Report of the Technical Consultation

on

Advances in Clinical Trial Design for Development of New TB Treatments

---

Glion-sur-Montreux, Switzerland

14-16 March 2018

END TB

World Health Organization

WHO/CDS/TB/2018.17

Suggested citation. Report of the Technical Consultation on Advances in Clinical Trial Design for Development of New TB Treatments, Glion-sur-Montreux, Switzerland, 14–16 March 2018. Geneva: World Health Organization; 2018 (WHO/CDS/TB/2018.17). Licence: CC BY-NC-SA 3.0 IGO).

Cataloguing-in-Publication (CIP) data. CIP data are available at http://apps.who.int/iris.

Sales, rights and licensing. To purchase WHO publications, see http://apps.who.int/bookorders. To submit requests for commercial use and queries on rights and licensing, see http://www.who.int/about/licensing.

Printed in Switzerland

Contents

This report is dedicated to

Professor Denis Anthony Mitchison

who participated in the design of the very first randomised controlled clinical trial for tuberculosis and devoted his whole life to improving treatment for tuberculosis.

Acknowledgements

Overall organization of the technical consultation

Christian Lienhardt & Payam Nahid

Chairman and Facilitators

Chairman: Payam Nahid. *Facilitators*: Lori Dodd, Michael Hoelscher, John Johnson, Carole Mitnick, Sumati Nambiar, Jim Neaton, Michael Rich, Rada Savic, Monique Surette, Andrew Vernon.

WHO Task Force on Introduction of New TB Drugs and Treatment Regimens

Payam Nahid (Chairman), Draurio Barreira, Grania Brigden, Geraint Davies, Kelly Dooley, Michael Hoelscher, Michael Kimerling, Alberto Matteelli, Norbert Ndjeka, Nguyen Viet Nhung, Michael Rich, Alena Skrahina, Andrew Vernon.

Technical consultation participants

Abdel Babiker (United Kingdom), Draurio Barreira (Switzerland), Nilesh Bhatt (Mozambique), Martin Boeree (Netherlands), Grania Brigden (France), Marco Cavaleri (United Kingdom), Tsira Chakhaia (Geaorgia), Edward Cox (USA), Geraint Davies (United Kingdom), Rodney Dawson (South Africa), Anne-Marie Demers (South Africa), Lori Dodd (USA), Kelly Dooley (USA), Kathy Eisenach (USA), Katherine Fielding (United Kingdom) , Mengqiu Gao (China), Jan Gheuens (USA) , Amita Gupta* (USA), Debra Hanna (USA), David Hermann (USA), Anneke Hesseling* (South Africa), Michael Hoelscher (Germany), Michael Hughes (USA), Amina Jindani (United Kingdom), John Johnson (USA), Soyeon Kim (USA), Michael Kimerling (Netherlands), Barbara Laughon (USA), Regine Lehnert (Germany), Vidya Mave (India), Alberto Matteelli (Italy), Carl Mendel (USA), Carole Mitnick (USA), Sachiko Miyahara (USA), Stella Mpagama (Tanzania), Sumati Nambiar (USA) , Norbert Ndjeka* (South Africa), Jim Neaton (USA), Nguyen Viet Nhung* (Vietnam), Andrew Nunn (United Kingdom), Bern-Thomas Nyang'wa (United Kingdom), Patrick Phillips (USA), Michael Rich (USA), Rada Savic (USA), Alena Skrahina* (Belarus), Jason Stout (USA), Monique Surette (Netherlands), Cherise Scott (Switzerland), Elin Svensson (Netherlands), Anete Trajman (Brazil), Andrew Vernon (USA) , Robert Wallis (South Africa). (*: remote participation)

WHO/HQ Secretariat

Christian Lienhardt (GTB/RTE), Dennis Falzon (GTB/LDR), Lorenzo Moja (EMP), Nicola Cocco (GTB/RTE), Piero Luigi Olliaro (TDR), Matteo Zignol (GTB/TME), Tereza Kasaeva (GTB).

Administrative and secretarial support

Lou Maureen Comia, Michael Tabiszewski

Meeting Report written by

Marjorie Imperial & Christian Lienhardt, with input from Payam Nahid and review by all participants

Financial support:

This consultation was financially supported by the Bill and Melinda Gates Foundation, grant Number: OPP1126615

List of Acronyms

ART	antiretroviral therapy
BMI	body mass index
CFU	colony-forming units
DOT	directly observed therapy
DS	drug susceptible
EBA	early bactericidal activity
EMA	European medicine agency
FDA	Food and Drug Administration
GLP	Good Laboratory Practices
GRADE	Grades of Recommendation Assessment, Development and Evaluation
GTB	Global Tuberculosis Programme
HFS	hollow fiber system
HIV	human immunodeficiency virus
HRZE	isoniazid-rifampin-pyrazinamide-ethambutol
LAM	Lipoarabinomannan
LTBI	latent tuberculosis infection
MAMS	Multi-Arm Multi-Stage
MBL	Mannose-binding lectin
MDR	multi-drug-resistant
MGIT	Mycobacterial Growth Indicator Tube
MIC	minimum inhibitory concentration
MRC	Medical Research Council
NNT	number needed to treat
OBR	optimized background regimen
PD	Pharmacodynamics
PET/CT	positron emission tomography/computed tomography
PK	pharmacokinetic
PK/PD	pharmacokinetic/pharmacodynamic
QTc	corrected QT interval
SMART	Sequential, Multiple Assignment, Randomized Trial
STEP	Selection Trial with Extended Post-Treatment follow up
TB	tuberculosis
TRPs	target regimen profiles
TSCC	time to stable culture conversion
TTP	time to positivity
WHO	World Health Organization
XDR	extensively drug-resistant

The Global TB Programme of the World Health Organization (WHO) convened a technical consultation on "Advances in Clinical Trial Design for Development of New TB Treatments" in Glion-sur-Montreux, Switzerland, from 14 to 16 March 2018. The consultation brought together researchers, academics, technical partners, TB drugs and regimens developers, trialists, regulators, guideline developers, programme managers, patient's representative and nongovernmental organizations.

Welcome, Introduction, and Objectives

Chair Dr. Payam Nahid (UCSF) and Dr. Christian Lienhardt (WHO, GTB/RTE)

In his opening statements, Dr Nahid recognized that the tuberculosis (TB) therapeutics field has reached a key time point wherein broad reflection on the contemporary TB trials of the last 15 years is warranted: what have we done correctly?, what were our mistakes?, and how can we improve? With the anticipated emergence of new drugs for TB, now more than ever, there is a need to revisit our approaches and define best practices for TB clinical trial design for the development of new regimens. Dr. Nahid noted that this is the first ever meeting to gather such a diverse group of stakeholders, including trialists, academia, research institutions, TB drug and regimen developers, contract organizations, regulators, guideline developers, non-governmental organizations and civil society to address these questions, and he thanked the Global Tuberculosis Programme (GTB) of the World Health Organization (WHO) for organizing and sponsoring this unprecedented consultation.

Dr. Lienhardt described the background and objectives of the consultation. The *Task Force on Introduction of New TB Drugs and Treatment Regimens*, established by the WHO/Global TB Programme (GTB), developed in 2016 a series of Target Regimen Profiles (TRPs) for new TB treatment through broad consultation with experts and stakeholders worldwide. The TRPs are intended to guide the development process towards anti-TB treatment regimen characteristics of critical importance to patients and programmes. To assist in the implementation of these TRPs, there is a need to guide the research community on optimal clinical trial designs and features for new anti-TB drugs and regimens, in consultation with relevant stakeholders in the field. The major challenges in the development of new TB treatments include the long developmental pathway to identify best regimens, the lack of direct readout of response and use of surrogate endpoints, and the lack of predictive quantitative relationships between Phase II and Phase III readouts. Clear and rationally justified approaches for the choice of drug combinations, trial design, selection of endpoints and analysis, are critically important, taking into account new developments in individual drug's PK/PD characteristics, microbiological aspects, use of biomarkers, standardization of approaches and data collection, and the effect in special populations.

Given the importance of the quality of evidence to be gathered, the **main objective** of this technical consultation was to outline, through expert consensus, the optimal characteristics of clinical trial designs for the development of new TB regimens that WHO would support for adoption in order to generate the highest quality evidence to inform policy guidance.

The **specific objectives** were:
(1) to review the respective strengths and limitations of current approaches for clinical development of new TB drugs and drug regimens; and
(2) to identify optimal practices and study designs to inform policy guidance on new drug regimens for the treatment of all forms of TB, taking into account recent developments in methods, tools and biomarkers.

Context and rationale:

Three presentations were made to provide context and set the stage for the consultation.

Key-Note 1: *Lessons learnt from the TB ReFLECT meta-analysis of fluoroquinolone-containing regimens for the treatment of drug-susceptible TB - Dr. Rada Savic (UCSF)*

Key-Note 2: *Lessons learnt on moving new drugs into new regimens for treatment of drug-susceptible and drug-resistant TB – Dr. Carl Mendel (TB Alliance)*

Discussants: *The point of view of the programme managers and end-users on the results of contemporary TB treatment trials - Dr. Nguyen Viet Nhung (National Lung Hospital, Hanoi, Vietnam), Dr. Alena Skrahina (Republican Research and Practical Center for Pulmonology and Tuberculosis, Minsk, Belarus), Dr. Norbert Ndjeka (National Department of Health Pretoria, South Africa).*

The findings from the TB ReFLECT meta-analysis of the three trials of fluoroquinolone-containing regimens for the shortened treatment of drug susceptible tuberculosis (DS-TB) are very helpful to provide context and set the stage for the consultation.[1] Although the trials independently failed to show non-inferiority of the 4-month experimental arms, 80% of patients were cured. These trials were preceded by several Phase IIB trials, some of these showing improved 2-month culture results after treatment with fluoroquinolone-containing experimental regimens, but these improvements in culture conversion did not translate to predicting long term clinical endpoints (e.g. durable cure versus relapse). The analysis found that patients with minimal disease, defined as low bacterial burden or absence of cavities and representing up to 47% of patients, are eligible for 4-month treatments. Conversely, patients with high baseline smear, cavitation, HIV co-infection, and low BMI (representative of malnutrition) defined “hard-to-treat” phenotypes that need more than the standard 6-month treatment duration in order to achieve the highest possible cure rates. In addition, incomplete adherence, independent of treatment duration, was the most significant risk factor for unfavourable outcome, highlighting the need for optimal tools for measuring and maximizing adherence. Other key lessons from the analysis were: (i) the urgent need for standardized and harmonized definitions of unfavourable outcomes to allow for better aggregation of data for pooled analysis across trials, (ii) the importance of collection of PK data to allow further investigation of drug-specific factors beyond adherence that influence outcomes, (iii) the urgent need for data in special populations (children, pregnant women, and HIV-coinfected patients), and (iv) the invaluable scientific benefits to the field of both data standardization and data sharing.

From the regimen developer's perspective, four aspect need to be taken into consideration. First, a new regimen must bring a value proposition, beyond efficacy or safety targets. Products with broader applications (eligible populations, etc.) gain in terms of delivery and scalability/distribution or cost, and can bring substantial impact and value that define the developmental pathway. In fact, obtaining approval for a drug and relying on real world use and data to figure out how best to use the drug is not tenable, and more investment will be needed by sponsors and donors to evaluate the needs of the market and develop programmes based on those needs. Second, volume of use matters and one solution to this is the unification of treatment for DS-, MDR- and XDR-TB with a single regimen. This approach can only be done with new chemical entities, but if successfully combined into safe and efficacious regimens, increased uptake and use of a new regimen can solve issues related to cost, supplies, and stock outs and increase efficiencies across heath care systems. Third, transition decisions from Phase II to Phase III continue to have significant uncertainty, and these limitations should be considered when designing Phase III trials. Lastly, the issue of the control groups most appropriate for a given trial situation needs careful consideration. It thus appears that each development programme needs to determine the most appropriate approach to design, depending on the situation and question to be addressed.

For the TB programme managers and end-users, expectations are for the next generation of regimens to have shortened durations (6-months or less for MDR/XDR TB and 2-4 months for DS-TB), be simple to administer (oral formulations), have minimal adverse events and drug-drug interactions, and to be accessible at low cost. TB programme managers, however, also noted the challenges involved with implementing new regimens that are recommended for use in a conditional way and endorsed by guidelines reliant on low quality evidence, calling for the conduct of high quality trials and provision of timely evidence to support recommendations.

- 0 – 0 – 0 – 0 – 0 –

*This document reports the proceedings of the consultation based on the key questions formulated and addressed by the technical consultation discussants and participants in the thematic sessions. For each of these, the discussion that took place is reported, and the outcomes of the discussion presented in synthetic Tables. The details of the presentations made by key-note speakers and discussants are provided, for each session, in the **ANNEX 2** of the Report.*

Session 1:
Pharmacokinetics/pharmacodynamics, microbiology and biomarkers.

Facilitator: *Dr. Rada Savic (University of California, San Francisco)*

Key-Note: *What can be learnt from PK/PD studies to advance the development of new regimens? – Dr. Kelly Dooley (Johns Hopkins University)*

Discussant 1: *TB regimen development: bridging translational gaps with quantitative pharmacology approaches and drug development tools – Dr. Debra Hanna (Critical-Path)*

Discussant 2: *What would be the most efficient framework for patient-level microbiology data to improve quantitative clinical PK/PD predictions and streamline model development? - Dr. Kathy Eisenach (TB or NOT TB Consulting)*

Top uncertainties/questions addressed at the Technical consultation:

1. *What is the importance of understanding PK/PD relationships by phase of regimen development?*
2. *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?*
3. *Can dynamic experiment-level in-vitro bacteriological assessments (i.e., HFS-TB) be integrated with patient-level bacteriological data to improve quantitative clinical PK/PD predictions and streamline model development?*
4. *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?*
5. *Should quantitative PK/PD models describing relevant bacteriologically-based covariates be used to guide dose finding and dose optimization in special populations during early development?*
6. *How to make use of PK/PD across clinical development phases to identify pharmacology-guided drug regimens?*

General agreement was reached on the critical importance of integrating PK assessments throughout the developmental pathway, from early through late stage development. Understanding PK/PD relationships will help avoid costly errors in decision making around dose and schedule selection and Phase III trial designs. It was expressed that success in achieving a robust understanding of PK-PD relationships will be most likely through integration of data across multiple studies rather than via isolated assessments within individual trials (see **Table 1**).

In early stage development, PK/PD models provide rationale to move forward with dosing strategies for component drugs as well as regimen composition by identifying PK target

concentration that maximizes efficacy, understanding the regimen make-up that best prevents emergence of resistance, and characterizing potential drug-drug interactions. In later stage development, population PK and PK/PD analysis can then provide an understanding of population variability and exposure-response relationships. From here, once PK/PD targets associated with satisfactory microbiologic activity are established, Monte Carlo simulations can be performed to predict target attainment in a simulated trial population that mirrors the distribution of exposures in a typical TB treatment trial. This approach can help define subpopulations that may require dose adjustments or for whom an experimental regimen is unlikely to be successful. The technical consultation also considered the question of whether PK/PD data can inform selection of regimen duration. It was noted that integrating PK as part of Phase III trials evaluating various regimens might provide data that could help answering this question, but it is still unclear to what extent current PK/PD approaches are fit for this purpose.

While consensus was reached around the critical importance of integrating PK sampling and analysis across the developmental pathway, it was acknowledged that collecting PK samples in Phase III trials can be resource intensive and expensive. Sparse sampling in all (with analysis of a subset of samples, such as those with unfavorable outcomes and a subset of those with treatment success), targeted intensive sampling, or performing targeted sampling in selected populations of interest were suggested as approaches to mitigate costs. Meeting's participants noted that the development of standardized templates for broad adoption by clinical trialists for the collection of PK samples, and the major data elements and parameters necessary across all stages of trials would help in this regard. It was noted that the WHO *Task Force on the pharmacokinetics and pharmacodynamics (PK/PD) of TB medicines* could consider developing a roadmap of minimum and optimal standards for PK/PD studies.

