR as the primary regimen to treat drug-resistant TB. The planned stepwise scale-up of the modified shorter treatment regimen for drug-resistant TB treatment is shown in Fig 3. In order to overcome challenges regarding drug importation for bedaquiline, the drug has now been registered in 2019 for compassionate use while the main regulatory process is underway.

Policy change in Vietnam requires a stepwise approach, utilising pilot projects with scale-up happening over a 3–4-year timeline. At the same time of implementing pilot projects, the NTP negotiates in-country drug registration processes. The involvement of the WHO country office with technical assistance and support for policy change has helped to minimise delays in these processes.

**Discussion**

WHO guidance strives to make recommendations that are based on the best and latest available evidence and that have applicability to diverse settings worldwide. The use of standardised evaluation methods like GRADE aims to assess study findings in a rigorous way but also ensure that due considerations for implementation are being addressed. However, WHO’s guideline processes cannot consider the nuances and sensitivities of the local socioeconomic, regulatory, and cultural conditions—this is left to the NTP when reviewing the guidance. As shown in the case studies described here, translating the research findings underlying new WHO guidance into programmatic guidance incurs substantial logistical challenges and delays for NTPs to mobilise the necessary resources and negotiate the regulatory framework. As in the three country examples, the process of adapting the recent WHO guidance on bedaquiline to the national situation is a multistage process, involving actors outside the NTP, such as donors and regulatory authorities, and is prone to delays.

The case studies highlight the challenges of introducing a new drug, particularly one with limited data on effectiveness and no long-term outcome data. The NTPs had to complete the necessary ethical, surveillance, and regulatory processes, and often pilot projects had to be
Dec 2015-2016 | 2017
---|---
3 sites | 19 sites

Xpert Test done

20,646 | 51,262

RR-TB

1407 (6.8%) | 2,275 (4.4%)

SL-LPA test done

523 (37%) | 1244 (55%)

FQ and SLI susceptible

298 (57%) | 767 (62%)

FQ and/or SLI resistant

108 (21%) | 179 (14%)

FQ Res (pre-XDR) | SLI - Res (pre-XDR)

68 (13%) | 124 (10%)

21 (4%) | 27 (2%)

FQ and SLI Res (pre-XDR)

19 (4%) | 28 (2%)

STR Enrolled

100 (34%)* | 81 (11%)

Bdq/ITR enrolled

99 (92%)* | 71 (40%)

*2015-2016 enrolments were limited by the enrolment target of 100 patients
undertaken to obtain real-life experience in the country, delaying the scale-up of the new drug (see Fig 4).

At the same time as new drugs were recommended to be added to the longer individualised regimen, WHO recommended a shorter standardised regimen for certain types of MDR-TB. NTP managers and staff had to work out how to implement the new drugs into their programmes as well as into a new treatment regimen, and this often required collecting data on efficacy and safety of both a new drug and a new regimen. Similarly, they had to ensure necessary funding not only to support the policy change process but also to procure the new drugs and the components of the standardised regimen, implement robust aDSM, and organise technical assistance or training for implementing the new policies. This required consideration of either national or donor resources, further adding to the implementation timeline, particularly for low- and middle-income countries that rely on the Global Fund and other donors to support their MDR-TB programmes. The recent update to the MDR-TB guidelines continues to recommend this dual approach of longer individualised regimens and more standardised shorter regimens [39].

To ensure that these new developments reach all relevant at-risk groups, the NTP needs to further engage with the national Ministry of Justice, Ministry of Migration, or other specific ministries. In countries that have placed TB high on the political agenda—such as Belarus, South Africa, and Vietnam—support for this engagement with other ministries may be easier than for other countries whose NTP may not have the support to engage with other ministries and national processes.

![Fig 2. Patient triage approach in Vietnam. Bdq, bedaquiline; FQ, fluoroquinolone; ITR, individualised treatment regimen; Res, resistance; RR-TB, rifampicin-resistant tuberculosis; SLI, second-line injectable; SL-LPA, second-line–line probe assay; STR, standardised shorter treatment regimen; XDR, extensively drug resistant.](https://doi.org/10.1371/journal.pmed.1002896.g002)

