

# Developing biomarker assays to accelerate tuberculosis drug development: defining target product profiles

Stephen H Gillespie, Andrew R DiNardo, Sophia B Georgioui, Wilber Sabiiti, Mikashmi Kohli, Ursula Panzner, Irina Kontsevaya, Norbert Hittel, Lieven J Stuyver, Jia Bin Tan, Reinout van Crevel, Christoph Lange, Thuong Nguyen Thuy Thuong, Jan Heyckendorf, Morten Ruhwald, Norbert Heinrich



Drug development for tuberculosis is hindered by the methodological limitations in the definitions of patient outcomes, particularly the slow organism growth and difficulty in obtaining suitable and representative samples throughout the treatment. We developed target product profiles for biomarker assays suitable for early-phase and late-phase clinical drug trials by consulting subject-matter experts on the desirable performance and operational characteristics of such assays for monitoring of tuberculosis treatment in drug trials. Minimal and optimal criteria were defined for scope, intended use, pricing, performance, and operational characteristics of the biomarkers. Early-stage trial assays should accurately quantify the number of viable bacilli, whereas late-stage trial assays should match the number, predict relapse-free cure, and replace culture conversion endpoints. The operational criteria reflect the infrastructure and resources available for drug trials. The effective tools should define the sterilising activity of the drug and lower the probability of treatment failure or relapse in people with tuberculosis. The target product profiles outlined in this Review should guide and de-risk the development of biomarker-based assays suitable for phase 2 and 3 clinical drug trials.

## Introduction

Tuberculosis drug development is hindered by the many limitations in the methods used to define outcomes in people with tuberculosis, especially the slow growth of *Mycobacterium tuberculosis* and the difficulty in obtaining suitable representative samples from patients throughout their treatment.<sup>1,2</sup> In the case of pretomanid, this journey took almost 20 years.<sup>3</sup>

The drug development process (figure) includes dose finding<sup>4</sup> and combination building,<sup>5</sup> which often occurs in short 2-week phase 2a early bactericidal activity trials.<sup>6–8</sup> Phase 2b studies extend the treatment period to 2 months, to assess the initial safety and efficacy of the regimens, with the most promising regimens then moving on to pivotal phase 3 studies.<sup>9</sup> During this period, innovative approaches such as multi-arm and multistage designs can accelerate the drug development process by testing multiple drugs, doses, or combinations simultaneously.<sup>10,11</sup> Developments in the methodology of drug trials have further expanded phase 2b studies to include end-of-treatment and longer-term outcome data (ie, relapse), with such trials being designated as phase 2c.<sup>9</sup> Each phase 2 study design includes a measure of treatment response (such as time to first negative culture) or change in bacterial load as the primary study endpoint. This endpoint could be supplemented by linking the bacterial response directly to the pharmacokinetics of the drug in trial.<sup>12,13</sup>

Several promising new and potent drugs against tuberculosis are now available for evaluation, and the accelerated development of these agents requires improved methods that allow for rapid testing. Unfortunately, the evaluation process for these new agents is slow and demanding,<sup>14</sup> including the determination of not only the optimal dose but also the most effective and safe combination of partner drugs over long study durations.

The length, arduousness, and resource-intensiveness of the process can partly be explained in terms of the need to monitor treatment response (ie, viability of *M tuberculosis*) as the trial endpoint for many patients with tuberculosis over long study durations. For this reason, drug trials require substantial investment in setting up and managing high-containment laboratories (biosafety level 3) and a strong cadre of scientists who can work in that environment. Furthermore, the growth-based reference standard is imperfect, further complicating the implementation of such techniques.<sup>15</sup> Bacterial culture, especially *M tuberculosis* culture, is compromised by phenotypic lag, which is the failure or delay of growth on artificial media,<sup>16–18</sup> and competition with rapidly growing organisms found in the sputum. Despite optimal decontamination procedures, the proportion of samples lost to contamination increases as the treatment progresses,<sup>2</sup> which leads to frequent loss of the crucial final sample to contamination at the end of the treatment, effectively nullifying the contribution of the participant to the trial.<sup>1</sup> The diagnostic methods available for treatment monitoring and their challenges are summarised in tables 1 and 2.

## Developing and evaluating more effective biomarkers

Biomarkers hold great promise in overcoming many of the challenges of culture-based endpoints (ie, time to result and contamination with other bacteria), and serve as early indicators of the efficiency of tuberculosis within the first 2 months of treatment and within long-term patient outcomes.<sup>2,19</sup> Despite this potential of biomarkers to accelerate the evaluation of novel drugs and regimens against tuberculosis, a few have been evaluated systematically for their ability to directly or indirectly correlate with the effect of early-phase treatment or with long-term outcomes in people with tuberculosis.<sup>2,19</sup>

Lancet Microbe 2024

Published Online  
[https://doi.org/10.1016/S2666-5247\(24\)00085-5](https://doi.org/10.1016/S2666-5247(24)00085-5)

Division of Infection and Global Health, School of Medicine, University of St Andrews, St Andrews, UK

(Prof S H Gillespie DSc, W Sabiiti PhD); Global TB Program, Texas Children's Hospital and Baylor College of Medicine, Houston, TX, USA (A R DiNardo PhD, C Lange MD);

Department of Internal Medicine and Radboud Center for Infectious Diseases, Radboud University Medical Center, GA Nijmegen, Netherlands (A R DiNardo, R van Crevel PhD); FIND, Campus Biotech, Geneva, Switzerland (S B Georgioui PhD, M Kohli PhD, J B Tan MSc, M Ruhwald PhD); Division of Infectious Diseases and Tropical Medicine, University Hospital, Ludwig-Maximilians-Universität München (LMU), Munich, Germany

(U Panzner MSc, N Heinrich PhD); German Center for Infection Research (DZIF), Partner Site Munich, Munich, Germany (U Panzner, N Heinrich);

Fraunhofer Institute for Translational Medicine and Pharmacology ITMP, Immunology, Infection and Pandemic Research, Munich, Germany (N Heinrich); Division of Clinical Infectious Diseases, Research Center Borstel, Borstel, Germany (C Lange); German Center for Infection Research (DZIF), Clinical Tuberculosis Unit, Borstel, Germany (C Lange);

German Center for Infection Research (DZIF), Partner Site Hamburg-Lübeck-Borstel-Riems, Borstel, Germany (I Kontsevaya PhD); Department of Infectious Disease, Faculty of Medicine, Imperial College London, London, UK (I Kontsevaya); Janssen Global Public Health R&D, Janssen Pharmaceutica NV, Beerse, Belgium (N Hittel PhD);

Otsuka Novel Products, Munich, Germany (L J Stuyver PhD); Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine, University of Oxford, Oxford, UK (R van Crevel); Oxford University Clinical Research Unit, Ho Chi Minh City, Viet Nam (T N T Thuong PhD); Respiratory Medicine & International Health, University of Lübeck, Lübeck, Germany (J Heyckendorf PhD, C Lange); Clinic for Internal Medicine I, University Hospital Schleswig-Holstein (UKSH) Campus Kiel, Kiel, Germany (J Heyckendorf)

