Report of the technical consultation on

Innovative Clinical Trial Designs for Development of New TB Treatments

Virtual meeting:
Thursday 20 August, Tuesday 1 September,
Tuesday 15 September, Tuesday 29 September,
Tuesday 13 October
2020

World Health Organization

Report of the technical consultation on innovative clinical trial designs for development of new TB treatments

ISBN 978-92-4-003086-2 (electronic version)
ISBN 978-92-4-003087-9 (print version)

Suggested citation. Report of the technical consultation on innovative clinical trial designs for development of new TB treatments. Geneva: World Health Organization; 2021. Licence: CC BY-NC-SA 3.0 IGO.

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

This report summarized the discussions held during the consultation and does not necessarily represent the decisions or policies of WHO.

Contents

Acknowledgments

Overall organization of the technical consultation
Christian Lienhardt and Saskia den Boon

Chair and Co-chairs
Chair: Christian Lienhardt. *Co-chairs:* Session 1: Rada Savic, Session 2: Morten Ruhwald, Session 3: Michael Hoelscher, Session 4: Carole Mitnick, Session 5: Payam Nahid.

Technical consultation participants
Draurio Barreira (Switzerland), Clifton Barry (United States of America), Martin Boeree (Netherlands), Grania Bridgen (France), Marco Cavaleri (Netherlands), Richard E. Chaisson (USA), Tsira Chakhaia (USA), Gavin Churchyard (South Africa), Daniela Cirillo (Italy), Angela Crook (United Kingdom of Great Britain and Northern Ireland), Geraint Davies (United Kingdom), Andreas Diacon (South Africa), Lori Dodd (USA), Kelly Dooley (USA), Kathleen Eisenach (USA), Daniel Everitt (USA), Amita Gupta (USA), Debra Hanna (USA), David Hermann (USA), Anneke Hesseling (South Africa), Jeremy Hill (Netherlands), Michael Hölscher (Germany), Robert Horsburgh (USA), Shitong Huan (China), Michael Hughes (USA), Marjorie Imperial (USA), Thomas Jaki (United Kingdom), Amina Jindani (United Kingdom), Soyeon Kim (USA), Christoph Lange (Germany), Barbara Laughon (USA), Christian Lienhardt (France), Vidya Mave (USA), Lindsay McKenna (USA), Dick Menzies (Canada), Carole Mitnick (USA), Stellah Mpagama (United Republic of Tanzania), Kissa Mwamwitwa (United Republic of Tanzania), Payam Nahid (USA), Sumathi Nambiar (USA), Norbert Ndjeka (South Africa), Eric Nuermberger (USA), Andrew Nunn (United Kingdom), Bern-Thomas Nyang'wa (United Kingdom), Nick Paton (Singapore), Patrick Phillips (USA), Helen Rees (South Africa), Morten Ruhwald (Switzerland), Rada Savic (USA), Ulrika Simonsson (Sweden), Alena Skrahina (Belarus), Mel Spigelman (South Africa), Kelly Stinson (USA), Eugene Sun (South Africa), Elin Svensson (Netherlands), Susan Swindells (USA), Elizabeth Talbot (Lebanon and USA), Guy Thwaites (United Kingdom), Francis Varaine (France), Andrew Vernon (USA), Nguyen Viet Nhung (Viet Nam), Yuhong Liu (China) Robert Wallis (South Africa), Nicholas Walter (USA).

Pharmaceutical industry
Alex Carlton (GSK, United Kingdom), Oscar Della Pasqua (UCL and GSK, United Kingdom), Florian von Groote (Evotec, France), Jeffrey Hafkin (Otsuka, USA), Gavin Koh (GSK, United Kingdom), Yongge Liu (Otsuka, USA), Robin Mogg (Gates MRI, USA), Alexander Pym (Janssen, J&J, United Kingdom), Charles Wells (Gates MRI, USA)

WHO/HQ Secretariat
Saskia den Boon, Dennis Falzon, Albert Figueras, Benedikt Huttner, Tereza Kasaeva, Corinne Merle, Lorenzo Moja, Fuad Mirzayev, Martine Penazatto, Vindi Singh, Matteo Zignol.

Administrative and secretarial support
Lou Maureen Comia, Jaqueline Lashley

Meeting Report written by

Marjorie Imperial, Christian Lienhardt and Saskia den Boon, with input from the co-chairs and review by all participants.

A summary of declaration of interest statements for all participants was provided by Dr. Saskia den Boon.

List of Acronyms

ACTG	AIDS Clinical Trials Group
BMGF	Bill & Melinda Gates Foundation
CDC	US Centers for Disease Control and Prevention
CFU	colony forming units
CMPH	Committee for Medicinal Products for Human Use
DR	drug resistant
DS	drug susceptible
EBA	early bactericidal activity
EMA	European Medicines Agency
FDA	US Food and Drug Administration
HIV	human immunodeficiency virus
IMPAACT	International Maternal Pediatric Adolescent AIDS Clinical Trial Network
LAM	lipoarabinomannan
MAMS	multi-arm multi-stage
MBL	molecular bacterial load
MBLA	molecular bacterial load assay
MDR	multidrug-resistant
MIC	minimum inhibitory concentration
NCEs	new chemical entities
NIH	National Institute of Health
PD	pharmacodynamic
PET-CT	positron emission tomography-computed tomography
PK	pharmacokinetic
PKPD	pharmacokinetic-pharmacodynamic
PLEG	Post licensing or post launch evidence generation
QSP	quantitative systems pharmacology
rDNA	recombinant deoxyribonucleic acid
RNA	ribonucleic acid
rRNA	ribosomal ribonucleic acid
TB	tuberculosis
TBTC	Tuberculosis Trials Consortium
TTP	time to positivity
XDR	extensively drug-resistant

Welcome, Introduction, and Objectives

Tereza Kasaeva (WHO Global Tuberculosis Programme) and Christian Lienhardt (Research Institute of Development, IRD, and University of Montpellier)

The Director of WHO Global Tuberculosis Programme, Dr. Tereza Kasaeva, opened the Technical Consultation by welcoming all participants. She thanked all participants for their commitment to combat tuberculosis (TB), particularly as the world faces the unprecedented COVID-19 crisis that may have serious impact on the progress of eliminating TB. This technical consultation with a large, diverse group of experts and stakeholders to address the design of innovative clinical trials for the development of new TB treatments is a prime example that will bring us another step forward to TB elimination.

Dr. Lienhardt introduced the background and objectives of the Technical Consultation. After decades of stagnation, research in TB therapeutics is experiencing a renaissance, with an increasing number of new and repurposed compounds undergoing evaluation as part of novel treatment regimens. The emergence of several new chemical entities (NCEs) raises the possibility that shorter, simpler, and safer regimens for all forms of TB may become available in the near future. Much insight has been gained over the past decade from large clinical trials evaluating treatment shortening potential for drug susceptible (DS) TB and from the development of new TB drugs and regimens to treat multidrug-resistant (MDR) and extensively drug-resistant (XDR) TB.[1] It is hoped that this experience will lead to novel ultrashort, safe, and effective regimens including newly developed and repurposed drugs. However, the development of new TB drugs is complex, lengthy, and costly, and the path to proven new TB treatment regimens is fraught with numerous obstacles and uncertainties. To address these issues, WHO organized in March 14–16, 2018, a technical consultation on '*Advances in Clinical Trial Design for New TB Treatments*' to identify and outline, through expert agreement, novel trial designs to inform policy guidance on new TB regimens. Building on the lessons learned from the rich history of TB clinical trials, this technical consultation reviewed the various research designs and tools used for the development of new TB treatments and made a series of proposals to advance these further. These proposals are described in detail in the consultation report, as well as in a series of papers assembled in a special PLOS Medicine Collection, '*Advances in Clinical Trial Design for Development of New Tuberculosis Treatments*', published from March 2019 to February 2020.[2,3]

Since then, remarkable efforts have taken place, and a number of NCEs have transferred from preclinical to early clinical phases, providing new hopes for a generation of better treatment

[1] Lienhardt C, Nunn A, Chaisson R, Vernon AA, Zignol M, Nahid P, Delaporte E, Kasaeva T. Advances in clinical trial design: Weaving tomorrow's TB treatments. PLoS Med 2020;17(2): e1003059. https://doi.org/10.1371/journal.pmed.1003059

[2] WHO. Report of the Technical Consultation on Advances in Clinical Design for Development of New TB Treatments. https://www.who.int/tb/publications/2018/clinical_trail_design_TB_treatments/en/

[3] PLOS Medicine Special Collection on Advances in Clinical Trial Design for Development of New Treatments for Tuberculosis. https://collections.plos.org/collection/novel-tb-treatments/

regimens, utilizing the new designs that were discussed in the 2018 WHO Technical Consultation. In parallel, further evidence was produced on the capacity to use novel regimens, legitimating an update of the situation regarding the novel clinical trial designs for new TB treatment and integrating the vast experience gained by industry and other groups performing TB drug R&D.

Based on these noticeable advancements over the last two years, the **main goal** of this technical consultation was to gain agreement on evidence-based approaches to trial designs and use of data to inform policy guidance on new regimens for the treatment of TB with the view to accelerate future regimen development.

The **specific objectives** were:
(1) to review latest evidence and assess the progress made in TB regimen development since the 2018 WHO Technical Consultation on '*Advances in Clinical Trial Design for Development of New Tuberculosis Treatments*';
(2) to review novel study designs with the view to streamline regimen development to test regimens with known drug composition, dosing and duration, taking into account recent developments in methods, tools and biomarkers;
(3) to reach consensus on evidence-based approaches to trial designs and use of data to inform policy guidance on new regimens for TB treatment.

This document reports the proceedings of the consultation based on a series of key questions addressed in five successive thematic sessions. For each of these, the main aspects of the discussions are reported, and key outcomes presented in synthetic Tables. The details of the presentations made by keynote speakers and discussants in each session, together with the comments made by the round table panel are provided in ANNEX 3.

Session 1: Translational PK-PD modelling techniques for bridging preclinical to clinical development phases

Co-chairs: Rada Savic (University of California, San Francisco, UCSF) and Christian Lienhardt (IRD and University of Montpellier)
Keynote: The role of Translational Platforms for Tuberculosis Drug Development- Rada Savic (UCSF)
Discussant 1: Evaluating the dose rationale for antimicrobials combinations - Oscar Della Pasqua (University College London and GlaxoSmithKline, GSK)
Discussant 2: Does translational modelling have the promise of de-risking and accelerating early phase clinical development in TB? – Kelly Dooley (Johns Hopkins University)
Round table panellists: Debra Hanna (Bill & Melinda Gates Foundation, BMGF), Florian Von Groote (Evotec), Elin Svensson (Radboud University Medical Center), Eric Nuermberger (Johns Hopkins University), and Ulrika Simonsson (Uppsala University).

The objectives of this session were:

- To examine approaches combining preclinical and clinical data to guide early drug development;
- To evaluate the role and use of translational modelling on clinical development decisions regarding appropriate drug combinations and predictions of duration of regimens, based on bacterial elimination rates or relapse experiments.
- To better understand the role of a 'translational space' to broaden our understanding of how preclinical models can help drive more efficient clinical development of novel TB regimens by facilitating the transition from pre-clinical to clinical phases.

Questions addressed at the technical consultation:

1. *Are we producing the right data from our preclinical experiments? What common elements should there be? Are there experimental tweaks that would produce more informative data?*
2. *Can we distinguish among drugs from the same class using preclinical and Phase I data alone?*
3. *How can we use preclinical pharmacokinetic-pharmacodynamic (PKPD) of a New Drug plus what is known about preclinical-clinical differences to make Phase 2 more efficient (or de-risk Phase 3)? What elements of a preclinical 'toolbox' of experiments can/should be employed before-the-fact rather than after-the-fact? When, where, and how?*
4. *What clinical trial data would help us refine emerging translational models? Do we collect the right information? Do we share it back? What is missing? What drug and combination-specific information can only come from clinical studies?*
5. *How can we as a TB research community come together to synthesize and make sense of preclinical/clinical data being produced across groups?*

There was general agreement on the use of *quantitative translational platforms* to bridge preclinical and clinical development phases in TB regimen development. Successful translation from *in vitro* and *in vivo* experiments to clinical trials involves a complex multiscale approach that requires data integration from experiments that investigate efficacy of single and multidrug regimens, considering the various effects of immunological response, lesion penetration, intra- and extra-cellular distribution, emergence of resistance and intra-bacterial drug transport. To address the preclinical to clinical knowledge gaps, a *translational toolbox* is proposed that accounts for plasma pharmacokinetic (PK) scaling, site-of-disease lesion PK and PKPD, host immune and bacteria interplay, drug action and bacterial state interplay (e.g. fast *vs* slow multiplying bacteria at early and late stages of treatment), PKPD in context of multi-drug therapy, emergence of resistance, and time and dynamics in regard to PKPD relationships. The role of such translational platform may be split into three applications: (1) prediction of early bactericidal activity, (2) prediction of long-term combination therapy outcomes and regimen ranking, and (3) prediction of optimal regimens for all patients. Experts agreed that confidence in clinical predictions from translational platforms will be contingent on rigorous testing and validation, facilitated by an iterative feedback process.

The availability of multiple competing translational tools requires a clear understanding of the questions to be addressed and how to integrate the tools into a cohesive computational platform to best describe the evolving treatment response in the host. Similarly, the broad range of integrated data and translational tools required to inform areas that are difficult to address in the clinical setting should be defined. Several participants shared the critical importance of preclinical studies and translational tools to study acquired resistance in order to prolong the utility of new drugs and regimens, particularly for the most potent drugs in the pipeline. Several new drugs and/or regimens will likely be needed in the future to treat TB due to the constant emergence of drug resistance and situations that require regimen adjustments (i.e. toxicity, co-morbidities, drug interactions, etc.). Tools and platforms are needed to understand the points of frailties in each regimen, such as the rapid development of resistance when there are periods of non-adherence and hetero resistance in microbial populations within the host. Data are also needed to understand the host-bacteria-drug interplay and the time component of PK-PD relationships as pathology, and possibly immune response, does evolve during treatment. Experts considered it critical that preclinical experimental designs, and consequently translational platforms, capture these evolving relationships.

Experts raised the importance of biomarkers, as they condition the conclusions from the translational space. Preference is currently given to liquid rather than solid culture conversion readouts to describe efficacy. It is not known, however, if inconsistent readouts from various preclinical assays are simply due to assay sensitivity or if they depict different PK-PD relationships that affect the drug contribution or rank order of regimens. The importance of data sharing and collaboration between preclinical scientists, trialists and modelers was reiterated, to identify and design preclinical experiments that bridge current knowledge gaps and align with new technologies that are evolving in the clinic. Several participants advocated for the evaluation of promising novel biomarkers in as many preclinical and clinical studies as possible to determine their value and role for predicting treatment response.

While there was agreement on the development and use of quantitative translational platforms to bridge preclinical to clinical development phases, there was concern from several participants over replacing Phase 2A early bactericidal activity (EBA) studies with these platforms. Safety is a key output of EBA studies, as these typically provide the first opportunity to obtain safety data that may not be predicted by preclinical models. *Translational platforms* should be used to *complement* EBA studies, and inform EBA study design, regimen rank order, combination selection and dose rationalization. Of note, regulatory agencies require understanding of each drug's contribution in combination regimens which may be only partially informed by translational platforms. Using translational platforms has the potential to accelerate the drug development process, yet it is essential to reflect on the various pros and cons of new methods at each stage of the developmental process. Indeed, advancements in translational platforms focus on efficacy outcomes but further investigation is needed for assessment of safety outcomes, interpatient variability in disease severity and treatment response, as well as adherence patterns. The relative scarcity of preclinical murine studies to look at variability and adherence was mentioned. Innovative approaches are needed to address these issues in high-throughput *in vitro* studies as there is currently low capacity to test all scenarios in animal models.

Overall, the currently evolving landscape offers a great opportunity to use quantitative translational modelling approaches to make evidence-based decisions and design clinical trials. Although there may be little data available for translation in some areas such as lesion/site of disease PK, differential drug effects on varying metabolic states of the bacteria, or the influence of the immune response, it was generally considered that, it is not 'the lack of data but the lack of data integration' that is hindering optimal translation between preclinical models and patient populations. With the collaboration between industry, academia, regulatory and government agencies promoted by this Technical Consultation, the importance for preclinical scientists, clinical trialists and modelers to work together is critical to produce the necessary high-quality data and fully gain understanding and confidence to bridge the knowledge gaps between preclinical models and clinical data. The adoption of translational modelling to inform dose rationalization, combination selection, regimen prioritization, and clinical trial designs, should be actively promoted among various stakeholders.

Session 2: Biomarkers to support and/or accelerate decisions on suitable regimens to be tested.

Co-chairs: Morten Ruhwald (FIND) and Christian Lienhardt (IRD and University of Montpellier)
Key-Note: Most advanced and promising biomarkers for treatment response that can be used in clinical development: a review - Morten Ruhwald (FIND)
Discussants 1-4: Insights from most promising biomarkers - Nick Walter (University of Colorado), Yongge Liu (Otsuka), Timothy McHugh (University College London) and Clifton Barry (NIH/NIAD)
Discussant 5: Using biomarkers in clinical development: practical issues - Kelly Stinson (Cultura Incorporated)
Discussant 6: PD versus surrogate: exploring future biomarkers - Debra Hanna (BMGF)
Round table panellist: Kathy Eisenach (Independent Consultant), Robert Wallis (The Aurum Institute), Elizabeth Talbot (Dartmouth College), Payam Nahid (UCSF)

The objectives of this session were:

- To examine the current portfolio of most advanced biomarkers for treatment response that can be used for TB drug development;
- To define readiness for inclusion of these novel biomarkers to support and/or accelerate go/no-go decisions in Phase 2/3 clinical trials.

Questions addressed at the technical consultation:

1. *What information do we need and what can we get from the various biomarkers?*
2. *Do we exploit the true potential of the biomarkers and what is needed to understand biomarker complementarity?*
3. *How do we define readiness for inclusion of novel biomarkers to support and/or facilitate go/no-go decisions along the clinical development pathway?*
4. *How to ensure that biomarkers are integrated in drug development and how to support developers using them?*

One of the most critical challenges in identifying better regimens for TB is to identify and validate reproducible and reliable biomarkers that provide quantitative data that is ideally orthogonal and that can differentiate potency and perhaps sterilizing capability better than current culture techniques. Biomarkers should ideally be easy to interpret and be portable across clinical phases and provide information in the preclinical space to inform rank ordering and prioritize regimens for clinical evaluation.
It was recognized that the TB community had not progressed much beyond experimental biomarkers because of the lack of investment and data generation that would allow moving promising biomarkers forward in development. Larger collaboration and openness in development of biomarker technologies is required, relying on effective gathering and sharing of data, protocols, know-how and other information on novel biomarkers in the preclinical and clinical stages. Because of the complex nature of TB, a single biomarker that captures all

components of disease and predicts each phase of clinical development is probably unrealistic. Participants agreed that *research should focus on identifying an array of integrated biomarkers that successfully select the most potent and promising comparative regimen(s) that should move forward through the development pathway*.

A critical goal of biomarker development would be to identify novel biomarkers with real-time properties to inform and expedite innovative clinical trial designs with adaptive protocols. Biomarkers should also be quantitatively robust, reproducible, not subject to microbial contamination, and less variable than traditional culture methods. Identifying biomarkers that detect the presence of organisms *that cause relapse*, thus providing a *qualitative* measure that complements quantitative measures, would be ideal. This is an important area of investigation with several interesting approaches. For instance, the "ribosomal ribonucleic acid (rRNA) synthesis (RS) ratio" and the "MPT64 release assay" are proposed markers which measure the impact of drugs and regimens on the physiologic state of mycobacteria, thereby providing orthogonal information to the traditional assays that enumerate mycobacterial burden (e.g., CFU).

The best way to investigate the value of potential biomarkers is to test them within regulatory TB drug trials. Experts agreed that Phase 2B/C trials are an ideal setting to evaluate investigational biomarkers because (i) they measure actual Phase 3 clinical endpoints at a reasonable time (usually 12 months) after initiating treatment in a small population; and (ii) they can be designed to allow correlation with early sputum culture endpoints (i.e. culture conversion or time to culture conversion), as well as primary clinical endpoints (i.e. non-relapsing cure 6 to 12 months post treatment). Participants suggested that FIND, Gates MRI and other organizations working on biomarker and diagnostics development collaborate with trial networks and consortium (e.g. AIDS Clinical Trials Group, ACTG; International Maternal Pediatric Adolescent AIDS Clinical Trials Network, IMPAACT; Tuberculosis Trials Consortium, TBTC, Unite4TB) to develop a clear strategy for biomarker integration and assessment in clinical trials and to standardize approaches and assurance of specimen quality. To promote this, a *Forum for biomarker evaluation, study coordination and stakeholder collaboration* would be critically impkortant and participants expressed that WHO would be ideally suited for organizing it. Further, communication should take place with community advisory boards for consideration of biomarker integration in future community clinical trials protocol reviews and for incorporation in advocacy efforts with research funders and sponsors.

Participants pointed at the need to distinguish biomarkers developed with the goal of predicting response at the individual *versus* at the trial level; the former would be evaluated in cohort studies on a single regimen, while the latter would be evaluated in a series of clinical trials investigating a broad range of treatment regimens and durations. Diagnostic trial platforms may be necessary where different markers would be evaluated in their capacity to predict relapse-free cure. One suggested example is a truncated trial where several biomarkers are evaluated in parallel and all patients are treated for a short 3-month duration with close monitoring of safety events and early relapse.

Participants also shared information on two data sources for biomarkers:

1. In the Predict TB trial (NCT02821832)[1,2], samples for biomarker evaluation are prospectively collected from whole genome sequence of confirmed relapsed or cured participants. These samples will be made available to the TB community and a Sampling Management Committee comprised of various stakeholders will evaluate and prioritize the samples that will be shared. The design of the Predict TB platform may serve as an example for embedding biomarker studies in future clinical trials and sharing data and information to further advance biomarker development.
2. The publicly available MARK-TB (Markedly Accelerating Research with Knowledge of Tuberculosis Biomarkers)[3] Biobank, sponsored by BMGF, Center for Disease Control and Prevention (CDC), US Food and Drug Administration (FDA), and National Institute of Health (NIH) and led by TB Alliance is also a model resource for data integration and sharing that can be improved and expanded. The dataset includes a cohort of clinical trial participants from the ReMOX (NCT00864383) and National Tuberculosis Programs (at the same sites ReMOX was conducted) with longitudinal samples for biomarker research and long-term clinical outcomes (relapse and cure).

[1] Chen RY, *et al*. Using biomarkers to predict TB treatment duration (Predict TB): a prospective, randomized, noninferiority, treatment shortening clinical trial. *Gates Open Res*. 2017 Nov 6;1:9. doi: 10.12688/gatesopenres.12750.1.

[2] https://predict-tb.com/

[3] MARK-TB was previously known as the Consortium for TB Biomarkers (CTB2); https://www.mark-tb.org/

Session 3: From Phase 2 to Phase 3: the role of adaptive and seamless designs to streamline clinical development

Co-Chairs: Michael Hölscher (Ludwig-Maximilians University) and Christian Lienhardt (IRD and University of Montpellier)
Keynote: Can innovations in phase 2 trial design facilitate rapid investigation of novel drug combinations in Phase 3 trial? – Geraint Davies (Liverpool University)
Discussant 1: How to evaluate in parallel several regimens with various duration, taking into account the severity of disease? (duration randomization) – Patrick Phillips (UCSF)
Discussant 2: Finding the right duration using multi-arm multi-stage designs – Thomas Jaki (Lancaster University)
Discussant 3: From phase 2 to phase 3: How to evaluate in parallel several regimens with various durations, taking into account the severity of disease? – Alex Carlton (GSK)
Round table panelist: Martin Boeree (Radboud University Medical Center), Daniel Everitt (TB Alliance), and Robin Mogg (Gates Medical Research Institute)

The objectives in this session were:

- To assess how innovation in clinical trials methodology can increase the confidence with which drug combinations can be tested early and brought to Phase 3 trial testing;
- To assess how these innovations can de-risk and accelerate early phase clinical development in TB, increasing the confidence that regimens selected for Phase 3 trials contain the right drugs at the right doses;
- To assess the possibility to evaluate in parallel several regimens with various durations.

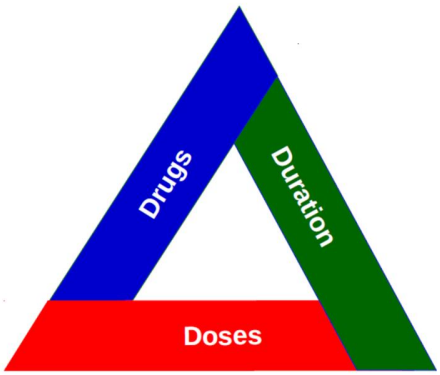
Questions addressed at the technical consultation:

1. *How best to address in a short, concise, and streamlined manner the key aspects of dose-finding, drug combination and duration selection for the identification and testing of best combined regimens for TB treatment?*
2. *Is it feasible to shorten the classical four-trial pathway? If yes, would three or two-trial pathways be desirable and feasible?*
3. *How can adaptation improve Phase 2A designs?*
4. *What is the minimum effect and sample size for Phase 2B studies?*
5. *Do Phase 2C designs adequately de-risk single duration Phase 3 trials, and does inflation of sample size in Phase 2C studies preclude efficient selection of combinations?*
6. *Are duration designs more applicable to a Phase 3 than a Phase 2 context or is information from both needed?*

Methods for transitioning TB drugs and regimens through Phase 2 to Phase 3 stages have evolved dramatically in the last decade, with the use of adaptive trial designs and the newly

proposed Phase 2C trial design, intending to facilitate the transition to the confirmatory Phase 3 trials. In addition, based on new PK-PD modelling and quantitative bacteriology approaches, Phase 2B studies can increasingly provide relevant data on the effect of drug doses and/or plasma concentrations on bacteriological response. Approaches combining preclinical and clinical data using Bayesian methods have the potential to guide early clinical development decisions in real time. All this conduces to consider clinical development from a pathway perspective, addressing the following questions: '*How much information can each stage provide?'* and *'How many trials are needed to confidently deploy a regimen?'* Together, the entire development pathway should consider each element of a trilemma (or triad) introduced by Dr. Davies, i.e.: dose-finding, combination selection and duration selection (Fig. 1). The pathways may differ according to the characteristics of the regimen and target population, as well as objectives and ultimate use (i.e. licensure for regulatory agencies vs. rapid deployment of regimens for programs) but overall, the information gathered in any pathway should collectively address each arm of the triad.