![Fig 3. Scale-up plans for STR and Bdq in Vietnam. Bdq, bedaquiline; Dlm, delamanid; MDR TB, multidrug-resistant tuberculosis; Std, standard; STR, standardised shorter treatment regimen.](https://doi.org/10.1371/journal.pmed.1002896.g003)
This policy update process needs to be repeated with the latest WHO guidance on MDR-TB [40], which has a number of significant changes for the NTP to consider. Bedaquiline scale-up and use will continue, as now bedaquiline is a group A drug (group A drugs are drugs that are strongly recommended for inclusion in a longer MDR-TB regimen) and as such is a key component of the new all-oral individualised long regimen [26]. The STR remains in the recommendations with a change in the injectable agent being used. With the welcome push for an all-oral regimen for MDR-TB, NTPs may want to consider operational research into the role of oral alternatives to the injectable agent in the STR, as has been done in South Africa, Belarus, and Vietnam. With another new drug, pretomanid [18], submitted for registration, and new regimens being recommended for latent TB infection (LTBI), the lessons learned implementing new or unregistered drugs and new regimens for MDR-TB will aid NTPs to ensure these new developments are adopted and scaled up, potentially using the pathways used for bedaquiline and the STR uptake.

Conclusion
The experience of Belarus, South Africa, and Vietnam suggests that intergovernmental collaboration and new guideline adoption and implementation are facilitated when TB has been placed high on the political agenda, in contrast to other countries where TB maintains a much lower profile. The pathways and tools developed by NTPs to implement the new TB drugs and regimens for MDR-TB can help ensure that the latest WHO guidance on MDR-TB and LTBI can be implemented and scaled up quickly. With strengthened programmes (including implementation of aDSM), NTPs can generate the evidence to show whether new drugs and regimens found to be effective in clinical trials will work in populations that need them most [40].

With the TB drug and regimen pipeline at its healthiest in over a decade, advances in all areas of TB care are expected in the next decade requiring national guidelines to adapt as a
priority. More updates to new guidance issued recently by WHO for the treatment of MDR-TB and LTBI are expected imminently as new drugs are submitted for registration, as well as results from new regimen studies being published in the coming years. A culture of change needs to be fostered and budgeted for and recognition needs to be given to countries that have supported their NTPs in this process. All actors in TB care, from international donors to national funding and regulatory agencies, need to support this approach to change, reacting promptly to and supporting new developments in TB therapeutics. The political attention to TB at the recent UN high-level meeting on TB [41] must be followed up with the appropriate funding and policy support so that NTPs are supported to rapidly review and adopt the best standard of care for people with TB. A systematic approach to evaluate how policies are used and adapted by countries and their impact—both as intended and inadvertent—would be a fruitful step in the feedback cycle that WHO and other professional bodies use when planning updates of new policy guidelines.

References


COLLECTION REVIEW

Advances in clinical trial design: Weaving tomorrow’s TB treatments

Christian Lienhardt1*, Andrew Nunn2, Richard Chaisson3, Andrew A. Vernon4, Matteo Zignol5, Payam Nahid6, Eric Delaporte1, Tereza Kasaeva5

1 Unité Mixte Internationale TransVIHMI, UMI 233 IRD–U1175 INSERM—Université de Montpellier, Institut de Recherche pour le Développement (IRD), Montpellier, France, 2 MRC Clinical Trials Unit at UCL, Institute of Clinical Trials and Methodology, University College London, London, United Kingdom, 3 Center for Tuberculosis Research, Johns Hopkins University School of Medicine, Baltimore, United States of America, 4 Division of TB Elimination, National Center for HIV, Viral Hepatitis, STD and TB Prevention, Centers for Disease Control and Prevention, Atlanta, Georgia, United States of America, 5 Global TB Programme, World Health Organization, Geneva, Switzerland, 6 UCSF Center for Tuberculosis and Division of Pulmonary and Critical Care Medicine, University of California San Francisco, San Francisco, California, United States of America

* christian.lienhardt@ird.fr

Summary points

- The development process for new tuberculosis (TB) regimens remains slow and costly. In this concluding paper of the PLOS Medicine Special Collection, we highlight the key suggestions made at a WHO Technical Consultation on “Advances in Clinical Trial Design for Development of New TB Treatments” held in 2018 to address this challenge.

- Pharmacokinetics and pharmacodynamics (PK-PD) properties of candidate drugs are critical for constructing effective combination regimens. Bridging PK-PD methods to the analysis of Phase II studies and integrating longitudinal culture results would help with clarifying dose–response relationships and to link drug exposure to bactericidal activity; this would provide valuable insights for the identification of the components of suitable regimens.

- New adaptive designs can accelerate Phase II and III trials and improve our ability to select regimens early for further investigation. Among these, the integration of extended posttreatment follow-up with collection of real-time treatment outcomes in the new hybrid Phase IIC design, with features drawn from both Phase II and Phase III trials, permits earlier identification of candidate regimens likely to succeed in Phase III.