Correspondence to: Prof Stephen H Gillespie, Division of Infection and Global Health, School of Medicine, University of St Andrews, St Andrews KY16 9TF, UK [shg3@st-andrews.ac.uk](mailto:shg3@st-andrews.ac.uk)

For UNITE4TB see <https://www.unite4tb.org>

To tap into this potential and accelerate drug development, UNITE4TB, a Europe-wide European Union-Innovative Medicines Initiative (EU-IMI)-funded research and industry consortium mandated to accelerate the development of novel drugs against tuberculosis,<sup>20</sup> prioritised research on developing tools for monitoring tuberculosis treatment. The diversity of the potential biomarkers (which measure either the host response or the pathogen), sample types, and approaches for assay development and clinical assessment render comparisons of biomarker suitability and performance for assessing the effectiveness of drug treatment challenging. To solve this dilemma, optimal and minimal criteria should be defined to evaluate the design and performance of biomarker-based assays for monitoring tuberculosis treatment, especially in the context of trials for drugs against tuberculosis.

We, therefore, sought to define key specifications such as intended use, performance, operational characteristics, and pricing in a consensus-based target product profile (TPP), to guide the development of biomarkers in alignment with the specific needs of tuberculosis drug trialists. Of note, the assay requirements for monitoring treatment efficacy in clinical drug trials differ from those for monitoring treatment efficacy in clinical practice. This work is, therefore, complementary to the recently published WHO TPPs for programmatic management of tuberculosis in high-burden settings.<sup>14,21</sup> In formulating the criteria for TPPs, we considered the possibility that the new developments might challenge current concepts regarding the design and performance of biomarker-based assays and therefore intentionally kept many of the criteria broad, to accommodate for future innovations while ensuring that the outlined characteristics would meet the needs of drug developers, clinical trialists, and reference laboratories involved in drug development and evaluation.

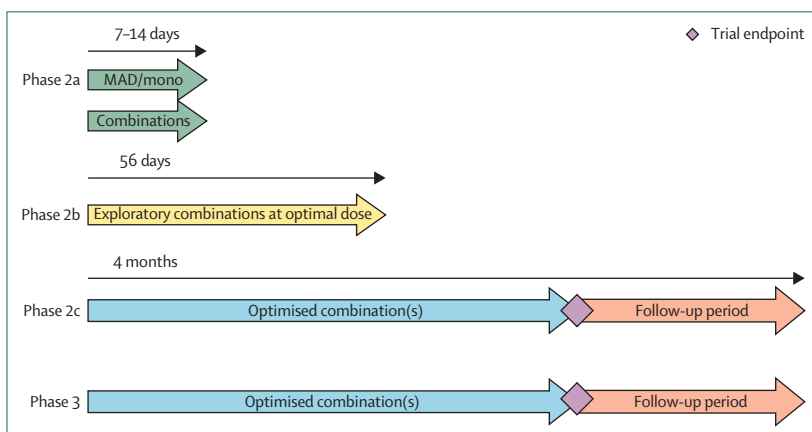
Although many studies have collected quantitative data by means of a range of techniques such as measurement of colony forming units,<sup>7,8</sup> Mycobacteria Growth Indicator Tube (MGIT) time to positivity,<sup>4,5</sup> Xpert,<sup>22</sup> or molecular

bacterial load assay,<sup>2</sup> this Review is the first attempt to draft a TPP to quantify treatment response in clinical trial settings. Here, we consolidate the evidence that can allow the community of trialists and diagnostics developers in the field of tuberculosis to engage in debating the next steps. The TPP can enable assay developers to adjust their assays to meet the needs of tuberculosis drug developers, while enabling the drug development community to shorten the duration of clinical trials and recruit fewer people with tuberculosis to achieve statistically significant results, thus reducing the costs of advancing novel drugs and regimens rapidly.

### Search strategy and selection criteria

To focus our efforts on the drug development process, we brought together a team of individuals representing academic institutions, research institutes, small-sized and medium-sized enterprises, public organisations, and pharmaceutical companies within the biomarkers work package of the UNITE4TB consortium. As part of the development of TPP drafts, the trials consortium discussed the overarching purpose of the assays to monitor tuberculosis treatment, the aims to be met, and the necessary operational and performance characteristics of tools for tuberculosis drug development. The first task was to review the relevant literature and scope the landscape, which was achieved by means of literature searches in PubMed for articles published between January, 1990, and June, 2023, with the following search terms: “tuberculosis”, “clinical trials”, “monitoring”, “biomarkers”, and “treatment response”. No language filters were applied. We also reached out to academic colleagues and those working in commercial entities actively engaged in biomarker development and evaluation. In addition, crucial discussions were held with trialists, statisticians, and regulatory representatives, to arrive at the optimal performance parameters taking into consideration the current reference standard (ie, growth-based culture methods).

For each category, we outlined the desirable minimal and optimal criteria for acceptable tuberculosis treatment monitoring tools. In parallel with the development of quantitative bacterial biomarkers, the number of immune and inflammatory response assays under development has substantially increased, and some of them are starting to come to the market.<sup>22,23</sup> The aim of these assays is to identify a measure of cure on the basis of normalisation of the immune and inflammatory status. The treatment response biomarkers that operate by this methodology and the relevant TPPs will be reviewed in a separate paper. We concluded, therefore, that there was a need to develop two TPPs, for tests that quantitatively assess viable tuberculosis bacteria, mainly for early-stage tuberculosis treatment response, and assess the late-stage tuberculosis treatment response that predicts the clinical long-term outcome of relapse-free cure in the context of trials for drugs against tuberculosis.



**Figure:** Tuberculosis drug development process

An overview of the timelines and assay methodology in the tuberculosis drug development process.

## Consultation and sense-checking

Draft 0 of the two TPP documents was distributed to 28 select individuals from the UNITE4TB consortium with experience in conducting tuberculosis drug trials or having backgrounds in tuberculosis biomarker research, or both, including individuals affiliated with ongoing tuberculosis efforts at Radboud University Medical Center (Radboudumc; Nijmegen, Netherlands), Ludwig-Maximilians-Universität München (LMU; Munich, Germany), Helmholtz Zentrum München Deutsches Forschungszentrum für Gesundheit und Umwelt (Neuherberg, Germany), KNCV Tuberculosis Foundation (KNCV; The Haag, Netherlands), Critical Path Institute Ltd. (C-PATH; Dublin, Ireland), The University of Porto (Porto, Portugal), The National University of Singapore (Queenstown, Singapore), GlaxoSmithKline Investigación y Desarrollo S L (GSK; Madrid, Spain), The Janssen Pharmaceutical Companies of Johnson & Johnson (Beerse, Belgium), Otsuka Novel Products (Munich, Germany), Oxford University Clinical Research Unit (OUCRU; Ho Chi Minh, Viet Nam), Research Center Borstel (Sülfeld, Germany), TASK Clinical Research Centre (Cape Town, South Africa), Instituto de Saude Publica da Universidade do Porto (ISPUP; Porto, Portugal), French National Research Institute for Sustainable Development (IRD; Marseille, France), Vita-Salute San Raffaele University (UniSR; Milan, Italy), and University College London (UCL; London, UK). The purpose of this initial internal review was to gather feedback on the outlined minimal and optimal performance criteria of biomarker-based assays for monitoring tuberculosis treatment in each TPP. All the feedback received was considered by the work package members, and the comments selected following several rounds of revision were incorporated into the final TPP documents. By the final revision, when a consensus was not reached between the work package members and internal reviewers on any performance criterion or metric, additional feedback and notes to explain any differences were added to the TPPs drafted.