Fig. 1: Regimen trilemma



While it was agreed that innovative pathways with fewer trials (e.g. three- or even two-trial pathway) would be preferable to the traditional four-trial pathway (Phase 1, 2A, 2B, 3), there is still a tension between de-risking Phase 3 trials and accelerating the pathway with fewer potentially 'overloaded' trials. How do we balance the proposed innovative and compressed Phase 2 approaches with conducing to Phase 3 shorter regimens for which we understand fully the independent contributions of each component with identified doses? A two-trial pathway could potentially overburden the trials with many objectives, consequently increasing the stakes of each trial and increasing the risk of failing to introduce a novel, shortened, well-supported combination regimen promptly. Clearly, shortened development pathways that include fewer trials will require careful consideration of potential trade-offs, and development teams will need to weigh the potential gains (e.g. accelerated development) against additional hurdles and complexities (e.g. addressing multiple components of the triad simultaneously in a single trial) of shortened pathways. When designing clinical development pathways, it is most

important to gather all the necessary information, not only for a regulatory perspective, but also for rapid deployment of the regimens in TB-affected communities. Ultimately, it will be up to the development team, in close consultation with affected communities, policymakers (e.g. World Health Organization, National TB Programs) and regulators, to determine if a shortened pathway can better serve those needs.

The question arose on the most suitable timing of dose-finding studies within the clinical development pathway, and it appeared that there was no clear answer. Although testing optimal doses in monotherapy can be different from combination therapy conditions, some argued that monotherapy dose-finding is essential, especially because there are two aspects that must be addressed with dose-finding: efficacy and safety. More innovative and creative approaches for monotherapy studies would be useful, such as incorporating adaptation in Phase 2A studies, similar to adaptive trial designs in Phase 1 dose-ranging oncology studies, as real-time biomarkers become viable. Certainly, doses do not need to be fixed after Phase 2A studies and there are possibilities to include dose-finding objectives in Phase 2B studies with slightly larger sample sizes, so the question *'When should dose-finding happen?'* could be modified as *'How extensive is dose-finding at each stage of the pathway?'*. Several participants were of the opinion to capture and measure as much variability and dose-response relationships as possible in early and intermediate stages of development because these stages are meant for learning and understanding.

Experts also discussed on the optimal stage to estimate duration-response relationships that can provide confidence on the shortest non-inferior or superior durations and provide recommendations on durations to take forward in development. More investigation is needed to define whether duration-randomization designs that can provide data to models and estimate duration-response relationship are more applicable to a Phase 2 than a Phase 3 setting, or if they are needed in both. Because estimating the shape of the duration-response relationship is critical for informing clinical trial designs as early as the middle development phase, the role of preclinical studies and translational platforms (described in Session 1) to provide suitable information on these relationships should be considered, particularly since observed duration-response relationships may be different between regimens. As an alternative to modelling and estimating the duration-response relationship curve, the *order restricted multi-arm multi-stage design*, presented by Dr. Jaki (see Annex 3), may provide a viable approach to ranking durations without making assumptions on the relationship. Overall, there are two possibilities: i) to determine a range of suitable durations in Phase 2C to take forward to Phase 3, or ii) to determine a single duration in Phase 2C to take forward to Phase 3 and use Phase 2C for selection of combinations with more limited data on long-term outcomes - combined with other methods such as prediction-based or meta-regression to inform possibly (risk-stratified) multi-duration Phase 3 trials.

The endpoint to be used in duration-randomization trials that simultaneously test regimens at various durations requires further investigation. Additionally, more research is needed on the best approach to combine and compare endpoints based on time since randomization versus endpoints based on time since completion of treatment, as well as on the choice of

noninferiority margins for multi-duration Phase 3 trials, especially when safety is of concern at longer durations.

To improve efficiency of the development pathway, particularly in Phase 2C studies, the pros and cons of enrichment strategies were discussed at length. The major advantage of enrichment strategies with 'hard to treat' patients is that there are more endpoints (i.e. relapses) and therefore greater power for regimen comparisons. Some participants expected that if a regimen is found successful in patients with 'hard to treat' disease, then they should also be successful in populations with 'easy to treat' disease. However, several participants expressed concern regarding inadequate data on populations with 'easy to treat' disease. Patients with 'easy to treat' disease constitute at least 2/3 of the global TB population, which can create a large knowledge gap and missed opportunities when moving from Phase 2 to Phase 3 trials since a regimen that does not work in patients with 'hard to treat' might work in the lower risk patients. Without information on the lower risk population from Phase 2 studies, it would be more difficult to justify shorter regimens to be tested in Phase 3 studies, hence requiring a multi-duration design in Phase 3 trials to account for the broad eligibility criteria. This may not be an attractive proposition for developers. There was also concern regarding drug types displaying various capacities to cure 'hard to treat' phenotypes, as this capacity is related to bactericidal activity, dosing, and penetration into lesion sites. This may have implications for the treatment of patients who might respond well to shorter treatments, such as children. Further, patients with cavitary disease may be hard to treat because the drugs in the regimen may not have good lesion penetration properties, or because the safe dose that is tested is at the minimal effective dose, so the immune system will need to play a bigger role, making immunocompromised patients (i.e. HIV-coinfected or diabetic) at higher risk. Ultimately, the current definition of patient phenotypes is primarily based on standard-dose (10 mg/kg) rifampicin-containing regimens and still needs to be confirmed across all other regimens.

As an attractive alternative to enrichment designs, risk stratification was proposed to define the best durations for patients with 'hard to treat' disease and shortest possible for populations with 'easy to treat' disease. Risk stratification would allow data to be gathered on the full breadth of phenotypes in clinical trials, with potential for making more informed decisions when moving forward along the development pathway. It was acknowledged, however, that bringing risk stratification into Phase 2C trials with multi-duration designs, where combination and duration selection typically occur, would add an additional layer of complexity. Still, populations with 'hard to treat' disease are a minority of the TB populations that are driving the poor outcomes, leading to failed trials of potent regimens that work in most patients (80%); it is crucial that research is led on how to incorporate innovative approaches beyond the standard 'one-size-fits-all' strategies.

The likelihood of the COVID-19 pandemic having a durable impact on the clinical development of TB regimens was recognized. Further discussion is required on how to handle the obstacles that are anticipated when designing trials and enrolling patients in the near future. One promising outcome of the COVID era, however, is that trials can be conceived, designed, and

implemented rapidly, highlighting the importance of working cooperatively and expeditiously to get combinations trials launched.

Session 4: New phase 3 trials and how they will facilitate ultimate regimen development

Co-chairs: Carole Mitnick (Harvard University) and Christian Lienhardt (IRD and University of Montpellier)
Innovative approaches in Phase 3 trial design:
Keynote 1: Setting the stage –Andrew Nunn (University College London)
Keynote 2: What are the most promising designs to accelerate regimen development? –Patrick Phillips (UCSF)
The place of adaptive design in current Phase 3 trials: design protocols and interim lessons learnt
Discussant 1: The end-TB trial- Carole Mitnick (Harvard University)
Discussant 2: DRAMATIC trial- Robert Horsburgh (Boston University)
Discussant 3: Adaptive trial designs used for COVID 19- Michael Hughes (Harvard University)
Discussant 4: Towards more personalized medicine: using treatment stratification strategies to enhance cure –Payam Nahid (UCSF)
Round table panellist- Andrew Vernon (CDC), Eugene Sun (TB Alliance), and Angela Crook (University College London)

The objectives in this session were:

- To describe innovative approaches in TB Phase 3 trial design
- To understand the necessary conditions to keep phase 3 TB trials relevant in face of rapidly changing conditions.

Questions addressed at the technical consultation:

1. *What is the role and place of Phase 3 randomized, internally controlled trials for evidence generation and what is the additional benefit of complementary observational research?*
2. *What can we do to accelerate Phase 3 trials in TB treatment development? What can be the role of adaptive trial designs and of platform trials ?*
3. *3. Certain decisions may differ for drug-susceptible (DS) vs drug resistant (DR) TB drug development pathways, particularly with regard to the choice of control, definition of non-inferiority margin, reliance on early markers, directness of comparisons, pragmatism, heterogeneity, stratification.*
   Is there more appetite for innovation in DR-TB?
4. *What is the space for innovation?*
   a. *Appropriateness to time and question*
   b. *Reflect/anticipate guidance*
   c. *Mitigating effects of time: efficiency vs. shortcut?*
   d. *Need for effective communication of results of innovative trials frameworks*
   e. *Standardization vs. specification*

The landscape of TB treatment has evolved considerably over the last 10 years, necessitating careful consideration of various trial aspects and characteristics to ensure that Phase 3 trials deliver high-quality evidence. These include: trial design (superiority or non-inferiority); use of adaptive designs; treatment stratification; choice of control (selection and use of standard of care (SOC)); trial conduct (study quality, treatment adherence, missing data); and data analysis (intention-to-treat; per-protocol analyses; estimation of treatment effect).

There were some differences in the understanding of what should be considered a 'Phase 3 trial'. The core understanding of Phase 3 trials was shared by all, including industry*: a phase 3 trial is a pivotal confirmatory trial that provides the **main**, though not unique, basis for regulatory authority approval by demonstrating robust safety and efficacy of a drug or regimen*. In general, substantial complexity in phase 3 trials should be avoided, although regulators insist that the best science is used in clinical trial design. Nevertheless, this re-emphasizes the need to undertake the bulk of exploration on drug development triad in Phase 2 trials to minimize uncertainties going into Phase 3 trials. Importantly, phase 3 trials produce evidence to establish, revise or change treatment guidelines and inform programmatic aspects – but it may take more than one Phase 3 trial to achieve this, and therefore 'Phase 3' can encompass more than just those trials designed to inform regulatory authority approval. Thus, Guideline development groups are often faced with operational and programmatic questions that phase 3 trials conducted for evaluation of efficacy and safety cannot fully address. To assess pragmatic issues for new regimens, additional phase 3 or phase 4 trials would be needed.

The critical role of randomized internally controlled trials to produce evidence to inform treatment practice was emphasized, whether complemented or not by observational research studies to collect additional evidence. In any case, endpoints must be clearly defined, and analysis methods should be appropriate to the question and data collected. A new framework for estimands and sensitivity analysis in clinical trials has been proposed in a recent addendum (Feb. 2020) to the ICH E9 Harmonized Tripartite Guideline on Statistical Principles for Clinical Trials.[1] This framework provides new definitions - estimands, estimators and estimates - to help articulate the treatment effects that are to be measured, address some of the issues with different analysis populations (e.g. intent-to-treat vs modified-intent-to-treat vs per-protocol), provide clear interpretation for different stakeholders with different perspectives (different estimands and potentially different estimates for different purposes), permit transparent definitions with feedback from the TB community prior to analysis and presentation of results, and facilitate cross trial analysis. This framework provides an opportunity for clearer specification of the questions of interest, the analysis and the interpretation.

[1] EMA. ICH E9 (R1) addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials. Feb. 2020. https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e9-r1-addendum-estimands-sensitivity-analysis-clinical-trials-guideline-statistical-principles_en.pdf

Experts discussed how to best reconcile the desire to gather information from early looks at Phase 3 trial data without jeopardizing the trial or jumping to premature conclusions that are not supported by full analysis- particularly when there is considerable pressure to improve treatment guidelines. A suggestion was to perform a retrospective analysis to assess how interim analysis, if implemented in studies of completed Phase 3 trials, would have affected decisions in the development of the regimens tested. With this type of retrospective analysis, one can ask: 'How often would we have gotten it right or wrong with an interim analysis?'. As an example, such an analysis was performed within the context of the RIFAQUIN trial.[2] It showed that, if a Phase 2C study with 90 to 100 patients per arm had been performed prior to the RIFAQUIN trial, then there would have been enough data to decide *not to proceed* with the short 4-month experimental regimen, which turned out to be inferior to the 6-month standard regimen - but it would have been reasonable to proceed with the 6-month experimental regimen, which appeared to be noninferior.

Platform trial designs were proposed as a potentially promising approach to accelerate Phase 3 trials and increase potential benefits. Beside improved efficiencies in recruitment, staffing, as well as regulatory and ethical approvals, this approach allows the comparison of multiple interventions against a common control with an adaptive design that provides opportunities for several agents to be tested under a single 'master protocol'. Such trial designs will require clear definition and standardization of endpoints, as well as clear pre-specification of stopping rules. Real-time biomarkers that could accurately predict long term outcomes of interest would be of considerable benefit to these platform trials in informing interim analyses for lack of benefit (futility) and future trial designs.

Regarding the choice of comparator, the risk of 'biocreep' in successive noninferiority trials was highlighted. In trials of DR-TB, there is limited evidence from randomized trials to support current guidelines, and ongoing trials are still anticipated to set the benchmark for standard regimens that will become the control in future randomized trials. The uncertainty about treatment effect of standard of care complicates the selection of NI margins for DR-TB. For DS TB trials there is a potential risk of biocreep if a short regimen that has reached noninferiority is used as the control regimen in the next trial. If, on the other hand, the older well-established control is used, there would be no direct comparison with the new non-inferior regimens. Including both as controls would increase the trial size but would allow between-regimen comparisons while avoiding the risk of biocreep.

Many participants urged 'thinking beyond noninferiority' trial designs and their standard interpretation; for example, using Bayesian analysis to assess posterior probability of non-inferiority, considering superiority in patient-relevant outcomes, like cost-effectiveness, or in a composite efficacy-safety-duration outcome after showing noninferiority for regulatory licensing. Additionally, superiority designs could examine the issue of current control regimens being unforgiving with regard to poor adherence. For example, in addition to maximizing

[2] Phillips PP *et al*. A new trial design to accelerate tuberculosis drug development: the Phase IIC Selection Trial with Extended Post-treatment follow-up (STEP). *BMC Med.* 2016 Mar 23;14:51. doi: 10.1186/s12916-016-0597-3.

adherence to show efficacy, as done in current phase 3 clinical trials, a pragmatic trial (phase 3 or phase 4) that tests regimens under routine conditions may be used to show superiority of more forgiving regimens over an unforgiving control regimen.

Incorporating stratified medicine principles into clinical trials may be another promising avenue. There is now robust evidence supporting the evaluation of stratified medicine approaches to treating people with TB.[3] In this respect, the newly proposed SPECTRA-TB (Stratified Patient-Centered Treatment Regimens for Active TB) trial design aims to compare the superiority of a stratified medicine strategy where duration is selected based on baseline patient risk factors, over the 'one-size-fits-all' fixed duration strategy (presented by Dr. Nahid, Annex 3). Alternatively, such design could evaluate a regimen in a population with 'hard to treat' disease but also include a small group with 'easy' and 'moderate to treat' disease. Then, during post-licensure, the shortest possible duration for the 'easy to treat' disease can be further investigated. However, several participants were concerned about focusing on regimens developed for 'hard to treat' disease, which is estimated to be present in only 20-25% of patients with TB: indeed, "overtreating" patients with less-severe disease may have important, negative implications for safety.

Lastly, consultation participants considered the importance of learning from contemporary COVID-19 trials when designing TB trials. A Viewpoint by Nicole Mather in reference to the UK's RECOVERY platform trial, was briefly discussed.[4] In that viewpoint, the author signalled how investigators '*repurposed infrastructure so that clinical trials could safely get data about more treatments from more patients more quickly'*. Further, involving regulatory authorities early in the design and development of these trials provided credibility with pharmaceutical companies and motivated their interest to be part of large-scale collaborations, which has been the foundation of many COVID-19 trials. With respect to TB, the increasing role and engagement of regulatory authorities was applauded. Participants, however, acknowledged that a major challenge persists with generating commercial engagement and collaborations with pharmaceutical companies. Multiple pleas were made to continue considering how to create the necessary large collaborations among multiple clinical trials networks and to adopt the approaches used in COVID-19 pandemic to address the TB epidemic.

[3] Imperial MZ, *et al*. A patient-level pooled analysis of treatment-shortening regimens for drug-susceptible pulmonary tuberculosis. *Nat Med.* 2018 Nov;24(11):1708-1715. doi: 10.1038/s41591-018-0224-2.
[4] Mather N. How we accelerated clinical trials in the age of coronavirus. *Nature.* 2020 Aug;584(7821):326. doi: 10.1038/d41586-020-02416-z.

Session 5: Real-world evidence and cohort data; special populations

Co-Chairs: Payam Nahid (UCSF) and Christian Lienhardt (IRD and University of Montpellier)
Keynote: Non-randomized data in policy development: what are their needs and roles in the production of evidence? –Payam Nahid (UCSF)
Discussant 1: The importance of safety data: can we accelerate regimen development without compromising safety? –Charles Wells (Gates Medical Research Institute)
Discussants 2 and 3: Conducting trials in key/vulnerable populations: what is the progress made and what are persisting challenges- Amita Gupta (Johns Hopkins University) and Anneke Hesseling (Stellenbosch University)
Discussants 4-8: The use of new drugs/regimens beyond registration:

1. *The view of sponsors –Alex Pym (Janssen Pharmaceutica) and Gavin Koh (GSK)*
2. *The view of the regulator –Marco Cavaleri (European Medicines Agency)*
3. *The view of the policy maker –Dennis Falzon (WHO)*
4. *The view of civil society –Lindsay McKenna (Treatment Action Group)*

Round table panelist: Susan Swindells (University of Nebraska), Vindi Singh (WHO), Sumathi Nambiar (FDA), and Kissa Mwamwitwa (Tanzania Medicines and Medical Devices Authority), and Francis Varaine (Médecins Sans Frontières)

The objectives in this session were:

- To determine the complementary activities and data to be collected in parallel or beyond pivotal Phase 3 clinical trials to assist in the development of policy recommendations on TB treatment
- To assess the need of non-randomized data and examine their complementary role in raising evidence for policy making;
- To assess the need of conducting trials in key populations.

Questions addressed at the technical consultation:

1. *What data are needed for the development of policy and/or treatment guidelines and what would be the optimal time to collect these data? What studies should be conducted to provide these data?*
2. *What is the role of observational data in policy making? What are the standards that observational data need to adhere to in order to be used?*
3. *How can evidence on PK, efficacy and safety be generated in key populations such as pregnant and lactating women, people living with HIV and children? How can these populations be included in clinical trials?*
4. *How can National TB Programmes, TB patients, and affected communities be engaged in research?*
5. *How can we as a TB research community ensure that all data required for policy and/or treatment guidelines are collected?*

Policy making on new TB treatments requires high quality data on whether interventions are optimal in terms of safety and efficacy, equally feasible in low-, medium-, and high- resource settings, and actionable in scaling up for global use. Participants agreed that there is a role for *real world evidence* in policy making, in addition to Phase 3 data. Experts recognized that, in recent years, treatment guidelines for DR TB have been predominately, but not exclusively, structured and based on observational data (e.g. individual patient data meta-analysis from programmatic data), primarily due to the lack of clinical trial data and pressure to update guidelines based on observational data while trials were still on-going. However, they considered that, moving forward, treatment guidelines and recommendations for the use of new drugs and regimens *should be based primarily on data from randomized clinical trials*, supplemented, as needs be, by data collected through rigorous pragmatic or operational research initiatives. Considering the experiences and lessons learned in the last decade, a strategic vision that goes beyond registration needs to be developed urgently to move regimens forward in a more efficient way.[5] Both the European Medicines Agency (EMA) and the U.S. Food and Drug Administration (FDA) are currently establishing frameworks that bring together regulators and recommending bodies to discuss the use of real-world evidence and the design of post-approval studies to inform regulatory and policy decisions. As an example, the European Network of Centres of Pharmacoepidemiology and Pharmacovigilance (ENCePP) has created a registry of all on-going studies in Europe that provide guidelines on how best to conduct a scientifically sound study. Late-stage clinical trials that serve together the objective of registration of a new TB drug or regimen and the development of public health guidelines would be ideal, with information collected on long-term, patient- and population-relevant outcomes, feasibility, acceptability, resource use, equity, and quality of life. In this context, *pragmatic effectiveness trials* that serve as a middle ground between pivotal efficacy studies and observational studies, need more consideration. In such trials, randomization is preserved,

[5] Lienhardt C, Vernon AA, Cavaleri M, Nambiar S, Nahid P (2019) Development of new TB regimens: Harmonizing trial design, product registration requirements, and public health guidance. PLoS Med 16(9): e1002915. https://doi. org/10.1371/journal.pmed.1002915

but a broader population is included, enhancing the generalizability of results and better reflecting the real-world conditions compared to pivotal trials.

Participants agreed that non-randomized, observational data and real-world evidence have a role to play in complementing Phase 3 trials by (i) addressing feasibility, acceptability, delivery strategies, and quality of life issues; (ii) generating data on special populations not customarily enrolled in clinical trials; (iii) evaluating safety and effectiveness in broader, more diverse populations; and (iv) providing additional data on post-marketing toxicities and adverse events. In contrast, these have limited relevance as the sole and primary source of evidence for policy making. Relative to randomized clinical trial data, non-randomized observational data, although large (e.g. > 12,000 patients in an MDR TB database[6]), are more complex and difficult to analyse and to interpret because they are extremely heterogenous, include data that are obsolete or irrelevant to current situations (e.g. regimens that do not include newest drugs for MDR-TB), and have complex missing data problems. While sophisticated data analysis approaches can control for inherent biases (to some degree), they are not sufficient to overcome all of them due to known and unknown confounding and can make results difficult to interpret. Ensuring standardization and harmonization of data collection may help simplify the analysis and increase the quality of data to inform policy making.

There was general agreement on the need to improve the evidence base for scientifically complex populations (e.g. pregnant and lactating women, people living with HIV) and paediatric populations; these should be integrated in clinical trials to acquire efficacy and safety data. Considerations are needed on how to include these populations in clinical trials, for example by using a staged approach for pregnant women and children, particularly to gather much-needed PK data. A standardized approach that would offer pregnant women an opportunity to continue in a study with the assigned experimental regimen or to switch to the control regimen but remain part of the trial (so follow up can be completed, allowing for contribution to the larger registry) would be suitable. However, the major limitation to this approach is the absence of insurance policies that would cover damages to the foetus or the mother while remaining in the trial. Regarding children, there is emerging consensus that extrapolation of efficacy from adults is generally acceptable, with only some paediatric questions requiring efficacy studies. A key priority is to generate PK data for representative ages and determine the appropriate doses, rather than conducting efficacy studies. However, children and adolescents should still be included in Phase 3 trials so that the results from these studies can be more generalizable.

When accelerating the regimen development pathway, there should be equal consideration for addressing knowledge gaps in efficacy and safety. In the last decade, safety concerns have slowed developmental pathways for new TB drugs and regimens - e.g. the uptake of bedaquiline and delamanid has been slowed down due to safety concerns, especially on their combination use, and the NixTB trial regimen, although formally approved by FDA and EMA, is

[6] Collaborative Group for the Meta-Analysis of Individual Patient Data in MDR-TB treatment–2017, *et. al*. Treatment correlates of successful outcomes in pulmonary multidrug-resistant tuberculosis: an individual patient data meta-analysis. *Lancet.* 2018 Sep 8;392(10150):821-834. doi: 10.1016/S0140-6736(18)31644-1.

still compounded with safety concerns. Adequate safety databases and continued surveillance post-approval is critical for regulatory decisions on benefit-risk and policy decisions. More discussion is needed on how protocols for observational studies and operational research can be standardized and aligned to capture suitable data that fulfil the needs of regulators and policy makers, particularly for establishing the safety profile.

A question was raised on how to get key evidence needed for policy making when sponsors are unlikely to collect these. For example, the DELIBERATE trial (NCT02583048 – see ANNEX 3) that assessed the potential risk of additive cardiotoxic effects of bedaquiline and delamanid was critical to moving these two drugs forward, yet the respective industry sponsors did not take on the project that was eventually sponsored by the NIAID AIDS Clinical Trials Group. Experts also agreed on the critical need to go beyond anecdotal and superficial approaches to generate high quality evidence for patient preferences and accessibility. Often, data collection on patient preferences is reserved for later stages of research, most commonly after issuing guidance based on assumptions in the absence of information. There is a need to include rigorously collected qualitative data as part of clinical trials early in the development pathway. TB treatment research consortia and sponsors should conduct complementary qualitative research to understand how TB-affected communities consider trade-offs in efficacy, safety, tolerability and duration; the results from such qualitative assessment would usefully inform trial design (e.g. for noninferiority margin selection).