- Once the efficacy of a regimen is demonstrated in high-quality randomized controlled trial(s), the assessment of its effectiveness under programmatic conditions may permit an estimation of the amount of nonadherence that could substantially alter the likelihood of treatment success.

- Vulnerable populations, such as children, pregnant women, and people with HIV infection, should be included in clinical trials from the outset, as these groups have unique characteristics regarding PK, safety, and efficacy, which necessitate special attention in drug and regimen development.

- We advocate here for a better systematization and harmonization of the approaches taken internationally to ensure that best practices and novel research designs are used to accelerate development of new TB regimens. By using all the creative approaches.
described in this Special Collection, we hope that the next generation of TB trials will bring the high-quality evidence for novel TB regimens that is required to meet the needs of the millions of new TB patients who become ill each year.

Introduction

In Homer’s Odyssey, each night, Penelope unwove the tapestries she made in the day to delay her marriage to one of the contenders for Ulysses’ throne while awaiting his return. In tuberculosis (TB) therapeutics, major advances 40 to 50 years ago established the current 6-month short-course chemotherapy (SCC) regimen that revolutionized TB treatment—but could we have made greater progress since [1,2]? Indeed, one can wonder if, over the past 40 years, like Penelope, the TB community has been weaving novel treatment regimens out of new and repurposed drugs, then unweaving them because of negative results and an unstructured strategy for advancing the field, despite an increasingly rich pipeline [3,4]. With new opportunities to test novel combinations to shorten TB treatment, a broad reflection on the way TB trials have been carried out over the past 40 years is legitimate, and in view of the many advances in microbiology, immunology, genetics, and pharmacology, it can help us draw from the lessons learnt to weave better TB treatments for tomorrow [5].

The most serious challenge we face in developing new TB therapeutics is our inability to identify optimal regimens early and efficiently. Limitations include the lack of direct measures of treatment response, unsatisfactory surrogate endpoints of treatment effects, and the lack of reliable predictors for Phase III clinical outcomes [6]. Identifying the optimal drug combinations and the most parsimonious trial designs to evaluate them requires critical insights incorporating recent developments in pharmacology, microbiology, biomarkers, and diagnostic assays. Given the long duration and high costs of medical development, and in view of the limited funding for TB research and development [7], it is crucially important to reassess the best practices for the development of the new TB treatment regimens of the future. In the current paper, we offer an assessment of challenges and dogma addressed by the WHO Technical Consultation on “Advances in Clinical Trial Design for Development of New TB Treatments” that took place in Glion-sur-Montreux, Switzerland, March 14–16, 2018 [8].

1. How can we bridge preclinical data into clinical trials and identify the pharmacokinetics and pharmacodynamics parameters that correlate best with bactericidal efficacy and toxicity in vivo?

The foundation of SCC rests on the evidence that complete sterilization of tuberculous lesions in the lungs requires at least 6 months of treatment [9,10] because of the presence of slow-growing or nonmultiplying bacilli, termed “persisters” [11]. The evidence is that these persisters are heterogeneous in nature, and their mechanism of formation results from multiple pathways [12]. Thus, although antibiotics are classically developed based on their activity against actively growing bacteria, drugs that kill the slowly or nonreplicating bacilli, like rifampicin, are critically important to shorten TB chemotherapy while retaining high efficacy [13].

Despite promising data from mouse models [14] and human studies of 2-month culture conversion rates [15,16] suggesting a potential for treatment shortening, 3 independent Phase III trials of fluoroquinolone-containing regimens for drug-susceptible tuberculosis failed to show efficacy in reducing the duration of treatment to 4 months, suggesting that treatment
shortening may not be feasible with the current drugs—or only in patients with limited risk factors and paucibacillary disease [17]. So, how can we ensure that new regimens with preclinical and in vitro promise will translate into sterilizing efficacy in humans?

In TB drug development, pharmacokinetics and pharmacodynamics (PK-PD) studies are generally carried out to assess the relationship between the blood and tissue levels of a new compound and the plasma or serum bactericidal activity of the compound against *Mycobacterium tuberculosis*. PK-PD modeling is now a routine component of preclinical studies [18]. PK-PD relationships are typically evaluated based on drug exposures in plasma, but this fails to consider the varying exposure of bacterial populations in diverse lesion compartments [19,20]. The ability of drugs to penetrate anatomic lesions and to kill both active and quiescent bacilli should be considered early in the drug development process, informing the rational combination of drugs with complementary activity against the bacterial subpopulations present in the lesions [21].