With regard to the value of microbiology-based PD measures (e.g. MIC) to inform drug development decision making, there was consensus on the importance of understanding these factors in greater detail. A key challenge for PK/PD studies in early and middle stage development is the reliance on microbiological measures that perform sub-optimally as surrogate markers for long term clinical endpoints. Nonetheless, as the most studied surrogate marker type for TB, modeling of the microbiologic relationships with PK have had a significant impact on decision making in recent regimen development programmes (e.g. TBTC rifapentine trials, HIGHRif trial, among others). It was noted that the large knowledge gaps in understanding how MIC can be used in analyses that link drug exposures and organism susceptibilities to outcomes is primarily due to minimal, if any, available data. Discussion ensued on the details of collection of isolates, timing of sputum collections (baseline or longitudinal), storage options, and types of assays that may be applied to the isolates. After considering the goal of this technical consultation, all agreed that standardization of tools and measurements on microbiological practices[2] is essential as we

gather data, but the technical details of defining the standardized approaches for collection are to be deferred to the *WHO Global Task Force on the PK/PD of TB medicines*. The meeting attendees agreed that additional research is urgently needed in TB biomarker development, ideally focused on culture-free, and if feasible, sputum-free, assays.

Table 1. ***Session 1: Pharmacokinetics/pharmacodynamics, microbiology and biomarkers***

Question	Consensus	Options	Research
What is the importance of understanding PK/PD relationships by phase of regimen development?	PK studies should be included throughout drug/regimen development phases, both in early and late stages of development. PK samples should be collected in all treatment trials with clear documentation of dosing history. 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*. Importance of PK in Phase II trials to allow understanding of dose-exposure-response relationships, for dose selection in definitive trials. Critical importance of PK-safety assessment in Phase II/III, to inform the need for dose/schedule adjustments. Particularly important for narrow therapeutic index drugs. Population PK modeling to understand sources of variability (e.g. sex, race, age, HIV status) in drug exposures and response Phase IIB/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.	Other PK studies should be performed in spirit of modern drug development, including: • Drug-drug interaction studies, especially with companion TB drugs or antiretrovirals • Evaluation of PK-toxicity relationships for key toxicity concerns (e.g. QTc). • Sparse PK collection in Phase III, to strengthen population PK modelling and to explore exposure differences in relevant subgroups including poor responders.	Optimal timing and frequency of PK sampling, by type of trial (e.g. Phase IIA, IIB, IIC) to yield most information in most efficient way. 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 II data to inform phase III design. Validation and refinement of translational tools and modeling activities (mouse model, HFS, systems pharmacology model) through data sharing. Biomarker (host, microbiology) explorations to find better ways to identify best regimens to carry forward from middle drug development.

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?	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. 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. Specific guidance from WHO PK/PD Task Force to provide details on standardized approaches for collection of isolates (what isolates? How to collect? How to store? When to collect? What type of assay would be needed?)	*M.tuberculosis* isolates should be stored, including at a minimum the baseline isolate and that of the last positive culture. Bacterial burden should be quantified longitudinally via collection of serial sputum samples.	Key research questions to answer by quantitative pharmacology by time of registration: - PK-PD underpinnings to support dose recommendations, including in hard-to-treat patients and special populations; - PK-toxicity relationships; - drug-drug interactions with companion TB and HIV drugs. Evaluation of value of MIC (static drug concentration in relevant medium) vs. dynamic susceptibility information in drug and regimen assessment.
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?		Investment in development of translational tools and modeling activities (mouse model, HFS, systems pharmacology model) that can inform regimen composition.	
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?	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.		Culture-free (and sputum-free) systems as alternatives to existing culture-based systems is urgently needed
Should quantitative PK/PD models describing relevant microbiology-based covariates be used to guide dose finding and dose optimization in special populations during early development?	Design of studies in special populations should be supported by Clinical Pharmacology principles (dosing regimen) and aided by model based design.		

Session 2:
Phase II to Phase III transition

Facilitator: *Dr. Michael Hoelscher (Ludwig-Maximilian's University)*

Key-Note: *From early Phase II to Phase III trials: a comprehensive review of the various clinical development phases and their seamless progression to move from early bactericidal activity of single drugs to pivotal trials of drug combinations – Dr. Gerry Davies (University of Liverpool)*

Discussant 1: *How could preclinical information influence design of Phase II studies to show efficacy of an individual drug, and are monotherapies in Phase II A studies (EBA) needed for assessment of bactericidal efficacy and dose-finding?- Dr. David Hermann (Bill & Melinda Gates Foundation)*

Discussant 2: *What information/markers/endpoints should be collected across Phase II studies to optimally select the appropriate combo regimens to move from Phase II to III? - Dr. Martin Boeree (Radboud University)*

Top uncertainties/questions addressed at the Technical consultation:

1. *What are the key preclinical and Phase I data and drug-drug interaction studies that provide adequate assurance of the safety of an individual agent prior to Phase II evaluation? What are the issues with interpretation of additional subsequent safety data from Phase II combination studies?*
2. *What alternatives exist to Phase IIA dose-finding and proof-of-concept studies using monotherapy over 14 days? To what extent and under what circumstances could preclinical information replace the need for such studies and contribute to demonstrating the contribution and value of individual agents in a regimen?*
3. *What methods are available for selection of combinations and dose-finding in Phase IIB studies (>14 days)? How could these be extended in efficiency and scope?*
4. *How predictive are different endpoints available in Phase IIB for predicting long-term outcome in Phase III (proportions versus time to event versus modelling, solid versus liquid versus molecular)?*
5. *What is the best method for determining likely duration for Phase III trials (Phase IIC, seamless designs, meta-regression methods)?*

The session facilitator, Dr. Hoelscher, urged the group to ***"think regimen"*** in place of the traditional single drug development pathways. With several dimensions in the process of regimen development that need to be considered, including dosing, durations, regimen composition, toxicities, treatment populations (hard to treat, DS *vs*. DR), and pharmacological compartments, it is not possible to test all combinations and therefore, it was emphasized that as a community we define the viable selection methods for regimens to move forward through development, while recognizing that each product development programme is unique because each

drug/regimen has its own inherent issues and therefore, timing and approach of each step in development programme might be different (see **Table 2**).

Very brief discussion revolved around the safety aspects of combining more than one new investigational drug. It was recognized that regulatory agencies provide clear guidelines under what circumstances two or more investigational drugs can be combined without preclinical combination toxicology studies. There was agreement among participants on consulting with regulators to discuss the minimum requirements for First-In-Human studies.

The value of the EBA component of Phase IIA studies were debated at length, specifically as a selection tool for regimen development, stemming from Discussant presentations in which the value of these early stage trials was questioned. The main argument for monotherapy EBA was that it would be the last chance of testing the drug as a single agent (before testing combinations) for safety, tolerability and to evaluate whether the drug has bactericidal activity against *M.tuberculosis* in humans. From a regulatory standpoint, it may be useful to have evidence on the value of each component of a regimen for approval - though proof of individual efficacy could already come from preclinical evidence. However, since activity with a single agent does not guarantee activity in combination therapy, additional testing of different combinations would be relevant to justify the regimen to be brought in later stages of clinical development. While HFS is good for evaluating PK-PD relationships in vitro, how it can predict EBA activity is unsure. Ultimately, one of the goals in regimen development is to streamline the process and one proposed approach was to shift questions traditionally addressed in EBA studies to earlier or conversely later stages – and skip this phase. However, it was acknowledged that there is reluctance to take large leaps from pre-clinical and Phase I studies directly into larger Phase IIB studies, with particular concern in safety, dose finding, and funding ($4 million EBA vs. $12-15 million Phase IIB). There was agreement among participants that Phase IIA studies address more than early bactericidal activity, but also address safety and pharmacokinetics, and to invoke innovation in Phase IIA studies, it was agreed that we should stop referring to Phase IIA studies as “Early Bactericidal Activity” studies. Further discussion is required on ways to innovate in Phase IIA to create a programme that integrates into the adjacent stages of development.

The question arose on whether we have enough safety data from preclinical and Phase I to go straight to combination studies in Phase II, and what regulators think about interpretation of safety data in Phase IIB without Phase IIA. It was proposed that EBA studies could be considered an optional pathway to regimen development and to accept alternative approaches since every development programme is unique. Proposed adaptions/alternative approaches included: 1) Incorporating combination EBAs that can better relate to later phases; 2) Use of the more standardized and reproducible quantitative TTP measure instead of CFU that can better connect

to later stages; 3) Longer phase IIA studies to reduce any information gaps between longer Phase IIB/C studies; 4) Use of multiple ascending dose trial design if safety is more of a concern than efficacy; and 5) Use of novel tools and designs (such as 14+14 SMART[a] and NEXTGen EBA[b]), to gain more information on testing combinations and drug mechanism of action. Lastly, to invoke innovation in Phase IIA studies, it was agreed that the term of Phase IIA should be used in the field as it is a more accurate than "Early Bactericidal Activity", or, in short, EBA studies, since the potential learnings are broader than bactericidal activity. More precise terms for Phase II studies were suggested, e.g. Phase II 14-day monotherapy study, Phase II 28-day combination therapy. Further discussion is required on ways to innovate in Phase IIA to create a programme that integrates into the adjacent stages of development.

With regard to the use of 8-week culture as an intermediate "surrogate" endpoint, caution was raised as the statistical definition of surrogacy is a marker that can replace clinical outcome and be reasonably likely to predict clinical benefit - in relation to which 8-week culture conversion has suboptimal linkage to ultimate clinical endpoints. Nevertheless, there was acceptance of the utility of culture conversion as an informative intermediate marker because of our extensive experience with its use and the numerous datasets available with a wide breadth of drugs/regimens. Analytical techniques using culture-based outputs, including time to positivity (TTP) and time to stable culture conversion, are being more routinely used and deserve further development. Additional research to identify culture-free (and sputum-free) biomarker endpoints as alternatives to existing culture-based ones is urgently needed. Until alternative biomarkers are identified and validated, it was proposed that intermediate studies continue to rely on more longitudinal and quantitative culture and microbiologic markers (i.e., time to stable culture conversion, TTP, among others) to allow better translation of results between drug development phases, and that better harmonisation of generation and collection of these data is desirable.

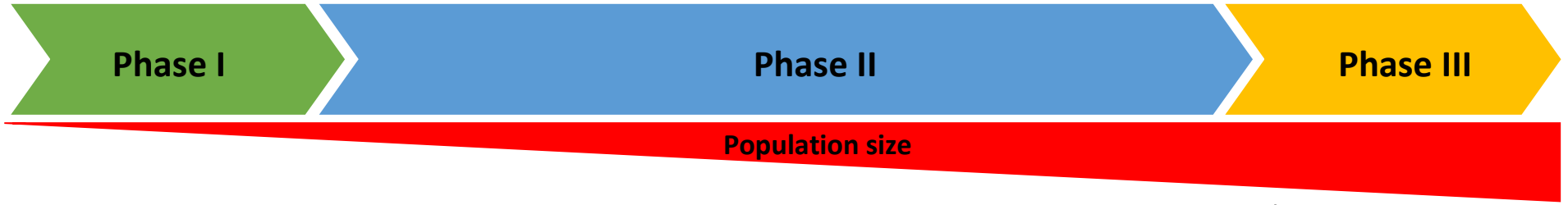
Based on experience from recent trials, Prof. Nunn expressed that stricter requirements linked to bolder targets may be warranted in selecting combinations that move forward to Phase III trials. For example, aiming for 100% culture conversion at 8-weeks instead of proportionally improving over conversion rates achievable through control regimens may allow for explorations to 3-month instead of 4-month regimens. With the anticipated emergence of several new drug compounds, we have the first opportunity to be more restrictive on choices and to aim for complete culture conversion at earlier time points, 4 or 6-weeks as compared to traditional 8

[a] Study design with 14-day monotherapy followed by 14-day combinatory therapy to assess sterilizing drug activity alone and in combination.

[b] Use of positron emission tomography/computed tomography (PET/CT) scans and immunological assays in addition to standard EBA methodology to assess sterilizing drug activity.

weeks. As contemporary regimens are hitting 95+% culture conversion rates at 8 weeks, the dynamic range at this time point is diminished, and therefore, we should shift to time-to-event or longitudinal endpoints.

Another area of discussion related to enrichment or restricted population designs for Phase IIB/C trials, i.e., enrolling only hard to treat populations, those with high smear grades and cavitation on baseline chest radiographs. Two enrichment approaches were discussed: 1. Enrichment, wherein the trial is limited to the population of interest and 2. Enrichment of population of interest as an adjunct to the populations customarily enrolled into trials. The rationale behind enriched designs in Phase IIB/C studies is that if a regimen is successful in hard to treat populations, then the regimen should also be successful in easy to treat populations who are characterized as having lower burden of disease. Consequently, if a hypothetical regimen achieves 100% week-8 culture conversion in hard to treat populations, there would be strong rationale to move forward to Phase III studies with a broader population of TB patients, perhaps using a more pragmatic trial design. However, several limitations were noted regarding enrichment or population restriction trial designs: 1) exclusion of any population from analysis necessitates more assumptions when moving forward to Phase III studies; 2) targeting of restricted population types can slow enrollments and require more funding and resources for screening activities; 3) benchmarks for potentially restricted populations of interest (i.e., cavitary disease) are not yet fully defined; and 4) uncertainties will still remain in selecting the most appropriate regimen and durations for other patient subtypes in later phase trials.



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

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

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 total of 12 months from randomization)

Population: Tuberculosis infected patients

Objective: To provide confirmatory evidence showing new treatment is safe and efficacious.

Primary endpoint: Proportion of participants experiencing bacteriological failure or relapse or clinical failure (composite unfavorable outcome) within 18 months after randomization.

Duration: At least 18 months (participants take full course of treatment, then followed for clinical outcomes for a total of at least 18 months from randomization)

Figure: ***The successive clinical trial phases in human development for TB drugs/regimens***

Table 2. ***Session 2: Phase II to Phase III transition***

Question	Consensus	Options	Research
What are the key preclinical and Phase I data and drug-drug interaction studies that provide adequate assurance of the safety of an individual agent prior to Phase II evaluation? What are the issues with interpretation of additional subsequent safety data from Phase II combination studies?	The regulatory authorities provide clear guidance for the GLP Toxicity programme. Consultations with regulators are necessary to discuss the minimum requirements for 'First-in-human' studies or later clinical studies. For combinations of novel agents the risk of overlapping toxicities should be considered and assessed as appropriate.		
What alternatives exist to Phase IIA dose-finding and proof-of-concept studies using monotherapy over 14 days? To what extent and under what circumstances could preclinical information replace the need for such studies and contribute to demonstrating the contribution and value of individual agents in a regimen?	The early Phase II activities in drug and regimen development encompass much more than mono-drug EBA studies. The more appropriate terminology for this phase of development should be 'Phase IIA'. Phase IIA monotherapy studies are useful. They are the first and last chance to test both safety and activity of a single drug in patients. Regulators require the demonstration of the contribution of each individual drug component of a regimen. Preclinical package and Phase IIA studies provide such data. Phase IIA studies confirm proof of concept in patients – however, preclinical data can inform the target exposures for Phase IIA, making their design more efficient. The need and aims of Phase IIA will vary depending on the characteristics of each individual drug candidate.	Adaptations to the traditional phase IIA design might be useful: - With more PKPD information upfront, studies can be more targeted and the monotherapy part can be shorter (e.g. 7 days); - There are novel designs such as 14+14 SMART or permutations of it, e.g. 7 plus 21/28, that should be considered with the goal of gathering more meaningful data for safety, activity and PK/PD when testing combinations; - Longer phase IIA studies (e.g. 4 weeks) can be useful as phase IIB/C become longer to reduce gap between developmental phases.	Novel markers are currently being investigated to be used in phase II studies (such as PET/CT Scan; Mannose-binding lectin-MBL; lipoarabinomannan-LAM), that might provide additional information that could be relevant to predict cure. These markers will need to be validated.

	MGIT TTD should replace CFU in order to facilitate efficacy connection to later stage phase II studies.		
What methods are available for selection of combinations and dose-finding in Phase IIB studies (>14 days)? How could these be extended in efficiency and scope?	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.	Possibility of supporting conclusions on efficacy with follow-up to long-term microbiology outcomes (Phase IIC).	Use of approaches that aim at increasing the discriminative power of phase IIB/C studies, such as enriching for hard-to-treat patients.
How predictive are different endpoints available in Phase IIB for long-term outcome in Phase III (proportions vs. time to event vs. modelling, solid vs. liquid vs. molecular)?	Culture markers remain the best and most studied intermediate markers as of now. Analytical techniques need to be improved and refined to obtain optimal outputs. To allow better translation of results between the developmental phases, overlapping readouts are required (e.g. TTP and TSCC in Phase IIA, Phase IIB/C and Phase III). Liquid culture systems provide more standardized and quantitative information than solid culture.	Demise of usefulness of 8-week culture conversion endpoint in unselected or enriched trials	Limited evidence supporting surrogate endpoints other than 8-week culture conversion. Limited data to support time-to-event measures due to reporting issues and sampling. Very limited data for independent Phase II 8-week combination studies based on quantitative bacteriology. Invest in research to identify culture-free (and sputum-free) surrogate biomarker(s).