Once all internal feedback was incorporated into the two TPPs, we selected an external panel of nine experts to provide another round of feedback on the draft documents through an online survey, with the goal of reaching a field consensus. Participants without declared or perceived conflicts of interest were selected for this exercise, while ensuring a broad representation of stakeholders with experience in tuberculosis biomarker research, including physicians and clinicians with experience in running treatment trials, public health practitioners with WHO and other policy experience, and experts from relevant settings with high disease burden. The final consensus was achieved through guided discussions with these external stakeholders on the available TPP drafts and the incorporation of survey results that defined needs with respect to the operational and performance characteristics.

Challenges	
<b>Microbiological samples</b>	
Overnight sputum*	High contamination in ambient temperature; difficulty in obtaining specimen(s); biohazard risk.
<b>Microbiological methods</b>	
CFU on solid agar†	Semi-quantitative; prone to contamination; absolute count reduced by the effect of selective medium and decontamination methods; long time to result; high day-to-day variation; requirement of technical expertise to perform CFU counts; BSL <sub>3</sub> laboratory facilities.
Becton Dickinson BACTEC MGIT‡	Semi-quantitative; prone to contamination; absolute count reduced by the effect of selective medium and decontamination methods; long time to result; correlation to CFU changes over the treatment, suggesting the need for assessment of different bacterial populations.
<b>Primary endpoint</b>	
Daily change TTP (day 0-14)	Compromised by sample loss: a contaminated sample results in the loss of a data point.
<b>Secondary endpoints</b>	
EBA (day 0-2)	Compromised by sample loss: a contaminated sample results in the loss of a data point.
EBA (day 2-14)	
EBA (day 0-7)	
EBA (day 7-14)	
BSL <sub>3</sub> =biosafety level 3. CFU=colony-forming unit. EBA=early bactericidal activity. MGIT=Mycobacteria Growth Indicator Tube. TTP=time to positivity. *Spot samples are a recognised alternative. †Lowenstein-Jensen medium was used as a diagnostic and antibiotic-containing agar-based media for quantitative counts. ‡Manufactured product incorporating automated growth detection.	
<b>Table 1: Summary of samples, methodologies, and endpoints used to measure early-stage outcomes in trials for drugs against tuberculosis</b>	

## Characteristics of the TPPs

During preparation of the TPPs, the following points were addressed: scope or intended use, pricing, expected performance, and operational characteristics of biomarker-based assays for tuberculosis drug development. Clinical trials of novel tuberculosis agents are usually performed in a high-burden setting in a low-income or middle-income country, which has a substantial effect on the choice and application of the biomarker tests; thus, this consideration was included. Both optimal and minimal criteria were defined for each characteristic of the proposed tests, to prevent perfect from being the enemy of good (or practical).

The TPP outlines the minimal criteria defined for a listed characteristic, considered the lowest acceptable specifications that could be applied for that characteristic in a clinical trial. Similarly, the optimal criteria were defined as the ideal specifications to move the field forward. In developing these criteria, we intended that the optimal and minimal criteria would define the range of acceptable assay performance and operational characteristics. Although assay developers would be expected to design their assays to meet the optimal criteria listed for each specification, an assay would still be expected to hold value even without meeting these criteria under the condition that the assay presents sufficient advantages, especially when compared with the current reference standard. Thus, we underlined the importance of focused research to replace or reduce the dependence on culture-based endpoints, as these endpoints are compromised by

		Challenges
Culture-based endpoints in phases 2b and 2c	Rate of change in MGIT TTP over the treatment (usually from day 0 to 56 or later).	Compromised by sample loss due to contamination and the consequent loss of data points; high day-to-day variation; no information on the effect of the treatment, after the culture converts to negative (goes below the limit of detection of MGIT).
	Time on treatment to sustained negative MGIT culture (sputum culture conversion, as defined by EMA <sup>14</sup> ).	Compromised by sample loss: a contaminated sample results in the loss of a data point; inability of patients to produce sputum later in the treatment; does not reflect treatment efficacy after the culture converts to negative; several ways to define this parameter, making inter-trial comparisons difficult.
	Proportion of patients with negative cultures after 2 or more months of the treatment.	Compromised by sample loss: a contaminated sample results in the loss of a data point (entire patient); binary endpoint, leading to larger sample sizes.
	Sustained cure vs relapse.	Outcomes occur late after randomisation and require long-term follow-up with patients, causing issues such as loss to follow-up and low clarity in dealing with intercurrent, mostly unrelated events such as deaths or premature terminations of study participation.
Clinical endpoints: long-term cure vs unfavourable outcome: duration ranging from phase 2c to phase 3	Relapse (bacteriologically confirmed).	Relapse might be defined as the return of microbiologically confirmed tuberculosis with the same strain that caused the first episode of the disease, identified on the basis of appropriate typing methods (genome sequencing); distinguishing relapse from new infection might be difficult <sup>15</sup> —EMA suggests that the case should then be counted as a relapse in the primary analysis of efficacy (re-infection would not be seen as an unfavourable outcome owing to unsuccessful treatment). However, this process could lead to an underestimation of the difference between the two treatment arms.
	Treatment failure (bacteriologically confirmed).	Treatment failure might be defined as deficient on-treatment sputum culture conversion at a pre-specified timepoint after the commencement of therapy or at the end of the treatment. The challenge is to distinguish persistent positivity (failure) from isolated single-positive cultures that often occur in patients responding well to treatment, but do not indicate failure of treatment; here, clinical observation on the resolution of signs and symptoms should be used. <sup>15,16</sup> Separately, patients whose symptoms and possibly radiological evidence indicate a recurrence with or without microbiological confirmation might be judged as a failure of the clinical treatment.

EMA=European Medicines Agency. MGIT=Mycobacteria Growth Indicator Tube. TTP=time to positivity.