Identifying the ideal synergistic use of bacteriostatic (i.e., growth inhibiting) and bactericidal antimycobacterial agents in combination as well as the timing and duration of their use across treatment phases remains a significant challenge in TB therapeutics. Mechanistic models and tools for regimen and dose optimization that evaluate the lesion-focused time course of drug levels following various drug combinations, doses, and schedules have been developed, which may lead to improved regimen selection [22]. Recently, artificial-intelligence-enabled parabolic response surface (AI-PRS) used in combination with in vitro high-throughput models has been proposed for identifying synergistic drugs to treat TB [23]. Additional complementary in vivo and clinical trials data are needed to determine whether these newer model-based techniques can facilitate the identification of maximally potent, safe, and tolerable shorter course regimens of the future.

Translational quantitative pharmacologic modeling provides an opportunity to identify preclinical and clinical PK-PD parameters that correlate best with bactericidal efficacy and toxicity [24] and to assess sputum culture results in early phase trials with clinically relevant endpoints in later-phase trials [25]. Learning from preclinical PK-PD properties of candidate drugs is critical for constructing effective combination regimens and providing an understanding of the contribution of specific agents to the entire regimen. Integrating microbiologic determinants, such as minimum inhibitory concentrations, with quantitative longitudinal culture results and PK-PD assessments should yield valuable insights during all phases of drug development. These quantitative PK-PD approaches will guide optimal drug dosing, as well as inform the assessment of drug–drug interactions [26]. This argues for the development of a standardized preclinical/clinical translational PK-PD modeling strategy for TB drug combinations with robust predictive features to guide rational selection of regimens to be moved forward into clinical development, support the selection of dose ranges to be studied, and provide quantitative predictions of clinical trial outcomes [27,28].

Based on this, the WHO Technical Consultation suggested that Phase II and Phase III trials systematically include PK sampling, so that PK-PD assessments linking drug exposures to bactericidal activity and treatment outcomes can be performed. Such analyses should account for other factors likely to affect outcomes, including disease severity and treatment adherence (Table 1) [8].

### 2. Do we have the best tools currently to identify relevant drug combinations to transition from Phase II to Phase III trials?

The early stages of clinical development should identify those drug combinations with the best safety, efficacy, and treatment duration profiles to bring to Phase III trials [29]. Phase IIA
studies assist in dose-finding and provide early evidence of antibacterial activity (Fig 1). Phase IIB studies test more stringently the efficacy of new drug regimens, using generally sputum culture conversion at 8 weeks as an endpoint [30]. Unhappily, as shown with the fluoroquinolone trials, Phase IIB studies are insufficient for predicting long-term outcomes and may fail to identify the degree of improved culture conversion necessary to achieve substantial treatment shortening [31].