Participants agreed that engaging National TB Programmes in operational research would facilitate translating research to policy and potentially accelerating intervention uptake. For example, National TB Programmes in 22 countries have shown high interest in supporting and adapting the ShORRT (Short, all-Oral Regimens for Rifampicin-resistant Tuberculosis) operational research package that aims to facilitate the conduct of operational research on all-oral shorter rifampicin-resistant TB treatment regimens with the aim to generate data that are harmonized across different implementation settings.[7]

Finally, with growth and sophistication of TB Community Advisory Boards in high burden countries, the importance of community engagement in TB research was underscored. Communities that feel connected to research are shown to be more likely to participate in clinical studies, with favourable effects on recruitment, enrolment and retention into studies, as well as contributing to reciprocal and lasting partnerships between communities and researchers. Communities that understand TB science are better positioned to advocate for continued research before governments and other funders and may increase uptake and adoption of TB research outputs.

Conclusions

Following on the essential aspects identified in the first Technical Consultation in 2018, this 2nd technical consultation on *"Innovative Tuberculosis Clinical Trial Design for Development of New*

[7] Resources available at https://www.who.int/tdr/research/tb_hiv/en/.

TB Treatments" has confirmed the remarkable advances made in TB drug development, from preclinical to clinical development phases, as well as in clinical trial designs, and use of real-world evidence to inform policy recommendations The thought-provoking presentations and the lively, rich, and productive discussions resulted in new suggestions to accelerate treatment development and further research areas.

ANNEX 1: Agenda

Agenda of WHO Technical Consultation on
Innovative Tuberculosis Clinical Trial Design for Development of New TB Treatments

Thursday 20 August, Tuesday 1 September,
Tuesday 15 September, Tuesday 29 September, Tuesday 13 October 2020

Background

After decades of stagnation, research in tuberculosis (TB) therapeutics is experiencing a renaissance, with an increasing number of new and repurposed compounds undergoing evaluation as part of novel treatment regimens. The development of new TB drugs is, however, complex, lengthy, and costly, and the pathway to proven new TB treatment regimens is fraught with numerous obstacles and uncertainties. With the emergence of several new chemical entities (NCEs) expected to transition into clinical testing in the next 5 years, there is a real possibility of shorter, simpler, safer regimens for all forms of TB becoming available. Much insight has been gained into clinical investigation of TB drugs from the development programs for new TB drugs and regimens for M(X)DR-TB over the past 5-10 years. Additionally, investigators in the field have learned much from recent large clinical trials evaluating treatment shortening potential for drug susceptible TB using existing tools. Taken together, it is hoped that this experience will lead to well-designed and conducted clinical trials evaluating the next generation of drugs and regimens leading to identification of ultrashort, safe, and effective regimens so desperately needed. To address these coherently, in March 14–16, 2018, the WHO organized a technical consultation on "Advances in Clinical Trial Design for New TB Treatments" to identify and outline, through expert consensus, the optimal characteristics of clinical trial designs to inform policy guidance for the development of new TB regimens. Building on the lessons learned from the rich history of TB clinical trials, the WHO technical consultation reviewed the various research designs and tools currently used in the conduct of clinical trials for development of new TB treatments and made a series of proposals to advance these further, seeking to move from evolutionary change informed by history to a bolder approach to innovation geared to the future. These were described in detail in the Consultation report, as well as in a series of papers that were assembled in a special PLOS Medicine Collection, "Advances in Clinical Trial Design for Development of New Tuberculosis Treatments", published from March 2019 to February 2020.

Over the last 2 years, remarkable efforts have been taking place, and NCEs have transferred from preclinical to early clinical phases, providing new hopes for the generation of better treatment regimens, utilising the new designs that were discussed in the WHO technical consultation. In parallel, further evidence was produced on the capacity to use novel designs, and further progress were being done, legitimating an update of the situation regarding the novel clinical trial designs for new TB treatment, integrating the vast experience gained by industry groups performing TB drug R&D.

Technical consultation objectives:
To gain large consensus on evidence-based approaches to trial designs and use of data to inform policy guidance on new treatment regimens for TB with the view to facilitate coordination of future regimen development efforts.

Specific objectives:

1. to review latest evidence and assess the progress made since the Technical Consultation on "Advances in clinical trial design for TB treatment" in 2018;
2. to review novel study designs with the view to streamline regimen development to test regimens with known drug composition, dosing and duration taking into account recent developments in methods, tools and biomarkers;
3. to reach consensus on evidence-based approaches to trial designs and use of data to inform policy guidance on new treatment regimens for TB.

Expected outcome:
A WHO position statement on the use of novel clinical trial designs for better TB treatments development.

Dates:
Considering that it is currently not possible to hold a face-to-face meeting, we will hold a series of 5 webinars, staged from August to October 2020. Each 3-hour webinar is designed to address one specific topic. The webinars will be spaced by 2-week periods. The webinars are organized by WHO/GTB, assisted by Dr Christian Lienhardt (IRD and University of Montpellier) who will co-chair each session.

Dates	***Session***	***Topic***	***Session co-chair***
20 August 15h to 18h CEST	Session 1	Translational PK/PD modelling techniques for bridging preclinical to clinical development phases.	Rada Savic
1 September 15h to 18h CEST	Session 2	Biomarkers to support and/or accelerate decisions on suitable regimens to be tested.	Morton Ruhwald
15 September 16h to 19h CEST	Session 3	From Phase II to Phase III: the role of adaptive and seamless designs to streamline development.	Michael Hölscher
29 September 15h to 18h CEST	Session 4	New Phase III trial designs and how they will facilitate ultimate regimen development	Carole Mitnick
13 October 15h to 19h CEST	Session 5	Real-world evidence and cohort data; special populations	Payam Nahid

Session 1: Translational PK/PD modelling techniques for bridging preclinical to clinical development phases. Thursday 20 August; 15h to 18h CEST

New translational models are being developed to inform regimen optimization and predict outcomes of late-stage clinical trials. These models, integrating mouse exposure-response data, information about disease pathology, and immune responses, linked with clinical PK, have been shown to predict the CFU counts in TB patients enrolled in clinical monotherapy EBA studies for a series of drugs. Thus, preclinical information could be a useful source of data for initial predictions of duration of *regimens*, based on bacterial elimination rates or relapse experiments. Researchers are now proposing to expand translational platforms to include combination regimens and shorter durations of treatment, integrating key data from PK-PD studies, drug-drug interactions, necrotic lesion penetration, and potentially the drug effect on bacterial persisters. Similar approaches combining preclinical and clinical data using Bayesian approaches have the potential to guide early phase clinical development decisions in real time. Such model-informed drug development may greatly improve the predictive accuracy of preclinical data and accelerate TB drug development. Prospective testing of these translational models is a critical step toward bridging the preclinical-clinical divide in more efficient and informative ways.

Objectives:

- To examine approaches combining preclinical and clinical data to guide early drug development;
- to evaluate the role and use of translational modelling to bear on clinical development decisions with regard to appropriate drug combinations and predictions of duration of regimens, based on bacterial elimination rates or relapse experiments.

Expected outputs: Better understanding of the role of a "**translational space**" to broaden our understanding of how preclinical models can help drive more efficient and effective clinical development of novel TB regimens by facilitating the transition from pre-clinical to clinical phases.

Session co-chair: Rada Savic

Time	**Item**	**Presenter**
5 min	Opening / housekeeping	Saskia den Boon (WHO/GTB)
10 min	Declarations of Interest Statements	Saskia den Boon (WHO/GTB)
10 min	Welcome	Tereza Kasaeva (WHO/GTB)
15 min	Introduction to the webinar series and objectives of session 1	Christian Lienhardt (IRD)
30 min	The role of Translational Platforms for Tuberculosis Drug Development	Rada Savic (UCSF)
15 min	Evaluating dose rationale for antimicrobial drug combinations	Oscar Della Pasqua (UCL/GSK)
15 min	Does translational modelling have the promise of de-risking and accelerating early phase clinical development in TB ?	Kelly Dooley (JHU)
45 min	Discussion 1. Virtual round table: 2. General discussion	Short comments (2 min) from -Debra Hanna (BMGF) -Florian Von Groote (Evotec) -Elin Svensson (Radboud Univ) -Eric Nuermberger (John Hopkins) -Ulrika Simonsson (Uppsala)
15 min	Summary and outstanding questions	Rada Savic/ Christian Lienhardt

Session 2: Biomarkers to support and/or accelerate decisions on suitable regimens to be tested.
Tuesday 1 September; 15h to 18h CEST

In order for regimens to be reliably selected for Phase 2 testing, investigators need to have reasonable confidence that the intermediate bacteriological endpoints on which they currently rely can be trusted to correctly 'predict' treatment effects on definitive long-term outcomes - such as treatment failure and relapse. Despite a wide series of studies and analyses, the traditional *8-week culture conversion rate* has not been clearly validated as a reliable surrogate marker for long-term treatment response in Phase 3 trials. Over the last decade, new approaches to Phase 2B studies based on longitudinal statistical modelling of quantitative bacteriology, time-to-positivity in MGIT or time-to-culture conversion data, have been used with more intensive sampling of sputum at earlier time points and reduced sample sizes. Because outcomes are measured on a continuous rather than binary scale, they offer longer-term advantages in terms of validation over the 8-week endpoint which is now approaching its ceiling as regimens improve in efficacy. Several RNA expression, cytokine, bacterial and radiological biomarkers have been proposed in the literature, but to date there has been neither comparison nor prospective validation of these biomarkers. Biomarkers that can predict relapse and guide treatment duration using innovative adaptive trial designs would greatly accelerate drug development in TB by enabling prioritised evaluation of the most promising regimens.

Objectives:

1) to examine the current portfolio of most advanced biomarkers for treatment response that can be used for TB drug development;
2) to define readiness for inclusion of these novel biomarkers to support and/or accelerate go/no-go decisions in Phase 2/3 clinical trials.

Expected output: Knowledge gained on most advanced biomarkers for treatment response that can be used for TB treatment development.

Session co-chair: Morten Ruhwald

Time	Item	Presenter
5 min	Opening / housekeeping	Saskia den Boon (WHO/GTB)
15 min	Recap Session 1 and objectives of Session 2	Christian Lienhardt (IRD)
20 min	Most advanced and promising biomarkers for treatment response that can be used in clinical development: a review	Morten Ruhwald (FIND)
40 min	Insights from most promising biomarkers: 1. An RNA ratio assay 2. Otsuka LAM assay 3. MBLA Assay 4. PET-CT: Highlights of NextGen EBA	Nick Walter (U. of Colorado) Yongge Liu (Otsuka) Timothy McHugh (UCL) Clifton Barry (NIH/NIAID)
20 min	Using biomarkers in clinical development: practical issues	Kelly Stinson (Cultura Incorporated)
15 min	PD versus surrogate: exploring future biomarkers	Debra Hanna (BMGF)
45 min	Discussion: 1. Virtual Round Table 2. General discussion	-Kathy Eisenach (Consultant) -Bob Wallis (Aurum) -Elizabeth Talbot (Dartmouth) -Payam Nahid (UCSF)
15 min	Summary and outstanding questions	Morten Ruhwald/Christian Lienhardt

Session 3: From Phase II to Phase III: the role of adaptive and seamless designs to streamline clinical development. Tuesday 15 September; 16h to 19h CEST

Approaches combining preclinical and clinical data using Bayesian methods have the potential to guide early phase clinical development decisions in real time and methods for transitioning TB drugs and regimens through Phase 2 to Phase 3 stage have evolved rapidly in the last decade. Thus, *Phase 2 EBA studies* are now proposed to be extended beyond 14 days to allow dosing of an experimental drug, within a multidrug therapy context (*14+14 design*). In this way, the safety of the investigational product can be directly assessed and drug–drug interactions can be measured to select the *optimal dosing of the candidate drug in the novel regimen*. Also, using PK/PD modelling approaches, *Phase 2B studies* can provide data on the effect of drug dose and/or plasma concentrations on bacteriological response. New approaches to Phase 2B studies based on longitudinal statistical modelling of quantitative bacteriology, (time to positivity in MGIT or time-to-culture conversion data) are increasingly adopted.

A major progress has been made with the use of *adaptive trial designs* in TB treatment development. Among these, the *multi-arm, multi-stage (MAMS) design* features early discontinuation of test regimens that fail to show promise at interim analyses. To facilitate the transition, s*eamless Phase 2/3 designs* are proposed in which the adaptive evaluation of regimens is followed by enrichment of the successful arm(s) with additional participants in the Phase 3 stage to achieve appropriate power for comparisons on the long-term outcomes in the selected arms. The key feature of seamless design is that Phases 2 and 3 are interwoven into a single trial with multiple stages. Further, the new *Phase 2C trial* design intends to facilitate the transition to the confirmatory Phase 3 trials: novel regimens are tested for their full intended duration of time to capture the fullest information about safety and microbiologic activity of the regimen, as well as give an *indication* on efficacy based on early relapse.

Objectives:

1. To assess how innovation in clinical trials methodology can increase the confidence with which drug combinations can be tested early and brought to Phase 3 trial testing;
2. To assess how these innovations can de-risk and accelerate early phase clinical development in TB, increasing the confidence that regimens selected for Phase 3 trials contain the right drugs at the right doses;
3. To assess the possibility to evaluate in parallel several regimens with various durations.

Expected outputs:

- *Knowledge on the means to accelerate early phase clinical development in TB, increasing the confidence in the selection of the regimens to be tested in Phase 3 trials.*

Session co-chair: Michael Hölscher

Time	**Item**	**Presenter**
5 min	Opening / housekeeping	Saskia den Boon
15 min	Re-cap session 2 and introduction to session 3	Christian Lienhardt (IRD)
30 min	Can innovations in phase 2 trial design facilitate rapid investigation of novel drug combinations in Phase 3 trial? (highlight alternative pathways)	Gerry Davies (Liverpool University)
45 min	How to evaluate in parallel several regimens with various durations, taking into account the severity of disease? (duration randomization)	-Patrick Phillips (UCSF) -Thomas Jaki (Lancaster University) - Alex Carlton (GSK)

45 min	Discussion: 1. Round Table 2. General discussion	- Martin Boeree (Radboud University) - Dan Everitt (TB Alliance) - Robin Mogg (GMRI)
15 min	Summary and outstanding questions	Michael Hölscher /Christian Lienhardt

Session 4: New Phase III trial designs and how they will facilitate ultimate regimen development. Tuesday 29 September; 15h to 18h CEST

The currently expanding pipeline of new drugs, new diagnostics, and new methods make possible real transformation in TB treatment (https://www.newtbdrugs.org/pipeline/clinical). The landscape of tuberculosis (TB) treatment has thus evolved considerably over the last 10 years, necessitating careful consideration of various trial aspects and characteristics to ensure that TB phase 3 trials deliver high-quality evidence. These include: trial design (superiority or non-inferiority; use of adaptive designs; treatment stratification), choice of control (with particular attention to the selection and use of SOC), trial conduct (study quality, treatment adherence, missing data) and data analysis (intention-to-treat; per-protocol analyses; estimation of treatment effect).

Objectives:

1. To describe innovative approaches in TB Phase III trial design
2. To understand the necessary conditions to keep phase III TB trials relevant in front of changing conditions

Expected outputs:
Knowledge on the main Phase 3 trial conditions and characteristics that are indispensable for delivery of most relevant and reliable evidence for effective and safe patients' treatment.

Session co-chair: Carole Mitnick

Time (+ questions)	**Item**	**Presenter**
5 min	Opening / housekeeping	Saskia den Boon (WHO/GTB)
15 min	Re-cap session 2 and introduction to session 3	Christian Lienhardt (IRD)
5 min	Introduction	Session chair
35 min	Innovative approaches in Phase III trial design: 1. Setting the stage 2. What are the most promising designs to accelerate regimen development?	Andrew J. Nunn Patrick Phillips
45 min	The place of adaptive design in current Phase 3 trials: design protocols and interim lessons learnt 1. The end-TB trial 2. DRAMATIC trial 3. Adaptive trial designs used for COVID 19	- Carole Mitnick (Harvard) - Bob Horsburgh (Boston Univ) - Michael Hughes (NIH-ACTG)

10 min	Towards more personalised medicine: using treatment stratification strategies to enhance cure.	Payam Nahid (UCSF)
45 min	Discussion 1. Virtual Round Table 2. General Discussion	- Andrew Vernon (CDC/TBTC) - Eugene Sun (TB Alliance) - Angela Crook (UCL)
15 min	Summary and outstanding questions	Session chair/Christian Lienhardt

Session 5: Real-world evidence and cohort data; special populations. Tuesday 13 October; 15h to 18h CEST

The development of shorter, simpler regimens combining new and/or existing drugs requires detailed information on their respective efficacy, safety and toxicity, their potential for drug–drug interactions, their propensity for development of drug resistance while on therapy, and their use in specific patient populations such as persons infected with human immunodeficiency virus (HIV), pregnant women, and children. Beyond efficacy, the capacity of the regimen to be delivered easily with no specific constraints of pre- or per-treatment monitoring will favour its effectiveness. In view of the increase in NCEs to be tested, and the enhanced use of novel trial designs in TB therapeutics, interactions are needed between researchers responsible for designing the next generation of TB trials, regulators, and policy makers for the development of subsequent policies on access to TB medicines. Late-phase clinical trial outputs that serve the objective of registration of a new TB drug or regimen can meet the needs of public health guidelines, provided that data on long-term, patient-relevant, and population-relevant outcomes are being collected. Additionally, public health factors such as feasibility, acceptability, resource use, equity, and quality of life should be part of data collection, as these are necessary when formulating public health recommendations. These can be obtained through non-randomized data gathered outside of a trial setting, mostly under programme conditions, or through the conduct of post-hoc pragmatic trials. Relevance of such data have to be discussed and agreed formally; in this, the dialogue between drug developers and regulators could advantageously include policy-makers to ensure best access and use of the new TB medicines or regimens. This would deter the risk of conflicting interpretation and/or messaging provided by investigators and policy makers that may hamper the access to- and use of- new regimens for TB treatment.

Objectives:

- To determine the activities to be carried out and the data to be collected in parallel or beyond Phase 3 clinical trials so as to assist in development of public health recommendations (i.e. bringing research into policy and practice);
- To assess the need and use of non-randomized data and examine their role in policy making;
- To assess the need of conducting trials in special key populations.

Expected outputs:

To gain knowledge on the type of *additional data* that need to be collected *in parallel* or *beyond* clinical development phases for policy development and implementation and understand their contribution in production of evidence.

To clarify on the need to conduct trials in special key populations.

Session co-chair: Payam Nahid

Time	Item	Presenter
5 min	Opening / housekeeping	Saskia den Boon (WHO/GTB)
15 min	Recap session 4 and introduction to session 5	Christian Lienhardt (IRD)
5 min	Introduction	Session chair
10 min	Non-randomized data in policy development: what are their needs and roles in the production of evidence?	Payam Nahid (UCSF)
20 min	The importance of safety data: can we accelerate regimen development without compromising safety?	Charles Wells (GMRI)
15 min	Conducting trials in key/vulnerable populations: what are the progress made and persisting challenges?	Amita Gupta (John Hopkins) Anneke Hesseling (Cape Town)
20 min	Discussion - I	Christian Lienhardt
60 min	The use of new drugs/regimens beyond registration: 1. The view of sponsors 2. The view of the regulator 3. The view of the policy-maker 4. The view of civil society	 Alex Pym / Gavin Koh (J&J / GSK) Marco Cavaleri (EMA) Dennis Falzon (WHO) Lindsay McKenna (TAG)
45 min	Discussion - II Round Table	- Sue Swindells (ACTG) - Vindy Singh (WHO/HIV) - Sumathi Nambiar (FDA) - Helen Rees (SAHPRA) - Kissa Mwamwitwa (TDMA) - Francis Varaine (MSF)
15 min	Summary and outstanding questions	Session chair/Christian Lienhardt
	Closure	

ANNEX 2: List of Participants

Name	Organization
Dr Draurio BARREIRA	UNITAID, Geneva, Switzerland
Dr Clifton BARRY III	NIH/NIAID, Bethesda, Maryland, USA
Dr Martin BOEREE	Radboud University, Nijmegen, Netherlands
Dr Grania BRIDGEN	The Union, Paris, France
Mr Marco CAVALERI	European Medicines Agency (EMA), Amsterdam, Netherlands
Richard E. (Dick) CHAISSON	Johns Hopkins University, Baltimore, Maryland, USA
Ms Tsira CHAKHAIA	Georgia State University, Atlanta, USA
Professor Gavin CHURCHYARD	Aurum Institute, Johannesburg, South Africa
Dr Daniela CIRILLO	San Raffaele Institute, Milan, Italy
Dr Angela CROOK	University College London, London, United Kingdom
Dr Geraint (Gerry) DAVIES	University of Liverpool, Liverpool, United Kingdom
Dr Andreas DIACON	PanACEA / University of Cape Town Lung Institute, Cape Town, South Africa
Dr Lori DODD	NIH, Bethesda, Maryland, USA
Dr Kelly DOOLEY	Johns Hopkins University, Baltimore, Maryland, USA
Dr Kathleen (Kathy) EISENACH	Independent consultant, USA
Dr Daniel (Dan) EVERITT	TB Alliance, New York, USA
Professor Amita GUPTA	Johns Hopkins University, Baltimore, Maryland, USA
Dr Debra HANNA	Bill & Melinda Gates Foundation, Seattle, Washington, USA
Dr David HERMANN	Bill & Melinda Gates Foundation, Seattle, Washington, USA
Professor Anneke HESSELING	Stellenbosch University, Stellenbosch, South Africa
Dr Jeremy HILL	KNCV Tuberculosis Foundation, The Hague, Netherlands
Professor Michael HOLSCHER	Ludwig-Maximilians University München, Germany
Professor Robert HORSBURGH	Boston University Boston, Massachusetts, USA
Ms Shitong HUAN	Gates Foundation, China Program, China
Professor Michael HUGHES	Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA
Ms Marjorie IMPERIAL	UCSF, San Francisco, USA
Professor Thomas JAKI	Lancaster University, Lancaster, United Kingdom
Professor Amina JINDANI	St George's University of London, London, United Kingdom
Dr Soyeon KIM	Frontier Science Foundation, Brookline, Massachusetts, USA
Dr Christoph LANGE	Research Center Borstel, Borstel, Germany
Dr Barbara LAUGHON	National Institute of Allergy and Infectious Diseases (NIAID), National Institutes of Health (NIH), Betheda, Maryland, USA
Dr Christian LIENHARDT	Institute for Research in Sustainable Development (IRD), University of Montpellier, France; and London School of Hygiene and Tropical Medicine, London, United Kingdom.
Ms Yuhong LIU	National Clinical Center on Tuberculosis, China CDC, Beijing, China
Dr Vidya MAVE	Johns Hopkins University, Baltimore, Maryland, USA
Ms Lindsay MCKENNA	Treatment Action Group (TAG), USA
Dr Dick MENZIES	McGill University, Montreal, Quebec, Canada

Dr Carole MITNICK	Harvard Medical School, Boston, Massachusetts, USA
Ms Stellah MPAGAMA	Kilimanjaro Christian Medical College, Moshi, Kilimanjaro, United Republic of Tanzania
Ms Kissa MWAMWITWA	Tanzania Medicines and Medical Devices Authority (TMDA), Dar es salaam, United Republic of Tanzania
Professor Payam NAHID	UCSF, San Francisco, California, USA
Dr Sumathi NAMBIAR	US Food and Drug Administration (FDA), Silver Spring, Maryland, USA
Dr Norbert NDJEKA	National Department of Health, Johannesburg, South Africa
Dr Eric NUERMBERGER	John Hopkins University, Baltimore, Maryland, USA
Professor Andrew NUNN	University College London, London, United Kingdom
Dr Bern-Thomas NYANG'WA	MSF, London, United Kingdom
Professor Nick PATON	National University of Singapore, Singapore
Dr Patrick PHILLIPS	UCSF, San Francisco, California, USA
Professor Helen REES	SAHPRA and University of the Witwatersrand, Johannesburg, South Africa
Dr Morten RUHWALD	FIND, Geneva, Switzerland
Professor Rada SAVIC	UCSF, San Francisco, California, USA
Professor Ulrika SIMONSSON	Uppsala University, Uppsala, Sweden
Dr Alena SKRAHINA	Republican Research and Practical Center for Pulmonology and Tuberculosis, Minsk, Belarus
Dr Mel SPIGELMAN	TB Alliance, Pretoria, South Africa
Dr Kelly STINSON	Cultura, Incorporated, USA
Dr Eugene SUN	TB Alliance, Pretoria, South Africa
Dr Elin SVENSSON	Radboud University, Nijmegen, Netherlands
Professor Susan SWINDELLS	University of Nebraska, Ohama, Nebraska, USA
Dr Elizabeth TALBOT	Dartmouth Geisel School of Medicine, Lebanon, New Hampshire, USA
Professor Guy THWAITES	University of Oxford, Oxford, United Kingdom
Mr Francis VARAINE	MSF, Paris, France
Dr Andrew VERNON	CDC, Atlanta, Georgia, USA
Dr Nguyen VIET NHUNG	Viet Nam National Lung Hospital, Hanoi, Viet Nam
Robert WALLIS	The Aurum Institute, Johannesburg, South Africa
Dr Nicholas (Nick) WALTER	University of Colorado Anschutz Medical Center, Denver, Colorado, USA

Pharmaceutical industry

Dr Charles WELLS	Gates Medical Research Institute, Cambridge, Massachusetts, USA
Dr Robin MOGG	Gates Medical Research Institute, Cambridge, Massachusetts, USA
Dr Gavin KOH	GlaxoSmithKline (GSK), London, United Kingdom
Ms Alex CARLTON	GlaxoSmithKline, London, United Kingdom
Professor Oscar DELLA PASQUA	University College London and GSK, London, United Kingdom
Dr Alexander PYM	Janssen Pharmaceutica, United Kingdom
Dr Jeffrey HAFKIN	Otsuka, Princeton, New Jersey, USA
Dr Yongge LIU	Otsuka, Rockville, Maryland, USA
Dr Florian VON GROOTE	Evotec, Lyon, France

World Health Organization

Dr Tereza KASAEVA	Global Tuberculosis Programme
Dr Matteo ZIGNOL	Global Tuberculosis Programme
Dr Saskia DEN BOON	Global Tuberculosis Programme
Dr Dennis FALZON	Global Tuberculosis Programme
Dr Fuad MIRZAYEV	Global Tuberculosis Programme
Dr Kerri VINEY	Global Tuberculosis Programme
Dr Lorenzo MOJA	Health Product Policy and Standards
Dr Corinne MERLE	Special Programme for Research and Training in Tropical Diseases
Dr Vindi SINGH	Global HIV, Hepatitis and STIs Programmes
Dr Martina PENAZZATO	Global HIV, Hepatitis and STIs Programmes
Mr Albert FIGUERAS	Health Product Policy and Standards
Dr Benedikt HUTTNER	Health Product Policy and Standards

ANNEX 3: Summary of Presentations

Session 1: Translational PK-PD modelling techniques for bridging preclinical to clinical development phases

Co-chairs: Rada Savic (UCSF) and Christian Lienhardt (IRD and University of Montpellier)

Introduction to Session 1- Christian Lienhardt (IRD and University of Montpellier)

Preclinical information could be a useful source of data for initial predictions of duration of regimens, based on bacterial elimination rates or relapse experiments. Researchers are now proposing to expand translational platforms to include combination regimens and shorter durations of treatment, integrating key data from PK-PD studies, drug-drug interactions, necrotic lesion penetration, and potentially the drug effect on bacterial persisters. Similar approaches combining preclinical and clinical data using Bayesian approaches have the potential to guide early phase clinical development decisions in real time. Such model-informed drug development may greatly improve the predictive accuracy of preclinical data and accelerate TB drug development. Prospective testing of these translational models is a critical step toward bridging the preclinical- clinical divide in more efficient and informative ways.