**Table 2: Summary of the samples, methodologies, and endpoints applied to late-stage outcomes**

sample loss due to contamination with faster growing organisms, which results in the loss of data points and wastage of the trialists' and patients' investments in the trial. Moreover, many people with tuberculosis are unable to produce a satisfactory sputum later in the treatment. Once a person with tuberculosis has culture-converted, the inability to produce a satisfactory sputum later in the treatment might not represent cure, and this limitation, together with the variability of defining culture positivity, can make inter-trial comparisons difficult.

Although sputum is the main sample used to define treatment outcome, the fact that some people, including women and most children, are unable to produce a satisfactory sample is less than ideal. Similarly, identification of a satisfactory sample for the evaluation of extrapulmonary tuberculosis, such as tuberculosis meningitis or osteomyelitis, is difficult. Reliance on sputum excludes those crucial populations from trials that have an urgent need for evidence of effective treatment regimens. Progress in this area is being achieved through the use of alternative specimen types for diagnosis, such as blood-based, stool-based, or urine-based tools.<sup>24,25</sup>

### Early-phase biomarkers

In the case of assays suitable for early-stage clinical trials for drugs against tuberculosis (table 3), the proposed criteria correlate with those accepted by regulators as phase 2b trial endpoints (table 1). In the context of phase 2 trials for the development of antimycobacterial compounds, the goal of the assay is to accurately quantify or reflect the number of viable bacilli in a patient sample as

an indication of the efficacy of a drug or drug regimen in a trial. Although sputum samples would be an acceptable sample type, a non-invasive clinical specimen was seen as preferable, given the difficulty in obtaining sputum from many patients, especially those from clinically vulnerable populations (eg, people living with HIV and children). The setting, target users, and target population, along with most of the outlined operational characteristics, would be consistent with the clinical laboratories and laboratory conditions associated with drug trials.

For pricing considerations, the price of the current reference standard (ie, MGIT) was considered, although in the context of clinical trials, a notably higher price was recognised as acceptable under the condition that time would be saved in the conduct of the study (eg, lower sample size) or the non-responders to treatment would be identified at earlier timepoints. However, depending on the performance and operational characteristics, the potential market can be expanded to the settings of clinical trials as well. The expected performance metrics were kept broad to accommodate the diversity of the potential assays, while ensuring that the test accurately quantifies or reflects the number of viable bacilli in patient samples and their decline with effective treatment, with a dynamic range reflecting that found in the target sample. The assay would be expected to show results with high intra-laboratory and inter-laboratory agreement.

A concordance correlation coefficient against the reference standard was recommended to be determined for quantitative assays. Ultimately, assays would be expected to

	Optimal	Minimal	Explanation
Intended use	In-vitro (laboratory) research-use-only test for detecting viability markers of <i>Mycobacterium tuberculosis</i> , to aid in the assessment of early efficacy during phase 2 trials for the development of antimycobacterial compounds.		This definition is open to any potential microbiological solution by means of any method and would include but not be restricted to previously published methods for assessed viability.
Goal of test	A marker or combination of markers to accurately quantify or reflect the number of viable bacilli in a sample obtained from a patient with tuberculosis, to indicate the efficacy of a drug or drug regimen in a trial.		
Target population	All patients with active pulmonary tuberculosis evaluated in an early-phase trial.	Adults with drug-susceptible pulmonary tuberculosis.	
Target user of test	Laboratory scientists with the ability to perform low-complexity to moderate-complexity assays.	Laboratory scientists with the ability to perform high-complexity assays.	For the definitions of low-complexity and moderate-complexity assays, refer to the terminology defined by WHO for tuberculosis tests. An example of a low-complexity assay would be Cepheid Xpert MTB/RIF Ultra; that of a moderate-complexity assay would be Roche Cobas MTB-RIF/INH; and that of a high-complexity assay would be assays similar to or more complex than line probe assays, including sequencing-based technologies. As the assays will be used in the context of a drug trial, the target users are expected to have training and experience appropriate for work in clinical laboratories.
Setting (health system level)	Well equipped laboratory with good clinical practice and EQA standards. Future development might include low-complexity assays that can be performed in high-burden remote settings.		Assays are expected to be used in clinical laboratories, as appropriate for drug trials, although the potential market can be expanded to the settings of clinical trials depending on the performance and operational characteristics.
Pricing of individual test (reagent costs only; at scale; ex-works)	<US\$20 per specimen.	<\$1000 per specimen.	Price corresponds to the cost of testing per specimen. Tests that directly replace quantitation of viable bacilli would ideally be in the lower part of this range.
Capital costs of instrumentation	<\$10 000	<\$100 000	Tests that directly replace quantitation of viable bacilli would ideally be in the lower part of this range. However, innovative solutions using more expensive capital equipment could prove to be cost-effective in the context of clinical trials but would need to be justified by means of economic costing and operational studies.
Tests to enumerate bacilli or replace quantitative culture for phase 2a trials and eventually phase 2b trials	The marker or combination of markers should quantify or reflect the number of viable bacilli in a clinical sample.		Given the range of biomarkers for monitoring the treatment of tuberculosis, in addition to the variability in clinical samples and imperfect reference standards, no specific performance metric is given. Developers and manufacturers should have data (for a sufficient number of samples and timepoints) in relation to EBA or culture, or both, to confirm that the test quantifies or reflects the number of viable bacilli in patient samples and shows decline over time of effective treatment.
Dynamic range of detection	For tests of viable bacteria, the dynamic range should mirror that found in the target samples from patients with tuberculosis. For example, for a sputum test, the range should be between $10^2$ CFU/mL and $10^7$ CFU/mL.		A concordance correlation coefficient should be determined for quantitative assays to assess the performance against a reference standard (EBA or culture, or both), as appropriate for the assay.
Repeatability and reproducibility	The marker or combination of markers should generate highly similar results when a sample is tested on multiple occasions in the same laboratory or in different laboratories (inter-laboratory agreement).		The methods and results of reproducibility determination should be acceptable to regulators.
Sample type	Non-invasive clinical specimens.	Sputum or non-sputum samples that are not more complex to obtain than sputum, or both.	Given that sputum samples can be difficult to obtain for specific populations (eg, people living with HIV and children), in addition to patients with tuberculosis, the test would ideally be performed in easy-to-obtain non-sputum samples, including blood, stool, exhaled breath, and urine samples.
Sample stability	The analyte should be stable for at least 8 h upon storage at 25°C and for up to 48 h upon storage at 5°C and tolerate at least two freeze and thaw cycles.	The analyte should be stable for same-day analysis upon storage at 5°C.	
Sample preparation and assay processing (total steps)	Integrated sample preparation and detection in a closed system with minimal technical input.	Fewer number of steps required for sample preparation and detection.	
Time to result	<24 h, with automated result interpretation.	<7 days	<7 days would still be considered a significant improvement over growth-based methods for early-phase studies. The optimal characteristic considers the time from patient sampling to the final result.
Result output	Automated result and interpretation. The results screen should be integrated into the technology and the ability to save and print the results should also be included; the device should have a commonly used port (eg, USB or USB-c) and the solution should be integrated with the existing laboratory management information system.	Results might require interpretation by means of software but should be available in 1 day. The results screen should be integrated into the technology and the ability to save the results should be included; the device should have a commonly used interface port (eg, USB or USB-c).	These criteria were defined considering the fact that laboratories at clinical trial sites are expected to have back-up generators for crucial pieces of equipment.