<table>
<thead>
<tr>
<th>Identified gaps</th>
<th>Proposed solutions</th>
<th>Additional comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>How can we bridge preclinical data into clinical trials and identify the PK-PD measurements that correlate best with bactericidal efficacy and toxicity in vivo?</td>
<td>Phase II and Phase III trials provide opportunities to collect rich and informative data on drug exposures and microbiological response over time. These trials should all include PK sampling, so that PK-PD assessments, linking drug exposures to bactericidal activity and ultimate treatment outcomes, can be performed; these analyses should account for other factors likely to affect outcomes, including disease severity and treatment adherence.</td>
<td>Development and validation of novel biomarkers should be integrated in all PK-PD activities to allow for rapid assessment of the biomarkers and properties of future potential surrogate for bacterial load.</td>
</tr>
<tr>
<td>Do we have the best tools currently to identify relevant drug combinations to transition from Phase II to Phase III trials?</td>
<td>Phase IIB/C studies, with arks testing different doses and duration and with collection of treatment outcomes, are likely to strengthen the process for identifying candidate regimens likely to succeed in Phase III.</td>
<td>More quantitative, longitudinal, and time-to-event measures (time-to-positivity on liquid media, time-to-stable culture conversion) are now in common use and are endorsed for broad uptake as viable alternatives to single time-point dichotomous endpoints. Adaptive approaches offer potential reductions in sample size.</td>
</tr>
<tr>
<td>How can we overcome the long duration, cost, and constraints of Phase III trials and simplify them without hampering validity and wider drug development?</td>
<td>Both noninferiority and superiority designs are relevant for studies of new TB regimens; their use depends on the indication (drug-susceptible or drug-resistant TB) and on the intended use and value proposition of the new regimen—e.g., better efficacy or shorter duration. New adaptive designs can accelerate Phase II and III trials and improve our ability to select regimens for further investigation.</td>
<td>Innovative, efficient designs (e.g., adaptive strategy designs) need to be further explored for TB drug and regimen development. Many have the potential to accelerate and enhance ability to learn.</td>
</tr>
<tr>
<td>What is the role of treatment adherence in development of new TB therapeutics?</td>
<td>Adherence remains an under-valued but important determinant of treatment success. More attention to this domain can help to address the global challenge of treatment default. High-quality data are needed to establish the efficacy and reliability of new methods to measure and sustain adherence.</td>
<td>Ensuring and measuring adherence in clinical trials are essential to correctly interpret results of the trials. Both explanatory and pragmatic trials are needed to answer questions about efficacy and safety and about expected effectiveness in programmatic conditions that includes assessment of adherence.</td>
</tr>
<tr>
<td>How can we include key populations, such as children, pregnant women, and people with HIV infection, in clinical trials from the outset, rather than as an afterthought?</td>
<td>More attention is needed to assure the provision of evidence relevant to key subgroups, including pregnant and breastfeeding women, young children, and persons with critical comorbidities such as HIV infection. Novel designs and approaches to integrated substudies would be useful.</td>
<td>The limited evidence base for the prevention and treatment of TB in pregnant women should be emphasized. More PK studies of first-line, second-line, and new anti-TB drugs in pregnant women are needed. Appropriate formulations of drugs for infants and young children should be developed during the early phases of regimen development and testing, whenever feasible. Drug–drug interaction studies between anti-TB and ARV drugs should be conducted as early as feasible within regimen development. Careful joint management of HIV and TB care is essential. In accordance with WHO guidelines, ART should be initiated as soon as possible for all HIV-infected participants with TB in clinical trials, and definitely within the first 8 weeks of TB treatment.</td>
</tr>
</tbody>
</table>

ART, antiretroviral therapy; ARV, antiretrovirals; PK-PD, pharmacokinetics and pharmacodynamics; TB, tuberculosis.

https://doi.org/10.1371/journal.pmed.1003059.t001
Innovations in recent years have enhanced the information gathered in these studies and have streamlined the selection of regimens. First, the application of PK-PD methods to the analysis of Phase IIA studies clarifies dose–response relationships and reinforces the validity of a shift of focus from single drugs to combinations of drugs [32,33]. New approaches to Phase IIB studies have been proposed, based on intensive sampling of sputum at various time points, with longitudinal statistical modeling of quantitative bacteriology, time-to-positivity in mycobacterial growth indicator tube (MGIT) system, or time-to-culture conversion data [34,35]. Because these outcomes are measured on a continuous rather than a binary scale, they are more sensitive to differences than the traditional 8-week endpoint.

Two new approaches have been proposed to enhance the capacity for early selection of relevant combinations to bring from Phase II to Phase III testing. The “multi-arm multi-stage” (MAMS) design allows testing of a broad range of combinations and dose levels without requiring a large sample size [36]. The second is a hybrid approach combining Phase II and Phase III features, the “Selection Trial with Extended Post-treatment follow-up” (STEP) Phase IIC, wherein limited long-term follow-up data on relapse are collected as well as data on culture conversion; this permits estimation of a Bayesian prediction interval for the likely results of a future Phase III trial [37]. Such Phase IIB/C studies, with arms testing different doses and duration and with collection of treatment outcomes, are likely to strengthen the process for identifying candidate regimens likely to succeed in Phase III [8] (Table 1).