Keynote: The role of Translational Platforms for Tuberculosis Drug Development- Rada Savic (UCSF)

In the 2018 WHO Technical Consultation, there was general agreement on the critical importance of quantitative PK-PD assessment throughout the whole developmental pathway. The proposed key research area of translational modelling and quantitative pharmacology to link all phases of regimen development has gained much interest and has substantially advanced over the last two years. Dr. Savic explained that successful translation from *in vitro* and *in vivo* experiments to clinical trials involves a complex multiscale approach that requires data integration from experiments that investigate efficacy of single and multidrug regimens, immunology, lung and lesion penetration, intra- and extra-cellular distribution, emergence of resistance and intra-bacterial drug transport. She strongly advocated for the development of quantitative translational platforms to optimize regimens, rank and prioritize regimens, and design clinical trials. These platforms should include a toolbox of methods and tools that integrates data across preclinical and clinical phases and describes plasma pharmacokinetic (PK) scaling, site-of-disease lesion PK, host immune and bacteria interplay, monotherapy PK-PD relationships, combination PK-PD relationships of multidrug regimens, emergence and impact of resistance, and biomarkers.

A workflow was proposed that integrates these tools with computational platforms to identify drug combinations that have the potential to accelerate sterilization, reduce relapse rates, and limit the emergence of resistance.[1] Dr. Savic highlighted the importance of accounting for the

[1] Ernest JP,*et al.* Development of New Tuberculosis Drugs: Translation to Regimen Composition for Drug-Sensitive and Multidrug-Resistant Tuberculosis. *Annu Rev Pharmacol Toxicol* 2021;61: 10.1146/annurev-pharmtox-030920-011143

host immune response when describing PK-PD relationships. She showed that murine PK-PD models accounting for TB infection and immune response have successfully predicted early bacterial activity from historical Phase 2A EBA studies. It was suggested that translational models accounting for these components have potential to replace monotherapy EBA studies for predicting efficacy but do not address safety objectives in these studies. For combination therapy EBA studies, with the goal of ranking regimens, empirical (non-pharmacological and data science driven) and mechanistic quantitative systems pharmacology (QSP) models that incorporated human immune response, granuloma lesions, multi-drug antimicrobial chemotherapy, and bacterial resistance, show promise in successfully predicting outcomes from historical clinical trials and ranking regimens based on clinical performance. Major advancements have also been made in the development of translational lesion-centric platforms from human lung resection data and the preclinical rabbit models, which are critical to describe the complex TB pathology and drug coverage in lesions. In summary, Dr. Savic categorized the role of translational platforms into three applications: (1) prediction of early bactericidal activity, (2) prediction of long-term combination therapy outcomes and regimen ranking, and (3) prediction of optimal regimens for all patients.

Discussant 1: Evaluating the dose rationale for anti-microbials- Oscar Della Pasqua (University College London and GSK)

Dr. Della Pasqua explained that dose optimization is critical for the success of future trials and requires robust characterization of the determinants of drug response and variability (PK-PD relationships). It is important to characterize the correlation between antibacterial activity, clinical cure and relapse but understanding the PK-PD relationships is critical for evidence-based dose rational decisions. Dr. Della Pasqua reflected on the strategies for dose and drug combination selections in recently failed and successful Phase 3 trials. In some clinical trial publications, dose rationale was presented but selected empirically without basis on quantitative PK-PD relationships, even in successful Phase 3 trials (e.g. STREAM Stage 1, ISRCTN78372190, NCT02409290; and Nix-TB, NCT02333799). Furthermore, drug combinations were commonly selected based on *in vivo* efficacy and currently approved doses without further treatment optimization. Therefore, individual drug contributions to overall treatment effect remains unclear, with little emphasis on dose optimization based on underlying PK-PD relationships.

Dr. Della Pasqua proposed an integrated translational pharmacology approach based on the use of drug-disease modelling that describe bacterial growth dynamics to prospectively select companion drugs and identify relevant doses/dosing regimens to be evaluated in humans in a systematic manner. The approach could be used to quantify the differences between preclinical models and clinical data to make more evidence-based decisions for dose optimization. He acknowledged the complexity of the system associated with translational pharmacology and suggested that simplification by dividing the system into smaller validated elements is needed to confidently scale and predict clinical outcomes prospectively. Overall, Dr. Della Pasqua explained that despite assumptions and potential limitations of quantitative translational approaches, the use of drug-disease modelling and simulation provides a robust framework for

dose selection, significantly increasing the probability of target attainment in patient populations.

Discussant 2: Does translational modelling have the promise of de-risking and accelerating early phase clinical development in TB? Kelly Dooley (Johns Hopkins University)

Dr. Dooley reviewed the role of current translational tools and the preclinical to clinical knowledge gaps that should be addressed to *de-risk* and accelerate early phase clinical development of TB regimens. She briefly discussed the knowledge gaps associated with:

- antagonism between drugs,
- over or under estimation of individual drug contributions,
- definition of PK-PD target(s) translated across species,
- omission of the time dependent PK-PD processes,
- extrapolation of individual drug activity in tested regimens to untested regimens,
- differences in drug and metabolite effects across species,
- differences in strains and minimum inhibitory concentration (MIC) variability across preclinical models and patient populations,
- extrapolation of optimal treatment duration from preclinical models to patient populations, and
- unaccounted human elements (adherence, tolerability and diversity).

To address these knowledge gaps, she summarized that translation platforms need to account for:

- plasma PK scaling (variability, protein binding),
- site-of-disease lesion PK and PKPD,
- host immune response and bacteria interplay,
- drug action and bacterial state interplay (e.g. fast *vs* slow multiplying bacteria at early and late stages of treatment),
- PKPD in context of multi-drug therapy,
- emergence of resistance, and
- time and dynamics in regard to PK-PD as TB pathology improves.

Dr. Dooley emphasized that it is 'not the lack of data but the lack of data integration' that is hindering optimal translation between preclinical models and patient populations. With the collaboration between industry, academia, regulatory and government agencies promoted by this Technical Consultation, the importance for preclinical scientists, clinical trialists, and modelers to work together is critical to produce the necessary data and fully gain understanding to bridge the knowledge gaps.

Round table panellist: Debra Hanna (BMGF), Florian Von Groote (Evotec), Elin Svensson (Radboud University), Eric Nuermberger (Johns Hopkins University), and Ulrika Simonsson (Uppsala University).

Drs. Hanna and Von Groote reinforced the critical importance and need to focus on the use of quantitative translational approaches, generation of high-quality data, and further validation

and refinement of current preclinical models. Dr. Hanna explained that the BMGF continues to strongly invest in this area with a new global collaboration launched in early 2020, the Project to Accelerate New Treatment for TB (PAN-TB collaboration), that focuses on the validation and refinement of the data-rich relapsing mouse model and *in vitro* hollow fiber system. There is also continued commitment to gathering and sharing the data as part of the Critical Path to Tuberculosis Drug Regimens effort, which has already led to several advancements in the design of preclinical experiments. It was acknowledged that the scarcity of clinical data on ultrashort regimens and the opportunity for drug combinations to have completely novel mechanism of actions will make predictions of clinical outcomes using translational platforms more challenging. Therefore, commenters emphasized the importance of optimally designing preclinical models, and getting regimens to the clinical stage to gather feedback data for informing and strengthening translational predictions.

Dr. Svensson acknowledged the difficult task of collecting valuable pulmonary lesion data in humans to inform preclinical models and advocated for more research on the reliability of rabbit preclinical models to confidently predict clinical data. She also recognized that the review of translational platforms presented by session speakers focused on the use of *in vivo* data with little emphasis on the *in vitro* hollow fiber system. More research is needed on the hollow fiber system's translational link since this *in vitro* system has several advantages (e.g. higher throughout, non-animal) compared to *in vivo* preclinical animal models.

Dr. Nuermberger welcomed the tremendous progress made and the inspiring new energy in the translational space of TB. The translational approaches presented here offer opportunities to move beyond single model analysis for more integrative quantitative approaches that complement an array of preclinical models and provide more evidence-based decision making. Translational approaches also provide the opportunity to bring a broader range of data into development process, particularly in areas that are more difficult to study in the clinical setting - for example, modelling of drug susceptibility across bacterial populations, impact of the emergence of drug resistance, treatment prevention therapies with long acting injectables and treatment of hard to treat populations. More research is needed on optimal experimental designs to inform translational modelling, which may differ from traditional experiments that are focused on other statistical endpoints. Dr. Nuermberger reiterated the critical importance of capturing evolving PK-PD relationships as pathology, and perhaps immune response, evolves during treatment. Lastly, Dr. Nuermberger made the case for promoting adoption of translational modelling exercises to inform clinical designs among various stakeholders.

Dr. Simonsson advocated that more research should be aimed at gaining a systematic understanding of the mechanistic role of classical PK-PD indices since it is currently included in several regulatory requirements. The regulatory agencies' view on the impact of PK-PD indices on regulatory requirements may change if a more systematic understanding of its role in relation to clinical outcomes can be provided. Additionally, more research at the preclinical stage is needed on the importance of PK interactions, which have been observed in Phase 1 studies, and the host immune response.

Session 2: Biomarkers to support and/or accelerate decisions on suitable regimens to be tested.

Co-Chairs: Morten Ruhwald (FIND) and Christian Lienhardt (IRD)

Introduction to Session 2 - Christian Lienhardt (IRD and University of Montpellier)

A key research question addressed at the 2018 WHO technical consultation was about the most efficient framework for identification and characterization of microbiology-based biomarkers to enable integration in modelling and simulation-based analyses. Since then, there have been several advancements in the area, particularly with research on novel biomarkers such as sputum LAM or the rRNA synthesis (RS) ratio, that open new perspectives for the design and conduct of clinical trials. Indeed, biomarkers that can predict relapse and guide treatment duration using innovative adaptive trial designs would greatly accelerate drug development by enabling prioritized evaluation of the most promising regimens. To select reliably regimens to be tested in Phase II, investigators need to have reasonable confidence that the intermediate endpoints on which they currently rely can be trusted to correctly 'predict' treatment effects on definite long-term outcomes - such as treatment failure and relapse.

At baseline, it is important to ensure that appropriate language is used to discuss biomarkers: A surrogate endpoint is defined as 'a biomarker intended to substitute for a patient-relevant clinical endpoint'. A biomarker is defined as 'a characteristic that is objectively measured and evaluated as an indication of normal biologic processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention'.[1] A 'surrogate marker' is defined as 'a laboratory measurement or physical sign that is used in therapeutic trials as a substitute for a clinically meaningful endpoint that is a direct measure of how a patient feels, functions, or survives and is expected to predict the effect of the therapy'.[2]

The primary difference between a biomarker and a surrogate marker is that a biomarker is a 'candidate' surrogate marker, whereas a surrogate marker is a test used, and taken, as a measure of the effects of a specific treatment. Although all surrogate endpoints are biomarkers, not all biomarkers are useful surrogate endpoints.[3] To be sure that the drug-induced effect on the surrogate predicts the desired clinical benefit, one has to understand the biological relationship between the surrogate and the clinical outcome. Furthermore, beyond knowing the biologic relationship between the surrogate and the clinical disease, we need to know the actions of the drug being investigated to conclude that the effect on the surrogate will translate into the beneficial clinical outcome desired. He described that a 'perfect' surrogate would fully capture the treatment effect on the definitive endpoint - and in the TB community we are still

[1] NIH Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clin. Pharmacol. Ther.* 2001 Mar. dol: 10.1067/mcp.2001.113989

[2] Temple R. Are surrogate markers adequate to assess cardiovascular disease drugs? *JAMA.* 1999 Aug 25; 282(8):790-5. doi: 10.1001/jama.282.8.790.

[3] Aronson JK. Biomarkers and surrogate endpoints. *Br J Clin Pharmacol.* 2005 May;59(5):491-4. doi: 10.1111/j.1365-2125.2005.02435.x.

in search for the *'perfect'* surrogate. Following the description by R. Katz[4] '*the ideal approach would be for a sponsor (in fact, the entire therapeutic community) to demonstrate that the effect on the surrogate and the effect on the clinical outcome are quantitatively similar across many drugs (and especially across many drug classes designed to treat a given clinical indication). In this way, one could be confident (…) that an effect seen on the surrogate for the next proposed treatment will have the expected and desired clinical effect. […] Such an approach is the only reliable method of demonstrating (…) that the effect seen on the surrogate, regardless of the treatment applied, will translate into the clinical benefit. Validated surrogate markers are those for which evidence has established that a drug-induced effect on the surrogate predicts (results in) the desired effect on the clinical outcome of interest'.* In summary, a surrogate endpoint should satisfy: 1) correlation with a definitive clinical endpoint (the treatment effect on surrogate corresponds to treatment effect on final outcome); 2) reproducibility / consistency of association between surrogate and final outcome; 3) clinical/biological plausibility of relation between surrogate and final outcome. [5]

Keynote: Most advanced and promising biomarkers for treatment response that can be used in clinical development: a review Dr. Morten Ruhwald (FIND)

To facilitate dosing and regimen selection decisions an ideal PD biomarker should be:

- sensitive and specific for clinically meaningful differences (e.g. relapse free cure);
- easy to sample, measure and report, preferably at bedside or point of care;
- on a continuous scale with a wide dynamic range and standard of care treatments in the middle of the scale to easily identify better or worse regimens;
- dynamic during early treatment in order to have shorter trials;
- translational between preclinical and clinical development stages; and
- robust, reproducible, and qualified.

According to the FDA Biomarker Qualification Program[6] a sponsor must:

- establish a reliable method to measure the biomarker and define a locked-down assay;
- provide analytical validation with detailed description of precision, accuracy, reproducibility, etc.; and
- show correlation between biomarker and outcome.

Based on these requirements, Dr. Ruhwald reviewed the current portfolio of the most advanced and promising biomarkers for treatment response that can be used in clinical development. There are two main categories of TB biomarkers: *host*- and *bacterial*- based biomarkers. Host-based assays measure inflammation or the host-immune and bacterial interplay, while bacterial-based assays can be separated into two main classes that measure either *M. tuberculosis* burden or *M. tuberculosis* fitness. Host-based biomarkers include

[4] Katz R. Biomarkers and surrogate markers: an FDA perspective. *NeuroRx.* 2004 Apr;1(2):189-95. doi: 10.1602/neurorx.1.2.189.

[5] Bucher HC, Guyatt GH, Cook DJ, Holbrook A, McAlister FA. Users' guides to the medical literature: XIX. Applying clinical trial results. A. How to use an article measuring the effect of an intervention on surrogate end points. Evidence-Based Medicine Working Group. *JAMA.* 1999 Aug 25;282(8):771-8. doi: 10.1001/jama.282.8.771.

[6] FDA. Biomarker Qualification Program. https://www.fda.gov/drugs/drug-development-tool-ddt-qualification-programs/biomarker-qualification-program

transcriptomic signatures, inflammation markers, host adaptive immune capacity markers, chest imaging, and clinicals signs and symptoms (i.e. clinical score, lung function). The most promising bacterial-based biomarkers include LAM concentrations in sputum or urine, MBL to measure *M. tuberculosis* burden and the RS ratio, culture free TB assay, and *M. tuberculosis* lineage and whole genome sequencing analysis to assess *M. tuberculosis* fitness. Based on these promising TB biomarkers, the current biomarker pipeline reflects different aspects of the disease, including host inflammation, host adaptive immune response, *M. tuberculosis* burden, and *M. tuberculosis* fitness and lineage. Dr. Ruhwald concluded that ideally, we would like a single biomarker that captures all of relevant aspects of disease, but in reality, an array of integrated biomarkers might be necessary.

Discussants 1-4: Insights from most promising biomarkers- Nicholas Walter (University of Colorado), Yongge Liu (Otsuka), Timothy McHugh (University College London) and Clifton Barry (NIH/NIAD)

On behalf of the Consortium for Applied Microbial Metrics, Dr. Walter discussed the RS ratio, a novel PD biomarker that describes the effect of drugs on rRNA synthesis, a fundamental bacterial physiological parameter. Suppression of the RS ratio appears to reflect the drugs' or regimens' sterilizing activity against the residual *M. tuberculosis* population[7]. In *M. tuberculosis*, the entire ribosomal operon includes stable mature rRNA, with 16S, 23S, and 5S subunits, and several interspersed unstable precursor rRNA (pre-rRNA) spacer sequences that are rapidly degraded, thereby serving as a biomarker of newly transcribed rRNA. The RS ratio is calculated as the relative amount of unstable, newly transcribed pre-rRNA spacers to stable mature rRNA, where higher ratio suggests more ongoing rRNA synthesis. Dr. Walter claimed that this biomarker provides a distinct new type of information on the physiological state of the *M. tuberculosis* populations, which can complement the information from existing PD markers that enumerate *M. tuberculosis* burden (i.e. CFU solid culture, TTP in liquid culture, and concentrations of LAM in sputum). In a study with the relapsing mouse model, the RS ratio recapitulated established rank order of sterilizing potency for standard of care and three novel regimens (PaMZ, BPaL, BPaMZ – Pa=pretomanid; M=moxifloxacin; Z=pyrazinamide; B=bedaquiline, L=linezolid), during the first weeks of treatment, at the end of treatment and at relapse assessment. Thus, this marker has promising ability to distinguish between regimens and serve as an early indicator of sterilizing activity. For this biomarker, the ultimate goal is not to recapitulate culture results, but rather to demonstrate its association with clinical endpoints so that it can ultimately help de-risk selection of regimens that move forward to late-stage development. This marker is also appealing for its quantitative nature and ability to be applied across all translational states (*in vitro*, *in vivo* and clinical studies), and should be evaluated and embedded in as many preclinical and clinical studies as possible to determine its true value and role for predicting clinical endpoints.

Dr. Liu discussed Otsuka's TB LAM ELISA assay and the potential for sputum LAM concentrations to be a biomarker of bacterial burden and PD response during treatment. The

[7] Walter ND, Born SEM, Robertson GT, *et al*. *Mycobacterium tuberculosis* precursor rRNA as a measure of treatment-shortening activity of drugs and regimen. Nature Communications 2021;12:2899

performance of TB LAM ELISA assay on clinical sputum samples have been examined in two diagnostic studies, two EBA studies, and a two-month culture conversion treatment study. Focusing on the analysis of data from one of the EBA studies (NexGen EBA, NCT02371681), Dr. Liu showed favourable correlation between sputum LAM concentrations versus CFU from solid culture and TTP from liquid culture during treatment and the consistent rank ordering of regimens with the use of LAM concentrations, CFU, and TTP. In addition, from the two-month treatment study, he showed that conversion of LAM concentrations to below lower limit of quantification and MGIT culture to negative were consistently aligned. Based on these favourable results, Dr. Liu explained that sputum LAM concentrations, as a measure of bacterial burden that is available near real-time, has potential to expedite clinical trials and inform innovate trial designs. However, collaborative research is still required before sputum LAM concentration would gain any regulatory endorsement as a biomarker in TB drug development.

Dr. McHugh discussed the molecular bacterial load assay (MBLA) that is based on RNA extraction from sputum and quantitative amplification of 16s rDNA. It has been validated that longitudinal profiles of MBLA in pulmonary TB patients reflect the bi-phasic decline observed in CFU and TTP from typical solid and liquid culture assays, respectively, during the first 14 days of standard TB therapy. However, a quantitative difference was observed between MBLA and solid culture assays, with MBLA detecting slightly higher CFU counts. This quantitative difference was replicated in the murine studies. More evidence is required to decipher whether this difference is due to assay sensitivity or, as suspected by Dr. McHugh, differing physiological states captured by each assay. Based on the experience gained from monitoring treatment response using MBLA in the PANBIOME (Pan-African Biomarker Expansion Programme) study, the major strengths of MBLA include robust and reproducible results across four African sites and reduced susceptibility to contamination that is often observed in culture assays. Further prospective evaluations are still required to confirm the interpretation of MBLA in preclinical studies and clinical trials and to understand where it belongs in the portfolio of TB biomarkers. MBLA is now being developed into a commercial platform by LifeARC and University of St. Andrews, but there are still practical considerations for operationalization of this tool.

Lastly, Dr. Barry briefly reviewed the NexGen EBA method for evaluating TB regimens. The NexGen EBA study was launched with the objective to learn about the early effects of different anti-TB drugs on microbiological, radiographic, and immunologic markers in people with TB. This Phase 2A EBA study conducted in Cape Town, South Africa, evaluated eight treatment arms in 160 drug-susceptible, HIV negative adults with smear-positive TB. In addition to the conventional EBA analysis with 14 days of inpatient monitoring and daily overnight sputum collection, the NexGen EBA method involved the collection of a baseline (pre-dose) and 14-day (after last dose) PET-CT scans and whole blood RNA signatures. The discussion focused on the PET-CT analysis to predict treatment response in the NexGen EBA trial. TB lesions generally follow the structure of the segments in the lung because most adult pulmonary TB emerges from bronchogenic spread of the disease and therefore lesions localized by intrinsic structures of the airways. Therefore, lesions were first analysed according to the lung segments in which they occur. It was noted that lesions were analysed individually within a patient because it was observed that individual lesions in most patients had different quantitative or qualitative

changes during treatment, where some lesions progressed and some regressed within the same patient. A labour-intensive computational extraction algorithm was developed to analyse the data. In total, over 1100 cavities and lesions were extracted from 160 patients. Based on the data from the PET-CT scans, similar results were observed when compared to the conventional EBA studies where rifampicin and moxifloxacin monotherapies had the greatest reduction in total lesion volume compared to isoniazid and pyrazinamide monotherapies and the combination therapies performed similarly well compared to the standard of care. In a second analysis that was motivated by the heterogenous responses to TB drugs within lesions, Dr. Barry showed that isoniazid and pyrazinamide combined showed a synergistic effect, but rifampicin and pyrazinamide combined showed a slight antagonistic effect. Interestingly, pyrazinamide response measured by PET-CT scans was shown to be related to baseline inflammatory status where lesions that are strikingly 'hot' at baseline show the strongest response (glycolytic activity or lesion mass). Dr. Barry claimed that this is consistent with Phase 3 data showing that pyrazinamide is only really active early in treatment and not as active if added after the initial intensive phase of treatment.[8]

Discussant 5: Using biomarkers in clinical development: practical issues - Kelly Stinson (Cultura Incorporated)

Dr. Stinson reviewed the practical issues of implementing biomarkers to accelerate clinical development. There has been a recent shift in surrogate markers used for regulatory approval, from sputum culture conversion in solid media in the 1950s, to sputum culture conversion in liquid media for approval of bedaquiline and delamanid as well as a composite outcome that included both clinical and bacteriological responses for the approval of pretomanid (as part of the bedaquiline, pretomanid, linezolid (BPaL) regimen) in 2019. Referring to the regulatory guidance documents for TB drug trials published by the FDA[9] and EMA[10], regulatory agencies are open to the use of new TB biomarkers in regulatory approval requests. To ensure consensus on how to evaluate performance and the role of new TB biomarkers, early and frequent discussions between drug developers and regulatory agencies, as protocols are developed and programs put into place, are recommended. She suggested that Phase 2B/C trials are an ideal setting to evaluate potential biomarkers because they can be designed to allow correlation with early sputum culture endpoints (culture conversion and time to culture conversion) and primary clinical endpoints (non-relapsing cure 6 to 12 months post treatment). Dr. Stinson emphasized that, although imperfect, sputum culture is here to stay for the foreseeable future,

[8] Fox W, Ellard GA, Mitchison DA. Studies on the treatment of tuberculosis undertaken by the British Medical Research Council tuberculosis units, 1946-1986, with relevant subsequent publications. *Int J Tuberc Lung Dis.* 1999 Oct;3(10 Suppl 2):S231-79.