(Table 3 continues on next page)



	Optimal	Minimal	Explanation
(Continued from previous page)			
Power requirements	Standard operating currents with a built-in UPS for utilisation in locations with variable power. In addition, battery or solar operation for locations with power supply disruptions.	Standard operating currents with a built-in UPS for utilisation in locations with variable power.	Although the test is anticipated to be used in a clinical laboratory with sufficient infrastructure, including UPS, a battery or solar-operated device would be ideal for operation in any setting.
Maintenance	Preventive maintenance at 1 year or >1000 samples; include maintenance alert. Mean time to failure of at least 18 months.	Availability of included supplies and maintenance teams to provide technical assistance as needed.	Some envisioned solutions might not require maintenance; the most important element will be the availability of supplies and maintenance teams to provide technical assistance as needed.
Calibration	Calibration should be done on site and the test should include all the necessary reagents and equipment.		Calibration should be simple to perform on site.
Operating temperature	Between +3°C and +40°C at 70–90% relative humidity.	Between +4°C and +30°C at 70–90% relative humidity.	
Spare supplies (not included in the kit)	None		
Internal quality control	Internal full-process positive and negative controls along with result standardisation across laboratories.		In addition to EQA.
Training and education needs	<1 day	3–5 days	Training times as appropriate for clinical laboratory scientists.

CFU=colony-forming unit. EBA=early bactericidal activity. EQA=external quality assessment. UPS=uninterrupted power supply. USB=universal serial bus.

**Table 3: Summary of the proposed criteria for a target product profile that can be used for assays that assess the number of viable tuberculosis bacteria in an early-stage tuberculosis clinical trial**

show advantages over the current culture-based reference standard in terms of the quantitative assessment of viable tuberculosis bacteria over the course of patient treatment, with focus on a quantitative output in correlation with treatment efficacy and a quicker time to result.

### Late-phase biomarkers

Developing a biomarker-based assay for use in phase 2c and pivotal phase 3 trials is most likely to be more challenging than developing one for use in the earlier stages of drug development, owing to the composite nature of the unfavourable outcome that serves as the study endpoint (table 4).<sup>1,26,27</sup> Traditionally, a combined endpoint is used in evaluative studies, which includes a marker of successful treatment (ie, culture negativity at the end of the treatment) and the absence of recurrent infection with the initial infecting strain (defined as relapse). Within this definition, a degree of complexity exists, which should be addressed by distinguishing recurrence from relapse by sequencing the initial and recurrent strains obtained from the patient and applying the predefined criteria, which requires long-term storage of these isolates.<sup>28–30</sup>

To ensure the suitability of assays for late-stage clinical trials for drugs against tuberculosis (table 4), these assays were expected to accurately quantify or reflect the number of viable bacilli in a sample, to provide an indication of the efficacy of a drug or drug regimen in a trial (ie, absence of live organisms, as a definition of cure). Alternatively, repeated measures of viable count could be used to predict the outcome by means of methods that include mathematical modelling. Given the absence of methodologies that can predict the outcome, our criteria were intentionally kept broad to allow for technologies that are not yet known.

Optimally, however, the assays should predict a relapse-free cure, replacing sputum culture conversion endpoints. As for early-stage assays, sputum samples would be an acceptable sample type, with non-invasive clinical specimens being more preferable, and the setting, target users, target population, and many of the outlined operational characteristics consistent with the settings associated with the drug trials.

### Price and cost

For pricing considerations, the price of the current reference standard (ie, MGIT) was again taken as a baseline. Quantifying and reflecting the number of viable bacilli were assumed to be cheaper methodologies to predict relapse-free cure. The expected performance metrics between biomarker-based assays enumerating bacilli and those predicting long-term outcomes would differ, with performance parameters of quantitative assays matching those of assays suitable for earlier-stage trials. Biomarker-based assays to identify the long-term cure in comparison with relapse in phase 2 or phase 3 trials would optimally show a good prediction of treatment outcome at earlier stages of treatment, as compared with culture in a per-trial arm analysis, with the minimal TPP criteria defined as the measurement at the end of the 12-month patient observation period being at least equivalent to culture and possibly having other advantages.<sup>31</sup> Otherwise, performance parameters were again kept broad to accommodate the diversity of potential assays, with the expectation that the assays would show advantages over the current culture-based reference standard in terms of prediction of treatment efficacy and show a shorter time to result than the current culture methods. Innovative solutions using

	Optimal	Minimal	Explanations
<b>Goal of test</b>	A marker or combination of markers to accurately predict relapse-free cure, which would be observed for up to 12–18 months after randomisation, or as required by regulatory agencies*.	A marker or combination of markers to quantify the number of viable bacilli in a sample obtained from a patient with tuberculosis, to be used as a replacement for sputum culture-conversion endpoints or endpoints that describe the change in bacterial load†.	The marker or combination of markers should address the limitations of the current gold standard technology mentioned in table 2, and in case of the optimal characteristics, predict relapse-free cure at an earlier timepoint than the microbiological reference. If the measurement of cure is not earlier, there would be a need for other operational advantages over culture.
<b>Target population</b>	All patients with active pulmonary tuberculosis evaluated in phase 2b, 2c, and 3 trials.	Adults with active, drug-susceptible pulmonary tuberculosis evaluated in phase 2b, 2c, and 3 trials.	
<b>Target user of test</b>	Laboratory scientists with the ability to perform low-complexity to moderate-complexity assays.	Laboratory scientists with the ability to perform high-complexity assays.	For definitions of low-complexity and moderate-complexity assays, refer to the terminology defined by WHO for tuberculosis tests. An example of a low-complexity assay would be Cepheid Xpert MTB/RIF Ultra, that of a moderate-complexity assay would be Roche Cobas MTB-RIF/INH, and that of a high-complexity assay would be assays that are similar to or more complex than line probe assays, including sequencing-based technologies. As the assays will be used in the context of a drug trial, target users are expected to have training and experience, as appropriate for work in clinical laboratories.
<b>Setting (health system level)</b>	Well equipped laboratory with good clinical practice and EQA standards.		Assays are expected to be used in clinical laboratories, as appropriate for drug trialling. Future development might simplify the methodology and allow for fewer demanding requirements.
<b>Pricing of individual test (reagent costs only; at scale; ex-works)</b>	<US\$20 per specimen	<\$1000 per specimen	Price corresponds to the cost of testing per specimen. Tests that directly replace quantitation of viable bacilli would ideally be in the lower part of this range. Innovative solutions using more expensive tests could prove to be cost-effective in the context of clinical trials, by shortening the time to result of a trial, allowing for more effective design of adaptive trials, reducing the number of non-evaluable patients, or otherwise lowering the sample size or time required, so their use would need to be justified by economic costing models.
<b>Capital costs of instrumentation</b>	<\$10 000	<\$100 000	Tests that directly replace quantitation of viable bacilli would ideally be in the lower part of this range. Innovative solutions using more expensive capital equipment could prove to be cost-effective in the context of clinical trials but would need to be justified in terms of economic costing and operational studies.
<b>Enumeration of bacilli or replacement of quantitative culture</b>	The marker or combination of markers should quantify the number of viable bacilli or other bacterial signatures in a clinical sample‡.		Given the range of biomarkers for monitoring of tuberculosis treatment, in addition to the variability in clinical samples and imperfect reference standards, no specific performance metric is given. Developers and manufacturers should have data (for sufficient number of samples and timepoints) in relation to the culture, to confirm that the test quantifies the number of viable bacilli in samples obtained from patients with tuberculosis and shows decline over time of effective treatment.
<b>Dynamic range of detection</b>	For tests of viable bacteria, the dynamic range should mirror that found in the target samples obtained from patients with tuberculosis. For example, for a sputum test, the range should be between $10^1$ and $10^7$ CFU/mL <sup>-1</sup> .		A within-specimen $R^2$ should be determined for quantitative assays, to define the performance against the reference standard (quantitative culture), as appropriate for the assay.
<b>Repeatability and reproducibility</b>	The marker or combination of markers should generate highly similar results when a sample is tested on multiple occasions in the same laboratory or in different laboratories (inter-laboratory agreement).		
<b>Timing of measurement or prediction</b>	Measurement during treatment or at the end of the treatment. Ideally, good predictive capability should be substantially earlier than at the end of the treatment.	Measurement at the end of the 12-month observation period post-randomisation.	If measurement is only attainable or predictive of cure late during the patient treatment, there should be other advantages to the use of the measurement that add benefit over current growth-based methods.