---

**Fig 1. The successive clinical trial phases in human development for TB drugs/regimens** [8]. TB, tuberculosis.

https://doi.org/10.1371/journal.pmed.1003059.g001

<table>
<thead>
<tr>
<th>Phase IIA</th>
<th>Phase IIB</th>
<th>Phase IIC</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Objective:</strong> To assess short term potency (i.e. early bactericidal activity, EBA) of experimental drug alone or in combination and identify optimal therapeutic dose(s).</td>
<td><strong>Objective:</strong> To assess intermediate efficacy, safety and tolerability of experimental tuberculosis regimens (3 or more tuberculosis drugs) to inform go/no-go decisions for phase III.</td>
<td><strong>Objective:</strong> To assess intermediate efficacy, safety and tolerability of experimental tuberculosis regimens (3 or more tuberculosis drugs) to inform go/no-go decisions for phase III.</td>
</tr>
<tr>
<td><strong>Primary endpoints:</strong> Rate of decline of in bacillary load over time</td>
<td><strong>Primary endpoints:</strong> Time to stable culture conversion, rate of decline in bacillary load over time, proportion of participants with negative sputum culture at specific time during treatment (e.g. 8 weeks)</td>
<td><strong>Primary endpoints:</strong> Time to stable culture conversion, rate of decline in bacillary load over time. <strong>Secondary endpoint:</strong> Proportion of participants experiencing bacteriological failure or relapse or clinical failure (composite unfavorable outcome) at 52 weeks (12 months), upon completing full treatment with experimental regimen</td>
</tr>
<tr>
<td><strong>Duration:</strong> Up to 2 weeks</td>
<td><strong>Duration:</strong> Up to 8 weeks</td>
<td><strong>Duration:</strong> 12 months (participants take full course of treatment to be studied in Phase III, then followed for clinical outcomes for a minimum of 12 months from randomization)</td>
</tr>
</tbody>
</table>

---

Population: Healthy volunteers (first-in-human studies)

**Objective:** To explore safety, tolerability, pharmacokinetics, and drug-drug interactions of experimental drug.

**Primary endpoints:** Pharmacokinetic profiles, safety and tolerability assessments

**Duration:** Variable

Population: Tuberculosis infected patients

**Phase IIA**

**Objective:** To assess short term potency (i.e. early bactericidal activity, EBA) of experimental drug alone or in combination and identify optimal therapeutic dose(s).

**Primary endpoints:** Rate of decline of in bacillary load over time

**Duration:** Up to 2 weeks

Population: Tuberculosis infected patients

**Phase IIB**

**Objective:** To assess intermediate efficacy, safety and tolerability of experimental tuberculosis regimens (3 or more tuberculosis drugs) to inform go/no-go decisions for phase III.

**Primary endpoints:** Time to stable culture conversion, rate of decline in bacillary load over time, proportion of participants with negative sputum culture at specific time during treatment (e.g. 8 weeks)

**Duration:** Up to 8 weeks

Population: Tuberculosis infected patients

**Phase IIC**

**Objective:** To assess intermediate efficacy, safety and tolerability of experimental tuberculosis regimens (3 or more tuberculosis drugs) to inform go/no-go decisions for phase III.

**Primary endpoints:** Time to stable culture conversion, rate of decline in bacillary load over time. **Secondary endpoint:** Proportion of participants experiencing bacteriological failure or relapse or clinical failure (composite unfavorable outcome) at 52 weeks (12 months), upon completing full treatment with experimental regimen

**Duration:** 12 months (participants take full course of treatment to be studied in Phase III, then followed for clinical outcomes for a minimum of 12 months from randomization)
3. How can we overcome the long duration, cost, and constraints of Phase III trials and simplify them without hampering validity and wider drug development?

Phase III confirmatory trials of TB treatment are high-cost undertakings, requiring large numbers of patients followed for long periods of time [38]. Innovative Phase III trial designs are needed for more efficient evaluation of a greater number of regimens with fewer patients and fewer resources, ensuring delivery of high-quality evidence for well-informed decisions by regulators, policy makers, healthcare providers, and patients.

Superiority trials provide robust evidence of benefit from a new drug or regimen when compared with a suboptimal standard of care. Developing new regimens for the treatment of drug-susceptible TB is, however, challenging because of the high cure rates achieved with current standard SCC under trial conditions. Noninferiority designs are more appropriate when new regimens may have practical advantages over current standard therapy (e.g., being shorter in duration or easier to adhere to) and thus may be preferred in real-life settings when such benefits may be advantageous even if the tested intervention is modestly less efficacious [39]. This margin of acceptance is defined by the noninferiority margin or delta. How narrow or wide this margin should be and how this translates into acceptable losses and desired gains is a matter of debate. A novel method is proposed that weighs potential gains and losses with the new regimen, which can then be translated into numbers of patients who would either benefit from, or be harmed with, the test intervention [40]. To minimize the possibility of biocreep (that occurs in noninferiority trials when a slightly inferior treatment is considered effective and becomes the active control in subsequent trials), the best available treatment should always be used as the control, and the margin of noninferiority should be determined on the estimated effect of the control, accounting for the variability and likely sources of bias in the estimate [41].