[9] Center for Drug Evaluation and Research Food and Drug Administration. Guidance for Industry Pulmonary Tuberculosis: Developing Drugs for Treatment. Draft Guidance. November 2013. http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htm

[10] European Medicines Agency. Addendum to the guideline on the evaluation of medicinal products indicated for treatment of bacterial infections to address the clinical development of new agents to treat pulmonary disease due to Mycobacterium tuberculosis. 2017. EMA/CHMP/EWP/14377/2008 Rev. 1

and therefore, quality and standardized sputum collection and handling is critical for comparing with new biomarker data.[11]

Discussant 6: PD versus surrogate: exploring future biomarkers - Debra Hanna (BMGF)
With the evolving TB development landscape, Dr. Hanna discussed the future of TB biomarkers and underscored the importance of gathering and sharing data and information on novel biomarkers in the preclinical and clinical stages, emphasizing the reliance on integrated databases as more efficient clinical trials are designed. One of the most critical challenges in identifying simpler, safer, and shorter regimens is the identification of a suitable biomarker that can provide information in the preclinical space for ranking and prioritizing regimens for clinical evaluation and that is portable across clinical phases. Dr. Hanna noted that the currently slow and costly clinical TB regimen development relies on a disconnected paradigm that measures culture conversion outcomes in Phase 2B then noninferiority tests in Phase 3 to evaluate regimens. She welcomes the alternative paradigm, enabled by biomarkers, where a Phase 2B/C study that measures actual Phase 3 clinical endpoints, one year after initiating treatment, in a smaller population is included in the development process. Such a paradigm would much benefit from novel biomarkers becoming available. Thus, in the PAN-TB collaboration, led by the BMGF, to increase the number of patients rapidly initiated on treatment and cured, there is commitment to include several biomarkers (including the Otsuka LAM assay, RS ratio assay and PET-CT scan) in all their studies. She acknowledged that embedding novel biomarkers into clinical trials is associated with many practical challenges but argued that the TB community has not progressed beyond experimental biomarkers because of the lack of investment and data generation that would allow informed decisions about moving promising biomarkers forward in development. In the context of the current TB regimen pipeline with 70+ possible regimen combinations that can be tested in Phase 2B/C designs, the critical importance of developing novel biomarkers is further highlighted. Lastly, Dr. Hanna echoed that novel biomarkers with real-time properties are necessary to expedite and inform innovative clinical trial designs with adaptive protocols.

Round table panellist: Kathy Eisenach (Independent Consultant), Robert Wallis (The Aurum Institute), Elizabeth Talbot (Darthmouth College), Payam Nahid (UCSF)
Dr. Eisenach described the evolution of TB research from the early 1990s when no new drugs were being developed to the early 2000s when WHO held a workshop on trial design for the anticipation of new TB drugs and then up to our current state with several new drugs in a dense TB pipeline. Over the years, there has also been an increase in funding and resources to support TB treatment and research, creating an ideal position to evaluate novel biomarkers. She reemphasized that to optimize success in biomarker research, the TB community must prioritize cooperation, coordination, and collaboration to make informed decisions on which biomarkers to test and the best approach for embedding them into studies. Finally, she encouraged the TB community to invest and consider the development and translation of simple, efficient, and reliable biomarker assays.

[11] See Mycobacteriology Laboratory Manual, Global Laboratory Institute; Mycobacteriology Laboratory Sourcebook For Harmonization and Support of Tuberculosis Clinical Trials (adopted by ACTG and CPTR)

Dr. Wallis expressed his concern for setting goals for TB treatment that are suboptimal. There is increasing evidence of long-term sequels despite successful TB treatment, reflecting persistent lung inflammation and injury. In view of this, new regimens will likely require both antimicrobials and host-directed therapies and clinical trials will require new endpoints and biomarkers that reflect host factors. Dr. Wallis also expressed his concern that PET-CT scans would not be a generalizable marker of lung inflammation and drug specific characteristics associated with PET-CT scans require further research.

Dr. Talbot was pleased to see that as a TB community there is general agreement in the urgent need for reproducible, reliable, and easy-to-interpret biomarkers that are validated and portable across the development pipeline. Yet, she acknowledged that there are still practical considerations that need to be considered when embedding the tools in clinical studies. She encouraged the use of resources to establish implementation strategies in parallel to the on-going research that are focused on the relationships between markers and outcomes. For future direction, she proposed that WHO produce Target Product Profiles for biomarker development to define the ideal surrogate marker, and that a group focused on development of biomarkers integrates the Stop TB Partnership Diagnostic Working Group. She also advocated for a regularly updated biomarker landscape analysis.

Dr. Nahid acknowledged the substantial advancements made over the last decade that are leading up towards achieving surrogacy. He noted that the key to further development of biomarkers will be knowledge integration, similar to what is needed for translational platforms described in Session 1. The results from the TBReFLECT (TB Re-analysis of Fluoroquinolones Executed Clinical Trials) analysis that used the Critical Path to Tuberculosis Drug Regimens data repository is an example of the types of transformative learning that can be gained from knowledge integration. He echoed the critical importance to embed biomarkers in regulated clinical trials. These clinical studies should be epidemiologically sound, with rigorous biomarker assessment in populations that capture the spectrum of TB disease in humans. Collecting PK and adherence information should also be standard in these studies, so that variability can be accounted for when analysing PD and biomarkers. More discussion is needed on the definition of microbiological outcomes, as Phase 3 trials are designed with composite definitions, and a robust microbiologically driven outcome is needed for biomarker assessment. More research is also needed on distinguishing between reinfection from relapse to fully understand biomarker results. Dr. Nahid encouraged the TB community to reflect on ways that biomarker research can be pushed forward and agreed that now is the time to produce a Target Product Profile for biomarker development.

Session 3: From Phase 2 to Phase 3: the role of adaptive and seamless designs to streamline clinical development

Co-Chairs: Michael Hölscher (Ludwig-Maximilians University) and Christian Lienhardt (IRD and University of Montpellier)

Introduction to session 3 - Christian Lienhardt (IRD and University of Montpellier)

Methods for transitioning TB drugs and regimens through Phase 2 to Phase 3 stages have evolved dramatically in the last decade, with the use of adaptive trial designs and the newly proposed Phase 2C trial design, intending to facilitate the transition to the confirmatory Phase 3 trials. In addition, based on new PK-PD modelling and quantitative bacteriology approaches, Phase 2B studies can increasingly provide relevant data on the effect of drug doses and/or plasma concentrations on bacteriological response. Approaches combining preclinical and clinical data using Bayesian methods have the potential to guide early phase clinical development decisions in real time.

Keynote: Can innovations in phase 2 trial design facilitate rapid investigation of novel drug combinations in Phase 3 trial? – Geraint Davies (Liverpool University)

Dr. Davies introduced a new perspective in clinical development: a pathway perspective, where the questions become how much information each stage can provide and how many trials are needed to confidently deploy a new regimen. Each development team is faced with a 'trilemma': dose-finding, combination selection, and duration selection. Information gathered in any pathway should collectively address each pillar in this trilemma for a given regimen, and each phase in the pathway should be designed in perspective of the entire development pathway. Pathways may differ according to objectives and audiences (i.e. licensure with regulatory agencies vs. rapid deployment of regimens in programs) as well as characteristics of the regimen and target population. Dr. Davies reviewed various pathways that may be appropriate depending on the objectives of development teams.[1] In the most conservative and traditional pathway, four trials (Phase 1, 2A, 2B, and 3) are involved where each typically provides one type of information. Innovation in recent years have led to consider pathways including fewer trials (for instance, as few as two trial pathways*)* where information on dose, duration, and combination selection are simultaneously gathered. These shortened and compressed pathways have, however, raised a series of issues that have been discussed extensively in the recent past, including: preference for monotherapy dose-finding, perceived need for reduction of number of combinations to be tested, lack of confidence in selection of intermediate endpoint, impact of delayed endpoint information on adaptation, imprecision of prediction of duration from preclinical to early clinical data, anxiety associated with interpretation of safety events in combination therapy, and the total pathway sample size and duration.

[1] Davies G, Boeree M, Hermann D, Hoelscher M. Accelerating the transition of new tuberculosis drug combinations from Phase II to Phase III trials: New technologies and innovative designs. *PLoS Med.* 2019 Jul 9;16(7):e1002851. doi: 10.1371/journal.pmed.1002851.

According to Dr. Davies, each pillar of the trilemma faces specific challenges that need to be considered carefully as indicated in Table A.1.

Table A.1 Challenges and possibilities in addressing dose-finding, combination selection and duration in the clinical development pathway.

	Challenges	Possibilities
Dose-finding	• Monotherapy vs combination therapy and use of preclinical and Phase 1 predictions to make decisions. • Limited scope for adaptation in Phase 2A studies, particularly with monotherapy. • Decisions are based on least evaluated or trusted biomarker (EBA). • Controversy on the usefulness of safety data from monotherapy studies when it is neither necessary nor sufficient for proof of concept from a regulatory perspective. • Controversy on the usefulness of extensive dose ranging and dose titration studies aimed to provide a full dose response curve for individual agent; information may not be needed to progress in development pathway (i.e. dispensable for regulatory purposes).	• View Phase 2A studies as a continuation of Multiple Ascending Dose (MAD) studies from healthy volunteers to patients. • Use of preclinical and Phase 1 data to define plausible dose range. • Use of adaptive randomization in Phase 2 with priors derived from Phase 1 or translational PK-PD modelling. • Use of real-time biomarkers for efficacy decisions. • Use of joint efficacy and safety criteria for adaptation. • Perform Phase 2A studies without internal control because it is not meaningful for proof of concept, and comparisons should be against zero (i.e. no treatment for 14 days).
Combination selection	• Constrained by preclinical selection and characteristics of component drugs. • Imperfect surrogacy of 8-week and time-to-event endpoints for relapse. • Superior power of intermediate versus definitive endpoints. • Opportunity for adaptation is dependent on recruitment rate and sample size. • 14-day combination studies are typically underpowered for formal comparisons. • Factorial comparisons much less efficient than unstructured comparisons. • In 14+14 design (14 days monotherapy + 14 days combination therapy), prior monotherapy may impact efficiency of subsequent selection of combinations limiting its utility for combination selection based on efficacy.	• Phase 2B is most appropriate format for combination selection and dose-finding. • Extend studies beyond 14 days (possibly even 56 days) can provide a more complete profile of the shape of the dose-response curve (at the expense of increasing variance in response measurements and increasing missing data requiring advanced statistical techniques). • With extended studies, acceptable power with sample sizes of 40 or possibly fewer, which may also be suitable for spaced dose-response studies. • More reliable estimates of elimination of bacilli if lower limit of detection and missing data handled correctly. • Extended studies facilitate use of time-to-event framework for intermediate outcomes
Duration selection	• Single duration Phase 3 trials possibly the riskiest aspect of clinical development. • Not directly informed unless *full intended regimens* are studied, and patients are	• Duration randomization design can inform selection of duration, where multiple durations of each candidate regimen are tested, and selection is based on the definitive endpoint.

	followed for determination of definitive endpoints. • Intermediate endpoints can provide some value, but only definitive endpoint can provide certainty. • Influence of duration on the definite endpoint is strong but there is also an effect related to different classes of drug with different potencies. • Development teams have different perceptions of potency which influence decisions on duration selection. • Predictions of duration from preclinical studies and meta-regression models have been informative but are imperfect with large prediction intervals. • Limited scope for adaptation on intermediate endpoints. • Prognostic stratification would imply a range of durations in unselected study populations.	• Explore trade-offs between adaptation and model-based estimation of duration in Phase 2C studies. • Multi-duration Phase 3 trials should not be discounted because more information can be collected, confidence in results will be greater, and it can potentially identify minimum durations: reinforces the use of Phase 2C to define the limited set of durations to be tested in Phase 3 trials. • Adaptive recycling of assigned duration in trial. • Exploit relationships between intermediate and definitive endpoints using predictive distributions.

Ultimately, it is essential to consider the context of the overall clinical development pathway when identifying novel combination regimens, as a number of trade-offs can be sought between innovation, time, and risk, when taking into consideration the design of each development stage. It is critical that the whole TB community think seriously about how innovations in clinical development can offer real opportunities to save lives.

Discussant 1: How to evaluate in parallel several regimens with various duration, taking into account the severity of disease? - Patrick Phillips (UCSF)

For the current standard of care, there is a clear duration-response relationship, with a monotonic increase that is approximately linear between 4 to 6 months and plateaus with longer durations.[2] This relationship can inform duration-randomization trial designs to be utilized in Phase 2 or 3 studies that investigate multiple durations of the same regimen (say, at least five durations for each regimen). With the expected monotonic increase in response as a function of duration, parametric models can be used to describe the duration-response relationship; for example, a fractional polynomial of 2 degrees is sufficiently flexible to describe the set of plausible shapes for the duration-response relationship.[3,4,5] To illustrate this trial design, Dr. Phillips provided examples where participants are randomized to at least five

[2] Fox W. Whither short-course chemotherapy? *Br J Dis Chest.* 1981 Oct;75(4):331-57. doi: 10.1016/0007-0971(81)90022-x.
[3] Horsburgh CR, Shea KM, Phillips P, Lavalley M. Randomized clinical trials to identify optimal antibiotic treatment duration. *Trials.* 2013 Mar 28;14:88. doi: 10.1186/1745-6215-14-88.
[4] Quartagno M, Walker AS, Carpenter JR, Phillips PP, Parmar MK. Rethinking non-inferiority: a practical trial design for optimising treatment duration. *Clin Trials.* 2018 Oct;15(5):477-488. doi: 10.1177/1740774518778027.
[5] Quartagno M, Carpenter JR, Walker AS, Clements M, Parmar MK. 25 Feb 2020 (pre-print). arXiv:2002.10962 [stat.ME] https://arxiv.org/abs/2002.10962

durations for several regimens, interim analysis are used to compare regimens to the control and to inform the response duration relationship by fractional polynomial modelling, and stopping rules are employed at interim analysis. He explained that duration-randomization trial designs have potential to provide confidence on the shortest non-inferior and superior durations and provide recommendations on durations for different regimens to take forward to Phase 3.

In the context of a Phase 2C platform, Dr. Phillips raised several considerations on the use of duration-randomization trial designs. First, the choice of primary endpoint since failure and relapse (clinical endpoints) and time to culture conversion (intermediate endpoints) are usually collected. Failure and relapse would be more meaningful because they directly capture the differences between durations, but large sample sizes (in the order of phase 3 trials) would be required for adequate power for between-regimen comparisons. Second, the ideal sample size of duration-randomization designs: a sample size of 60-80 participants *per duration* would provide adequate power for duration-response modelling using relapse and failure but would be massively over-powered for time-to-culture conversion (300-400 per regimen). Conversely, if a sample size of 60-80 participants *per regimen* is used, there will be reasonable power to compare time to culture conversion but limited power for duration-response modelling using relapse and failure. The sample size consideration will likely depend on the number of promising regimens that would be available. Third, the duration randomization approach is focused on *estimation* of the duration-response curves and not well suited for *decision making*. The Multiple Comparison Procedure-Modelling (MCP-Mod) statistical approach[6], a hybrid approach that combines hypothesis testing and curve estimation originally for Phase 2 dose-ranging studies with the purpose of finding suitable dose(s) for confirmatory Phase 3 trials, was presented as an option and work is needed to evaluate how it can be applied and adapted for the proposed duration-randomization studies in the context of TB.

To improve the representativeness and efficiency of Phase 2C trials, the possibility of using patient enrichment or patient risk stratification approaches was reviewed. Regarding the former, an option would be to enrich for high-risk participants, increasing relapse rates and therefore increasing power for regimen comparisons (smaller sample sizes) but the major challenge would be in translating the Phase 2C relapse rates to confirmatory Phase 3 studies where eligibility would be much broader. Since readily available markers of disease burden can be used to stratify patients into groups by risk of relapse, a second option would be to have durations, and possibly drugs, modified by patient risk strata, but this would create complex adaptive design stopping decisions that are now dependent on regimen, durations, and risk strata, in addition to reducing power in risk-strata subgroups. The last option would be to ignore risk stratification. However, this may result in a waste of resources since enrolling low risk participants with high probability of cure, irrespective of regimen and duration, would not generate any new information about the optimal duration of regimens. This option may also transfer risk back to Phase 3 trials where chance of noninferiority is dominated by high-risk strata therefore being potentially counter-productive for de-risking Phase 3 trials. In that

[6] Information on MCP-Mod statistical approach available at http://go.usa.gov/x3jPR

situation, small 'Phase 2B' trials could act as a 'gatekeeper' to prevent very poor regimens from entering Phase 2C, collect further information on safety events, and inform the choice of durations to be tested in phase 2C.

Discussant 2: Finding the right duration using multi-arm multi-stage designs –Thomas Jaki (Lancaster University)

In a typical multi-arm multi-stage (MAMS) study design, treatments are compared to a common control by quantifying the difference in a pre-specified test statistic, and interim analyses are included in the trial to determine if each treatment is satisfactory (e.g. non-inferior to control), or requires more data, and if so, the trial continues to collect data and another interim analysis is performed. The MAMS design assumes that arms are independent – this is a major limitation to compare multiple durations of a single (or multiple) regimen(s) because it is expected that longer durations will result in improved outcomes. To address this, Dr. Jaki proposed the use of an *order restricted MAMS design* that accounts for ordering among treatment effects when computing the boundaries for decisions rules at each interim analysis. Based on simulation studies, this type of design has potential to reduce the expected sample size by 25-30% compared to a fixed sample size design (no interim analysis) and 10% compared to a typical MAMS design that treats arm independently (ignores ordering). Accordingly, the benefits of the ordered restricted MAMS design are the utility of a very familiar framework, the potential to use short term endpoints for duration selection, and the opportunity to re-allocate patients if shorter durations are stopped. However, the major drawback of this design is that extrapolation between durations and advanced data sharing/modelling would not be possible. Dr. Jaki concluded that this design cannot be considered as a 'silver bullet' to addressing the duration question, as sample size requirements are larger than for modelling approaches but can be a viable option with a familiar framework that controls traditional statistical properties, such as type I error.

Discussant 3: From phase 2 to phase 3: How to evaluate in parallel several regimens with various durations, taking into account the severity of disease? –Alex Carlton (GSK)

Ms. Carlton presented an example study design for a Phase 2C trial that builds upon the topics discussed by former presenters. Assuming the dose-finding and safety questions are addressed prior to the Phase 2C study, a parallel-arm duration-ranging design was proposed where the only varying component of the design is duration. In this open-label design, multiple durations are tested for each regimen, a standard of care is included to ensure study validity, and randomization into each arm is equal. A sample size of 60-70 participants per arm with drug susceptible TB with or without HIV-coinfection was proposed. Additionally, the study population was proposed to be enriched with high-risk phenotypes, with the rationale that this population are of higher clinical priority and are more difficult to treat and the expectation that regimens that are likely to fail will fail sooner in this population, and regimens that are effective in this population should also be effective in the lower risk populations. Regarding follow-up, participants are to be followed for 12 months post-randomization, regardless of duration, to ensure safety. However, the primary endpoint will be the percentage of patients with unfavourable outcomes 6 months post treatment, which would provide a direct link with unfavourable outcomes evaluated in Phase 3 trials but at an earlier timepoint. The 6-month

timepoint was chosen because a majority of relapse occur within this time[7] and the point from post-treatment rather than from randomization was chosen to avoid bias against shorter regimens. It was recommended to have staggered interim analysis with pre-specified stopping rules to drop arms that are non-efficacious and ensure patient safety. This design is anticipated to provide the data necessary to describe and model the duration-response relationship.

Round table panellist - Martin Boeree (Radbound University), Daniel Everitt (TB Alliance), and Robin Mogg (Gates Medical Research Institute)

Dr. Boeree noted that the duration randomization design is very promising for duration selection, but risk stratification in these designs need careful and critical consideration.

Many of the questions raised in this session are exactly what TB Alliance has been confronted with as they plan how to efficiently develop new combination treatments for TB. Dr. Everitt argued that monotherapy EBA studies are important because it is the one time to show proof of concept and an independent effect at the dose that is taken forward. He explained that both US and European regulators find this information very important, particularly for a new drug in development. However, EBA studies should not be the only basis for dose finding since it is only 14 days and the performance of the full regimen is ultimately the most important information to be gathered. The 14+14 design may be attractive for economizing patients, but one caveat is that as a drug is combined into regimens there can be difficulty in discriminating the effect of different doses because monotherapy may have already reduced bacterial load substantially. While some designs in the Phase 2B/2C stages may allow for dose-finding, the duration-randomization designs should presuppose that the right dose has been selected based on prior studies.

Dr. Mogg expressed that middle stage development is still caught between limited power to make decisions based on definitive clinical endpoints and limited confidence to make decisions based on intermediate endpoints. She outlined two goals that the BMGF is focused on when considering Phase 2B and Phase 2C clinical development programs. The first goal is to de-risk regimens that have high probability of success to show noninferiority in Phase 3 trials. The best way to make decisions for de-risking is to use definitive clinical endpoints at an earlier stage of development since there is a lack of confidence in intermediate endpoints. However, since definite endpoints is difficult to use in middle development, the second goal is to proactively find ways to identify and clinically validate early biomarkers and show their association with definitive clinical endpoints. Dr. Mogg reiterated the promising opportunities of using established statistical methodologies, like MCP-Mod and Bayesian statistical frameworks, to de-risk regimens, and more research should explore the utility of these approaches in the context of optimal duration finding in TB. Because there is less assurance of favourable outcomes as shorter regimens or enrichment strategies with high-risk patients are tested, she also suggested j at interim analyses in these situations to limit the number of patients being exposed to ineffective therapies

[7] Nunn AJ, Phillips PP, Mitchison DA. Timing of relapse in short-course chemotherapy trials for tuberculosis. *Int J Tuberc Lung Dis.* 2010 Feb;14(2):241-2.

Session 4: New phase 3 trials and how they will facilitate ultimate regimen development

Co-chairs: Carole Mitnick (Harvard University) and Christian Lienhardt (IRD and University of Montpellier)

Introduction to session 4- Session co-chairs: Carole Mitnick (Harvard University) and Christian Lienhardt (IRD and University of Montpellier)

The currently expanding pipeline of new drugs, new diagnostics, and new methods make real transformation in TB treatment possible. The landscape of TB treatment has evolved considerably over the last 10 years, necessitating careful consideration of various trial aspects and characteristics to ensure that Phase 3 trials deliver high-quality evidence that can inform treatment practice. These include: trial design (superiority or non-inferiority); use of adaptive designs; treatment stratification, choice of control (selection and use of standard of care SOC), trial conduct (study quality, treatment adherence, missing data); data analysis (intention-to-treat; per-protocol analyses; estimation of treatment effect). Further, as more innovative approaches are introduced, interpreting and communicating the results from trials may be more challenging, and standardization of some aspects of the trial (e.g. use of the same definitions of exposure and outcomes) may be needed to allow aggregating and pooling of results and individual patient data to inform guidance. Subsequently, this session focused on new Phase 3 trial designs and how they will facilitate ultimate regimen development.

Innovative approaches in Phase 3 trial design:

Keynote 1: Setting the stage –Andrew Nunn (University College London)

Compared to how drug development was conducted 50 years ago (in the very first African short course trial led by the British Medical Research Council unit), drug development today requires much longer timelines, and Dr. Nunn discussed several ways to accelerate Phase 3 trials. First, platform trials have potential to improve efficiency with smaller sample sizes, reduced requirements for approvals by using a common protocol, and reduced time and cost. These trials provide an opportunity to compare multiple regimens, multiple arms, and possibly even multiple doses (although, some argue that the dose questions should be dealt with in earlier phases) to a common control and use adaptive designs to drop or add new arms as the trial progresses. However, identifying real-time biomarkers that could accurately predict long term outcomes of interest would maximize the benefit afforded by these platform trials, in addition to accelerating Phase 3 trials overall. An early look at data from current trials may help accelerate Phase 3 trials, but this early look approach could result in problems arising from biased management decisions during treatment and biased assessment of potential relapses. As an example, in the STREAM (Standard Treatment Regimen of Anti-tuberculosis Drugs for Patients with MDR-TB) Stage 1 trial (ISRCTN78372190, NCT02409290), the data monitoring committee recommended an early look at the trial data partly due to occurrence of adverse events attributed to the high-dose fluoroquinolone. As a consequence, the results were made public approximately 12 months earlier than anticipated, and with incomplete follow-up for some participants. Because non-inferiority was not demonstrated at this point, many people wrongly concluded that the intervention regimen was inferior, although it proved noninferior to

the standard of care after follow-up was completed. Biases created by early looks at the data could be mitigated by making data only available to a selected group. In addition, it is important to consider the timing of early looks at data. There may be a case for sharing the results of interim analysis with other trial groups, particularly when results are unexpected.

One of the main challenges in drug-resistant trials has been the changes in recommendations of standard of care during the trials. For example, as widespread adoption of a "new" standard of care occurs, the number of participants that are in the "old" standard of care comparator arm will be limited and this will compromise any comparisons of intervention arms with the "old" standard of care comparator arm. Dr. Nunn asked what should happen when a successful new regimen is identified and established as the "new" standard of care. In the next trial should the "old" standard of care be retained, or would a comparison with the old standard of care by means of a network meta-analysis be sufficient?