Session 3:
New trial designs and how they may facilitate regimen development

❖ *Sub-Session 3.1: Novel trial designs*

Facilitators: *Dr. Carole Mitnick (Harvard Medical School) & Dr. Jim Neaton (University of Minnesota)*

Key-Note: *Where have we come from and have we made any mistakes? (High-level) review of different trials and trial designs used to date in TB phase III trials, and how designs have progressed over time – with discussion of respective strengths and weaknesses) – Prof. Andrew Nunn (University College London)*

Discussant 1: *Have non-inferiority trials served us well? Given challenges in interpretation and limitations of non-inferiority phase III trials, what are the recommendations for the future of non-inferiority design in TB therapeutics? – Dr. Piero Olliaro (WHO/TDR)*

Discussant 2: *What are the appropriate controls for rifampin susceptible and rifampin resistant TB trials? How do we manage the challenges posed by the various designs on the selection of appropriate control groups, including the issue of changing standards of care? – Dr. Ed Cox (US Food and Drug Administration, FDA)*

Discussant 3: *Where are we now and where should we be going? Presentation of new ideas that are being proposed for ongoing and future trials – with discussion of strengths and weaknesses. – Dr. Patrick Phillips (UCSF)*

Top uncertainties/questions addressed at the Technical consultation:

1. *What novel trial designs are appropriate for TB phase III trials? Which are most efficient? Are there ways that non-inferiority can be avoided? Can and should clinical trials be designed to show superiority?*
2. *What are appropriate controls for MDR/XDR-TB regimens? (Superiority trial design with add-on to optimized background regimen (OBR) compared to placebo has shown its limitations. Should we be comparing the effect of one treatment regimen to another rather than the potential added benefit of a new single agent?)*
3. *In what patient populations, trial designs and under what circumstances would the use of historical controls be acceptable to WHO and regulatory bodies??*
4. *Should we have different trial considerations for registrational trials compared to trials conducted to address public health needs? Can trials be designed to address both the public health and registration needs?*
5. *When the standard of care changes while a trial is underway, how can the scientific integrity of the trial be preserved, particularly with regard to choice of control(s)?*

Non-inferiority and superiority trials

The merits and challenges of non-inferiority and superiority trial designs were discussed at length (see **Table 3**). Many participants considered the superiority trial design as the ideal approach to show robust evidence of benefit of a new regimen. However, due to high efficacy of control regimens, it is difficult to show superiority unless very large sample sizes are enrolled. In some cases, however, the sample size for non-inferiority designs will be larger than for superiority designs, depending on the set margin, and whether it is reasonable to assume that the true rate for the experimental treatment is more favorable or equal to that for the control.

Some questioned and others advocated for superiority designs that restrict to enrolling hard-to-treat populations. The concept of nested superiority in non-inferiority trials where non-inferiority results are the primary outcome and superiority results in a subset of the population are the secondary outcome was discussed. The challenges of using superiority designs in TB, including in subgroup populations (hard to treat), considering the current lack of evidence to suggest superiority of treatments, were acknowledged. The importance of having a strong understanding of, and rationale for, acceptable non-inferiority margins, not only for regulatory perspectives but also for WHO policy decisions was clearly emphasized. In the case of many TB trials, a non-inferiority design is relevant because one is studying, for instance, a shorter regimen which is easier to take than current standard, so one would be willing to give up some efficacy defined by the margin. Defining an acceptable non-inferiority margin is not only based on a new regimen showing superiority over placebo but also on clinical judgement. For example, narrower non-inferiority margins may be justified if the new regimen does not provide large benefits over the standard e.g. same duration and similar safety/drug interactions profile. Conversely, wider margins could be justified when a new regimen has a significantly shortened duration, say to 2 months.

The importance of understanding patient and provider perspectives in regard to trial designs and endpoints was discussed. Best practice includes a formal assessment of these preferences. It was recognized that quantitative assessment, supported by qualitative data, would be helpful to understand these preferences. In addition, the interests of the TB programmes also need to be considered. Such information should provide additional supportive elements for the selection of the non-inferiority margin. In this regard, there was consensus that developers be more transparent in describing their approach and rationale for selecting non-inferiority margins because these aspects can highly influence the interpretation of results. Therefore, protocols should clearly state the clinical judgments that played any role in the decision of the non-inferiority margin. A study that has a set and justified margin, which is then ignored or adjusted post-hoc, would be unacceptable. Further discussion is needed on defining standards for expressing results from non-inferiority tests, as interpretation may vary across regulatory and

public health point of views. It was suggested that a new metric that is easy to interpret and accounts for public health consequence be developed because of the current challenges with interpreting non-inferiority trials. It was suggested that a working group of statisticians be formed to further explore the possibility of this approach.

Pragmatic trials

Trials that define "public health superiority" and test for the public health benefits of a new regimen as an alternative to the usual non-inferiority explanatory trials were discussed extensively. Another option would be to follow-up on Phase III non-inferiority trials with a more pragmatic trial that may use either superiority or non-inferiority design. Pragmatic trials are usually more inclusive, assess a number of relevant clinical outcomes, have minimal data collection, and include diverse sites as well as diverse study participants. They often seek to mimic "real-world" use of regimens, yet retain randomization as a key design feature. Such pragmatic trials would allow transition from specialized explanatory *efficacy* to *effectiveness* of the new regimen, which would increase the generalizability of the findings. It was acknowledged that no single trial can address all clinical and public-health relevant research questions at once. If pragmatic trials are pursued, design features that balance pragmatism against classical explanatory trial approaches will require careful consideration for testing each experimental regimen.

Composite outcomes

There is significant variability in the composite definitions used for unfavorable outcomes in Phase III clinical trials. Recently used composite outcomes have included a variety of components including death, relapse (and sometimes reinfection), treatment failure, lost to follow up, discontinuation due to adverse event, a change in the drug components of the regimen, retreatment without bacteriological evidence of failure, among others. Dr. Lienhardt explained that the historical MRC trials only included relapse as a strict bacteriological outcome, but in the 2000's the definition moved toward aiming for success in the *programmes* thereby considering a favourable vs. unfavourable outcome, the latter including treatment failure and relapse, but also lost to follow-up, deaths of uncertain cause, and changes in therapy. He suggested that moving back to more granularity with TB-specific outcome would be appropriate, thus disaggregating the components of the composite outcome. Dr. Phillips noted that the problem of the composite outcome is really a "missing" data problem that forces all outcomes into a binary group. He recommended applying missing data methods instead, such as data imputation or censoring (time to event analysis). In fact, events included in the composite outcome should be of equal or near equal significance so that the trial results are not determined by events of minor importance. Both EMA and FDA representatives expressed openness for discussion around composite outcomes for TB. They noted, however, that those outcomes commonly included in composite endpoints have relevance for TB therapeutics. They also stressed that having a pre-

planned approach to analysis of study participants lost to follow-up (missing data) is important, and that all efforts should be made to retain all participants in the study. They acknowledged that it is difficult to address all questions and endpoints at once, but noted that regulatory agencies are understandably conservative when it comes to selecting primary analyses for clinical trials to support initial drug or regimen approval. That said, after approval, it is possible to conduct additional trial to assess additional key endpoints in a pragmatic setting.

It was noted that no other outcome holds a severity or significance as high as mortality, and yet other outcomes are combined with it in endpoint analyses. As an alternative, the participants discussed the potential for individual components of composite outcomes to be assigned different weights or hierarchy, using a risk-benefit approach (e.g. death > relapse ≥ treatment failure). It was acknowledged, however, that the rationale for such a ranking system would need to be well justified. Additional topics discussed regarding endpoint definitions included: understanding the effect of co-morbidities on endpoints; variability in endpoint definitions in different populations (i.e., children, DS vs MDR/XDR); optimal analytic approaches to handling multiple events (i.e., when there is both retreatment and subsequent death); the importance of continuing follow up after first event; the misconception that adding components to the composite endpoints will improve power; and the statistical importance of non-TB related outcomes in the setting in which regimen efficacy is high.

Overall, there was an agreement that when composite outcomes are used for Phase III clinical trials, they need to be clearly defined, with rationale provided for the components included. Additionally, consensus was reached that composite outcomes should be collected and reported using standardized approaches and in a manner that would allow for subsequent disaggregated analyses.

Controls for MDR/XDR Trials

In terms of selection of controls for MDR/XDR TB trials, participants first expressed concerns regarding recent trial designs in MDR/XDR TB that added-on a medicine to an "optimized background regimen" (OBR), comparing outcomes to an OBR with placebo. This approach was viewed as expensive, inefficient, resulting in multiple and variable composition of control regimens being included in the comparator arm. From a regulatory point of view, it was re-emphasized that a control arm ought to be included whenever feasible and that it should represent the standard of care[a]. It was noted that use of historical controls may be acceptable if adequately justified and only under exceptional conditions; examples include: 1) ability to establish comparability of the treatment and historical control group; 2) ability to reproduce

[a] http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E10/Step4/E10_Guideline.pdf

inclusion/exclusion criteria and efficacy and safety endpoints in historical controls; 3) a large treatment difference expected to compensate for possible bias; 4) ancillary treatments similar in treated patients and historical controls; and 5) reliable estimates of outcomes in the historical control group. The merits and challenges involved with using the WHO-endorsed longer duration (up to 24 months) regimen for MDR-TB as a control arm were discussed. It was acknowledged that the use of a control regimen with variability across sites in terms of composition of drug components, the inherent complexities of a control regimen requiring injectables, and the long overall duration of this control regimen with the implications its use has on total follow-up duration of follow-up are notable areas of concern. It was noted that in the endTB trial (NCT02754765), some participating countries are using the shorter-course regimen for a subset of patients, following WHO revised guidelines (2016).[a] Overall, meeting participants agreed that retaining flexibility in control arm selection is reasonable and may allow for designs that are more pragmatic in terms of strategy comparisons to the experimental regimens, however, the implications of such approaches on the trial statistical considerations as well as the final interpretation of the trial results need to be carefully considered from the outset. Until there are improved regimens identified for MDR-TB, or new recommendations being issued, the WHO-endorsed 9-month shorter course regimen was proposed as a reasonable control to use in trials when populations eligible for that regimen are studied. It was acknowledged that the preliminary analyses of the STREAM Stage 1 trial could not confirm non-inferiority of the shorter-course 9-12 month regimen when compared with the standard 24-month regimen. It was noted, however, that the shorter regimen still achieved relatively high cure rates in STREAM Stage 1, that it is already used in many settings as a standard of care, and that the implications of not meeting protocol-specified non-inferiority may differ across the perspective of patients, clinicians, policy makers and trial investigators, as noted in the recent WHO position statement on this regimen.[b]

[a] World Health Organization. WHO treatment guidelines for drug-resistant tuberculosis – 2016 update (WHO/HTM/TB/2016.04). Geneva, World Health Organization, 2016. http://apps.who.int/iris/bitstream/handle/10665/250125/9789241549639-eng.pdf?sequence=1

[b] WHO. Position statement on the continued use of the shorter MDR-TB regimen following an expedited review of the STREAM Stage 1 preliminary results. http://www.who.int/tb/areas-of-work/drug-resistant-tb/treatment/WHOPositionStatementShorterRegimensSTREAMStage1.pdf

Table 3. ***Session 3.1: Novel trial designs***

Question	Consensus	Options	Research
What novel trial designs are appropriate for phase III trials in TB drug/regimen development? Are there ways that non-inferiority can be avoided? **Can and should clinical trials be designed to show superiority?**	Innovative, efficient designs (e.g., adaptive strategy designs) should be further explored for TB drug and regimen development. Many have the potential to accelerate and enhance ability to learn. Additional expertise will be required to support interpretation for guidance and implementation. Non-inferiority & superiority designs both remain relevant for studies of new regimens. Whether studies should be designed as non-inferiority or superiority studies depends on the characteristics of the drugs and regimens and on their intended use and value propositions, e.g. shortened duration of treatment. Non-inferiority *margins* should be clinically and programmatically meaningful. Justifications for their selection should be clearly stated in the protocol. The protocol should include assessment of constancy assumption, variability of control rates based on historical data, justification for and power implications of assumptions for true rates for experimental and control groups. Pre- and post-engagement of providers and patients would be helpful to understand their	Status of per-protocol (PP) analysis in non-inferiority trials: modified intention-to-treat (MITT) as primary population instead of co-primary MITT and PP ? Can follow-up be shortened for subset of patients who are the last enrolled (i.e., while nearly all participants will be followed for full follow-up [e.g., 1 year] and the participants enrolled in the last 6 months will be followed for less time [e.g., 6 months]) ? How to control for bias?	How to integrate innovative designs: • adaptive designs, including seamless phase II/III designs and adaptive randomization? • study designs involving multiple treatments that vary different factors, and methods of analysis for assessing the optimal outcome across the response surface of the outcome of interest for the factors varied. Methods for summarizing findings from non-inferiority trials, including use of posterior probabilities to estimate probability of a range of outcomes. How to avoid “bio-creep” in non-inferiority trials? Can being explicit about factors that inform choice of non-inferiority margin sufficiently inform interpretation and help to avoid it? Assess the use of “number of patients needed to treat” to try and define optimal margin of non-inferiority.

	preferences and values, to assist in the determination of the non-inferiority margin. When composite outcomes are used, information on all their components should be collected, disaggregated and reported. In addition, final outcomes should be reported for study participants who leave a trial early. Additional efforts should be made to prevent missing data. Alternative methods should be used for handling missing data in analysis so as to avoid bias, and/or inconclusive results. Examples include: • avoid simply excluding patients considered to have "unassessable endpoints" (e.g., non-adherence, lost to follow-up, discontinuation of treatment); • anticipate loss to assessable population by increasing sample size; • use statistical methods to consider impact of missing data (e.g., multiple imputation, censoring) and carry out sensitivity analyses. Where possible, ensure in the trial design the ability to separate the effects of a new combination of drugs from the effect of the variation in the duration of treatment, i.e. include different durations.	Components to consider in composite outcomes: safety and efficacy or just efficacy? Limit composite outcomes to TB-related components – how to best do that considering co-morbidities? Components to consider in composite outcomes: safety and efficacy or just efficacy? Limit composite outcomes to TB-related components – how to best do that considering co-morbidities? Should non-TB-related deaths or all deaths be considered as part of trial endpoints?	How to weight or define a hierarchy of the components of a composite outcome and construct a test statistic for comparing groups? When are co-primary endpoints appropriate for trials as an alternative to a composite outcome with multiple components? How to incorporate observed non-adherence into the analysis of non-inferiority trials? How the non-inferiority margin should be determined using data from earlier trials if the primary composite endpoint includes components that relate to efficacy, safety, and study conduct (e.g., losses to follow-up)". When to consider time to event versus binary outcomes? How to incorporate measures of lung function into trial endpoints especially for trials studying shortening or treatment duration? How can interim results be used to inform subsequent research without causing bias and impacting the integrity of the trial?

What are appropriate controls for MDR/XDR-TB regimens? Should we be comparing the effect of one treatment regimen to another rather than the potential added benefit of a new single agent?	Seek alternatives (e.g., substitution) to trials that add-on a single new drug to a background regimen, as this latter design has shown its limitations.		
In what patient populations, and under what circumstances would the use of historical controls be acceptable to WHO and regulatory bodies?	Phase III trials should include randomized controls. Use of historical controls may be acceptable if adequately justified and under *exceptional* conditions. Regulatory bodies provide guidance on such situations.[a]		
Should we have different considerations for registrational trials compared to trials conducted to address public health needs? Can trials be designed to address both the public health and registration needs?	*A single Phase III trial (on a regimen or drug) cannot answer <u>all</u> relevant questions.* Both explanatory and pragmatic trials are needed to answer questions about efficacy and safety (explanatory) and about expected effectiveness in programmatic conditions (pragmatic). - endpoints should be specific to the purposes. - composite endpoints should be disaggregated and reported. - endpoint definitions should be harmonized to allow for pooled analyses.	The number of sites involved in trials should be increased (this will decrease time to results and need for adjustment to changes in standard of care).	How to take into account via modeling the potential public health benefit of an intervention in helping defining the non-inferiority margin?