(Table 4 continues on next page)

	Optimal	Minimal	Explanations
(Continued from previous page)			
<b>Surrogacy on trial level, to predict long-term outcome</b>	The marker should predict the long-term outcome that is substantially superior to culture in a per-trial arm analysis.	The marker should predict a long-term outcome that is at least as good as culture in a per-trial arm analysis.	The reference standard for long-term cure vs failure or relapse needs to be defined in accordance with regulatory (EMA or FDA) guidance, requiring isolate sequencing and confirmation of relapse by matching the sequence with that of the initial isolate.
<b>Sample type</b>	Non-invasive clinical specimens (including blood).	Sputum or non-sputum samples that are not more complex to obtain than sputum samples (eg, urine or blood), or both.	Given that sputum samples can be difficult to obtain for specific populations (eg, people living with HIV and children) and patients with tuberculosis who are undergoing treatment, the test would ideally be performed in easy-to-obtain non-sputum samples, such as aerosol and urine.
<b>Sample stability</b>	Analyte should be stable for at least 8 h upon storage at 25°C and for up to 48 h upon storage at 5°C and tolerate at least two freeze and thaw cycles.	Analyte should be stable for same-day analysis upon storage at 5°C.	
<b>Sample preparation and assay processing (total steps)</b>	Integrated sample preparation and detection in a closed system with minimal technical input.	Fewer number of steps required for sample preparation and detection.	
<b>Time to result</b>	<24 h, with automated result interpretation.	<7 days.	<7 days would still be considered a substantial improvement over growth-based methods for early-phase studies. The optimal characteristic considers the time from patient sampling to result.
<b>Result output</b>	Automated result and interpretation. The results screen should be integrated into the technology and the ability to save and print results should be included; the device should have a commonly used port (eg, USB or USB-c) and the solution should be integrated with the existing laboratory management information system.	Results might require interpretation through software but should be available in 1 day. The results screen should be integrated into the technology and the ability to save results should be included; the device should have a commonly used interface port (eg, USB or USB-c).	These criteria were defined with the expectation that the laboratories at the sites of the clinical trials have back-up generators for crucial pieces of equipment. Interfacing and data transfer to laboratory information management systems for electronic data transfer into the trial database is desirable.
<b>Power requirements</b>	Standard operating currents with a built-in UPS for utilisation in locations with variable power. In addition, battery or solar operation for locations with power supply disruptions.	Standard operating currents with a built-in UPS for utilisation in locations with variable power.	Although the test is anticipated to be used in a centralised clinical laboratory with sufficient infrastructure, including uninterrupted power, a battery or solar-operated device would be ideal for operation in any setting.
<b>Maintenance</b>	Preventive maintenance at 1 year or >1000 samples; should include maintenance alert. A mean time to failure of at least 18 months.	Availability of spare supplies and maintenance teams to provide technical assistance, as needed.	Some envisioned solutions might not require maintenance; the most important element will be the availability of supplies and maintenance teams to provide technical assistance, as needed.
<b>Calibration</b>	Calibration should be done on site and the test should include all the necessary reagents and equipment.		Calibration should be simple to perform on site.
<b>Operating temperature</b>	Between +3°C and +40°C at 70–90% relative humidity.	Between +4°C and +30°C at 70–90% relative humidity.	
<b>Spare supplies (not included in the kit)</b>	None		
<b>Internal quality control</b>	Internal full-process positive and negative controls, along with results standardisation across laboratories.		In addition to EQA.
<b>Training and education needs</b>	<1 day	3–5 days	Training times, as appropriate for clinical laboratory technicians.

CFU=colony-forming unit. EMA=European Medicines Agency. EQA=external quality assessment. FDA=US Food and Drug Administration. UPS=uninterrupted power supply. USB=universal serial bus. \*With the availability of shorter and more potent regimens, this time period is likely to be shortened and this characteristic will need to be updated. †Removing the need for culture would substantially simplify the trial practice. ‡The definition of “other marker” is open to any biologically plausible methodology for which adequate data are available.

**Table 4: Summary of the proposed criteria for use of a target product profile in late-stage clinical trials, including phase 2b, phase 2c, and pivotal phase 3 clinical trials, on tuberculosis**

expensive tests such as imaging could be cost-effective in the context of a clinical trial, but their use would need to be justified by means of economic costing models and operational studies.