Dr. Nunn also discussed the interpretation of non-inferiority trials and suggested avoiding undue emphasis on the binary question of "Was noninferiority achieved or not?". A Bayesian analysis has been proposed to calculate the posterior probability that one regimen is worse than another by a range of percentage differences using different prior probabilities; this Bayesian analysis was used in the STREAM publication.[1]

Keynote 2: What are the most promising designs to accelerate regimen development? –Patrick Phillips (UCSF)

Dr. Phillips provided the "Nuts and Bolts" that need to be considered as Phase 3 trials are designed. At baseline, he noted that randomized controlled trials are necessary, but not sufficient, to provide high-quality evidence; if trial quality is compromised, the evidence is downgraded. Any acceleration of Phase 3 trials has to be done with caution: fast evidence is rarely high-quality evidence. He reiterated that not all parts of the drug development trilemma need to be addressed in phase 2 studies; instead, it may very well be appropriate to take forward multiple durations and multiple regimens to Phase 3, as done with multi-arm Phase 3 TB trials (e.g. STAND NCT02342886; TB-PRACTECAL, NCT02589782; endTB, NCT02754765; STREAM, ISRCTN78372190, NCT02409290, ISRCTN18148631; TBTC S31/A5349, NCT02410772).

Dr. Phillips returned to the concept of 'platform trial' using the example of the STAMPEDE (Systemic Therapy in Advancing or Metastatic Prostate Cancer: Evaluation of Drug Efficacy; NCT00268476) trial that has been ongoing since 2005. The trial started with five interventions and a control and, over the years, various arms have been stopped, while others have been added. This trial has proved to be very efficient for evaluating many different interventions, and the major advantages include multiple interventions compared against a common control, single 'master protocol', adaptive designs that provide opportunities for several agents to be tested, and improved efficiencies in recruitment, staffing, as well as regulatory and Institutional

[1] Nunn AJ, et. al. A Trial of a Shorter Regimen for Rifampin-Resistant Tuberculosis. *N Engl J Med.* 2019 Mar 28;380(13):1201-1213. doi: 10.1056/NEJMoa1811867. Epub 2019 Mar 13.

Review Board approvals. In order to fully capitalize on platform trials designs for TB regimen development, the need for better biomarkers of treatment response was reemphasized.

He then made the case for standardised methods and endpoint definitions. He mentioned an on-going systemic review of protocols and statistical analysis plans from recent DS and DR-TB Phase 3 trials that showcased the varying range of outcome definitions and analysis methods - making it difficult to clearly interpret results across the trials.[2] Dr. Phillips presented a new framework for estimands and sensitivity analysis in clinical trials that was proposed in a recent addendum (Feb. 2020) to the ICH E9 Harmonized Tripartite Guideline on Statistical Principles for Clinical Trials.[3] Briefly, *the estimand* is determined to provide a specification of the target of estimation (i.e. what is it that we want to estimate in the clinical trial?), *the estimator* is determined to provide the method of estimation (i.e. how do we go about estimating the estimand?), and lastly, *the estimate* is determined to provide information of the treatment effect. With this new framework, there can also be sensitivity estimators for a single estimand, where multiple estimators can be used to provide estimates. There is also a possibility to have multiple estimands, but they should always be prespecified in the protocol prior to the start of the study. The ICH E9 documents provide five attributes for how to define the estimand, providing a basis on how to define what needs to be estimated to address a specific scientific question of interest. The five attributes are:

- The treatment: intervention, and comparator
- The population of patients targeted by the specific question
- The endpoint to address the scientific question
- The specification of how to account for intercurrent events (events that prevent or affect measurement of the primary outcome, e.g. non-TB death, treatment changes, treatment withdrawals, etc.)
- The population-level summary for the endpoint which provides a basis for a comparison between treatment conditions.

This framework provides a standardized language to help articulate the treatment effects that are to be measured, address some of the issues with different analysis populations (e.g. intent-to-treat vs modified-intent-to-treat vs per-protocol), provide clear interpretation for different stakeholders with different perspectives (different estimands, and potentially different estimates, for different purposes), permit transparent definitions with feedback from the TB community prior to analysis and presentation of results, and facilitate cross trial analysis. Ideally, the goal is to move towards standardization in outcome and estimands definition, but at the very least this framework provides an opportunity for clear specification of the question of interest, the analysis, and the interpretation.

---

[2] NK Hills, J Lyimo, P Nahid, R Savic, C Lienhardt, PPJ Phillips. A systematic review of endpoint definitions in late phase tuberculosis therapeutic trials, 29 April 2021, PREPRINT (Version 1) available at Research Square [https://doi.org/10.21203/rs.3.rs-397643/v1]

[3] EMA. ICH E9 (R1) addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials. Feb. 2020. https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e9-r1-addendum-estimands-sensitivity-analysis-clinical-trials-guideline-statistical-principles_en.pdf

Noninferiority Phase 3 trials are recognized as a pathway to regulatory licensing, which have been successful. However, noninferiority trials have a number of limitations: they do not address inherent benefit(s) of intervention, the abstract notion of the 'margin of non-inferiority' is difficult to interpret, the chance of showing noninferiority is highly dependent on the control arm event rate and arbitrary margin, and noninferiority trials are easier to manipulate than superiority trials through trial conduct and choice of analyses. Dr. Phillips explained that superior interventions are needed to bring the TB epidemic under control, and it is time to 'think beyond noninferiority': although noninferiority trials should not be dispensed with altogether, attendees were encouraged to consider and explore alternatives; for example, considering superiority in patient-relevant outcomes, like cost-effectiveness, or in a composite efficacy-safety-duration outcome after showing noninferiority for regulatory licensing. Additionally, patient-centered (e.g. the SPECTRA-TB design described by Dr. Nahid below) and pragmatic trials (e.g. BEAT TB trial, NCT04062201), perhaps with adaptive designs and structured decision making, may be other options that go beyond noninferiority.

Lastly, the risk of 'biocreep' in informing non-inferiority trial designs was discussed. In the context of DR TB, there is limited evidence from randomized trials to support the treatment effect for the Standard of Care which means that the NI margin is not firmly established. On-going trials are anticipated to set the benchmark for standard regimens that will become the control in future randomized trials. In contrast, in the context of DS TB, there is a real risk of biocreep, which will be critical to consider when designing phase 3 trials. With a short regimen that has reached noninferiority, Dr Phillips re-invoked the question raised by Dr Nunn 'What should the control be in the next trial?'. If the new noninferior regimen is the control, there is high risk of biocreep, but if the older well-established control is used, there would be no direct comparison with the new noninferior regimens. On the contrary, if both the noninferior and well-established regimens are used as controls, the trial would be larger but biocreep can be avoided and between-regimen comparisons can be made.

The place of adaptive design in current Phase 3 trials: design protocols and interim lessons learnt

Discussant 1: *The end-TB trial- Carole Mitnick (Harvard University)*

Improvement in scientific and economic efficiency and ethics are the main perceived advantages of using adaptive design clinical trials, which have been more commonly used in the last 30 years with regulatory guidance now available for their use. The end-TB trial (NCT02754765) is an on-going Bayesian response-adaptive, randomized, controlled, open label, clinical trial for fluoroquinolone-susceptible, rifampicin-resistant TB. The trial compares to the current WHO standard of care, five 9-month, all oral regimens, each containing one of the new TB agents, bedaquiline or delamanid, and at least one of the repurposed agents, linezolid and/or clofazimine. The overall sample size is 750 participants with a non-inferiority margin of 12%, 80% power and one-sided alpha of 0.025. The primary endpoint is at 73 weeks, with interim endpoints at 8 and 39 weeks and a final endpoint at 104 weeks post randomization. Dr. Mitnick provided a high-level overview of the Bayesian response-adaptive randomization design used in the endTB trial. The perceived potential drawbacks of Bayesian adaptive-randomization designs include sample size required, operational complexities, and poor performance characteristics. However, it was shown that in sample size simulations the Bayesian adaptive-

randomization design was 10-25% more efficient in overall sample size compared to the balanced (fixed) randomization design, and the efficiency largely relies in the ability to have early results, hence at 8 or 39 weeks. Additionally, the Bayesian adaptive-randomization design allocated greater proportions of patients to noninferior arms than balanced randomization design. With regard to operational complexities and performance characteristics (e.g. limited data to define endpoint relationships *a priori* and delay in information for adaptation, missing or contaminated culture results), the relationships among early and later endpoints, although initially based on limited data, are updated according to interim analysis and power has proved to be robust to delays in culture reporting. Another potential challenge with the Bayesian adaptive-randomization design is the risk of inflating type I errors when there are changes in population or treatment response over time (e.g. including mostly salvage patients at first then with more confidence in the regimen including a broader range of patients). To manage this risk, Dr. Mitnick suggested to not invoke stopping rules or have very strict stopping rules, to perform subgroup or stratified analysis, to assess the time trend, and/or to bootstrap to control type I errors. Overall, the Bayesian response-adaptive randomization design can offer scientific, economic, and ethical advantages and in the context of DR TB trials, the operational challenges are minor and performance characteristics are manageable but require careful attention to trial integrity. Careful consideration is still needed to determine if and how the data and information from these trials can be aggregated for additional analysis with other individual patient data.

Discussant 2: DRAMATIC trial- Robert Horsburgh (Boston University)

Dr. Horsburgh introduced the DRAMATIC (Duration Randomized Anti-MDR-TB and Tailored Intervention Clinical) trial (DMID protocol 20-0022) that uses a duration randomization design with the objective to describe the duration-response relationship of an experimental regimen in MDR TB patients. This trial aims to describe the relationship between baseline prognostic risk strata and sustained cure, evaluate the association between RS ratio and sustained cure, and identify the shortest duration of the study regimen with acceptable safety and efficacy for a Phase 3 clinical trial. Patients (N=220) with pulmonary fluoroquinolone-susceptible rifampicin-resistant TB will be randomized in Vietnam and the Philippines into one of four arms with the same regimen combination (delamanid, levofloxacin, clofazimine, bedaquiline, and linezolid) but different durations: 16, 24, 32, or 40 weeks. On-going MDR TB trials test 9- and 6-month treatment regimens, but the DRAMATIC trial will help describe treatment outcomes for regimens as short as 16 weeks, and hopefully provide information on the duration when response plateaus. Because the objective of the study is to describe the duration-response curve, the study will have fixed randomization into intervention arms with various durations and will not include a control arm. The participants' disease will be categorized into easy-, moderate- and hard-to-treat disease groups at baseline to assess if the shape of the duration-response curves depends on disease phenotypes. There will be intensive monitoring of safety in the first 36 patients because of the concern with QT prolongation of three drugs in the regimen. It is expected that testing the four durations will provide the full duration-response curve, but if the durations tested are still in the plateaued region of the curve, the trial would still provide valuable information, although not on the shortest duration of the regimen.

Discussant 3: Adaptive trial designs used for COVID 19- Dr. Michael Hughes (Harvard University)
With several international trials for vaccines, preventive interventions, and treatments for COVID-19, the common feature of these trials is the use of platform (master protocol) designs to facilitate rapid evaluation of multiple different interventions. The Accelerating COVID-19 Therapeutic Interventions and Vaccines (ACTIV) initiative in the US is based on the collaboration of the NIH, pharmaceutical companies, and the FDA.[4] Within the ACTIV initiative is ACTIV-2 platform trial (NCT04518410) for outpatient treatment of COVID-19 that is designed and led by investigators in ACTG, including Dr. Hughes. The key objective of the ACTIV-2 trial is to rapidly and efficiently screen (Phase 2) and evaluate (Phase 3) multiple potential therapeutics for reducing hospitalization or death among adults with symptomatic confirmed SARS-CoV-2 infection, with an aim to evaluate about 10 agents with different modalities within 18 months. The key features in the trial include:

- a randomized platform allowing agents to be added and evaluated within the same trial infrastructure,
- potentially seamless transition from Phase 2 into Phase 3 evaluation if 'graduation' criteria met,
- ability of agents to enter at different times and patients randomized to agents available at a given time or to control arm,
- a trial design focus initially on comparisons of each agent to placebo control, not to compare agents,
- in Phase 3, reasonably aggressive futility boundary to discard agents which are unlikely to show effect on hospitalization or death,
- increased speed in evaluating agents needs balancing against increased complexity in trial conduct.

Dr. Hughes reviewed the complexities in the design of the ACTIV-2 platform trial. First, the identification of the appropriate control group, because agents have different modalities of administration (e.g. infusions, injections, oral, nasal sprays), some participants may not be eligible for all agents, and not all sites are enrolling participants to all agents. The solution was to have a partially blinded approach, but the implementation would be complex with a caveat that when combining placebo controls there would be a mix of different modalities of administration. The use of historic control information was also considered but will highly depend on the stability of participant characteristics entering the trials over time. A second complexity is the control of type I errors in Phase 3 trials. The major intent of the trial is to focus initially on comparisons of each agent to placebo control, to identify a group of agents that have therapeutic benefits in preventing hospitalization or death - and not to compare agents head-to-head. Consequently, type I error is controlled separately for each agent vs control comparison and not controlled across multiple comparisons of agents to control (i.e. family-wise or experimental-wise type I error rate not controlled). This is viewed as a potential source of controversy but has been accepted by the FDA. The third complexity is the inclusion of a 'seamless' Phase 2 to Phase 3 transition in the trial that depends on markers with huge

[4] See https://www.nih.gov/research-training/medical-research-initiatives/activ

uncertainty on their ability to predict the effect of an agent on hospitalization or death. Because little is known about risk factors of hospitalization or death amongst patients, the trial uses an array of markers (e.g. virologic, symptomatic, or oxygen saturation) from biological and clinical samples (e.g. pharyngeal nasal swaps, anterior nasal swabs, blood samples, saliva samples, etc.) to evaluate graduation criteria to move an agent from Phase 2 to Phase 3. Still, graduation is also allowed for agents that show some effect on hospitalization or death in Phase 2 without an effect on other markers of interest, and reasonably aggressive futility boundaries are in place to discard agents unlikely to show an effect. This leads to complex decisions for graduation of agents from Phase 2 to Phase 3 but provides opportunities to identify strong predictors. There were concerns with multiplicity and inflation of type I error rates with the use of several markers, but simulation studies suggested they are protected when looking for a signal with a sizeable effect for any Phase 2 outcome. Finally, to mitigate concerns of a seamless Phase 2 to Phase 3 design with the potential for wasted enrollment, an interim analysis was introduced with the idea of pausing enrollment at the end of Phase 2, and focusing on enrollment into other agents, unless the graduation criteria is met at the interim analysis.

Dr. Hughes further discussed the anticipated evolution of the graduation criteria as information on the link between markers and clinical outcomes of interest is gathered. In the ACTIV-2 trial, there is an element of biomarker or clinical marker discovery that will need to feedback into the choice of Phase 2 outcomes for graduation evaluation. In the shorter term, the idea is to have the graduation criteria evolve and adapt to focus on the stronger predictors (or composite predictors) of hospitalization or death. The collection of many different types of samples in Phase 2 will allow research to identify the particular sample(s) that are more predictive than others, which may eventually lead to trial simplification. In the longer term, with the evaluation of a large number of agents, the goal may shift to focus on predictors (or composite predictors) for which differences between randomized arms best predict differences in proportion of hospitalizations or deaths. This would involve a meta-analysis concept for the evaluation of surrogate endpoints. Overall, based on the experiences with the ACTIV-2 trial, Dr. Hughes agreed that increased speed in evaluating agents requires a balance against increased complexity in trial conduct.

Discussant 4: Towards more personalized medicine: using treatment stratification strategies to enhance cure –Payam Nahid (UCSF)

As of today, the WHO endorses the most effective treatment programmes that: i) are linked to early detection and accurate diagnosis and staging, ii) adhere to evidence-based standard of care, and iii) are provided in an equitable and sustainable way. To provide perspective, Dr. Nahid described the oncology program view of cancer treatment worldwide which uses a stratified medicine approach that is endorsed by the WHO for ensuring quality treatment for cancer. In contrast, the current standard TB treatments aim to treat all patients of all phenotypes and all severities with a *one-size-fits-all* regimen, even though patients are clearly ranging from, say, mild modular infiltrates in the right middle lobe through extensive bilateral disease with a lot of parenteral destruction. The success rates have consistently been shown to be inadequate for a curable disease using *one-size-fits-all* regimens that are considered the

standard of care for TB treatment, and stratified medicine principles have potential to change the narrative around how people with TB are treated and how trials for TB are conducted.

Dr. Nahid discussed the TBReFLECT (TB Re-analysis of Fluoroquinolone Executed Clinical Trials) program (sponsored by WHO, Critical Path to TB Drug Regimens Program, and BMGF) that led to a patient level pooled analysis of four phase 3 trials for drug susceptible TB with nearly 4000 participants and six regimens.[5] The key findings from this analysis were that subgroups defined by minimal disease severity (evidenced by non-cavitary disease or low smear grade) when treated with 4-month fluoroquinolone substitution regimens had non-inferior outcomes compared to the 6-month standard control regimen. On contrary, *harder to treat* subgroups defined by high smear grade, cavitary disease, or low BMI indicating malnourishment, were not non-inferior and favoured the 6-month standard control regimen. More sophisticated analysis and modelling has since been performed to develop data-driven risk stratification algorithms and clinical simulation tools.[6] It was shown that an integrated suite of baseline and on-treatment markers predict TB-related outcomes and can be used to identify risk strata with more precision. Based on a target cure rate of 93%, a low, moderate, and high-risk group was identified with a breakdown of approximately 25%, 50%, and 25% of the pooled population, respectively. It would be anticipated that the low and moderate risk groups can be treated with shorter regimens but when interventions and designs are focused on curing all patients with a one size fits all regimen, we are really only targeting the high-risk group comprising of approximately 25% of the population who drive the unfavourable outcomes. In the context of MDR-TB, there are fewer phase 3 trials compared to drug-susceptible TB to inform stratified medicine approaches. However, individual patient data from a cohort of over 12000 MDR-TB patients[7], significant risk factors that identified MDR-TB patient subgroups that may be eligible for shorter treatment durations were consistent with those identified in the TBReFLECT analysis for drug susceptible TB. Furthermore, 78% of the population categorized into the low or moderate risk group while 22% into the high risk group, again, consistent with the drug susceptible TB population.[8] This supports the idea on stratified medicine for TB care, which is a paradigm shift in overall objectives in TB care and is a patient-centred approach that enhances cure rate for the most severe TB cases, while reducing duration, toxicity, and cost to programs and patients for the less severe TB cases.

[5] Imperial MZ, *et al*. A patient-level pooled analysis of treatment-shortening regimens for drug-susceptible pulmonary tuberculosis. *Nat Med.* 2018 Nov;24(11):1708-1715. doi: 10.1038/s41591-018-0224-2.
[6] Imperial MZ, Phillips PPJ, Nahid P, Savic RM, Predicting optimal treatment durations for tuberculosis patients: risk stratification algorithm and simulation tool [abstract]. In: TBScience Conference; 2020 Oct 20-21; Virtual Conference.
[7] Collaborative Group for the Meta-Analysis of Individual Patient Data in MDR-TB treatment–2017, *et al*. Treatment correlates of successful outcomes in pulmonary multidrug-resistant tuberculosis: an individual patient data meta-analysis. Lancet. 2018 Sep 8;392(10150):821-834. doi: 10.1016/S0140-6736(18)31644-1.
[8] Garcia-Cremades, et al. Risk Stratification for Identifying Optimal Treatment Duration in All MDR-TB Patients [abstract]. In: Conference on Retroviruses and Opportunistic Infections; 2021 Mar 6-10; Virtual Conference.

These results and further epidemiological modelling[9] have led to the design of the SPECTRA-TB (Stratified Patient-Centered Treatment Regimens for Active TB) trial, a randomized, open-label controlled phase 3 superiority strategy trial, currently in development by the TBTC. Based on baseline risk factors, drug susceptible TB patients are randomized to either the 6-month standard regimen control-arm or to the SPECTRA-TB strategy that allocates patients into low, moderate, or high-risk groups with durations adjusted accordingly. This trial compares the superiority of the strategy over the one-size-fits-all fixed duration for all patients. Dr. Nahid concluded that a robust evidence base exists supporting the evaluation of stratified medicine approaches to treating people with TB and that continuing development of one-size-fits-all regimens carries the risk of failing to meet non-inferiority, driven by participants with the highest severity of disease.

Round table panellist- *Andrew Vernon (CDC), Eugene Sun (TB Alliance), and Angela Crook (University College London)*

Dr. Vernon raised several points that illustrated his concerns as a trialist. First, with respect to the rapid development pathway that the British Medical Research Council unit led 50 years ago, regular review of the evolving data was critical to helping move new trials forward, and new trials started before on-going trials were completed based on those early looks. However, the early looks raised several problems, particularly biased investigator management of patients, so more discussion is needed on how this type of bias can be managed when there are early looks at the data to accelerate Phase 3 trials. Second, he explained the critical importance of improving communication and interpretation of clinical trial design and results. He favoured the use of multiple estimands and sensitivity analysis in clinical trials, especially to mitigate the risk of misrepresentation and to provide a more complete understanding of the trial results, and advocated to build better understanding around the use of innovative approaches (e.g. Bayesian response-adaptive randomization design), especially for non-statisticians and non-experts. Lastly, he pointed out the large variability in response observed in Phase 3 trials. Some has already been associated with identified risk factors but there is still high variability that, in part, may be due to the high variability observed in exposure between patients and populations. Yet, there is no way to rapidly consider this source of variability in clinical trials, so more research is needed in this area.

Dr. Vernon pointed out that the large-scale collaborations in drug development for COVID-19 that Dr. Hughes described is also absolutely critical to accelerate drug development for TB. A prime example is the TBTC S31/A5349 trial that was conducted three times faster because of the powerful, effective collaboration between TBTC and ACTG networks. Acknowledging that each network has their own goals, Dr. Vernon encouraged networks to simultaneously participate in large-scale collaborative trials, but still pursue specific trials of interest to the network.

[9] Kendall EA, *et al*. Priority-Setting for Novel Drug Regimens to Treat Tuberculosis: An Epidemiologic Model. *PLoS Med.* 2017 Jan 3;14(1):e1002202. doi: 10.1371/journal.pmed.1002202.

Dr. Sun interprets a Phase 3 trial as a study that provides the main, although not only, basis for regulatory authority approval by demonstrating robust safety and efficacy of a drug or regimen. Specifically, he asserted that Phase 3 trials should guide product labelling, addressing the questions 'In whom the drug or regimen should be used?' and 'How it should be used?'. Although it is a component of an overall development plan, it is ultimately a major component of a submission package to regulatory authorities and should be designed in this context. Therefore, he maintained that *the bulk of the exploration on the drug development trilemma should be in Phase 2 and not in Phase 3 studies*. Because Phase 3 studies always represent an increase in variability and heterogeneity and TB is already burden by multiple complexities that other diseases do not encounter, he claimed it is best to avoid introducing additional complexities to Phase 3 trials. Overall, he urged staying in the Phase 2 space, as long as necessary, to sufficiently minimize the risk of failed Phase 3 studies. Although the technical consultation's prominent theme has been on the application of novel and innovative technologies for optimizing efficacy outcomes, he encouraged the TB community to not neglect the safety component of regimens, which are just as important for regulatory approvals. Notably, he explained that the sample size of Phase 3 trials, may be influenced by the requirements of the safety database, and not only the efficacy database.

As a trialist, Dr. Crook was pleased to see that trial conduct was considered in the presentations as implementation is the most difficult part of any new platform. With the burst of COVID-19 treatment and vaccine trials in the last 6 to 9 months, she encouraged the group to consider how to potentially leverage and apply the learnings gained from the rapid implementation of clinical trials for COVID-19 to TB regimen development. Dr. Crook also agreed that it is time to look beyond standard noninferiority designs and consider trials with new patient-centered outcomes.

Session 5: Real-world evidence and cohort data; special populations

Co-Chairs: Payam Nahid (UCSF) and Christian Lienhardt (IRD and University of Montpellier)

Introduction to session 5 –Christian Lienhardt (IRD and University of Montpellier)

Once new treatment regimens have shown their efficacy in randomized controlled trials, factors of treatment uptake and delivery need to be assessed to evaluate their implementation in programmatic and clinical conditions, often referred to as "real-world evidence". In view of the increase in new chemical entities to be tested, and the enhanced use of novel trial designs in TB therapeutics, it is likely that a range of new regimens become available for TB treatment in the near future. To prepare for optimal development of public health policy guidance, early interactions are needed between product developers, researchers, regulators and policy makers.[1] Late-phase clinical trial outputs that serve the objective of registration of new TB drug or regimen can meet the needs of public health guidelines, provided that data on long-term, patient-relevant, and population-relevant outcomes are being collected. Public health factors such as feasibility, acceptability, resource use, equity, and quality of life should be part of data collection, as these are necessary when formulating public health recommendations. This session focused on read world evidence and cohort data and on special populations.