If there is a choice, superiority trials will always be preferred to noninferiority trials, whether in drug susceptible or drug resistant TB. It is better to be able to conclude that a new intervention is significantly better than standard treatment than that the new treatment is as good within certain limits, which is less persuasive in terms of benefit and subject to the somewhat arbitrary choice of noninferiority margins. Noninferiority becomes the design of choice when the control regimen is likely to perform to a very high level of success, a setting in which exceptionally large numbers would be required to demonstrate significant benefit. This situation is more likely to pertain in drug susceptible TB than in drug resistant TB, which, in most settings, has suboptimal results. A noninferiority design was used in the STREAM Stage 1 trial in which the proposed intervention was of substantially shorter duration, which, if demonstrated, would be of considerable benefit to patients and health services, irrespective of whether it was found to be noninferior or superior to the long-duration control [42]. Future trials in drug resistant TB in the next few years may be either of superiority or noninferiority design depending on assumptions regarding the control regimen and the benefits expected from the intervention regimen(s).

Adaptive strategies, using the MAMS design [43] or employing adjusted randomization so that more patients are enrolled into the more promising arms (so-called Bayesian adaptive randomization) [44], can improve flexibility in trial conduct by allowing changes to defined features after the trial has begun, provided that these potential changes are prespecified in the protocol.

In short, both noninferiority and superiority designs are relevant for studies of new TB regimens, depending on the indication (drug-susceptible or drug-resistant TB) and on the intended use of the new regimen, as well as on sample size considerations. New adaptive
designs can accelerate Phase II and III trials and improve our ability to select regimens for further investigation. (Table 1).

4. What is the role of treatment adherence in development of new TB therapeutics?

In a meta-analysis of the trials of fluoroquinolones-containing 4-month regimens for drug-susceptible TB, modest nonadherence was associated with significantly increased risk of unfavorable outcome, in both experimental and control regimens [17]. This underscores the importance of the quality of execution in noninferiority trials, as differential adherence across treatment arms could lead to erroneous conclusions about treatment efficacy [39,45].

To evaluate how adherence influences outcomes for specific regimens, per-protocol analyses should assess a range of thresholds for acceptable adherence (e.g., 95%, 90%, 80%) [46]. A stronger analytic approach might evaluate the effect of baseline prerandomization variables associated with poor adherence on trial outcomes [47]. Defining “significant” nonadherence is difficult; it depends on multiple factors specific to each trial, including the PK of the individual drugs, the dosing schedule, and other risk factors and comorbidities that could influence the risk of treatment failure/relapse. Once the efficacy of a regimen is demonstrated in controlled trials, the assessment of its effectiveness under conditions close to programmatic reality, for instance, through the conduct of observational studies or pragmatic trials [48], could permit estimation of the amount of nonadherence that would substantively alter the likelihood of treatment success. Such an approach was applied in trials assessing various methods of directly observed therapy, as well as in treatment of DR-TB [49–51].

From the above, it appears that adherence remains an under-valued but important determinant of treatment success. Therefore, within clinical trials, it is necessary to measure adherence carefully in order to know the extent to which a regimen might be vulnerable to reduced adherence particularly under program conditions. More attention to this domain will help address the global challenge of treatment default. High-quality data are needed to establish the efficacy and reliability of new methods to measure and sustain adherence (Table 1).

5. How can we include vulnerable populations, such as children, pregnant women, and people with HIV infection in clinical trials from the outset, rather than as an afterthought?

Populations such as pregnant or breastfeeding women and very young children have been excluded from trials (or at best, grossly under-represented) because of the potential risks of new drugs. These key populations, together with HIV-infected patients, form a substantial proportion of the global TB burden and have unique characteristics regarding PK, safety, and efficacy, which necessitate special attention in drug and regimen development [52].

Concerns of potential harm from TB therapeutics to mother and fetus have led to exclusion of pregnant women from most trials of TB therapies [53]. As a result, evidence for TB treatment during pregnancy or breastfeeding has come mainly from case reports and small series [54]. Including pregnant women in TB trials would provide more rigorous evidence of safety and activity than post-marketing surveillance [55]. TB trials should include experts in maternal-fetal medicine and the care of pregnant women who can determine reasonable approaches for risk/benefit assessment in this population.

Children account for approximately 10% of all TB cases, and the effects of age and weight on drugs’ PK are most pronounced and challenging to predict in this population. Inclusion of children in TB drug development requires specific attention to trial design, including the definition of trial outcomes, timing of inclusion, and ethical considerations [56]. The inclusion of
children (or development of integrated substudies) must be carefully considered and encouraged for new TB regimens [52].