### Regulatory considerations

To replace current tuberculosis treatment regimens with improved, shorter, effective, all-oral regimens incorporating new drug compounds, the costs should be lowered and the clinical evaluation of new drugs and drug regimens

accelerated. Although some progress has been made in this field following the introduction of new tuberculosis diagnostics, including Xpert MTB/RIF Ultra (Cepheid, Sunnyvale, CA, USA)<sup>32</sup> and BACTEC MGIT (BD MGIT, Becton Dickinson, Franklin Lakes, NJ, USA),<sup>15</sup> these technologies have not noticeably decreased the time or resources needed to conduct the drug trials, and the tuberculosis treatment monitoring methods currently in use have modestly improved the conduct of clinical trials, when compared with those of trials conducted 50 years



ago.<sup>33</sup> Of note, the requirements for monitoring treatment response in clinical trials are different from those in clinical practice; thus, the TPPs presented in this paper should complement those recently published by WHO.<sup>14,21</sup>

To accelerate drug development and offer improved therapeutic options to people with tuberculosis, there is a need to move beyond the accepted reference standard biomarker-based assay for tuberculosis detection and treatment monitoring—ie, sputum culture conversion. Culture has the advantage of being recognised as a US Food and Drug Administration-regulated technology with an in-vitro diagnostic use label (eg, bioMérieux BacT/ALERT 3D–BD MGIT–Thermo Scientific VersaTREK). Culture conversion can serve as a surrogate endpoint, a marker that itself does not provide a direct measurement of the clinical benefit but either predicts or is reasonably likely to predict the clinical benefit, in clinical trials.<sup>15</sup> Thus, the increase in time to positivity is regularly used as a surrogate of the bacterial load.<sup>26,34,35</sup> Similarly, a decline in the signal from DNA-based PCR methodologies such as Xpert MTB/RIF Ultra correlates with the bacterial load,<sup>36</sup> although the relationship between the bacterial count and the signal measured by means of PCR and culture is known to diverge as the tuberculosis treatment progresses.<sup>37,38</sup>

The absence of an unequivocal definition of cure for people with tuberculosis is the biggest challenge in developing a suitable biomarker for monitoring treatment in tuberculosis trials. Although methods to quantify bacterial load in early-stage trials are straightforward and several mathematical methods are available to analyse these data,<sup>13,39</sup> for pivotal phase 3 trials, bacteriological cure is defined differently, relying on culture negativity, while disregarding the fact that relapse can occur at any time after the completion of therapy.<sup>1,26,27</sup> To address this discrepancy, a post-treatment period has been defined, although any isolate should be sequenced to differentiate relapse from re-infection, and mixed infection with another more resistant strain can also preclude accurate result interpretation.<sup>30,40</sup> Similarly, in case an unequivocal definition of the immunological signature of cure is not available for host-dependent biomarkers, relying upon a pattern of markers associated with cure might provide a more likely probability of a favourable outcome.<sup>22,23</sup> Such an approach has not yet been recognised by regulatory authorities.

## Conclusions and future work

The TPPs that we have developed and presented for biomarker-based assays that can be used to monitor response to experimental therapy in drug trials are intended to stimulate discussions in the clinical trial community, provide a better picture of the tools necessary to improve trialling methodology, and ultimately accelerate regimen development and evaluation. Notably, suitable biomarker-based assays meeting the TPP criteria are anticipated to at least match the current standard of serial measurements of MGIT time to positivity, to monitor the response to therapy, considering that these methods are prone to contamination with

oral bacteria that cause false positive results and prevent tuberculosis detection. Thus, a major goal of a new biomarker could be to obviate the need for sputum culture. Over the past 20 years, several technologies have been proposed for this purpose, including serial measurements of specific mRNA, 16S ribosomal RNA, or mycobacterial lipoarabinomannan.<sup>2,27–29,41,42</sup> Some of these techniques are in the late stages of development and are undergoing regulatory trials for tuberculosis treatment monitoring.

These TPPs reflect the current perspectives of diagnosticians, clinical trialists, microbiologists, and others in setting aspirational criteria for monitoring the effect of tuberculosis treatment and accelerating clinical trials for drugs against tuberculosis. In addition to considering the valuable inputs from experts in this field, we strove to keep some performance criteria sufficiently broad to remain mindful of current and future inventions and developments that we might not be aware of currently. Ultimately, the two TPPs (suitable for early-stage and late-stage drug trials) were closely aligned with one another with regard to the target population (ie, drug trial participants with pulmonary tuberculosis), target users and setting (laboratory scientists in well equipped laboratories), pricing, repeatability and reproducibility, dynamic range, sample type, and operational performance characteristics as suitable for clinical laboratories participating in trials for drugs against tuberculosis. Notably, the TPPs differed in their ultimate goal, with the optimal criterion for the TPP for late-stage tuberculosis treatment response markers being an accurate prediction of relapse-free cure, whereas that for early-stage tuberculosis treatment response markers being accurate quantification of the number of viable bacilli in a sample obtained from the participant.

These TPPs represent consensus criteria that could be used to guide and de-risk the development of biomarker-based assays suitable for phase 2 and 3 clinical drug trials. Tools meeting the outlined minimal and optimal criteria would be expected to identify regimens with higher sterilising activity and lower probability of treatment failure or relapse in people with tuberculosis, thus revolutionising the conduct of trials for drugs against tuberculosis. We believe that these TPP criteria will be useful to diagnostic developers and clinical trialists, stimulating innovation and further research leading to the launch of new and improved tools for the development of drugs against tuberculosis.

### Contributors

All authors contributed to the development of the TPPs. SHG wrote the manuscript. SHG, ARD, and IK prepared the manuscript for publication. All authors had full access to all the data in the study and had final responsibility for the decision to submit for publication. All authors have seen and approved the final manuscript.

### Declaration of interests

SHG and WS developed the tuberculosis bacterial load assay and are collaborating with LifeArc to bring the assay to market. IK is participating in studies evaluating the feasibility of the TB22 signature developed by her research group as a treatment monitoring tool. JH holds patent shares for the TB22 signature. LJS is an employee of Johnson & Johnson and reports employee stock and stock options from

Johnson & Johnson. NHe is an employee of the University Hospital (LMU) and reports grants to the institute from Beckman Coulter, European and Developing Countries Clinical Trials Partnership (EDCTP; EU Horizon 2020), and DZIF for research and evaluation of new tuberculosis diagnostics. NHi reports grants or contract and consulting fees from Otsuka Novel Product Muenich (ONPG) and having detailed information on the LAM sputum biomarker. RvC reports Integrative omics and aspirin response to define the Therapeutic targets for Pediatric TBM grant (R01AI165721) from the National Institutes of Health and participating in the Data Safety Monitoring Board for three randomised controlled trials focused on tuberculosis meningitis. SBG, MK, MR, and JBT are employees of FIND, the global alliance for diagnostics, a not-for-profit foundation that supports the evaluation of publicly prioritised tuberculosis assays and implementation of WHO-approved (guidance and prequalification) assays using donor grants. FIND has product evaluation agreements with several private-sector companies that design diagnostics for tuberculosis and other diseases. These agreements strictly define FIND's independence and neutrality with regard to these private-sector companies. SBG and MR are also members of the New Diagnostics Working Group (NDWG) Stop TB Partnership. CL is a patent holder on biomarkers for monitoring of tuberculosis treatment. All other authors declare no competing interests.