Keynote: Non-randomized data in policy development: what are their needs and roles in the production of evidence? –Payam Nahid (UCSF)

Dr. Nahid conveyed that there is a need for observational data and data collected under programmatic conditions, for the development of public health recommendations, in addition to Phase 3 data. However, only two randomized Phase 3 trials (the Otsuka 242-09-213 trial, NCT01424670; and the STREAM Stage 1 trial, ISRCTN78372190 and NCT02409290), and a third non-randomized Phase 3 trial (the Nix-TB trial, NCT02333799), have contributed to the development of recent treatment guidelines on drug resistant TB. Thus, the latest 'WHO Consolidated Guidelines on Drug-Resistant Tuberculosis Treatment'[2], 'WHO Rapid Communications Updates'[3]; and the 'American Thoracic Society, US Center for Disease Control and Prevention, European Respiratory Society, and Infectious Disease Society of America Clinical Practice Guideline'[4] were all predominately, though not exclusively, structured and based on observational data (e.g. individual patient data meta-analysis from programmatic data[5]), as mainly observational data were available. Relative to clinical trial data, observational

---

[1] Lienhardt C, Vernon AA, Cavaleri M, Nambiar S, Nahid P. Development of new TB regimens: Harmonizing trial design, product registration requirements, and public health guidance. *PLoS Med.* 2019 Sep 6;16(9):e1002915. doi: 10.1371/journal.pmed.1002915.

[2] WHO consolidated guidelines on drug-resistant tuberculosis treatment. 2019 update. ISBN: ISBN 978-92-4-155052-9. https://www.who.int/tb/publications/2019/consolidated-guidelines-drug-resistant-TB-treatment/en/.

[3] WHO Rapid Communication: Key changes to the treatment of drug-resistant tuberculosis. 2019. WHO/CDS/TB/2019.26. https://www.who.int/tb/publications/2019/rapid_communications_MDR/en/.

[4] Nahid P, *et al*. An Official ATS/CDC/ERS/IDSA Clinical Practice Guideline. *Am J Respir Crit Care Med.* 2019 Nov 15;200(10):e93-e142. doi: 10.1164/rccm.201909-1874ST.

[5] Collaborative Group for the Meta-Analysis of Individual Patient Data in MDR-TB treatment–2017, *et. al*. Treatment correlates of successful outcomes in pulmonary multidrug-resistant tuberculosis: an individual patient data meta-analysis. Lancet. 2018 Sep 8;392(10150):821-834. doi: 10.1016/S0140-6736(18)31644-1.

data, although large (e.g. > 12,000 patients in MDR TB database[5]), are much more complex and difficult to analyse and interpret, requiring sophisticated approaches to attempt to control for inherent biases. Further, with several uncertainties associated with observational data, the quality of evidence is considered low, whereas data from randomized trials are considered high quality evidence.[6] Recommendations based on very low quality and certainty in evidence, can often only be conditional, which may create confusion for programs. In a number of recent occasions, WHO had to make strong recommendations based on very low quality and certainty in evidence which created a disconnect between the need for guidance and the quality of available data, sparking much debates and requiring involvement of many stakeholders for policy making. As a result, it is not uncommon to have policies that vary between regions. It must also be recognized that policy decision makers face considerable pressure to update their guidelines, as shown with the push for updates on the use of bedaquiline in August 2019, leading WHO to call for individual patient data on the treatment of DR TB.[7] Why weren't these questions addressed in clinical trials? In fact, clinical trials were ongoing that included collection of such data, and this plea for international data shows that policy decision makers face pressure to update guidelines while trials are still ongoing. For example, while STREAM Stage 2 trial (NCT02409290) was on-going, WHO approved the use of new all-oral shorter rifampicin-resistant TB treatment regimens by National TB Programmes using data collected under operational research conditions. In line with this, an operational research package has been recently released to provide a standardised methodology for conducting such operational research so that the data generated are harmonised across different implementation settings.[8]

Summarized below are the opportunities and challenges associated with the use of observational data.

Opportunities	**Challenges**
Data on the "real world" use of the drug/intervention in TB programmes.	Data controlled by parties who developed the registry, whether it be public (local, state, federal government, academic institutions), or private (manufacturer, academic, etc.) institutions. Non-standardized reporting of adverse outcomes, without clear attribution, leads to lingering uncertainty about safety
Data on the safety of the drug/intervention (e.g., long-term safety, newly emergent events, and rare adverse events).	
Reflect the patient population treated in a practice setting, beyond the confines of the inclusion and exclusion criteria utilized by RCTs.	Patients are not prospectively randomized to an intervention. Multivariate analyses can help adjust for variables but cannot exclude unknown factors that would be addressed by a larger randomized analysis.

[6] WHO Handbook for guideline development, 2nd Ed. https://apps.who.int/iris/handle/10665/145714

[7] WHO Public call for individual patient data on treatment of drug-resistant tuberculosis on use of bedaquiline for longer than 6 months as part of longer treatment regimens; use of all oral bedaquiline-containing regimens of 9-12 month duration; concurrent use of bedaquiline and delamanid; and use of bedaquiline-containing regimens in pregnant women. https://www.who.int/tb/features_archive/Public-call-individual-patient-data-treatment-drug-res/en/.

[8] ShORRT (Short all-Oral Regimens for Rifampicin-resistant Tuberculosis) Research Package. Resources available at https://www.who.int/tdr/research/tb_hiv/en/.

Measure outcomes with a long indolent course (e.g. pulmonary function, emphysema, quality of life, disabilities, delayed toxicities, etc.) relevant to TB disease/treatment that may not be captured by RCTs.	Risk of bias is greater: treatment and patient selection, patient attrition.
Additional measures such as treatment practices (diagnosis, severity of disease, application of treatment guidelines) in the community setting.	

In conclusion, observational data have a potential role in:

- addressing feasibility, acceptability, delivery strategies, and quality of life issues;
- generating data on special populations not customarily enrolled in clinical trials;
- evaluating safety and effectiveness in broader, more diverse populations; and
- providing additional data on post-marketing toxicities and adverse events.

However, observational data cannot:

- durably serve as the primary source of evidence and data for policy decision making;
- adequately overcome biases inherent to observational studies simply through sophisticated analyses techniques. The absence of randomization results in imbalance of groups such that differences in outcomes cannot be attributed to the therapy being evaluated;
- speed-up the process for accessing and scaling up new therapies;
- answer questions best addressed by randomized clinical trials or become *de facto* approach to addressing other critical questions post approval (conditional or otherwise).

Additionally, observational studies do not generally cost less when considering totality of cost, burdens, and delays in scale up, when the uncertainty is transferred to countries, programs, local and international policy decision makers.

Discussant 1: The importance of safety data: can we accelerate regimen development without compromising safety? –Charles Wells (Gates Medical Research Institute)

Although the drug development trilemma introduced by Dr. Davies (Session 3) is a great way to consider the efficacy side of drug development, Dr. Wells argued that safety should also be kept at the forefront of regimen development. Based on a brief review of recently approved drugs (bedaquiline, delamanid) and regimen for drug resistant TB, Dr. Wells underscored the importance of adequate safety databases, particularly for new drugs or regimens with several safety knowledge gaps. In addition, it is essential to consider the downstream requirements and perspectives of National TB Programmes, including for example, the types of safety monitoring that might be required for new regimens.

Thus, with approval of bedaquiline and delamanid and the high likelihood that these drugs were going to be combined in new treatment regimens, and the knowledge that both medicines may cause QT prolongation, safety assessment was needed. This led to the DELIBERATE Trial (DELamanId BEdaquililne for ResistAnt TubErculosis, NCT02583048) sponsored by the NIH to assess the QT effects of bedaquiline or delamanid alone or in combination in MDR TB patients. The study found that the QT interval (corrected by Fridericia

formula -QTcF) of bedaquiline and delamanid combined was not more than additive and no participants experienced QTcF prolongation of Grade 3 or higher. These results provided critical safety data for the use of both drugs together. Similarly, in the 1990s, three large trials (total N ~ 3400) compared the use of rifampicin and pyrazinamide to isoniazid monotherapy for treatment of latent TB infection and found no difference in adverse events or mortality. This evidence led to the recommendation of the rifampicin and pyrazinamide regimen as an alternative to isoniazid monotherapy for latent TB infection. However, in a national survey, rates of liver injury, hospitalization, and death associated with rifampicin and pyrazinamide therapy exceeded rates with isoniazid monotherapy, leading to the withdrawal in the use of the rifampicin and pyrazinamide, and it was determined that the randomized trials lacked the adequate power to detect fatal events. This example shows that even if there is substantial safety data gathered in clinical trials, further data and evidence is needed to continuously build on the safety profiles because rare events that were not anticipated may still occur. Such experiences should be considered when accelerating regimen development so that safety is not compromised.

Conducting trials in key/vulnerable populations: what is the progress made and what are the persisting challenges?

Discussant 2: *Pregnant and lactating women- Amita Gupta (Johns Hopkins University)*

TB is at its peak incidence in a woman's life during reproductive age. Women may get pregnant during clinical trials for treatment and prevention, even if contraception is recommended. Despite the fact that there are an estimated 216,000 cases of TB in pregnancy per year, pregnant women are excluded from TB clinical trials.[9] Challenges related to TB and pregnancy are: a lower sensitivity of commonly used TB diagnostic tests; variable drug safety assessment and outcome according to trimester of exposure; potential differential drug dosing, treatment responses and drug-drug interactions; and potential pregnancy-related immunologic and physiologic dynamic changes. Special studies and efforts are needed to counter the automatic exclusion of pregnant and lactating women that currently pervades the TB trial landscape.

Dr. Gupta discussed ways to include pregnant women in TB trials.[10] When a woman becomes pregnant while participating in a TB trial, she could be offered, with shared decision-making, to re-consent and continue on the study drug when sufficient data on reproductive toxicity exist. Many of these women would have already had 1st trimester exposure and there may be sufficient equipoise to continue them on a drug, particularly in the case of MDR or XDR TB. Further, there are opportunities for PK studies, either stand-alone or nested within larger parent trials, to gather more information from pregnant women. If the Phase 3 trial represents the first time a drug is being used to demonstrate efficacy and safety, one possibility is to first conduct the study without pregnant women and then determine if there is enough equipoise to include them at a later point in the study. This approach was used in the PHOENIx trial

[9] Sugarman J, *et al*. Tuberculosis in pregnancy: an estimate of the global burden of disease. *Lancet Glob Health*. 2014 Dec;2(12):e710-6. doi: 10.1016/S2214-109X(14)70330-4.

[10] Gupta A, *et al*. Inclusion of key populations in clinical trials of new antituberculosis treatments: Current barriers and recommendations for pregnant and lactating women, children, and HIV-infected persons. *PLoS Med.* 2019 Aug 15;16(8):e1002882. doi: 10.1371/journal.pmed.1002882.

(NCT03568383). Overall, Dr. Gupta encouraged the clinical trial community to continually strive to improving the evidence base for scientifically complex populations, including pregnant and lactating women. An FDA guidance on 'Pregnant Women: Scientific and Ethical Considerations for Inclusion in Clinical trials published in 2018' and the Task Force on Research Specific to Pregnant Women and Lactating Women (PRGLAC) provides key recommendations on how best to conduct research in pregnant women and include them in clinical trials.[11,12] Additionally, establishing a TB pregnancy registry, similar to what has been done for antiretrovirals[13], would be a powerful tool to learn more about therapeutics for pregnant women with TB. It would allow for pooled analysis across different trials and studies, which is especially important because sample sizes for pregnant women are generally small and would help decipher the signal to noise ratio.

Key research questions on TB therapeutics in pregnancy and postpartum include:

- Do first line TB drugs need to be dose modified or not used?
- Is rifampicin interaction more of a concern with newer antiretrovirals (integrase inhibitors) in pregnancy?
- What is the optimal timing to initiate isoniazid preventive therapy?
- Can we safely use shorter regimens for LTBI?
- How should pregnant women with MDR- or XDR-TB be treated?
- How should pregnant women who are contacts of MDR-TB patients be managed?

Discussant 3: Paediatrics- Anneke Hesseling (Stellenbosch University)

Dr. Hesseling discussed advances made in paediatric clinical trials, and the key remaining challenges. She challenged the perception that paediatric TB is not relevant in terms of disease burden. Even though children (0-14 years) only represent 10% of the global disease burden and contribute little to the transmission of disease, they do contribute substantially to the severe morbidity and high mortality of TB. Children often have paucibacillary disease causing challenges for diagnosis, trial entry points, and definition of treatment response and endpoints. Additionally, they have a broad disease spectrum, with children below the age of 5 having the highest incidence and the highest risk to develop the most severe forms of TB. Nonetheless, the majority of children with TB develop pulmonary/intrathoracic TB, typically a lymph node disease that is uncomplicated, so there is an opportunity to treat children with regimens that are shorter than those for adults. In this respect, the Phase 3 SHINE (Shorter Treatment for Minimal Tuberculosis in Children, ISRCTN63579542) trial showed that a 4-month regimen was non-inferior to the standard 6-month regimen in children with 'non-severe' clinically diagnosed TB disease, highlighting the feasibility of stratification in children.[14]

---

[11] FDA. Pregnant Women: Scientific and Ethical Considerations for Inclusion in Clinical Trials. 2018. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/pregnant-women-scientific-and-ethical-considerations-inclusion-clinical-trials.

[12] Task Force on Research Specific to Pregnant Women and Lactating Women (PRGLAC). https://www.nichd.nih.gov/about/advisory/PRGLAC

[13] Antiretroviral Pregnancy Registry. http://www.apregistry.com/

[14] Main findings of the SHINE trial presented by Dr. Eric Wobudeya (Makerere University, Johns Hopkins Research Collaboration, Kampala, Uganda) at 51st virtual Union World Conference on Lung Health.

The need for paediatric PK studies to inform dosing in children is critical because children eliminate drug faster than adults, achieving lower concentrations of TB drugs at the same mg/kg dose. Therefore, higher mg/kg doses are required to achieve adult target exposures which ideally would correlate with efficacy. Paediatric PK studies should include children of representative ages (especially young with consideration of developmental pharmacology), as well as HIV co-infected and malnourished children who tend to have poorer treatment outcomes. A major advancement in this area is the adoption of state-of-the-art population PK methods that allow for paediatric studies to be more efficient, clinically relevant, innovative, adaptive, and informative in dosing. For example, in the IMPAACT Study P1108 (NCT02906007), the use of population PK modelling informed dosing in the younger cohorts with PK modeling in semi-real-time and safety monitoring in real-time. Other major advancements are the emerging consensus that extrapolation of efficacy from adults to children is generally acceptable, with only specific questions, such as those on shortening of treatment regimens for paediatric populations, requiring efficacy studies (e.g. SHINE trial), and that age de-escalation is obsolete. The IMPAACT Study P1108 showed that enrolling older and younger cohorts of children in parallel accelerated the access to bedaquiline for younger children and generated the much-needed PK and safety data more rapidly. Of course, there are exceptions, for example, when higher toxicity is expected in younger groups or if there is no formulation available for younger groups.

Drug formulation has remained a challenge in paediatric studies, together with acceptability and palatability that influence treatment adherence. Dr. Hesseling advocated for a standardized caregiver questionnaire to be included in every paediatric trial to gather information on acceptability and palatability. Timely investment in predictable, dispersible, and palatable child-friendly formulations is critical, particularly for young children. There is now emerging consensus that paediatric formulation should be developed at the time that adult Phase 2 trials are initiated and there are strong initiatives that focus on paediatric drug formulation development – e.g. PADO TB[15] and GAP-f[16].

Lastly, Dr. Hesseling highlighted the challenges in monitoring safety and clinical assessments in children. A novel body mapping tool has been successfully developed to help describe how regimens were tolerated by adolescents treated for MDR TB[17], and similar innovative tools are needed for children. She strongly advocated for inclusion of children in Phase 3 trials to increase generalizability. Paediatric TB trials are about access; if paediatric therapeutic trials are not ramped up, children will lose the much-needed access to new treatments.

---

[15] Meeting report: Reaching UNGA HLM on TB targets for ending TB in children and adolescents: First Paediatric Antituberculosis Drug Optimization Meeting (PADO-TB 1). http://www.stoptb.org/wg/dots_expansion/childhoodtb/assets/documents/PADO-TB1_Meeting_Report_FINAL_v8March2019.pdf

[16] Global Accelerator for Paediatric Formulations (GAP-f). http://gap-f.org/

[17] Zimri K, *et al.* A novel approach for eliciting adolescent MDR-TB treatment tolerability: qualitative data from South Africa. *Int J Tuberc Lung Dis.* 2020 Jan 1;24(1):43-47. doi: 10.5588/ijtld.19.0207.

The use of new drugs/regimens beyond registration:
1. The view of sponsors
Discussant 4: Alex Pym (Janssen Pharmaceutica)

Dr. Pym provided a brief overview of the bedaquiline clinical development program. Bedaquiline had accelerated approval in the US at the end of 2012 and conditional approval in the EU in 2013. Approval was based on three phase 2 trials (a 7-day EBA study, TMC207-C202, NCT00523926; a placebo-controlled study, TMC207-C208; NCT00449644; and a cohort study, TMC207-C209, NCT00910871). The post-marketing commitments included conducting a Phase 3 trial, a paediatric PK study, and establishing a multi-country registry for real world evidence. The first WHO guidelines for use of bedaquiline in MDR-TB treatment in 2014 mentioned that bedaquiline was 'not generally recommended for routine use due to the lack of evidence'.[18] With newly available evidence, in 2016, bedaquiline was recommended as an 'add on agent',[19] and in 2018, as one of the 'medicines to be prioritized in long-course regimens' at the same time that injectables were de-prioritized.[20] Ultimately, in 2019, it was included in the shorter regimens.[21] Real-world evidence has mainly supported this evolution in the adoption of bedaquiline, with the first data coming from South Africa as it was an early adopter of bedaquiline, followed by the work from the Collaborative Group for the Meta-Analysis of Individual Patient Data in MDR-TB treatment.[22] Yet, there is much debate on how much impact real-world evidence should have on policy decision, and Dr. Pym questioned whether it is worth waiting for randomized clinical trial data (indeed, safety data from these trials are critical) before making guideline changes. He emphasized the challenging task of designing Phase 3 trials in the context of dynamic treatment guidelines for MDR-TB.

Then the relationship between approval, adoption, and uptake of bedaquiline was discussed. In 2014, the Global Drug Facility Distribution and the United States Agency for International Development (USAID) Bedaquiline Donation Program were established, leading to the incorporation of bedaquiline into the South African TB Program a year later. Between 2016 and 2018 there was enough real-world evidence for bedaquiline to be adopted in the South African treatment guidelines (2018) and the WHO treatment guidelines (2019). These adoptions generated programmatic uptake of bedaquiline but it took approximately 7 to 8 years between the initial approval and substantial uptake of bedaquiline. Could have this been reduced? In hindsight, randomized clinical trials may have helped accelerate this process. Nevertheless,

---

[18] WHO. Interim policy guidance on the use of bedaquiline in the treatment of multidrug-resistant tuberculosis. 2013. WHO/HTM/TB/2013.6. https://www.who.int/tb/publications/mdrtb-treatment-guideline/en/

[19] WHO treatment guidelines for drug-resistant tuberculosis. 2016 update. WHO/HTM/TB/2016.04. https://apps.who.int/iris/bitstream/handle/10665/250125/9789241549639-eng.pdf?sequence=1

[20] WHO treatment guidelines for multidrug- and rifampicin-resistant tuberculosis. 2018 update. WHO/CDS/TB/2018.15.

[21] WHO consolidated guidelines on drug-resistant tuberculosis treatment. 2019 update. ISBN: ISBN 978-92-4-155052-9. https://www.who.int/tb/publications/2019/consolidated-guidelines-drug-resistant-TB-treatment/en/.

[22] Collaborative Group for the Meta-Analysis of Individual Patient Data in MDR-TB treatment–2017, *et. al*. Treatment correlates of successful outcomes in pulmonary multidrug-resistant tuberculosis: an individual patient data meta-analysis. *Lancet.* 2018 Sep 8;392(10150):821-834. doi: 10.1016/S0140-6736(18)31644-1.

there is a role for real-world evidence to get medicines to patients quickly but more needs to be done to understand how to efficiently utilize this type of evidence.

Finally, Dr. Pym discussed the post-marketing commitment to establish a multi-country registry for bedaquiline. The objective of the registry was to gather data on safety and describe patient characteristics, bedaquiline utilization, and emergence of resistance. The registry collected data between 2015 to 2017 from MDR-TB patients initiating a bedaquiline-containing treatment regimen (N ~3000 patients) or treatment regimen not containing bedaquiline. This required a multi-stakeholder collaboration that was supported by the South Africa Department of Health. One of the major challenges in developing the registry was that regulatory authorities were very focused on the quality of the real-world evidence; therefore, it was critical to invest in updating the Electronic Drug-Resistant TB Registry (EDRWeb), ensure collection of appropriate data, and collaborate with several institutions who provided IT support, data collection and analysis expertise. Overall, to generate the quality real-world evidence data that regulatory authorities would accept, large investments, many resources, and strong collaborations were required. It was suggested that drug developers have early dialogue with policymakers and stakeholders, preferably pre-approval, to design a unified development and implementation plan for real-world evidence studies. Dr. Pym called for a much-needed global infrastructure and network for collecting high quality, complete, and standardized real-world evidence and data that would be appropriate for both regulators and policy makers. Lastly, he highlighted the importance of global surveillance of emergent resistance that needs to be put on the forefront of the real-world evidence agenda to protect and preserve the longevity of the most potent agents with new mechanisms of action for treatment of TB.

Discussant 5: Gavin Koh (GSK)

Dr. Koh underscored the critical importance of safety data. Indeed, from a drug developer's point of view, the real struggle is about balancing safety and efficacy. A developer aims to push the dose high enough and the duration long enough to achieve appropriate efficacy without compromising the safety of patients. For most indications, regulators require a safety database on the final dose and final duration on at least a few hundred patients, but the exact numbers are subject to discussion with the regulators prior to starting the Phase 3 programme. For DR TB, is it likely that the regulators will require smaller numbers of subjects, but for DS TB, much larger safety database will likely be required. Although efficacy may be demonstrated with adequate statistical power when including 50 participants through the use of innovative methods described in this webinar, and complete enrolment can be achieved at a single site in 1 month, regulators may still insist on a 500-participant study across multiple WHO regions so that an adequate safety database can be generated from a wide range of relevant populations.

For drug developers, the purpose of a Phase 3 or a registration trial is to provide definitive data on the safety and efficacy of a new drug or regimen in the final study population, for a specified indication, at the final dose, using the final manufacturing process, in the final formulation, for the final duration, against the most clinically relevant endpoint for that disease. Clinical data that omit any of these points will likely be required by the regulators to submit additional data to fill these gaps. The additional data may be in the form of one or more additional clinical

trials, up to and including the possibility of completing a full Phase 3 trial. Thus, while there is a clear need to test efficacy in drug development, the proposals cannot focus on efficacy alone. If safety is not included as a key consideration in these discussions, then there is a danger that the efficiencies in the developmental pathway may turn out to be false economies.

2. The view of the regulator:

Discussant 6: *Marco Cavaleri (European Medicines Agency, EMA)*

Dr. Cavaleri reminded the participants that the EMA has an explicit legal mandate, not only to authorize medicinal products, but also to follow up on what occurs after registration. In particular, the EMA works with the sponsors to agree on developing a Risk Management Plan (that outlines the follow-up procedures and post-authorization studies on safety (mainly) and efficacy of approved medicinal products), collecting safety data through Adverse Drug Reactions, and monitoring benefit-risk of approved medicinal products. The regulators can also impose post-authorization safety and efficacy studies to address gaps in the registration package. If the balance between benefit and risk changes post-authorization, the regulators can impose on sponsors variation, suspension, or revocation. The regulators have an obligation to be transparent in publishing opinions on the status of post-authorized medicinal products and maintain a database of all products introduced in the EU market.

EMA is active in a number of real-world evidence strategies and initiatives. Criteria for acceptability of real-world evidence for regulatory use include demonstrated quality, transparency, internal and external validity, consistency across data sources and countries, and adequacy in terms of the measure for the population investigated. There are several publications that discuss the agency's view on conducting real-world evidence studies in the context of international registries and future real-world evidence framework.[23,24] The Data Analysis and Real World Interrogation Network (DARWIN EU) project is one of the activities that has emerged from an European Union (EU) perspective.[25] It aims to create a secure platform to access and analyse EU healthcare data to support decision-making through the product lifecycle. This project benefits both regulators and recommending bodies because it provides a continuous option of monitoring the benefits and risks of medicinal products in the post-approval, real-world setting. It also complements clinical trials and supports development, authorization and supervision of medicines, thereby supporting patients' access to safe and effective medicinal products. This is currently a European project but can be extrapolated and used in a global context. Dr. Cavaleri advised that to develop this type of project globally, one must start with a global network with integrated databases that could result into meaningful collection and merging of data.

---

[23] Cave A, Kurz X, Arlett P. Real-World Data for Regulatory Decision Making: Challenges and Possible Solutions for Europe. *Clin Pharmacol Ther.* 2019 Jul;106(1):36-39. doi: 10.1002/cpt.1426. Epub 2019 Apr 10.

[24] Eichler HG, *et al*. Are Novel, Nonrandomized Analytic Methods Fit for Decision Making? The Need for Prospective, Controlled, and Transparent Validation. *Clin Pharmacol Ther.* 2020 Apr;107(4):773-779. doi: 10.1002/cpt.1638.