[a] Examples include:
- ability to establish full comparability of the treatment and historical control groups;
- ability to reproduce inclusion/ exclusion criteria and efficacy and safety endpoints in historical controls;
- a large treatment difference expected to compensate for possible bias;
- ancillary treatments similar in treated patients and historical controls; and
- estimates of outcomes in the control group are reliable and unlikely to have changed over time.

	- pragmatic trial designs could consider cluster randomization and appropriate endpoints.		
When the standard of care changes while a trial is underway, how can the scientific integrity of the trial be preserved, particularly with regard to choice of control(s)?	Anticipate changes in standard of care during planning and execution of a trial and have a plan for responding to such changes if warranted. The number of sites involved in trials should be adequate to allow timely completion of the trial. This will reduce the time needed for patient enrollments and to obtaining results as well as the risk of needing to adjust control arms in accordance with changes in standard of care.		

❖ *Sub-session 3.2: Measuring and maximizing adherence.*

Facilitators: *Dr. Andrew Vernon (CDC) & Dr. Lori Dodd (NIH)*

Key-Note: *The importance of maximizing adherence in clinical trials and its practical relevance in explanatory and pragmatic trials. Are we paying enough attention? - Dr. Rada Savic (UCSF)*

Discussant: *Are alternative techniques and digital health approaches (e.g. video-observed treatment) ready for use in place of traditional in clinic directly observed therapy (DOT) both for measuring and optimizing adherence and patient support? - Dr. Katherine Fielding (London School of Hygiene and Topical Medicine)*

Top uncertainties/questions addressed at the Technical consultation:

1. *Should adherence be considered similarly for various types and phases of trials? (e.g., need to discriminate between trials whose role is to define regimen efficacy vs. trials focused on regimen effectiveness in practical implementation; approach to "per protocol" analysis in the context of non-inferiority trial designs).*
2. *What methods are currently available to ensure adherence in clinical trials? What methods have been shown to be reliable in which populations?*
3. *In what ways should TB trials seek to incorporate and contribute to development/validation of electronic methods for measuring adherence? How to link information on adherence to individual PK/PD and to individual outcomes?*
4. *Until more information on valid electronic methods is available, what is the role of traditional in-person DOT? How should it be implemented to optimize the validity of trial data?*
5. *Do currently available data allow determination of "significant" adherence failure? If not, how might such data be obtained?*
6. *What evidence supports the concern that non-adherence contributes to acquisition of drug resistance? What measures might be incorporated into trials to help resolve this question?*

The critical importance of treatment adherence in relation to regimen efficacy and effectiveness was fully acknowledged and discussed (see **Table 4**). Given that the TB therapeutics field is seeking treatments of shorter duration, the importance of *measuring* and *ensuring* adherence is further accentuated. It was noted that the necessary degree of rigor in ensuring adherence may vary in relation to the phase of drug development. For example, in Phase I and Phase II studies complete adherence is imperative to understand the anti-mycobacterial, safety and tolerability characteristics of a drug or regimen under idealized conditions, whereas in Phase III studies, the specific objective, and questions of efficacy vs. effectiveness, could influence decisions on how rigorously adherence is maintained, with the imperative that adherence be rigorously measured. The existence and importance of different patterns of poor adherence (e.g., irregularity of timing

of a daily dose vs sporadic missed doses vs multiple sequential missed doses) was acknowledged[a]. How these various patterns might affect treatment outcomes for a given regimen, and in a given phase of treatment, is, however, not well understood. It was suggested that adherence should be measured as precisely as possible so that outcomes can be quantified relative to this important variable. It was argued that adherence to therapy is especially important in the context of "Per Protocol" analyses. Discussion addressed the threshold for defining "per-protocol" (i.e., highly adherent) populations. The point was made that a common threshold for defining "protocol-correct" adherence used in contemporary clinical trials (i.e., that *at least 75-80% of doses* are completed within a defined period of time) is both arbitrary and misleading. The pooled analyses in TB-ReFLECT indicate that adherence is strongly associated with outcome. This suggests that data analyses ought to be pursued for all experimental regimens to determine the impact of varying levels of reduced adherence, and the thresholds needed to achieve target efficacy. In addition, it was suggested that differing patterns of non-adherence should also be considered when defining "protocol correct" completion of therapy. Several experts noted that adherence is a post-randomization observation; its inclusion violates the balance introduced by randomization, and potentially biases results. It was suggested that one might adjust for adherence using baseline (i.e., pre-randomization) covariates predictive of adherence, thereby respecting the randomization principle while controlling for this important element. Prof. Neaton pointed out that adherent participants may differ fundamentally from those who are non-adherent, in terms of risk for poor outcome; as one example, he cited data from the Coronary Drug Project trial, in which mortality was lower in adherent vs non-adherent participants in both the test and control arms.[3]

The evaluation and integration of novel digital technologies, such as those presented by Dr. Fielding, as part of future trials were generally endorsed by meeting participants, since such tools can provide highly granular adherence measurements (e.g. daily electronic documentation tools) for analyses. However, it was noted that many of the digital technologies are still in testing stages, and their role in improving adherence and influencing outcomes as part of larger Phase III trials remains uncertain. Overall, the meeting participants supported greater rigor in measuring and analyzing adherence in clinical trials. The choice of methods for collecting and enhancing adherence, inclusive of digital or other non-classical options, should be left to developers.

[a] see Blaschke et al., Annu.Rev.Pharmacol.Toxicol 2012; 52:275-301

Table 4. ***Session 3.2: Measuring and maximizing adherence***

Question	Consensus	Options	Research
Should adherence be considered similarly for various types and phases of trials? **(e.g., need to discriminate between trials whose role is to define regimen efficacy vs. trials focused on regimen effectiveness in practical implementation);**	Ensuring and measuring adherence in clinical trials is essential to correctly interpret results of the trials. Improvements in adherence monitoring and reporting are imperative. New, better methods for assessing adherence are becoming available. These will help improving our (as yet) limited understanding of the impact (on outcomes) of varying adherence patterns. The degree of strictness on ensuring adherence to treatment depends on the phase of drug development. For example, in Phase I and Phase II studies, strict adherence is imperative to make decisions on regimens to move forward to late stage development. In Phase III studies, the objectives of the trial would drive the decision on adherence implementation and measurement. In any case, it is imperative to measure and report adherence to understand performance. Efforts to enhance adherence should follow feasibility in each setting.	Differing views on per-protocol analysis and role of adherence, especially in the context of non-inferiority trials. Caution is advised when adjusting for post randomization variables like adherence to define per-protocol populations (e.g. Coronary Drug Project [3]).	Differing views on efficacy vs. effectiveness. Some felt it important to assess efficacy in highly adherent patients, while others felt that this was well controlled through randomization. Concern about implementing effectiveness research in a non-inferiority setting was expressed.

What methods are currently available to ensure adherence in clinical trials? What methods have been shown to be reliable in which populations?			A large number of novel methods for assuring adherence have been developed fairly recently, but none has been convincingly validated either as a way to measure adherence or as a means for improving it. The group strongly supported further research to evaluate these new tools and methods to improve and monitor adherence to treatment. Clinical trials offer an excellent platform for sub-studies in this area.
In what ways should TB trials seek to incorporate and contribute to development/ validation of electronic methods for measuring adherence? How to link information on adherence to individual PK/PD and to individual outcomes?			The group strongly encouraged further studies on how TB trials should seek for linking information on treatment adherence to individual PK/PD and individual outcomes.
Until more information on valid electronic methods is available, what is the role of traditional in-person DOT? How should it be implemented to optimize the validity of trial data?		Differing views on application of in-person DOT in trials and use of various types of adherence support. Better means to measure adherence and its association with outcomes would contribute usefully to this discussion.	
Do currently available data allow determination of "significant" adherence failure? If not, how might such data be obtained?		The group recognized the difficulty to define "significant" adherence: this would depend on multiple factors specific to each trial setting, including the component drugs of the regimen, the dosing schedule, the PK of the individual drugs, and other risk factors and co-morbidities which could influence the risk of treatment failure/relapse.	Possible methodological approaches to adjust for variable adherence using pre-randomization variables, considering the notion of "forgiveness".

		Embedded in this discussion was consideration of the concept of "*forgiveness*" of a regimen, which would be defined by the types and amount of *non-adherence that would not substantively alter the likelihood of treatment success*. Some felt this aspect should be reflected in the regimen's efficacy and required no other adjustment. Others suggested this aspect should be taken into account in the determination of the non-inferiority margin.	
What evidence supports the notion that non-adherence contributes to acquisition of drug resistance? What measures might be incorporated into trials to help resolve this question?			Further research is needed as the association of non-adherence with acquired drug resistance is recognized in publications, but the mechanisms underlying the association are largely speculative.

❖ *Sub-session 3.3: Addressing special populations*

Facilitators: *Dr. Monique Surette (European & Developing Countries Clinical Trials Partnership) & Dr. John Johnson (Case Western Reserve University)*

Discussant 1: *Inclusion of special populations in clinical trials: current recommendations and barriers to implementation – with a focus on children. Dr. Anneke Hesseling (Stellenbosch University).*

Discussant 2: *Considerations for investigation of treatment in pregnant women- Dr. Amita Gupta (Johns Hopkins University).*

Discussant 3: *What would make the inclusion of special populations easier for researchers – with a focus on HIV affected populations? - Dr. Michael Hughes (Harvard University)*

Top uncertainties/questions addressed at the Technical consultation:

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

There was general consensus on the value and importance of including special populations in TB treatment trials whenever possible. Some special populations such as pregnant and lactating women and children have been under-represented or excluded from trials due to perceived risks, acceptability to Institutional Review Boards, and limited funding. Specific aspects are discussed below (see **Table 5**).

Pregnant and lactating women

The limited amount of data available to guide the care of TB in pregnant and lactating women is a major shortcoming in TB therapeutics. However, it was agreed that in order to inform how best to include pregnant women routinely in clinical trials, more data from non-trial conditions would be helpful. One approach to gathering real-world data in this population is through a global registry where a set of pre-determined, baseline data are collected from pregnant women treated for TB by routine programmes. Dr. Gupta expressed her desire to have such a registry but noted that such an effort requires long-term support and funding that is currently unavailable.

Another approach is to collect data from women who become pregnant while participating in a clinical trial and receiving study drugs. Current trial practice is to discontinue study drugs at the time a participant is identified as being pregnant and define them as 'unassessable'[a]. Instead, Dr. Gupta proposed that newly pregnant participants be consented again, reviewing all current information about the drug or regimen during pregnancy including any shifts in risk-benefit balance, thereby offering them the option to continue study drug treatment unless it is known that the drugs in the regimen are teratogenic. Examples of such secondary consent forms for participants who become pregnant have already been developed and used in some clinical trials.[4] Dr. Gupta led a discussion around the justification and ethical basis that have historically been used for excluding pregnant women from clinical trials in general. Concerns of potential harm to the women and/or fetus were raised, including an increased frequency or severity of adverse events or new adverse events occurring in pregnant women, and the potential unpredictability of these adverse events in pregnant women compared to other populations. Meeting participants asked what efficacy, PK outcomes and information would be feasible and important to collect in pregnant women? Dr Gupta acknowledged that studies involving pregnant women will likely be inadequately powered to detect differences in efficacy in a separate group analysis but this can possibly be included as a secondary objective. However, she advocated that both PK and outcome data be collected for pregnant women so that all data, whether from trials or registries, can be pooled for analysis once sufficient data have accumulated. She encouraged the TB community to engage and commit to making changes to transform regimen development for the prevention and treatment of TB during pregnancy. She proposed that protocols under development be shared for review by experts in the care of TB in pregnant women and maternal-fetal medicine specialists, bioethicists who can further comment on the risk and benefits of potentially including pregnant women in the trial during its planning stage, local IRBs, as well as engaging pregnant women from the community.

Children

Despite the fact that children are being more routinely included in contemporary clinical trials as compared to prior decades, meeting participants acknowledged that there is room for improvement in how and when to optimally include children in TB trials. Reference was made to an international consensus statement from clinical investigators, funders and regulatory authorities on the earlier inclusion of children in TB trials, which highlights key aspects of trial design, the timing of inclusion of children and practical and ethical considerations regarding children's inclusion in TB trials[b]. If children are to be included in adult trials, additional and different inclusion and exclusion criteria may be needed due to the differing clinical features and

[a] Note: the outcome of the pregnancy is usually followed and reported.

[b] Nachman S, Ahmed A, Amanullah F, et al. Towards early inclusion of children in tuberculosis drugs trials: a consensus statement. *Lancet Infect Dis.* 2015 Jun;15(6):711-20

diagnostic challenges of TB in children. In addition, the definitions of *unfavorable outcome* used in determining study endpoints require careful consideration as they are likely to differ from adults. It was agreed that adolescents (12 years and older), who have disease characteristics similar to adults, should be routinely considered for inclusion in adult phase IIb and III trials. Legal requirements for the participation of children in clinical trials vary by country. It was agreed that, where feasible and justified through consultations with pediatricians, ethics committees, trial methodologists, and regulatory bodies, the inclusion of children should be carefully considered and supported during protocol development. For very young children, meeting participants acknowledged that large clinical trials may not be feasible, but noted the immense value of having access to PK data in these subgroups to inform policy guidance on new drugs and regimens deemed to be safe and efficacious in adolescent and adult populations.

Persons living with HIV

The care of HIV-infected patients and the optimal use and timing of initiating ART during TB treatment has dramatically evolved in recent years, based on high-quality evidence from large international clinical trials. As HIV treatment has improved, the impact of HIV-co-infection on treatment outcomes during trials of new anti-TB drugs and regimens is less certain. Indeed, treatment outcomes in HIV-co-infected patients may be highly dependent on the specifics of the management of ART. It is important to understand whether mortality or other poor outcomes during TB treatment in HIV co-infected patients is related to HIV or TB. If feasible, stratification based on HIV status is encouraged, recognizing that this approach results in larger sample sizes if the goal is to evaluate outcomes in each stratum. There was clear consensus on the critical importance of integrating HIV care into TB treatment programmes, realizing that HIV is managed separately in many TB clinical trials sites and that the two programmes are not optimally linked. It was proposed that more training and resources be committed to assure that clinical sites are capable of providing high quality care for both TB and HIV. It was also suggested that protocol teams should have members with HIV expertise to help guide trial design decision-making and to help identify any special protocol implementation considerations in the context of HIV co-infection.