The care of HIV-infected TB patients and the optimal use and timing of ART during TB treatment has dramatically evolved in recent years [57]. Treatment outcome in HIV-infected patients is highly influenced by proper management of ART, including a recognition of potential interactions between some antiretrovirals and TB drugs, particularly the rifamycins [58]. Carefully designed drug–drug interaction studies are a major element of clinical research on TB therapeutics in HIV-infected persons that should be conducted early in drug development. Within a clinical trial, provision of expert clinical management for patients with coinfection is extremely important.

We believe that more attention is needed to provide evidence relevant to important subgroups, including pregnant and breastfeeding women, young children, and persons with HIV infection (Table 1).

**Conclusion: Weaving better TB treatments for tomorrow**

After years of neglect, more than 30 human trials are currently testing various drugs or drug combinations for the treatment of TB [59]. At least 10 of these trials investigate shorter treatments for drug-susceptible TB, and a further 10 test new combinations for shorter and less toxic treatment of drug-resistant TB. Although this renaissance in TB therapeutics research is welcome, the overall structure of the field remains an uncoordinated and fragmented effort by numerous research groups and consortia pursuing their own goals, with dependence on access to products, funding, and enrollment capacity. The current process appears less systematic than the stepwise approach taken by the British Medical Research Council (MRC) 50 years ago. With the current uncoordinated approach, are we helping to finish Penelope’s tapestry (after Ulysses’ return), or are we just unwaving it again? In this respect, it is important to note that pretomanid, a new chemical entity, has been recently approved by the US Food and Drug Administration (FDA) for the treatment of adult patients with extensively drug-resistant pulmonary TB, in combination with bedaquiline and linezolid, based on a single arm, noncontrolled, nonblinded study in 109 people [60]. Although the study achieved a major step in the treatment of this very difficult-to-treat condition, it is noteworthy that it also bypassed some of the normal requirements for randomized controlled trials of new drugs or drug combinations. This may be considered acceptable given the absence of successful standard treatment of extensively drug-resistant TB (XDR TB) and the consequent high probability of death (26%), and not randomizing patients can be reasonable when the intervention arm is likely to confer a potential benefit and when the health condition under study does not have any cure (e.g., Ebola virus disease [61]). However, for the investigation of new treatments of TB and multidrug-resistant (MDR)-TB, conditions for which a reasonable standard of care exists, the use of full Phase III randomized controlled trials should be the rule. Moreover, even though high-quality programmatic, observational data can be invaluable for understanding the performance of regimens in field conditions and for policy decision-making, such data cannot replace the need for high-quality randomized controlled trials to evaluate the efficacy, safety, and tolerability of a new treatment regimen that are key for policy development.

The articles in this *PLOS Medicine* Special Collection describe promising innovations in the search for new TB treatments. These have the potential to improve the rational identification of regimens that can be swiftly brought from early to late clinical development phases, reduce development risk, and accelerate clinical progress in TB therapy, increasing our confidence that the regimens selected for Phase III trials contain the right drugs at the right doses without deleterious drug–drug interactions. Through the use of appropriate research designs and
selection of adequate endpoints, we can produce high-quality evidence for the transformation of TB treatment that is essential for the development of guidelines and ensuring strength of recommendations [62]. This requires close interaction between researchers designing the next generation of TB trials, regulators, policy makers, and advocacy groups to achieve best harmonization of the research pipeline and the subsequent policies on use and access to TB medicines [63].

The End TB Strategy calls for the introduction of new tools by 2025 in order to reach the 2030 targets of a 90% reduction in TB deaths and 80% reduction in TB incidence compared with 2015 levels [64]. Achieving these targets requires the development and introduction of new tools, in addition to ensuring universal access to existing technologies, including shorter, safer, and more effective treatment for all forms of TB. Making progress toward this goal requires maintenance of a robust pipeline of new compounds and improvements in treatment of drug-susceptible and drug-resistant TB using novel combinations of new and repurposed drugs [65].

We advocate here for the international adoption of a better harmonized approach to regimen development to ensure that best practices are used to accelerate development of new TB regimens. By using all the creative approaches described in this Collection, we hope that the next generation of TB trials will yield high-quality evidence for novel regimens that meets the needs of the 10 million persons who become ill with TB each year. Such an approach should help us to reduce these numbers more rapidly by together weaving a tapestry of highly effective, safe, and accessible TB treatments.

Acknowledgments

The authors would like to thank all experts who participated in and contributed to the WHO Technical Consultation on Advances in Clinical Trial Design for Development of New TB Treatments in Glion-sur-Montreux, Switzerland, March 14–16 2018.

Disclaimer

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the authors’ affiliated institutions nor the Centers for Disease Control and Prevention.

References