[25] EMA. HMA-EMA Joint Big Data Taskforce Phase II report: 'Evolving Data-Driven Regulation'. 2019. EMA/5484203/2019. https://www.ema.europa.eu/en/documents/other/hma-ema-joint-big-data-taskforce-phase-ii-report-evolving-data-driven-regulation_en.pdf

Dr. Cavaleri reviewed examples of real-world evidence in the context of TB that have been utilized by EMA. For pretomanid, the approval of a fixed-dose combination in a restricted patient population with high unmet need was based on a single pivotal study, Nix-TB, with approximately 100 patients. The data was considered premature; therefore, confirmatory data was needed to provide a full evaluation of the benefit-risk ratio. However, the Committee for Medicinal Products for Human Use (CMPH) at EMA considered the Nix-TB data was sufficient for granting conditional marketing authorization with an obligation to provide comprehensive efficacy and safety data, in addition to what was already submitted. Additional data on clinical pharmacology were also required, along with information on special populations (patients with renal or hepatic impairment, safety in patients over 65 years old, use in pregnancy) and children and adolescents.[26] The ZeNix study and completed Nix-TB dataset are aimed to provide data on these aspects, which potentially makes it possible to convert the conditional authorization into full market authorization. Overall, the CHMP considered that the benefit risk-balance was positive for pretomanid (in fixed combination) and the superior outcomes and simplified shorter duration regimen for a highly unmet medical need led to conditional marketing approval even though additional data are still required. This example shows that substantial data is still needed post approval.

The conditional marketing authorization for delamanid also required post-marketing commitments. Two specific obligations were imposed on the sponsor: i) to complete and report the Phase 3 Otsuka 242-09-213 trial (NCT01424670) and ii) to conduct a new study to compare 200 mg/day versus 400 mg/day. Otsuka 242-09-213 was conducted to show favourable efficacy of delamanid but failed to do so, mainly because the standard of care with the use of a potent fluoroquinolone, which was never studied in a randomized trial, performed better than expected. The results were beneficial for the standard of care but detrimental to delamanid, the drug of interest, leading to the dilemma on how to move forward. It was decided that because the endTB trial (a pragmatic study, NCT02754765) was already in the planning phases, it would be a good study to assess the role of delamanid use in all-oral combinations to assess if delamanid had indeed any role in the treatment of MDR TB.

The importance of a prospective dialogue with regulators and recommending bodies on pre- and post-licensure evidence was highlighted. EMA has established the Postlicensing or postlaunch evidence generation (PLEG) focus group, a pilot that brings together regulators and recommending bodies to discuss the design of post-approval studies, including studies on real-world evidence that could inform both regulatory decisions on monitoring benefit-risk and policy decisions for recommendations and reimbursements.[27] PLEG is part of the continuum of evidence development for a medicinal product, complementing earlier evidence, facilitating

[26] In Europe there is an obligation to have a paediatric investigation plan submitted and approved by the paediatric committee.

[27] Moseley J,*et al*. Regulatory and health technology assessment advice on postlicensing and postlaunch evidence generation is a foundation for lifecycle data collection for medicines. *Br J Clin Pharmacol.* 2020 Jun;86(6):1034-1051. doi: 10.1111/bcp.14279.

further elucidation of a product's benefit/risk profile or value proposition, and exploring broader aspects of disease management and provision of healthcare.

Dr. Cavaleri advocated for more collaborative work, similar to the PLEG focus group, to ensure that the data generated from randomized trials or observational studies is fit for the purpose of both regulators and policy makers. Clearly, there is an increased appetite for the use of real-world evidence compared to the past. Still, there are confounders and issues that may refrain regulators from making decisions based on real-world evidence. Establishment of real-world evidence registries locally and globally may be a good next step forward. Dr. Cavaleri hopes that the WHO and other organizations can help in setting up these endeavours which can be a basis for continuous data generation and collection that could inform decisions by regulators and public health authorities. Finally, he suggested that the EMA experience with Health Technology Assessment bodies and in the context of Article 58 (Regulation (EC) No 726/2004)[28] collaboration with WHO for vaccine, HIV prevention, and other neglected disease should be extended to other public health areas such as TB.

3. The view of the policy-maker:
Discussant 7: Dennis Falzon (WHO)

Dr. Falzon provided the perspective of WHO, as a normative body, on the role of clinical trials and observational studies to provide the evidence needed to inform global treatment policy. According to the WHO Constitution (1946), "In order to achieve its objective, the functions of the Organization shall be: (a) to act as the directing and coordinating authority on international health work; … (n) to promote and conduct research in the field of health; … (u) to develop, establish and promote international standards with respect to food, biological, pharmaceutical and similar products; …".[29] This highlights some of the differences between WHO, regulators and other technical partners.

Dr. Falzon first described the process for developing recommendations and the different types of WHO technical documents: guidelines, rapid communications, operational handbooks, technical reports, and target product profiles/preferred product characteristics. One of the most important evolutions in the last 12 years at WHO has been the establishment of a system for evidence collection, summarization, presentation, and discussion.[30] The system is based on the 'Grading of Recommendations, Assessment, Development, and Evaluations' (GRADE) framework to rate the quality of evidence for set outcome(s) and translate evidence into recommendations. The PICO backbone is used to design a clinical or public health question that frames the challenge for which guidance is sought and is formulated on four critical elements: the *population* of concern, the *intervention* proposed to address the issue, the current intervention or *comparator* (should one exist), and the expected *outcome*. The decision framework used to develop recommendations considers questions related to the priority of the

[28] Regulation (EC) No 726/2004 of the European Parliament and of the Council of 31 March 2004. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32004R0726
[29] Constitution of the World Health Organization. 1946. https://apps.who.int/gb/bd/
[30] WHO Handbook for guideline development, 2nd Ed. https://apps.who.int/iris/handle/10665/145714

problem, the desirable and undesirable effects, the certainty of the effects, the influence on cost and resource requirements, the impact on health equity, the acceptability to key stakeholders and the feasibility to implement. Transparency, as well as the choice and screening of experts for potential interests and mitigation of conflict, are also important elements of the system.

Given the shortage of trial data, since 2010, there has been an increased use of observational data for drug resistant TB treatment policy making. While such studies do not provide high-quality evidence, they closely reflect contemporary practices under programmatic conditions, including in resource constrained settings.[31,32] The availability of individual patient records has allowed for more refined analysis than study-level meta-analysis. The consolidation of individual patient data has motivated more contributors to share treatment experiences over the years. However, these recent trends do not diminish the importance of clinical trials in TB treatment policy, to improve the evidence of efficacy and safety and to address critical research knowledge gaps. For example, in 2016, with mounting evidence from different settings of the effectiveness of the new shorter 9-month MDR-TB regimen, observational data were used by WHO to develop its first recommendation for this regimen. An assessment of the STREAM trial data could only be conducted two years later, after its completion, which reinforced the continued recommendation of the regimen and raised the certainty in the evidence supporting its use from 'low' to 'moderate'.

Several lessons have been gained from recent clinical trials, including the following:

- control arms in MDR-TB clinical trials (STREAM trials, NCT02409290; and Otsuka 242-09-213 trial, NCT01424670) performed far better than the same regimens observed under routine programmatic conditions, suggesting that, together, supportive treatment, prevention of treatment interruption and robust pharmacovigilance could improve patient outcome;
- trials testing a new agent on top of an optimized background regimen may mask the best assets of the tested agent and do not mirror how that agent would be used by programmes to improve outcomes;
- achieving non-inferiority limits is not valued in the same way by guideline users compared with other stakeholders, particularly when the intervention carries clear benefits (e.g. shorter durations, less toxicity, lower costs);
- in settings with very high mortality, an intervention that generally improves survival may still have value even if there are unresolved issues on its safety;
- changes to TB treatment policy while a trial is ongoing may question the ethics of continued enrolment in the control or intervention arm on a regimen that is no longer recommended; and

[31] WHO consolidated guidelines on drug-resistant tuberculosis treatment. 2019 update. ISBN: ISBN 978-92-4-155052-9. https://www.who.int/tb/publications/2019/consolidated-guidelines-drug-resistant-TB-treatment/en/.
[32] Nahid P, *et al*. An Official ATS/CDC/ERS/IDSA Clinical Practice Guideline. Am J Respir Crit Care Med. 2019 Nov 15;200(10):e93-e142. doi: 10.1164/rccm.201909-1874ST.

- newly emerging data may challenge common beliefs about efficacy and toxicity of regimens and early communication of trial results may influence the legitimacy of concurrent trials.

Dr. Falzon further discussed standardized vs stratified approaches. Although stratification may sound logical, there may have unintended influence on equity that need consideration if the added value of customized regimens is outweighed by additional resource needs. Identification of individual risk groups and optimal differentiated care would require simplification, and further implementation research is needed about feasibility of departures from standard approaches.

Lastly, Dr. Falzon explained that the WHO perspective differs from the one of regulatory agencies. WHO treatment guidelines are primarily driven by the demand for improved options or existing uncertainties over practice, while the regulatory process is focused on the submission for approval of an agent. For policy development, options are assessed by expert panels on a balance of potential benefits to risks and the use of drugs can be considered beyond their primary indication. This off-label use is common practice in drug-resistant TB treatment policy where treatment options are often limited. While stringent regulatory authorities are national or regional, WHO has a global span and caters to settings that differ in TB prevalence, resources, health structure, and capacity of drug-safety monitoring. While WHO seeks evidence from large studies with broad geographical spread, the availability of data is often influenced by the regulatory requirements. In summary, the desirable features of a study from a policy maker perspective are those that provide the elements needed to make strong recommendations. Alongside evidence for efficacy and safety, WHO highly values studies on feasibility, resource requirements, and acceptability. Time is key in the development of recommendations because the process cannot be compressed, hence the importance of sharing preliminary results with WHO early.

4. Discussant 8: Lindsay Mckenna (Treatment Action Group)

Ms. McKenna presented the civil society perspective on off-label use of drugs and regimens for TB beyond registration. Based on their product labels, bedaquiline, delamanid, and pretomanid, are limited to use in specific populations: bedaquiline and delamanid are reserved for MDR TB when effective regimens cannot otherwise be composed and pretomanid for XDR or treatment intolerant/non-responsive MDR TB. However, according to latest WHO treatment guidelines,[33] bedaquiline is a group A drug, i.e. a core component of the recommended MDR TB regimen and is no longer reserved for use 'when an effective regimen cannot otherwise be composed' – hence its use beyond the existing label. In contrast, delamanid is a group C drug and is being used as indicated on the label. Technically, both agents are used off-label for extrapulmonary TB and beyond 6 months. For pretomanid, the guidelines and label do align. All three new drugs were approved under special accelerated pathways, but for guideline development and practice, more data is needed beyond what was available at the time of regulatory approval

[33] WHO consolidated guidelines on drug-resistant tuberculosis treatment. 2019 update. ISBN: ISBN 978-92-4-155052-9. https://www.who.int/tb/publications/2019/consolidated-guidelines-drug-resistant-TB-treatment/en/.

under these special pathways. Regarding repurposed medicines, moxifloxacin, levofloxacin, linezolid and clofazimine are always used off label in the context of indication and duration (beyond 28 days).

There are several knowledge gaps pertaining to best use of new drugs, particularly information that is critical for clinical use/policy making (and to patients). A number of these have been preliminarily filled with non-randomized or observational data. However, the use of observational data to inform policy change beyond what was recommended at the time of regulatory approval is not rapid (contrary to popular belief), ranging from 6 to 8 years. Randomized data provides higher quality evidence and takes 6-10 years post-authorization to address remaining key questions and knowledge gaps on the use of new drugs and regimens.

Ms. McKenna advocated for re-establishing evidence standards in TB treatment research and answering questions important to TB programs and patients and to promoting equity at earlier stages of the development pathway. In TB, there is a cycle where new drugs are approved on limited data and guidelines are iteratively updated based on real-world evidence preceding clinical trials end results. The pre-existing standard of care for MDR TB treatment, based on expert opinion and limited evidence, was long, toxic and poorly performing, and major changes to guidelines made over the last 5 years should be considered exceptional since emerging program data mainly influenced the benefit-risk ratio. However, moving forward, recommendations for the use of new drugs and major changes to the standard of care should be based primarily on data from randomized clinical trials, supplemented by data collected through rigorous pragmatic and operational research initiatives. Questions on optimal use (i.e. combination and duration), drug-drug interactions, safety, and use in special populations should be addressed at earlier stages in developmental pathways. This would help expedite and expand uptake, promoting equity and reducing need for off label use.

Often, data collection on patient preferences is reserved for later stages of research, after an intervention is already available, or most commonly after issuing guidance based on assumptions in absence of information. There is a need to normalize the inclusion of rigorous qualitative acceptability work as part of clinical trials much earlier in the development pathway. TB treatment research consortia and sponsors should establish (and fund) complementary qualitative research focused on acceptability to understand how TB-affected communities consider trade-offs in efficacy, safety, tolerability and duration. The results from such qualitative assessment should also inform trial design (e.g. noninferiority margin selection) and be documented in protocols and publications. Researchers, product sponsors, regulatory authorities and policymakers need to consider patients' preferences and priorities of affected communities before research begins - ideally at the stage where the primary research question(s) for a new treatment and/or regimen are being determined. There is a critical need to go beyond anecdotal and superficial approaches to high evidence standard for patient preference and accessibility work.

Ms. McKenna underscored the importance and value of engaging communities in TB research. People have a right to participate in medical research as more than trial participants (Denver

Principles, Declaration of Helsinki, human rights law). Communities that feel connected to research are more likely to participate in clinical studies, with favourable effects on recruitment, enrollment, and retention into studies and reciprocal and lasting partnerships between communities and researchers. Additionally, communities that understand TB science are better positioned to advocate for continued research before governments and other funders and may increase the likelihood of the uptake and adoption of TB research outputs. The TBTC S31/A5349 trial offers a positive example of benefits of early and continued community engagement in research – the Community Research Advisors Group (CRAG) advocated for and ultimately won the inclusion of people with HIV at lower CD4 counts and adolescents in the trial, expanding the generalizability of the results and populations ultimately able to benefit from access to shorter regimens.

Round table panelist: Susan Swindells (University of Nebraska), Vindi Singh (WHO), Sumathi Nambiar (FDA), Kissa Mwamwitwa (Tanzania Medicines and Medical Devices Authority), and Francis Varaine (Médecins Sans Frontières)

With the TB development pathway being too long, several panellists strongly advocated for data that inform public health recommendations to be collected in parallel to Phase 3 clinical trials. For example, Dr. Swindells illustrated a staged approach to overcome the major concern of potential drug-drug interactions with antiretroviral therapies when including people living with HIV in clinical trials. First, only a certain number of people living with HIV are enrolled, and PK data can be collected and analysed to assess potential drug-drug interactions. Then, if appropriate, inclusion can be expanded to include more people living with HIV. Panellists also advocated to end the exclusion of women who become pregnant on studies. Similar to the HIV example, options for keeping pregnant and lactating women in trials should be built into studies, particularly to gather the much-needed PK data. Regarding children, the panellists recognized the challenging task of addressing the formulation issues, and although PK studies in children may be more expensive and complicated, a more inclusive dataset from clinical trials would still be valuable and increase generalizability. Further, excluding these scientifically complex populations from clinical trials leaves policy makers to make decisions without evidence, therefore relying on extrapolation.

Dr. Nambiar explained that even if a medicinal product is approved based on a favourable benefit-risk assessment, there are often gaps and uncertainties that the development program could not address because of the need to balance feasibility and sound scientific principles in the assessment of efficacy and safety. When development programs are streamlined to address an unmet need, there are likely to be more gaps and uncertainties and sponsors will likely be required to collect additional information post-approval. It would be helpful to work with sponsors and policy makers during the pre-approval stage to discuss and determine the prioritization of collecting in clinical trials the key information useful to policy makers.

Dr. Nambiar referred to the FDA draft guidance documents that outline many principles relevant to TB drug development including extrapolation of efficacy and need of concurrent cohorts: Development of Anti-Infective Drug Products for the Pediatric Population, published in

June 2020[34]; Pregnant Women: Scientific and Ethical Considerations for Inclusion in Clinical Trials, published in 2018[35]; and Post-approval Pregnancy Safety Studies, published in 2019[36]. She also suggested to review recent approvals for anti-infective products for an understanding on the kinds of post-marketing studies that have been required regarding pregnant women. Like the EMA, FDA also provides a real-world evidence program and framework[37]. The framework considers whether the real-world evidence data is fit for use, the study used to generate real-world data can provide adequate scientific evidence, and the conduct of the study meets FDA regulatory requirements. A subcommittee within the agency oversees this work and provides advice and recommendations on all proposals.

Although most countries adopt newly approved drugs into TB treatment guidance based on safety and efficacy information from drug developers, the safety database is usually inadequate. Therefore, Dr. Mwamwitwa urged sponsors and developers to ensure continued safety monitoring and surveillance of their products in all populations, especially children and pregnant women, and in various regimen compositions, so TB treatment guidance can be continually informed. She echoed the need for more collaboration between regulators, sponsors, and policymakers from pre- to post-approval to ensure the generation of high-quality data for regulatory and guidance purposes.

Dr. Varaine described issues that are inherent to the use of real-world evidence and their risk of systemic bias. He noted that unfortunately real-world evidence is currently the main basis for DR-TB treatment guidelines because of the dramatic lack of RCTs in that field; WHO guidelines are largely based on the IPD meta-analysis and recommendations on key questions such as classification of drugs, composition of treatment regimen, and duration of treatment are based on very low-quality evidence. He noted that often there is a misconception that large datasets combined with advanced and sophisticated statistical methods can overcome systemic bias, but there are always confounders and unknowns. Because bias is difficult to assess, the approaches and tools required are complex, and consequently, the analysis and results can become difficult to interpret, especially by policymakers and clinicians. Dr. Varaine listed a few questions that the TB community should consider when using real world data: Should there be standards for observational studies? Could analysis methods be improved, and limitations clarified? How can we improve real world evidence so that it is of better quality? Could there be an agreed upon approach to handling deaths, lost to follow-up, or other missing data types? Should protocols for observational studies be registered to ensure that the negative, positive, and non-conclusive

[34] FDA. Development of Anti-Infective Drug Products for the Pediatric Population, Guidance for Industry. 2020. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/development-anti-infective-drug-products-pediatric-population

[35] FDA. Pregnant Women: Scientific and Ethical Considerations for Inclusion in Clinical Trials, Guidance for Industry. 2018. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/pregnant-women-scientific-and-ethical-considerations-inclusion-clinical-trials

[36] FDA. Postapproval Pregnancy Safety Studies, Guidance for Industry. 2019. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/postapproval-pregnancy-safety-studies-guidance-industry

[37] FDA. Frameworks for FDA's Real-World Evidence Program. 2018. https://www.fda.gov/media/120060/download

results are published and available? He finally recalled that conclusions from observational studies are sometimes refuted by RCTs (in this respect the observational cacophony versus the randomized harmony has been recently mentioned for the COVID-19)[38].

[38] Califf RM, Hernandez AF, Landray M. Weighing the Benefits and Risks of Proliferating Observational Treatment Assessments: Observational Cacophony, Randomized Harmony. JAMA. 2020;324(7):625-626. doi:10.1001/jama.2020.13319.

ANNEX 4: Declaration of interest statements

The following meeting participants disclosed no interests related to the topic of the meeting:

Clifton Barry, Martin Boeree, Grania Bridgen, Marco Cavaleri, Tsira Chakhaia, Angela Crook, Norbert Djeka, Lori Dodd, Amina Gupta, Debra Hanna, Michael Holscher, Robert Horsburgh, Shitong Huan, Marjorie Imperial, Thomas Jaki, Yuhong Liu, Vidya Mave, Lindsay McKenna, Stellah Mpagama, Patrick Phillips, Helen Rees, Morten Ruhwald, Alena Skrahina, Guy Thwaites.

The following participants declared interests that were judged not to conflict with the objectives of the meeting:

Dick Chaisson declared consulting for Sanofi.

Gavin Churchyard declared that his employer received grants from Sanofi, BMGF and EU2020.

Daniela Cirillo declared receiving a grant to perform MIC for Bedaquiline as SRL in Janssen Pharmaceutica, as well as a grant by TBA to study MIC distribution for new drugs. Cepheid: all of the tasks were committed by FIND. Presently working with FIND on specific projects not specifically related. Co-coordinator of WP5 (tuberculosis) of EU-Pearl IMI 2018. Member of consortium UNITED for TB applying to IMI second stage.

Geraint Davies was the academic co-ordinator of the PreDiCT-TB consortium, a public-private partnership funded by the European Union Innovative Medicines Initiative and the European Federation of Pharmaceutical Industries and Associations. Although this role involved engagement with industrial partners (GSK, Sanofi, Janssen) in pre-competitive areas of research into TB drug development, these activities were fully supported by public funding from the EU and neither himself nor his research institution received any funding from EFPIA or from the individual industrial partners. His employer, the University of Liverpool, is also an academic partner to a current Stage 2 application to the innovative Medicines Initiative call UNITE4TB, the outcome of which is expected in December 2020.

Andreas Diacon declared being the CEO and founder of the non-profit organization TASK Foundation, specializing in clinical research with a special focus on TB. The organization receives funding from various networks based in the US and Europe as well as pharmaceutical companies. He does not receive direct payment from any sponsor, commercial institute or network.

Kelly Dooley declared being protocol co-chair for the AIDS Clinical Trials Group (ACTG) study A5343 assessing use of delamanid and bedaquiline among patients with MDR-TB for which they received drug donation by Otsuka and Janssen and ViiV. Support and funding for this trial are provided by NIH Division of AIDS. She is not receiving salary support from drug companies for this project. She also declared being the principal investigator for the Assessing Pretomanid for

Tuberculosis (APT) trial, assessing pretomanid (previously PA-824, investigational drug) for treatment of drug-sensitive TB. Support and funding is from the US Food and Drug Administration (FDA). Drug donation from Global Alliance for TB Drug Development (pretomanid) and Pfizer (rifabutin). She does not receive funds from drug companies for this project.

Jeremy Hill declared employment and consulting with the TB Alliance, as well as receiving research support to his employer from the TB Alliance.

Barbara Laughon is an employee of the US National Institute of Health and provides a salary for seeking out the best approaches to the cure and prevention of tuberculosis worldwide. Travel costs to attend scientific meetings and WHO consultations have been paid by the US National Institute of Health. In addition, she serves as co-chair for the STOP TB Partnership's Working Group on New TB Drugs.

Payam Nahid declared receiving a federal CDC contract to support clinical trial units in University of San Francisco, California, USA and Hanoi, Viet Nam. He also declared serving as a member of the DSMB for an MDR-TB clinical trial, TB-PRACTECAL. In addition, discussions are underway with MSF for future potential participation of Viet Nam clinical trial units in Hanoi and HCMC in ENDTB MDR-TB clinical trials, but no contracts or agreements were offered, signed or formalized at the time of the meeting.

Sumathi Nambiar is an FDA employee and as such has represented the Agency's position on clinical trials for drugs being developed for the treatment of tuberculosis.

Eric Nuermburger declared consulting for Sanofi and receiving research support from Janssen, Shionogi, TB Alliance and Gates Medical Research Institute.

Andrew Nunn declared employment and receiving a grant from Janssen Pharmaceuticals. The STREAM trial on which he is co-chief investigator is partly funded by Janssen Pharmaceuticals and part by USAID and covers part of his salary.

Bern-Thomas Nyang'wa declared employment and consulting with Medecins sans Frontieres, TBPRACTECAL trial

Nick Patton declared receiving donations of rifapentine for TRUNCATE-TB trial from Pfizer, donation of linezolid for TRUNCATE-TB trial from Sanofi, a grant for a clinical trial (HIV) from Janssen and payment for being a member of the advisory board from Sanofi.

Ulrika Simonsson declared a scientific grant from EU as part of the consortium ERA4TB to Uppsala University.

Mel Spigelman declared research support, including grants, collaborations, sponsorships, and other funding from various sources.

Eugene Sun declared employment with the TB Alliance and receiving research support from BMGF and other sources.

Susan Swindells declared receiving public statements and research support from US NIH, as well as salary and travel support from the US NIH as protocol chair of the BRIEF-TB clinical trial, and travel support to serve as a member of the US NIH Adult & Adolescent Antiretroviral Treatment Guidelines Panel, TB Section.

Andrew Vernon declared working in the Division of TB Elimination at CDC in the United States, where he directs a branch that conducts clinical trials in tuberculosis treatment and prevention. He has participated in the development of US guidelines, and on rare occasions presented information about their group's work to FDA. His group has worked with several commercial collaborators for over the past 25 years. Most recently, they have collaborated with NIH and Sanofi on the conduct of a multinational phase 3 trial of TB treatment using daily rifapentine. Sanofi provided medications for the trial, and supported costs of pharmacokinetic testing. From 2007-2017, Sanofi made contributions to the CDC Foundation for the support of CDC research on rifapentine. These funds supported costs for study supplies and materials, for pharmacokinetic testing, and for 2-3 contract staff who worked to support these trials. Neither he nor any other CDC employee, nor any family member, received any benefits from these funds. Moreover, these funds were only a small proportion of overall trial costs, the vast majority of which were borne by CDC as the trials' sponsor. Their group has conducted trials in treatment of latent TB infection and began a new LTBI trial last summer. Sanofi is providing medication for that LTBI trial but is not involved in the design or conduct of that trial. Total contribution to the CDC Foundation over 10 years was ~$3million. Cost of medications and PK testing for current phase 2 active TB trial is not clear. The trial enrolled over 2500 persons and has completed all treatment.

Robert Wallis declared receiving research grant support from the Gates Foundation, H2020 and NIH

Nick Walter declared a patent application belonging to the University of University of Colorado Anschutz Medical Campus

The following participants work for the pharmaceutical industry and were exempted from declaring any interests:

Alex Carlton, Jeffrey Hafkin, Gavin Koh, Yongge Liu, Robin Mogg, Oscar Della Pasqua, Alexander Pym, Charles Wells.

9789240030862