Table 5. ***Session 3.3: Addressing special populations***

Question	Consensus	Options	Research
Aside from use of well-designed trials based on solid pre-clinical data, conducted under the protections outlined in existing regulations, what are the biggest barriers to including special populations in clinical trials? What approaches or measures might stimulate greater inclusion of special populations in trials, including greater community engagement and awareness?	The safety of new drugs administered during pregnancy is problematic. Data on the safety of first and second line anti-TB drugs and new agents is limited. More information is needed through: - a systematic review of the current knowledge about the prevention and treatment of TB in pregnant women; - establishment of an *international registry* to systematically collect and report data on maternal and fetal outcomes for pregnant women treated for TB. Women who become pregnant while on clinical trials should be followed up for maternal & fetal outcomes & adverse events. Many are now switched to standard therapy. The alternative of re-consenting study participants who become pregnant after fully informing them about any additional risks & benefits and providing them the option to continue on study treatment & follow-up (i.e. dual consent) should be further explored. Regardless of whether study drugs are continued or participants are switched to standard therapy, female participants who become pregnant should be fully followed-up according to the trial protocol schedule of events, unless the participant withdraws consent.		Additional studies of the PK & safety of first-line, second-line, and new anti-TB drugs in pregnant women are needed. More studies are needed to determine the optimal time to start treatment of LTBI during pregnancy and the best drugs and regimens to use. Further studies of the treatment of MDR- and XDR-TB in pregnant women and how to treat pregnant women and their children who are close contacts of patients with MDR- and XDR-TB are greatly needed. Further phase III studies of shorter treatment of LTBI in children, whether contacts of DS or MDR-TB cases, are needed. For children, separate efficacy studies are required for the treatment of less frequent forms of disease where the response or safety is expected to differ from adults. Regulatory guidance and algorithms for such situations exist. More sensitive diagnostic tools using sputum-free systems for

What would make the inclusion of special populations easier for researchers?	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. Formal ethics consultation should be considered early in drug development and trial planning to advise investigators about the inclusion of pregnant women in TB clinical trials. Experts in obstetrics and neonatology should be consulted about what information is needed and data collected to inform inclusion of pregnant women in TB trials Appropriate formulations of drugs for infants & young children should be developed during the early phases of regimen development and testing, whenever feasible. For HIV-infected individuals, drug-drug interaction studies between anti-TB drugs & relevant ARV drugs should be conducted earlier than Phase III, whenever feasible. This can be done either separately early in drug/regimen development or nested within trials. There needs to be flexibility within a trial and the product development plan to assess the drug-drug interaction of new ARVs (e.g. dolutegravir) that are introduced after a trial has begun.		paucibacillary TB for children are needed.
What special considerations need to be taken into account to include special populations into trials? Can they be included as an additional	PK and safety studies are needed to inform clinicians, programmes and regulatory authorities about the use of new drugs and regimens for the treatment of TB in children. These studies may be initiated	Stratify randomization by HIV-, HIV+/high CD4, HIV+/low CD4 to avoid imbalances between randomized arms in these risk groups.	

arm of a study, or as part of a larger patient group?	earlier – as soon as adequate information is available about the safety and required dose-exposure in adults. It is important that there is expertise in management of HIV and other co-morbidities at sites and in protocol teams. Good management of both HIV & TB care is essential. This requires good TB and HIV laboratories, and close collaboration between HIV and TB caregivers during TB trials. 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.		
At what phase is it most appropriate to include special populations?	PK & safety studies provide sufficient evidence to inform the treatment of limited forms of TB in children with new drugs or regimens. These trials should be initiated earlier (e.g. phase IIb) - as soon as adequate information is available about the safety & needed dose-exposure in adults. In general, TB disease and drug handling in adolescents are similar to adults. Adolescents (> 12 years) should be included in contemporary TB clinical trials. Appropriately treated HIV+ patients with TB or LTBI should be enrolled in all TB clinical trials.	Inclusion of pregnant women in phase III clinical trials is recommended but concerns are expressed about trial insurance/indemnification & whether this would be acceptable to local institutional review boards & sponsors.	
Areas where special populations are included should be prioritized based on burden. What are these priority areas and what are the requirements for each population?		Consider enrolling pregnant women in current phase III trials if they are intolerant to - or have major contraindications for currently used regimens for treatment of TB during pregnancy	

Session 4:
The interplay between trials and guidelines: the importance of sound evidence to inform policy guidance and clinical practice.

***Facilitators**: Dr. Sumathi Nambiar (FDA) & Dr. Michael Rich (Partners in Health)*

***Key-note:** Do the trial considerations that serve the objectives of registration meet the needs for development of public health guidance? – Dr. Christian Lienhardt (WHO/GTB)*

***Discussant 1:** Outcome definitions in clinical trials – should they vary to fit regulatory and programmatic decision-making needs? - Dr. Andrew Vernon (US CDC)*

***Discussant 2:** The point of view of the programme managers and end-users on the use and translation of WHO guidelines into national strategy plans Programme Managers: Dr. Nguyen Viet Nhung, Dr. Alena Skrahina, Dr. Norbert Ndjeka*

***Discussant 3:** The point of view of the regulator – Dr. Marco Cavaleri (EMA)*

***Discussant 4:** Which evidence and criteria should guide selection of medicines candidate to WHO guideline recommendations and the Model List of Essential Medicines (EML) - Dr. Lorenzo Moja (WHO/EMP)*

Top uncertainties/questions on this topic addressed at the Technical consultation:

1. *How will emerging outcomes, from measures/endpoints used for supporting trial adaptation, to registration endpoints be integrated into the standards for policy making?*
2. *What clinical trial outcomes are required to inform regulatory and programmatic decision-making need to be prioritized for prospective implementation in novel trial designs?*
3. *How can current/novel clinical trial endpoints that are intended to support regulatory decisions be subsequently translated to support programmatic implementation?*
4. *What are the gaps in the application and structure of current standards and how should this be addressed?*
5. *Should the assessment of clinical trial outcomes be updated for harmonization across regulatory and programmatic objectives, and if yes, how?*
6. *How to ensure that trial data at individual patient level can be pooled for enhanced meta-analysis when reviewing evidence for policy making by WHO and other professional bodies?*

The interplay between trial objectives and designs with the evidence base required for policy decisions and clinical practice guidance development was discussed (see Table 6). Participants discussed how trial designs and results that are used as the basis for a drug approval may not be sufficient for policy making decisions. Dr. Davies noted the difficulty of aligning endpoints and patient populations with pragmatic implications for public health authorities. Caution was also advised on attempting to address too many questions in a single trial. There was, however,

general agreement that more collaboration between drug and regimen developers, regulators and policy makers is needed to better inform regulatory and public health objectives and assist in decision-making. It was proposed that the 'pre-specified multiple analyses' mentioned by Dr. Vernon be standardized in this regard. Many noted concerns about the risks of multiple testing and over-interpretation of clinical trial data but most agreed that pre-specification and use of hierarchical approaches when performing multiple analyses are viable options. It was noted by Dr. Nahid that multiple, secondary analyses are already performed by the WHO when individual patient-level data are requested and disaggregated for re-analysis (according to WHO defined outcomes of interest) as part of the GRADE based approach to formulation of recommendations. As such the value of harmonizing endpoints definitions and ensuring that composite endpoints can be disaggregated across TB clinical trials was again emphasized. Such harmonization would better 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.

The value of conducting operational research both as a means to generate additional evidence beyond clinical trials and to aid translation of WHO guidance into practice was discussed. Dr. Rich encouraged the group to support programmes that conduct operational research. Programme managers agreed, and noted that the severity and disease burden seen at their settings already compels them to conduct operational research. It was proposed that the group endorse development of standardized approaches and methodologies for the conduct of operational research.

Dr. Nahid and others noted that policy recommendations, even in situations when certainty in the evidence is very low, are generally considered as defining the standard of care. As a consequence, the design and conduct of trials that would address the relevant knowledge gaps left with this uncertainty and that would improve the evidence base become challenging - including convincing donors to fund such research and for ethics committees to approve it. There was general agreement that WHO has responsibility to clearly state for every recommendation based on very low or low certainty in the evidence that significant knowledge gaps remain and that additional research is needed. Dr. Lienhardt proposed that conditional recommendations be associated with a statement on the need for further research – for example: *"considering that, after careful assessment of available evidence, recommendation is conditional due to low-certainty in the quality of evidence, further funding for and conduct of research is encouraged to enhance availability of high-quality evidence*". There was general agreement that data standardization and data sharing using existing and endorsed standards are imperative in this regard. This will ensure that trial data at individual patient level can be pooled for enhanced meta-analysis.

There was discussion around the possibility of WHO offering the option of consultation during protocol development, perhaps through the *Task Force on Introduction of New TB Drugs and Treatment Regimens,* or other mechanisms. It was argued that offering consultation without explicit commitments brings value in aligning interest for post-approval and programmatic studies. Discussion in prior sessions noted the immense value of more dialogue and engagement across regimen developers, regulatory and policy making bodies. However, how best to engage in this domain remained unclear, particularly if a sponsor faces divergent opinions between regulatory and policy recommending bodies in regard to protocol design features. The participants proposed exploring ways to seek general input and comment from policy recommending bodies, particularly for post-approval studies and to increase interaction with regulatory authorities.

Table 6. ***Session 4: The interplay between trials and guidelines: the importance of sound evidence to inform policy guidance and clinical practice***

Question	Consensus	Options	Research
How will emerging outcomes from measures/endpoints used for supporting trial adaptation to registration endpoints be integrated into the standards for policy making?			Treatment success outcomes in recent trials of MDR-TB were much higher than that reported in prior trials and across programme settings. Additional research is needed to better understand the performance of the standard of care for rifampicin-susceptible and resistant TB in various conditions and settings to aid in the design of future studies.
What clinical trial outcomes are required to inform regulatory and programmatic decision-making and need to be prioritized for prospective implementation in novel trial designs?	A single clinical trial cannot fully address all relevant regulatory and policy/public health questions. Explanatory trials, novel adaptive trials, pivotal trials for licensure need to be followed up with pragmatic trials to understand the optimal use of new drugs and regimens.	Consider post-authorization studies to answer some of the questions that cannot be addressed in the registrational trial(s) to help bridge gaps in knowledge.	
How can current/novel clinical trial endpoints that are intended to support regulatory decisions be subsequently translated to support programmatic implementation?		Operational research in parallel with clinical trials can aid to translate clinical trial outcomes to WHO guidance and add evidence for better programmatic implementation. Often patients enrolled in trials are not reflective of the general population; consider ways to make trial population more reflective of the population of patients who will be receiving treatment in real life. Also consider pragmatic studies for better evidence on programmatic implementation	
What are the gaps in the application and structure of current standards and how should this be addressed?			

Should the assessment of clinical trial outcomes be updated for harmonization across regulatory and programmatic objectives, and if yes, how?	Communication between drug/regimen developers, regulators and recommendation bodies is essential and should be encouraged and facilitated as early as possible at design stages.	Approaches to collecting clinical outcomes data that can potentially address assessment of safety and efficacy of the product and answer questions that are important from a programmatic perspective should address the followings: • secondary/exploratory analyses are an option – but caution in over-interpreting the data. • sample size implications if multiple primary analyses considered. • Importance of pre-specifying analyses. • caution about trying to answer all questions from a single trial. Consistent definitions across different trials is needed; limitations of using surrogate endpoints (eg. culture conversion) for development of guidelines.	
How to ensure that trial data at individual patient level can be pooled for enhanced meta-analysis when reviewing evidence for policy making by WHO and other professional bodies?	Data should be collected using standard definitions and use of data standards for trial data are essential; Clinical trial data should be made available for sharing to conduct individual patient-level analyses. Such databases are used by WHO and other recommending bodies in policy making. GRADE method should be well understood by all stakeholders	As data quality improves, recommendations based on lower quality data should be re-examined. A relevant process to address this should be established.	

Technical consultation Wrap-up

Proposal for the consensus document

All facilitators presented the proposals for the consensus document based on the presentations and discussions from each session, and further final discussion ensued. Group consensus was reached in a number of points, summarized in **Tables 1** through **6** above. These tables present, for each session, the pre-established questions that were posed to the group and the areas in which consensus was reached among participants. For the questions and areas in which additional debate and alternative options still exist, or for which further research is still needed, these are indicated under the respective "alternative option" and "research" columns.

Closing statements

Dr. Nahid thanked the participants for their role in this unprecedented technical consultation that brought together many stakeholders representing various interests and groups. The presentations, reflections and discussions on recent clinical trials were impactful and will move clinical trial design and approaches to TB regimen development forward. He noted that the discussions, many of which resulted in areas of consensus, were highly engaging, also identifying numerous areas for additional research. Dr. Lienhardt seconded this conclusion and thanked all participants for an extremely stimulating and thought-provoking technical consultation. He then presented the next stages and timeline, and proposed that a special journal issue be prepared that summarizes the technical consultation proceedings for publication by early 2019.

The Director of WHO Global TB Programme, Dr. Tereza Kasaeva, closed the technical consultation by thanking all participants and organizers of the technical consultation. She acknowledged the challenging but impressive objective of developing expert consensus on advances of clinical trial design for TB regimen development.

References

1. Imperial M, Nahid P, Phillips PPJ, et al. A Patient-Level Pooled Analysis of Treatment Shortening Regimens for Drug-Susceptible Pulmonary Tuberculosis. To be published. *Nature Med* 2018
2. Eisenach K, Demers A-M, Jones F. Mycobacteriology Laboratory Sourcebook For Harmonization and Support of TB Clinical Trials. Version 1.0. http://resources.psmile.org/resources/process-control/section-specific-information/microbiology/mycobacteriology-tb-source-book/. Published 2018.
3. Coronary Drug Project Research Group. Influence of Adherence to Treatment and Response of Cholesterol on Mortality in the Coronary Drug Project. *N Engl J Med*. 1980;303(18):1038-1041. doi:10.1056/NEJM198010303031804
4. Gupta A, Mathad JS, Abdel-Rahman SM, et al. Toward Earlier Inclusion of Pregnant and Postpartum Women in Tuberculosis Drug Trials: Consensus Statements From an International Expert Panel. *Clin Infect Dis*. 2016;62(6):761-769. doi:10.1093/cid/civ991
5. Merle CS, Fielding K, Sow OB, et al. A Four-Month Gatifloxacin-Containing Regimen for Treating Tuberculosis. *N Engl J Med*. 2014;371(23):1588-1598. doi:10.1056/NEJMoa1315817
6. Gillespie SH, Crook AM, McHugh TD, et al. Four-Month Moxifloxacin-Based Regimens for Drug-Sensitive Tuberculosis. *N Engl J Med*. 2014;371(17):1577-1587. doi:10.1056/NEJMoa1407426
7. Jindani A, Harrison TS, Nunn AJ, et al. High-Dose Rifapentine with Moxifloxacin for Pulmonary Tuberculosis. *N Engl J Med*. 2014;371(17):1599-1608. doi:10.1056/NEJMoa1314210
8. Conde MB, Mello FCQ, Duarte RS, et al. A Phase 2 Randomized Trial of a Rifapentine plus Moxifloxacin-Based Regimen for Treatment of Pulmonary Tuberculosis. *PLoS One*. 2016;11(5):e0154778. doi:10.1371/journal.pone.0154778
9. Burman WJ, Goldberg S, Johnson JL, et al. Moxifloxacin versus ethambutol in the first 2 months of treatment for pulmonary tuberculosis. *Am J Respir Crit Care Med*. 2006;174(3):331-338. doi:10.1164/rccm.200603-360OC
10. Rustomjee R, Lienhardt C, Kanyok T, et al. A phase II study of the sterilising activities of ofloxacin, gatifloxacin and moxifloxacin in pulmonary tuberculosis. *Int J Tuberc Lung Dis*. 2008;12(2):128-138.
11. Dorman SE, Johnson JL, Goldberg S, et al. Substitution of moxifloxacin for isoniazid during intensive phase treatment of pulmonary tuberculosis. *Am J Respir Crit Care Med*. 2009;180(3):273-280. doi:10.1164/rccm.200901-0078OC
12. Phillips PPJ, Mendel CM, Burger DA, et al. Limited role of culture conversion for decision-making in individual patient care and for advancing novel regimens to confirmatory clinical trials. *BMC Med*. 2016;14(1):19. doi:10.1186/s12916-016-0565-y
13. Shenai S, Ronacher K, Malherbe S, et al. Bacterial Loads Measured by the Xpert MTB/RIF Assay as Markers of Culture Conversion and Bacteriological Cure in Pulmonary TB. Neyrolles O, ed. *PLoS One*. 2016;11(8):e0160062. doi:10.1371/journal.pone.0160062
14. Dharan NJ, Amisano D, Mboowa G, et al. Improving the sensitivity of the Xpert MTB/RIF assay on sputum pellets by decreasing the amount of added sample reagent: a

laboratory and clinical evaluation. *J Clin Microbiol*. 2015;53(4):1258-1263. doi:10.1128/JCM.03619-14

15. Mauri L, D'Agostino RB. Challenges in the Design and Interpretation of Noninferiority Trials. Drazen JM, Harrington DP, McMurray JJV, Ware JH, Woodcock J, eds. *N Engl J Med*. 2017;377(14):1357-1367. doi:10.1056/NEJMra1510063
16. Sugarman J, Colvin C, Moran AC, Oxlade O. Tuberculosis in pregnancy: an estimate of the global burden of disease. *Lancet Glob Heal*. 2014;2(12):e710-6. doi:10.1016/S2214-109X(14)70330-4

ANNEX 1
Agenda of WHO Technical consultation
Advances in Clinical Trial Design for Development of New TB Treatments

Hotel Victoria, Glion-sur-Montreux
14-16 March 2018

Background

The Global Tuberculosis Programme (GTB) of the World Health Organization (WHO) established in 2012 a "Task Force on the development of policies for introduction of new TB drugs and treatment regimens" to guide and assist in the preparation of strategic plans and frameworks to develop recommendations for the responsible introduction of new drugs/regimens in various settings. In 2016, the Task Force developed Target Regimen Profiles for TB treatment (TRPs) intending to guide the development process towards important anti-TB treatment regimen characteristics, through wide consultations with experts and stakeholders worldwide. In line with its mandate of advising on the development and introduction of new TB drugs and regimens, the Task Force recommended that the WHO develop a consensus document to guide the research community on optimal clinical trial designs for new anti-TB drugs and regimens, in consultation with pertinent stakeholders in the field. Based on these recommendations, WHO/GTB is organising a technical consultation on Clinical Trial Design for New TB Treatments with the aim to get consensus on identifying the best trial designs to inform policy guidance on new treatment regimens for TB. Starting from the shared importance to be given to quality of evidence, the main questions the technical consultation will address is what should be the optimal characteristics of clinical trial designs for the development of new regimens for the treatment of TB that WHO could recommend, taking into account important aspects such as individual drugs PK/PD characteristics, microbiological aspects, use of biomarkers, and the effect in special populations.

Objective of the Technical consultation:
To develop expert consensus on evidence-based approaches to trial designs and use of data to inform policy guidance on new treatment regimens for TB.

Specific Objectives:

1. to review the respective strengths and limitations of current approaches for clinical development of new TB drugs and drug regimens;
2. to identify optimal practices and study designs to inform policy guidance on new drug regimens for the treatment of all forms of TB, taking into account recent developments in methods, tools and biomarkers.

Expected outcome
A Consensus Document that will indicate, with proper information and review, the optimal, agreed-upon approaches to trial design for TB treatments development.

Audience: TB drugs and regimens developers, trialists, academia, research institutions, contract research organizations, regulators, guideline developers, programme managers, nongovernmental organizations and civil society.

*: Remote presentation

Wednesday, 14 March 2018

Welcoming and Introduction		
08:00 – 8:30	Registration	
08:30 – 08:40	Welcome	Chair – Payam Nahid
08:40 – 09:00	Objectives of the meeting	Christian Lienhardt
09:00 – 09:20	Lessons learnt from the TB ReFLECT meta-analysis of Fluoroquinolone-containing regimens for the treatment of drug-susceptible TB – broad overview	Rada Savic
09:20 – 09:40	Lessons learnt on moving new drugs into new regimens for treatment of drug-susceptible and drug-resistant TB	Carl Mendel
09:40 – 09:55	The point of view of the programme managers and end-users on the results of contemporary TB treatment trials	Nguyen Viet Nhung* Alena Skrahina Norbert Ndjeka*
Session 1: PK/PD, microbiology and biomarkers – Facilitator: Rada Savic		
09:55 – 10:15	Key-Note: *What can be learnt from PK/PD studies to advance the development of new regimens?*	Kelly Dooley
10:15 – 10:30	Discussant 1: *TB Regimen Development: Bridging Translational Gaps with Quantitative Pharmacology Approaches and Drug Development Tools*	Debra Hanna
10:30 – 11:00	*Coffee/Tea break*	
11:00 – 11:15	Discussant 2: *What would be the most efficient framework for patient-level microbiology data to improve quantitative clinical PK/PD predictions and streamline model development?*	Kathy Eisenach
11:15 – 12:00	Discussion	All
12:00 – 12:30	Wrap-up: Areas of consensus / Areas of uncertainty	Facilitator
12:30 – 13:30	*Lunch*	
Session 2 : Phase II to Phase III transition – Facilitator: Michael Hoelscher		
13:30 – 14:00	Key-Note: *From early Phase II to Phase III trials: a comprehensive review of the various clinical development phases and their seamless progression to move from early bactericidal activity of single drugs to pivotal trials of drug combinations*	Gerry Davies
14:00 – 14:15	Discussant 1: *How could preclinical information influence design on Phase II studies to show efficacy of an individual drug, and are the monotherapies in Phase II A studies (EBA) needed for assessment of bactericidal efficacy and dose-finding?*	Dave Hermann
14:15 – 14:30	Discussant 2: *What information/markers/endpoints should be collected across Phase II studies to optimally select the appropriate combo regimens to move from Phase II to III?*	Martin Boeree
14:30 – 15:30	Discussion	All
15:30 – 16:00	Coffee/Tea break	
16:00 – 16:30	Discussion (cont'd)	All
16:30 – 17:15	Proposals for the consensus document	Facilitator
17:15 – 18:00	Recap and next stages	Chair

Thursday, 15 March 2018

08:45 – 09:00	*Recap of Day 1*	*Chair*
Session 3: New trial designs and how they may facilitate regimen development		
Sub-session 3.1: New trial designs – Facilitator: Carole Mitnick & Jim Neaton		
09:00 – 9:20	Key-Note: *Where have we come from and have we made any mistakes?* (High-level) review of different trials and trial designs used to date in TB phase III trials, and how designs have progressed over time – with discussion of respective strengths and weaknesses).	Andrew Nunn
9:20 – 9:35	Discussant 1: *Have non-inferiority trials served us well?* Given challenges in interpretation and limitations of non-inferiority phase III trials, what are the recommendations for the future of non-inferiority design in TB therapeutics?	Piero Olliaro
09:35 – 09:50	Discussant 2: *What are the appropriate controls* for RS and RR TB trials? How do we manage the challenges posed by the various designs on the selection of appropriate control groups, including the issue of changing standards of care?	Ed Cox
09:50 – 10:15	Q&As	All
10:15 – 10:45	*Coffee/Tea break*	
10:45 – 11:00	Discussant 3: *Where are we now and where should we be going?* Presentation of new ideas that are being proposed for ongoing and future trials – with discussion of strengths and weaknesses.	Patrick Phillips
11:00 – 12:15	Discussion : *Strengths and weaknesses of different trial designs – specifically in the context of generating evidence to inform public health guidelines (directly or indirectly), including the role of internal/historical controls, and the handling of changing standards of care.*	All
12:15 – 12:45	Wrap-up: Areas of consensus / Areas of uncertainty	Facilitators
12:45 – 13:45	*Lunch*	
Sub-session 3.2: Measuring and maximizing adherence - Facilitator: Andrew Vernon & Lori Dodd		
13:45 – 14:05	Key-Note: *The importance of maximising adherence in clinical trials and its practical relevance in explanatory and pragmatic trials. Are we paying enough attention?*	Rada Savic
14:05 – 14:20	Discussant: *Are alternative techniques and digital health approaches (e.g. video-observed treatment) ready for use in place of traditional in clinic DOT both for measuring and optimizing adherence and patient support?*	Katherine Fielding
14:20 – 15:30	Discussion	All
15:30 – 16:00	Wrap-up: Areas of consensus / Areas of uncertainty	Facilitators
16:00 – 16:30	Coffee/Tea break	

Sub-session 3.3: **Addressing special populations – Facilitator:** Monique Surette & John Johnson		
16:30 – 16:50	Discussant 1: *Inclusion of special populations in clinical trials: current recommendations and barriers to implementation – with a focus on children*	Anneke Hesseling*
16:50 – 17:00	Discussant 2: *What would make the inclusion of special populations easier for researchers – with a focus on HIV affected populations?*	Michael Hughes
17:00 – 17:10	Discussant 3: *Considerations for investigation of treatment in pregnant women*	Amita Gupta*
17:10 – 17:40	Discussion	
17:40 – 18:00	Proposals for the consensus document	Facilitators

Friday, 16 March 2018

08:30 – 08:45	*Recap of Day 2*	*Chair*
Session 4: The interplay between trials and guidelines: the importance of sound evidence to inform policy guidance and clinical practice – Facilitators: Sumathi Nambiar & Michael Rich		
08:45 – 09:00	Key-note: *Do the trial considerations that serve the objectives of registration meet the needs for development of public health guidance?*	Christian Lienhardt
09:00 – 09:15	*Outcome definitions in clinical trials – should they vary to fit regulatory and programmatic decision-making needs?*	Andrew Vernon
09:15 – 09:35	*The point of view of the programme managers and end-users on the use and translation of WHO guidelines into national strategy plans*	Nguyen Viet Nhung* Alena Skrahina Norbert Ndjeka*
09:35 – 09:50	*The point of view of the regulator*	Marco Cavaleri
09:50 – 10:10	*The point of view of the policy maker* *(1) Assessment of evidence for WHO guidelines development* *(2) Evidence needed to update the essential medicines list and to obtain WHO pre-qualification.*	Lorenzo Moja
10:10 – 10:40	*Discussion*	All
10:40 – 11:00	*Coffee/Tea break*	
11:00 – 12:00	Discussion: *The importance to produce sound evidence to inform guidelines, clinical practice and TB control programmes; will new trial designs produce better and more relevant evidence for public health needs (in addition to regulatory requirements)?*	All
12:00 – 12:30	Wrap-up: Areas of consensus / Areas of uncertainty	Facilitators
12:30 – 13:30	*Lunch*	
Technical consultation wrap-up		
13:30 – 15:30	Proposals for the consensus document	Session facilitators – Sessions 1 to 4
15:30 – 15:50	Recap on technical consultation outputs	Chair – Payam Nahid
15:50 – 16:00	Next steps	Christian Lienhardt
16:00 – 16:15	Closing	Tereza Kasaeva* Director GTB

ANNEX 2
Summary of Presentations

Context and Rationale

Key-Note 1: Lessons learnt from the TB ReFLECT meta-analysis of fluoroquinolone-containing regimens for the treatment of drug-susceptible TB- Dr. Rada Savic (UCSF)

To provide context and set the stage for the consultation, Dr. Savic outlined recent findings from the TB ReFLECT meta-analysis of fluoroquinolone-containing regimens that aimed to shorten drug susceptible tuberculosis (DS-TB) treatment to four months.[1] Although the trials independently failed to show non-inferiority of the 4-month experimental arms, 80% of patients were cured.[5–7] These trials were preceded by several Phase IIB trials, some of these showing improved 2-month culture results after treatment with fluoroquinolone-containing experimental regimens, and these improved intermediate culture results were recapitulated in the Phase III trials, but the improvements in culture conversion did not translate to predicting long term clinical endpoints (e.g. durable cure versus relapse).[8–12]

The meta-analysis aimed to: 1) identify patient groups that would be eligible for 4-month treatments, 2) profile "hard-to-treat" patient populations, 3) identify drug-specific factors predictive of unfavourable response, 4) provide direction for future clinical trials, and 5.) provide data driven evidence for potential impact on TB global goals. The analysis found that patients with minimal disease, defined as having low bacterial burden or absence of cavities and representing up to 47% of patients, are eligible for 4-month treatments. Conversely, patients with high baseline smear grade, cavitation, HIV co-infection, and low BMI (representative of malnutrition) defined the "hard-to-treat" phenotypes. Dr. Savic challenged the 'one-size-fits-all' approach to selecting TB treatment and proposed a shift in trial design to stratified medicine approaches in which patient subgroups are allocated "time-on-treatment" with greater precision. In addition, incomplete adherence, independent of treatment duration, was the most significant risk factor for unfavourable outcome, highlighting the need for optimal tools for measuring and maximizing adherence. Dr Savic further suggested that pragmatic trials that test effectiveness under field conditions will be particularly important for regimens that are less "forgiving" of reduced adherence – as observed with rifampicin-containing regimens. Other key lessons from the analysis were: (i) the need for standardized definition of unfavourable outcomes to allow better aggregation of data for pooled analysis, (ii) the importance of collection of PK data to allow further investigation of drug-specific factors beyond adherence that influence outcomes, (iii) the need for data in special populations (children, pregnant women, and HIV-coinfected patients), and (iv) the invaluable scientific benefits to the field of both data standardization and data sharing.

Key-Note 2: *Lessons learnt on moving new drugs into new regimens for treatment of drug-susceptible and drug-resistant TB – Dr. Carl Mendel (TB Alliance)*

Dr. Mendel listed the perspectives gained at TB Alliance while bringing forward new regimens for TB:

1. A new regimen must bring a value proposition, beyond efficacy or safety targets. Products with broader applications (eligible populations, etc.) gain in terms of delivery and scalability/distribution or cost, and can bring substantial impact and value that define the developmental pathway. Obtaining approval for a drug and relying on real world use and data to figure out how best to use the drug (approach with delamanid and bedaquiline) is not tenable, and more investment will be needed by sponsors and donors to understand the needs of the market and develop programmes based on those needs.
2. Volume matters and one solution is moving toward unification of treatment for DS-, MDR- and XDR-TB globally. This approach can only be done with new chemical entities, but if successfully combined, increased uptake and use of a new regimen can solve issues related to cost, supplies, and stock outs and increase efficiencies across heath care systems.
3. Transition decisions from Phase II to Phase III continue to have significant uncertainty. These limitations should be taken into account when designing Phase III trials. Large high quality trials can lead to improved translational understanding of intermediate endpoints, and new designs that integrate Phase III outcomes into Phase II trials are being developed. Other proposals have suggested skipping Phase II studies and/or combining them with Phase III. Overall, Dr. Mendel concluded that all these options have merits and that each development programme, by virtue of being unique, will need to determine the most appropriate approach to transition from Phase II to III.
4. What are the most appropriate control groups for each particular situation? Dr. Mendel used case studies to describe the four common controlled clinical trial designs: (i) randomized, placebo-controlled, double blind; (ii) randomized, partially blinded or unblinded; (iii) non-randomized with concurrent control; and (iv) non-randomized with historical control and concluded that the choice of design is dependent on the situation and question to be addressed.

Discussants: *The point of view of the programme managers and end-users on the results of contemporary TB treatment trials - Dr. Nguyen Viet Nhung (National Lung Hospital, Hanoi, Vietnam), Dr. Alena Skrahina (Republican Research and Practical, Center for Pulmonology and Tuberculosis, Minsk, Belarus), Dr. Norbert Ndjeka (National Department of Health Pretoria, South Africa)*

The programme managers presented their expectations for drug and regimen developers: 6-months or less for MDR/XDR-TB and 2-4 months for DS-TB, combinations of 4 drugs or less, pan-TB regimens, oral formulations and regimens with easy implementation, effective

extrapulmonary TB treatments, regimens with minimal adverse events and drug-drug interactions, treatments for underrepresented and special populations, and low cost regimens. In addition, they explained that guidelines relying on low quality evidence and issuing conditional recommendations make it difficult to implement new regimens, despite the urgent need for improvements in TB care. Therefore, they emphasized the need for high quality and timely evidence to support recommendations.

Session 1: Pharmacokinetic/pharmacodynamic, microbiology and biomarkers

Facilitator: Dr. Rada Savic (UCSF)

Key-Note: *What can be learnt from PK/PD studies to advance the development of new regimens? – Dr. Kelly Dooley (Johns Hopkins University)*

Dr. Dooley reviewed the framework for pharmacology-informed regimen development and listed the key knowledge needed to move forward to Phase III trials: 1) clinical PK, 2) PK/PD relationships, 3) PK-toxicity relationships (to define therapeutic margins), 4) drug-drug interactions (especially with antiretroviral treatment (ART)), 5) identification of hard-to-treat populations and quantification of PK/-PD in these patients, 6) PK/safety for dose selection in special populations and 7) PK variability (requiring dose adjustment in target patient subgroups). The optimal approach and relative importance of integrating PK/PD is defined in part by the phase of drug and regimen development. In pre-clinical and Phase I studies the aim is to get initial information on pharmacokinetics (including food effect), metabolism and induction or inhibition effects of drugs (via human hepatocyte studies), drug-drug interactions and maximum tolerated dose. In early Phase II development, semi-intensive PK sampling is performed to gain more information on PKPD with different doses and dosing frequencies. Later stage Phase II studies generally employ sparse or semi-intensive PK sampling, and in these studies we begin to learn more about hard to treat populations, sources of PK variability and impact of covariates on PK (e.g.) effect of weight on clearance (to address weight banding). Finally, in large Phase III studies, wherein there is often broad geographic diversity of sites, greater variability in adherence, and inclusion of harder to treat populations, the inclusion of sparse PK sampling accompanied with detailed dosing histories can provide invaluable data on real-world variations in drug pharmacokinetics, but more importantly, can be used to help us understand reasons for trial failures or less-than-optimal treatment responses in relevant populations. Dr. Dooley urged the TB therapeutics community to broadly integrate PK sampling schemes into clinical trials as a means to fully understand the clinical pharmacology of drugs and regimens that in turn would allow for optimized regimens and doses, inform subsequent clinical trial designs, and maximize overall efficiency of regimen development programmes.

Discussant 1: *TB Regimen Development: Bridging Translational Gaps with Quantitative Pharmacology Approaches and Drug Development Tools – Dr. Debra Hanna (Critical-Path)*

Dr. Debra Hanna underscored that with new chemical entities on the horizon, the field faces an impracticable challenge in evaluating all possible new combinations of drugs in the TB developmental pipeline (1000+ combinations), considering the large translational knowledge gaps that exist throughout the developmental process. Translational approaches that leverage PK/PD data to bridge the gaps are needed to inform regimen composition and dose selection in the early stages of development (preclinical to Phase 1) and to optimize decision making on regimens to move forward in the later stages of development (a much larger knowledge gap than the former). It was emphasized that integration of data and collaborations between several institutions is critical to development of the tools necessary to bridge the gaps. Dr. Hanna provided examples of pre-clinical tools (e.g. hollow fiber system -HFS-, murine models, physiologically-based pharmacokinetic (PBPK) and translational PK/PD models) that can improve predictions and inform decision-making in early stages of development. Lastly, a seamless Phase II/III development strategy utilizing adaptive clinical trial approaches (with a "real-time" biomarker of bacterial burden) was proposed to improve late phase efficiency.

Discussant 2: *What would be the most efficient framework for patient-level microbiology data to improve quantitative clinical PK/PD predictions and streamline model development? - Dr. Kathy Eisenach (TB or NOT TB Consulting).*

Dr. Eisenach discussed the use of microbiological data to improve quantitative clinical PK/PD predictions. First, she reviewed the areas of caution for performance and interpretation of minimum inhibitory concentration (MIC) data, namely that results vary by assays/methodologies, that there is imprecision in measurements on a single assay, that specialized laboratories are needed resulting in delays, and finally the absence of quality assessment panels. Nonetheless, MIC data are highly informative and the concept of longitudinally performing MIC testing post drug-exposure and more frequently during late phase trials was raised for consideration. In EBA studies, on- or post-treatment re-measurement of MIC following antibiotic exposure may provide information on the emergence of resistance (or reductions in susceptibility) of drugs of interest in the cocktail, and in later stage development may provide more information on reasons for failure or relapse. More sophisticated MIC testing approaches were proposed with consideration of a checkerboard MIC testing approach to study synergy, antagonism, and additive/indifferent effects of drugs in combination and the use of time-kill dynamic curves combined with PD parameters. Dr. Eisenach also discussed culture and Xpert approaches that can generate quantitative data that may reflect sputum bacterial load. Both culture (MGIT TTD) and Xpert at baseline and longitudinally have been shown to have association with treatment response.[13,14] The critical importance of attaining high-quality specimens using standardized and harmonized SOPs for respiratory specimens processing was reinforced.

Session 2:
Phase II to Phase III transition
Facilitator: Dr. Michael Hoelscher (Ludwig-Maximilians University)

Key-Note: *From early Phase II to Phase III trials: a comprehensive review of the various clinical development phases and their seamless progression to move from early bactericidal activity of single drugs to pivotal trials of drug combinations – Dr. Gerry Davies (University of Liverpool)*

Dr. Davies provided a comprehensive review of the various clinical development phases and the progression from early bactericidal activity of single drugs to pivotal trials of drug combinations. Based on a meta-analysis performed in support of the technical consultation, Dr. Davies concluded that there is reasonable support for using intermediate bacteriological endpoints such as 8 week culture conversion and median time-to-culture conversion for decision-making on which regimens to move forward to Phase III studies, though drug class-specific effects still need to be considered. He pointed out that, in practice, there has been widespread adoption of longitudinal and time to event methods in Phase IIB analysis in the last ten years, but acknowledged that further calibration of effect sizes is still needed. Even so, methods for prediction of duration of regimens from intermediate results based on meta-regression models or Phase IIC designs are more rational than current empirical approaches and reduce risk to future Phase III trial participants. Furthermore, these newer analytical approaches make simple frequentist adaptation of trials feasible though the impact of adaptation can be limited by feedback on endpoints and speed of recruitment. Dr. Davies also suggested the possibility of replacing Phase IIA monotherapy trials by using data from preclinical and Phase I studies to contribute more formally to proof of concept and planning of Phase II trials of combination therapy.

Discussant 1: *How could preclinical information influence design on Phase II studies to show efficacy of an individual drug, and are the monotherapies in Phase II A studies (EBA) needed for assessment of bactericidal efficacy and dose-finding?- Dr. David Hermann (Bill & Melinda Gates Foundation)*

Dr. Hermann reviewed early and middle TB drug and regimen development phases and submitted that in Phase IIA monotherapy EBA studies, the objectives are to determine whether safety or tolerability data preclude continued development given the TRPs; and to define the lowest, maximally effective dose to be taken forward for subsequent development. The possible translational disconnect between selecting a dose based on a monotherapy, two-week EBA exposure-response slope versus selecting dose in the context of longer-term combination trials was acknowledged. Dr. Hermann pointed out that in vitro HFS appears to correlate well with EBA results and systems pharmacology models are emerging to translate murine pre-clinical data to predict clinical response (also described in Dr. Hanna's presentation), raising the question of whether monotherapy EBA studies are required as proof of efficacy. Regulators require proof of efficacy or contribution of each individual drug comprising a regimen. Till now EBA studies were the standard mechanism to prove bacteriological efficacy, however preclinical models are evolving as as a plausible alternative. If the only goal of the EBA study were to address bacteriological efficacy (i.e., does the drug work?), then EBA monotherapy may not be necessary

because it has been shown that it can be addressed in preclinical models and the HFS qualification provides a proof of concept defined as monotherapy bacteriological activity. If the goal is to address dose finding, then EBA studies are informative, addressing safety and tolerability questions, refining exposure response relationships defined in preclinical development, and notably, to address any uncertainties identified pre-clinically for drugs with new mechanisms of action. Dependent on the objectives and nature of a given programme, Dr. Hermann suggested differing objectives can be addressed in Phase IIA EBA studies.

Discussant 2: *What information/markers/endpoints should be collected across Phase II studies to optimally select the appropriate combo regimens to move from Phase II to III? - Dr. Martin Boeree (Radboud University)*

Dr. Boeree reviewed the data, markers, and endpoints that should be collected across Phase II studies. Dr. Boeree argued that EBA studies have limited use for understanding efficacy but should be used to collect information on dose selection. For middle development phase studies, bacteriological endpoints on culture status at 8 weeks and possibly 12 and 16 weeks was recommended, with analytic emphasis on time to event endpoints (e.g. time to stable culture conversion, TSCC). In this regard, he noted the importance of standardization of collection time points and culture procedures across trials. Ultimately, Dr. Boeree recommended that trial designs incorporate and collect as many specimen types and biomarkers as can be feasibly integrated and afforded, particularly novel assays that measure bacterial load. In phase IIC studies, it was noted that long term relapse data is collected, and when combined with PK data and patient characteristics for PK/PD analysis, key knowledge gaps described in Session 1 can be better addressed through this approach. Lastly, Dr. Boeree recommended that an adaptive trial design combining Phase II and III trials with a feedback mechanism can be used to select the most promising regimens to move forward to Phase III (e.g. Multi-Arm Multi-Stage (MAMS) design, Phase IIC Selection Trial with Extended Post-treatment follow-up (STEP). These are designs in which an iterative optimization approach is used to test different drugs, different doses, and different core combinations.

Session 3:
New trial designs and how they may facilitate regimen development

❖ *Sub-Session 3.1: Novel trial designs*
Facilitators: Dr. Carole Mitnick (Harvard Medical School) & Dr. Jim Neaton (University of Minnesota)

Key-Note: *Where have we come from and have we made any mistakes? High-level review of different trials and trial designs used to date in TB phase III trials, and how designs have progressed over time – with discussion of respective strengths and weaknesses – Prof. Andrew Nunn (University College London).*

Prof. Nunn provided a high-level review of conducted phase III trials, with a focused discussion on how designs have evolved and any mistakes that were made. He started with a review of the

British Medical Research Council (MRC) era trials, a time period in which regimen development moved more quickly, with trials starting based on results of interim analyses of prior trials still ongoing. He also commented that decisions on regimens were not based as they are now on meeting non-inferiority margins, which he observed have been increasing over time. Widening margins reduces sample size requirements, resulting in cheaper and faster studies, but the potential for the "bio-creep" phenomenon should be considered. He expressed concern regarding the high variability in interpretation of results of non-inferiority trials by developers, regulators, and policy makers. To address this interpretation issue, he noted a recently proposed approach using Bayesian analysis whereby posterior probabilities that one regimen is worse than another is examined for a range of effect sizes using different prior probabilities. Regarding follow-up in phase III trials, Prof. Nunn referred to analyses of 15 MRC Tuberculosis and Chest Diseases Unit short course trials, wherein it was shown that 80% of relapses occurred within 6 months of stopping treatment and 90% within 12 months, providing evidence to reconsider the traditional requirement for 24 months post-randomization follow-up for relapses in Phase III trials. Using the STREAM Stage 1 (NCT02409290) trial as an example, Prof. Nunn raised the issue of improving standard of care over time and implications for assumptions used in power/sample size calculations (particularly important in STREAM where sample size assumed a higher favorable outcome in experimental compared to control arm); he questioned whether the STREAM trial would have been possible today given the current MDR-TB standard of care and whether the shorter-regimen would have been found clearly non-inferior to the WHO control if a more pragmatic trial had been undertaken. In conclusion, Prof. Nunn stated that much had changed over 50 years, and conducting phase III TB trials has become more difficult. Lessons learnt, or still to be learnt, include: the need to properly define trial outcome(s); the need to take into account the effect of HIV co-infection; the possibility of shortening trial duration by limiting follow-up to 6m post treatment for latter enrollees; advantages and disadvantages of increased non-inferiority margins need to be carefully assessed. Prof. Nunn questioned whether early results from phase III trials could be used to inform the design of the next trial without damaging the integrity of the current trial? Lastly, he suggested that the TB therapeutics field should consider pursuing 3-month regimens or shorter, fully recognizing that there will also be some patients (HIV-infected or other hard to treat population) that may require extended treatment durations and possible retreatment. More research is needed also into measuring and improving adherence.

Discussant 1: *Have non-inferiority trials served us well? Given challenges in interpretation and limitations of non-inferiority phase III trials, what are the recommendations for the future of non-inferiority design in TB therapeutics? – Dr. Piero Olliaro (WHO/TDR).*

Dr. Olliaro referred to the published article, "Challenges in the Design and Interpretation of Non-inferiority trials"[15], and summarized key assumptions in non-inferiority trial designs in accordance with the CONSORT statements: the current standard of care should be used as the

active control arm; the control arm performance should be assessed to determine if outcomes are as expected; endpoints should be clinically relevant; composite endpoints should be avoided due to complexities in interpretation[j]; and the choice of non-inferiority margin needs to be fully justified. He noted the challenges of defining "meaningful" non-inferiority margins (i.e., meaningful statistically, clinically, from patient perspectives, or public health/programme perspectives. Also, data on efficacy of control may not be with same endpoint and/or is from an observational study for the target population of interest). He proposed that non-inferiority results be translated into an appropriate metric that is meaningful in terms of practicality and proposed the use of the "number needed to treat" (NNT) as a more informative means of reporting otherwise binary outcomes from randomized clinical trials. He provided an example from recent fluoroquinolone trials showing that a one-third reduction in treatment duration (4 months regimen with a fluoroquinolone replacing ethambutol) will translate to one more failed treatment (compared to standard of care) for every 47 (best case) to 9 (worst case) patients treated. With such a quantified approach, the focus moves to what number of additional unfavorable outcomes is acceptable from the patient and programme perspective and what factors should contribute to this decision.

Discussant 2: *What are the appropriate controls for rifampin susceptible and rifampin resistant TB trials? How do we manage the challenges posed by the various designs on the selection of appropriate control groups, including the issue of changing standards of care? – Dr. Ed Cox (US Food and Drug Administration, FDA).*

Dr. Cox acknowledged that developing drugs for treatment of TB is challenging because of the nature of the pathogen, the disease, the need for treatment with multiple drugs and the highly effective cure rates achieved with current standard treatment in drug-susceptible disease. He referred to FDA guidance on "Pulmonary Tuberculosis: Developing drugs for treatment"[k] and "Codevelopment of Two or More New Investigational Drugs for Use in Combination"[l] to illustrate various clinical trial designs and settings in which they can be considered. Due to the complexity of designing trials for TB treatment, he advised that developers consult with the regulatory agencies early on, during protocol development. A variety of trial designs are acceptable, but trials should be designed to differentiate effective drugs from those with less or no effect. It is also critical to understand that a single trial will not answer all patient, programme and regulatory important questions. Dr. Cox emphasized that the use of concurrent control groups and

[j] Note: These points are also true for superiority studies. A problem potentially unique for non-inferiority studies is the "constancy" assumption - the trials from which historical data show that the control is effective should be similar to setting of proposed trial, i.e., the endpoint as well as inclusion criteria, etc. In general, a reliable estimate of the control event rate is very important in a non-inferiority trial.

[k] Food and Drug Administration. Guidance for Industry Pulmonary Tuberculosis: Developing Drugs for Treatment. (2013)
.https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM373580.pdf

[l] Food and Drug Administration. Guidance for Industry Codevelopment of Two or More New Investigational Drugs for Use in Combination. (2013). https://www.fda.gov/downloads/drugs/guidances/ucm236669.pdf

randomization in clinical trials remains critically important for controlling bias and provided examples of concurrent internally controlled trials in which use of external controls could have altered interpretation of efficacy and safety results. Quoted from the ICH E10 guideline[m], *"Inability to control bias is the major and well-recognized limitation of externally controlled trials and is sufficient in many cases to make the design unsuitable […]. The groups can be dissimilar with respect to a wide range of factors, other than use of the study treatment, that could affect outcome, including demographic characteristics, diagnostic criteria, stage or severity of disease, concomitant treatments, and observational conditions (such as methods of assessing outcome, investigator expectations).*" Nevertheless, there are select situations where an external control may be acceptable, including when the effect of the experimental treatment is dramatic, the usual course of the disease is highly predictable, endpoints are objective, and the impact of baseline and treatment variables on the endpoint is well characterized (ICH E10). Overall, Dr. Cox emphasized that high quality evidence facilitates well informed decisions by healthcare providers, patients, policy makers and regulators. Any proposed change in the standard of care is best achieved with provision of high quality evidence.

Discussant 3: *Where are we now and where should we be going? Presentation of new ideas that are being proposed for ongoing and future trials – with discussion of strengths and weaknesses. – Dr. Patrick Phillips (UCSF).*

Dr. Phillips reviewed existing and emerging clinical trial designs taking an actual ongoing or planned trial to illustrate the design. He described the 'reference standard' non-inferiority design (the example being TBTC/ACTG S31/A5349), designs that evaluate numerous ($\geq$3) regimens in a single trial (e.g. endTB trial and a seamless phase II/III platform), designs that evaluate different durations of the same regimen (CLO-FaST and the Durations design), designs that address the heterogeneity of patient outcomes (TRUNCATE-TB and CURE-TB), and the continuum between explanatory and pragmatic trial designs (STREAM Stage 1 and pragmatic stepped wedge design. He highlighted that the objective of a phase III trial is to generate clear unambiguous evidence of superior efficacy and safety of the intervention (compared to no treatment or to standard of care) that can convince a broad assortment of stakeholders, and that the choice of trial design is informed by the trial objective and setting. He echoed comments by Dr. Nunn about the potential benefits of phase III trial designs that allow for rapid reporting without substantially undermining internal validity (e.g. STREAM Stage 1, which lasted 7-8 years) and the place of Bayesian adaptive trials with 6-monthly interim analyses that report out, using a range of priors (skeptical, flat, enthusiastic) (e.g. endTB, NCT02754765).

[m] INTERNATIONAL CONFERENCE ON HARMONISATION OF TECHNICAL REQUIREMENTS FOR REGISTRATION OF PHARMACEUTICALS FOR HUMAN USE CHOICE OF CONTROL GROUP AND RELATED ISSUES IN CLINICAL TRIALS E10. (2000).
http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E10/Step4/E10_Guideline.pdf

❖ *Sub-session 3.2: Measuring and maximizing adherence Facilitators: Dr. Andrew Vernon (CDC) & Dr. Lori Dodd (NIH)*

Key-Note: *The importance of maximizing adherence in clinical trials and its practical relevance in explanatory and pragmatic trials. Are we paying enough attention? - Dr. Rada Savic (UCSF)*

Dr. Savic presented on the importance of measuring and maximizing adherence of TB treatments. Recent analyses of TB-ReFLECT patient-level data have re-affirmed that lower adherence (even less than 90%) is significantly associated with worse outcomes. This finding suggests the value of efforts to optimize tools and approaches to measure adherence in trials, so that we might understand better how adherence patterns influence outcomes for specific regimens. Dr. Savic noted that the term "forgiveness" is sometimes used to describe a treatment regimen that maintains a high efficacy in the face of incomplete adherence.[n] Many factors may influence this feature, including - in the case of TB - the pharmacology of the drug(s), the specific microbial disease, and the mechanisms involved in achievement of cure. In this context, she noted that the gap between the efficacy of a regimen and its effectiveness can be larger for regimens which are less "forgiving." Further illustrating the potential differential impact of adherence when studying efficacy versus effectiveness of a regimen, Dr. Savic described two different approaches relating to adherence to interventions in clinical trials. One approach is to get as close to *efficacy* as possible, using rigorously applied directly observed therapy throughout treatment. The alternative approach is to use minimal or no adherence-enhancing interventions to get as close as possible to *effectiveness* (i.e., programmatic reality) in the context of a randomized controlled trial while precisely measuring adherence. Overall, she concluded that less than 100% completion of a regimen is the rule, rather than the exception, in clinical trials and in the field, and that our approaches to regimen development and trial design should carefully consider the impact of variations in adherence on key outcomes.

Discussant: *Are alternative techniques and digital health approaches (e.g. video-observed treatment) ready for use in place of traditional in clinic directly observed therapy (DOT) both for measuring and optimizing adherence and patient support?- Dr. Katherine Fielding (London School of Hygiene and Topical Medicine)*

Dr. Fielding reviewed a variety of emerging tools proposed for measuring and optimizing adherence in programme settings, many of which are being tested in on-going clinical trials. She noted that whereas the potential of digital technologies to improve TB care and adherence is large, the evidence base to-date is rather limited. Ongoing studies and trials on "mobile health" in the TB field will provide new data soon; however, their applicability will likely be setting-specific. She concluded that these newer digital health tools are not yet ready for full global

[n] Forgiveness can be defined as how long drug action continues to be above therapeutically effective concentrations at its site or sites of action after a last-taken dose.

implementation, but should be studied further and considered in specific trial designs and settings.

❖ *Sub-session 3.3: Addressing special populations*
Facilitators: Dr. Monique Surette (European & Developing Countries Clinical Trials Partnership) & Dr. John Johnson (Case Western Reserve University)

Discussant 1: *Inclusion of special populations in clinical trials: current recommendations and barriers to implementation – with a focus on children. Dr. Anneke Hesseling (Stellenbosch University).*

Dr. Hesseling summarized current approaches, barriers, and opportunities for conducting TB trials in children. About 10% of all TB cases occur in children, with significant TB-related morbidity and mortality, including from severe forms of TB in young children like TB meningitis. Currently, most TB treatment recommendations for children including regimens and dosing, are extrapolated from adults. However, this model is not always optimal and should be revisited for both current and future drug and regimen development. With regard to TB regimen development for children, Dr. Hesseling listed a number of key characteristics of children with TB that should be considered, namely that: 1) there is a wide spectrum of disease phenotypes by age, as well as variability in the risk for progression from infection to active TB; 2) most children have intrathoracic disease (~75%); 3) paucibacillary disease is much more common as compared to adults; 4) severe and disseminated TB are frequent in young children; 5) treatment outcomes for both DS-and DR-TB are generally good in children, provided treatment is initiated early; 6) The PK of drugs differs in children compared to adults (lower concentrations and faster elimination, particularly in the youngest children, which has led to under-dosing of several key anti-TB drugs in children, e.g. rifampicin and levofloxacin); and 7) The diagnosis of TB and treatment outcomes are often dependent on clinical measures because classical sputum bacteriologic data (smear and culture) are frequently less available for infants and younger children who often cannot spontaneously produce sputum for examination. According to regulatory authorities[o], when disease progression and response to treatment in children are comparable to adults, PK studies aimed at achieving drug levels similar to those for adults in conjunction with safety assessments may be sufficient for approving a new drug or regimen (e.g. Opti-Rif Kids and IMPAACT P1108, NCT02906007). Small and efficient PK and PK modeling studies have been extremely useful in informing better treatment of TB in children. Small blood volume sampling methods and sparse sampling strategies have been validated for conducting drug PK studies in infants and children. Conversely, for forms of TB where similar efficacy between

[o] Food and Drug Administration. Guidance for Industry Exposure-Response Relationships — Study Design, Data Analysis, and Regulatory Applications. 800–835 (2003). https://www-fda-gov.ucsf.idm.oclc.org/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM072109.pdf

children and adults cannot be assumed (e.g. MDR-TB or minimal forms of pulmonary TB in children where many children will be over-treated resulting in unnecessary risk of toxicities), then separate efficacy studies in children are needed (e.g. SHINE for treatment shortening of DS-TB - ISRCTN63579542, TB-CHAMP for MDR-TB prevention - ISRCTN92634082). The reliance on clinical definitions for disease and for determining outcomes, the importance of testing child-friendly formulations, and the need for integrated strategies that allow early inclusion of children in the same trials as adults (e.g. inclusion of adolescents in Phase IIb studies dependent on initial safety data) are a few of the unique considerations when designing trials involving children.

Discussant 2: *Considerations for investigation of treatment in pregnant women- Dr. Amita Gupta (Johns Hopkins University).*

Dr. Gupta noted that the key barrier towards obtaining data and developing an evidence base for TB treatment during pregnancy and lactation is the frequent exclusion of pregnant and lactating women from clinical trials. Current evidence for TB treatment during pregnancy and lactation is based on small case reports and case series.[16] There are numerous knowledge gaps, including the reproductive toxicity of drugs, transport of drugs across placenta and breast milk passage, pharmacokinetics and safety in pregnancy, the optimal time to initiate prophylaxis, and the benefits and risks of TB treatment in pregnancy. Dr. Gupta challenged the field to shift the standard approach from *assumed* exclusion to *presumed* inclusion. She described the ethical and scientific foundation for inclusion of pregnant women into TB clinical trials and reviewed selected key recommendations from an international expert panel[3]. These include, specifically: 1) pregnant/postpartum women should be eligible for Phase III MDR TB trials unless there is a compelling reason for exclusion; 2) drug companies should be encouraged to complete reproductive toxicity studies before beginning Phase III trials; 3) trials of shortened treatment regimens for LTBI should be designed to improve completion rates and reduce the risk of progression to active TB during pregnancy and the postpartum period; 4) targeted PK studies should be nested in all studies when evidence is lacking; and 5) an international pregnancy registry should be created to accumulate data on maternal and infant outcomes of women treated for LTBI or active TB during pregnancy. Dr. Gupta reiterated that the lack of data has hindered the treatment of pregnant and postpartum women. Well-designed studies that evaluate maternal, fetal, and infant outcomes are urgently needed to gather information to assess drug safety and efficacy and relative risks and benefits and ultimately to inform evidence-based practice guidance for the care of pregnant women with TB or LTBI.

Discussant 3: *What would make the inclusion of special populations easier for researchers – with a focus on HIV affected populations? - Dr. Michael Hughes (Harvard University).*

Worldwide, about 11% of all new TB cases occur in HIV-infected persons and TB is a leading cause of death in these individuals. Dr. Hughes summarized the rationale and challenges for including HIV-infected populations in clinical trials. Including HIV-infected persons not only allows trial

findings to be more generalizable, but also provides key data for the treatment of co-infected populations, which broadens labeling possibilities for drug developers. In addition, it is key to evaluate treatments for high-risk populations in order to be able to achieve goals of the WHO END-TB Strategy. However, there are also challenges related to accessing and diagnosing HIV-infected populations with TB, such as the potential for differences in outcome rates by HIV infection and disease status and antiretroviral treatment use (as noted in the OFLOTUB trial, STREAM Stage 1 and other trials), important drug-drug interactions, complexities in the clinical management of co-morbidities, adverse events, and immune reconstitution inflammation syndrome (IRIS). Dr. Hughes highlighted that although substantial trial infrastructure now exists worldwide in high TB and HIV burden settings, it is essential that TB and HIV care programmes be well-integrated in those settings for optimal patient management. In addition, he emphasized the importance of having TB and HIV expertise both on the protocol team and among the groups of local investigators to manage the complexities of caring for TB and HIV co-infected participants. Finally, he noted the importance of access to quality TB diagnostics, including molecular and culture systems, since HIV-infected populations are frequently smear negative, paucibacillary and often present with extrapulmonary TB. To address the possibility of differential outcomes by HIV status and ART use, he suggested stratified randomization into three categories (HIV-negative, HIV-positive with high CD4 cell count, and HIV-positive with low CD4 cell count) and requiring protocols to specify the concomitant use of appropriately selected ART initiated within 4 to 8 weeks of starting TB treatment according to international guidelines. Lastly, he advocated for conducting drug-drug interaction studies with ART early during the evaluation of new drugs and regimens. This might best be achieved through nested drug-drug interaction studies within phase II and III trials or by having an ongoing "master protocol" which could rapidly incorporate and evaluate new drug-drug combinations. There should also be flexibility within a trial and the product development plan to assess the drug-drug interaction of new ARVs (e.g. dolutegravir) that are introduced after a trial has begun.

Session 4:
The interplay between trials and guidelines: the importance of sound evidence to inform policy guidance and clinical practice.
Facilitators: Dr. Sumati Nambiar (FDA) & Dr. Michael Rich (Partners in Health)

Key-note: *Do the trial considerations that serve the objectives of registration meet the needs for development of public health guidance? – Dr. Christian Lienhardt (WHO).*

Dr. Lienhardt described the WHO procedures for the development of guidelines and the nuances involved in moving from clinical trial results into making policy decisions for public health. The WHO uses the "Grades of Recommendation Assessment, Development and Evaluation (GRADE)

System", which provides an explicit and transparent approach to: 1) assess the quality of evidence across studies and outcomes and 2) translate evidence to recommendations. Guideline development incorporates multiple processes to minimize bias and optimize usability and involves transparency in all judgements and decision making. To formulate evidence-based recommendations, four factors are taken into account: 1) magnitude of benefits and harm, 2) consideration of resource use, feasibility, acceptability, and equity, 3) certainty ("quality") of evidence, and 4) Patients' values and preferences. Ultimately, the main aspect for providing recommendations is: *what is the best available evidence that can be brought about that ultimately benefits patients?* Dr. Lienhardt concluded that late phase clinical trial outputs that serve objectives of registration of a new drug or regimen can meet the needs for development of public health guidelines, especially if the registration trials collected data on long-term, patient-relevant and population relevant outcomes. He emphasized, however, that additional public health factors such as feasibility, individual and population level benefits and harms, acceptability, resource use, equity and quality of life are also considered when formulating practice guideline recommendations.

Discussant 1: *Outcome definitions in clinical trials – should they vary to fit regulatory and programmatic decision-making needs? - Dr. Andrew Vernon (US Center for Disease Control and Prevention, CDC).*

Dr. Vernon reflected on the definition of clinical endpoints in trials and presented three considerations. First, he noted that trials are increasingly using composite endpoints in which all unresolved events (e.g., lost to follow-up; deaths of uncertain cause; changes in therapy) are grouped together as "unfavorable outcomes." While acknowledging the possible utility of such a conservative approach, he noted that this can influence critical outcome assessments in non-inferiority trials. He advocated that trials carefully pre-specify targeted analyses (e.g., separate analyses of failure/relapse in per protocol patients; e.g. separate analyses of all-cause mortality) which will more specifically permit assessment of efficacy *vs.* effectiveness. To this end, Dr. Vernon emphasized the importance of careful and individual recording and presenting of *all* unfavorable events, such as relapse, failure, lost to follow up, drop-out, death, even if for the clinical trial purposes they are combined into composite endpoints for analyses. Secondly, he endorsed the adoption of high quality, rigorous and standardized procedures for the real-time evaluation of patients suspected of having a poor treatment response during the trial, as an alternative to prior approaches that relied on an expert "Endpoint Review Committee" evaluating the outcome at the end of the trial. Such a committee, while valuable, cannot correct for a mismanaged event that lacks adequately clear and/or sufficient information to classify an event with high confidence. Lastly, he re-emphasized the importance of assuring high rates of adherence to study medications in trials.

Discussant 2: *The point of view of the programme managers and end-users on the use and translation of WHO guidelines into national strategy plans Programme Managers: Dr. Nguyen Viet Nhung (National Lung Hospital, Hanoi, Vietnam), Dr. Alena Skrahina (Republican Research and Practical, Center for Pulmonology and Tuberculosis, Minsk, Belarus), Dr. Norbert Ndjeka (National Department of Health Pretoria, South Africa)*

Drs. Skrahina, Ndjeka and Nhung, as TB programme managers, highlighted the immense value and importance of WHO guidelines in determining national policies for the care of TB patients. They expressed their appreciation for the evidence-based and transparent approaches used by the WHO for making recommendations. However, the complexity of adapting procedures to local needs and the numerous rapid updates to guidelines, often based on low or very low quality of evidence, can lead to challenges in the implementation of new tools and therapies. In such situations, adoption and implementation of recommendations often need further reflection. The TB programme managers urged the global TB research community to design and conduct clinical trials that will result in the generation of high-quality evidence at the patient and population levels using relevant TB outcomes, as such data can lead to the most robust and highly implementable public health recommendations possible for TB care.

Discussant 3: *The point of view of the regulator – Dr. Marco Cavaleri (EMA)*

Dr. Cavaleri presented on the perspective of regulatory agencies on the interplay between trials and guidelines. Regulatory agencies, including the EMA 'Committee for Medicinal Products for Human Use' are charged with assessing all available data and form an opinion based on whether or not quality, safety and efficacy requirements are met for approval of a drug, and if there is a positive risk-benefit balance to support the use in the claimed indication. Dr. Cavaleri acknowledged subtle but important distinctions in the needs of the regulatory and policy recommending bodies. For example, regulators need to establish that a drug submitted for licensure is safe and effective for the proposed use, while recommending bodies need to define how to use the drug within a regimen in a way that addresses the public health needs. He acknowledged the strong desire to design trials in ways to address both the needs for licensure and for determining optimal use in the public health domain. He highlighted, however, that significant challenges arise when planning and designing clinical trials for both licensure and recommending bodies, in part due to important gaps in knowledge in TB, e.g. the lack of relevant data regarding the efficacy of standard of care for MDRTB. Regulatory agencies have tools for early approval for new drugs that address unmet need according to specified criteria, e.g. the conditional marketing authorization pathway in the EU where the benefit-risk balance of the new drug is such that the immediate availability outweighs the limitations of less comprehensive data than normally required. The implications would be that, while awaiting for further data to be generated post-approval, there is less data than normal immediately available to support policy recommendations. He urged more collaboration between regimen developers, regulators and

policy recommending bodies to better inform pre-licensure pivotal studies and to define post-approval studies.

Discussant 4: *The point of view of the policy maker: Which evidence and criteria should guide selection of medicines candidate to WHO guideline recommendations and the Model List of Essential Medicines (EML) - Dr. Lorenzo Moja (WHO)*

The same criteria underpin the logic of recommendations in WHO guidelines and selection of essential medicines: a careful selection of those interventions that are relevant, effective and cost-effective. Both follow a transparent process and rely on extensive systematic evidence synthesis and appraisal, along with the assessment of comparative cost-effectiveness, in addition to feasibility and acceptability. The treatment of tuberculosis faces specific challenges, including the need to use a combination of at least four different antibacterial medicines. Since there is an urgent need to shorten and simplify therapy for both drug-sensitive and drug-resistant TB, results from recent pivotal trials or from non-inferiority trials were evaluated. These evaluations led to conditional recommendations in WHO guidelines, and thereafter (i.e. for bedaquiline and delamanid) inclusion on additional therapeutic options into the EML. For the policy maker involved in the evaluation of TB treatments, non-inferiority trials provide the way to evaluate new treatments that have approximately the same efficacy as the current standard of care, but may offer other benefits such as shorter duration or better safety profile. Early phase studies such as PK/PD studies have an important informative role but it is difficult to attribute to them an absolute decisional weight which might tip the balance in favour or against an intervention. Evidence from these sources is important to identify which molecules are promising and should be prioritized for additional research, to optimize current regimens and to extend indications to children and pregnant women. More interaction between researchers responsible of designing the next generation of TB trials and policy makers is warranted to achieve better harmonisation between the research pipeline and policies on access to TB medicines.