#### Acknowledgments

The study was supported by the UNITE4TB consortium. UNITE4TB has received funding from the Innovative Medicines Initiative 2 Joint Undertaking (JU) under grant agreement no 101007873. The JU receives support from the EU's Horizon 2020 research and innovation programme and European Federation of Pharmaceutical Industries and Associations (EFPIA), Deutsches Zentrum für Infektionsforschung e. V. (DZIF), and LMU. European Federation of Pharmaceutical Industries and Associations/Associated Partners (EFPIA/AP) contributed to 50% of the funding; the contribution of DZIF and the University Hospital (LMU) has been granted by the German Federal Ministry of Education and Research. CL is supported by DZIF under grant TTU-TB 02.709. The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. We thank Bob Wallis, Clifton Barry, Stella Mpagama, Nyanda Ntinginya, and Claudia Denking for reviewing the TPPs.

#### References

- Gillespie SH, Crook AM, McHugh TD, et al. Four-month moxifloxacin-based regimens for drug-sensitive tuberculosis. *N Engl J Med* 2014; **371**: 1577–87.
- Sabiiti W, Azam K, Farmer ECW, et al. Tuberculosis bacillary load, an early marker of disease severity: the utility of tuberculosis molecular bacterial load assay. *Thorax* 2020; **75**: 606–08.
- Conradie F, Bagdasaryan TR, Borisov S, et al. Bedaquiline–pretomanid–linezolid regimens for drug-resistant tuberculosis. *N Engl J Med* 2022; **387**: 810–23.
- Boeree MJ, Diacon AH, Dawson R, et al. A dose-ranging trial to optimize the dose of rifampin in the treatment of tuberculosis. *Am J Respir Crit Care Med* 2015; **191**: 1058–65.
- Diacon AH, Dawson R, von Groote-Bidlingmaier F, et al. 14-day bactericidal activity of PA-824, bedaquiline, pyrazinamide, and moxifloxacin combinations: a randomised trial. *Lancet* 2012; **380**: 986–93.
- 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; **16**: e1002851.
- Gillespie SH, Gosling RD, Uiso L, Sam NE, Kanduma EG, McHugh TD. Early bactericidal activity of a moxifloxacin and isoniazid combination in smear-positive pulmonary tuberculosis. *J Antimicrob Chemother* 2005; **56**: 1169–71.
- Jindani A, Aber VR, Edwards EA, Mitchison DA. The early bactericidal activity of drugs in patients with pulmonary tuberculosis. *Am Rev Respir Dis* 1980; **121**: 939–49.
- Phillips PPJ, Dooley KE, Gillespie SH, 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; **14**: 51.
- Bratton DJ, Phillips PP, Parmar MK. A multi-arm multi-stage clinical trial design for binary outcomes with application to tuberculosis. *BMC Med Res Methodol* 2013; **13**: 139.
- Boeree MJ, Heinrich N, Aarnoutse R, et al. High-dose rifampicin, moxifloxacin, and SQ109 for treating tuberculosis: a multi-arm, multi-stage randomised controlled trial. *Lancet Infect Dis* 2017; **17**: 39–49.
- Zimmer AJ, Lainati F, Vasquez NA, et al. Biomarkers that correlate with active pulmonary tuberculosis treatment response: a systematic review and meta-analysis. *J Clin Microbiol* 2022; **60**: e0185921.
- Hills NK, Lyimo J, Nahid P, Savic RM, Lienhardt C, Phillips PPJ. A systematic review of endpoint definitions in late phase pulmonary tuberculosis therapeutic trials. *Trials* 2021; **22**: 515.
- Gupta-Wright A, den Boon S, MacLean EL, et al. Target product profiles: tests for tuberculosis treatment monitoring and optimization. *Bull World Health Organ* 2023; **101**: 730–37.
- Phillips PPJ, Mendel CM, Burger DA, et al. Limited role of culture conversion for decision-making in individual patient care and for advancing novel regimens to confirmatory clinical trials. *BMC Med* 2016; **14**: 19.
- Mukamolova GV, Turapov O, Malkin J, Woltmann G, Barer MR. Resuscitation-promoting factors reveal an occult population of tubercle bacilli in sputum. *Am J Respir Crit Care Med* 2010; **181**: 174–80.
- Faraj A, Clewe O, Svensson RJ, Mukamolova GV, Barer MR, Simonsson USH. Difference in persistent tuberculosis bacteria between in vitro and sputum from patients: implications for translational predictions. *Sci Rep* 2020; **10**: 15537.
- Lipworth S, Hammond RJH, Baron VO, Hu Y, Coates A, Gillespie SH. Defining dormancy in mycobacterial disease. *Tuberculosis (Edinb)* 2016; **99**: 131–42.
- Heyckendorf J, Georghiou SB, Frahm N, et al. Tuberculosis treatment monitoring and outcome measures: new interest and new strategies. *Clin Microbiol Rev* 2022; **35**: e0022721.
- Boeree MJ, Lange C, Thwaites G, et al. UNITE4TB: a new consortium for clinical drug and regimen development for TB. *Int J Tuberc Lung Dis* 2021; **25**: 886–89.
- MacLean EL, Zimmer AJ, Boon S den, et al. Tuberculosis treatment monitoring tests during routine practice: study design guidance. *Clin Microbiol Infect* 2024; published online Jan 3. <https://doi.org/10.1016/j.cmi.2023.12.027>
- Friedrich SO, Rachow A, Saathoff E, et al. Assessment of the sensitivity and specificity of Xpert MTB/RIF assay as an early sputum biomarker of response to tuberculosis treatment. *Lancet Respir Med* 2013; **1**: 462–70.
- DiNardo AR, Gandhi T, Heyckendorf J, et al. Gene expression signatures identify biologically and clinically distinct tuberculosis endotypes. *Eur Respir J* 2022; **60**: 2102263.
- Heyckendorf J, Marwitz S, Reimann M, et al. Prediction of anti-tuberculosis treatment duration based on a 22-gene transcriptomic model. *Eur Respir J* 2021; **58**: 2003492.
- Carratala-Castro L, Ssenooba W, Kay A, et al. A stool based qPCR for the diagnosis of TB in children and people living with HIV in Uganda, ESwatini and Mozambique (Stool4TB): a protocol for a multicenter diagnostic evaluation. *BMC Infect Dis* 2024; **24**: 233.
- Musisi E, Sessolo A, Kaswabuli S, et al. High *Mycobacterium tuberculosis* bacillary loads detected by tuberculosis molecular bacterial load assay in patient stool: a potential alternative for non-sputum diagnosis and treatment response monitoring of tuberculosis. *Microbiol Spectr* 2022; **10**: e0210021.
- Dorman SE, Nahid P, Kurbatova EV, et al. Four-month rifampentine regimens with or without moxifloxacin for tuberculosis. *N Engl J Med* 2021; **384**: 1705–18.
- Paton NI, Cousins C, Suresh C, et al. Treatment strategy for rifampin-susceptible tuberculosis. *N Engl J Med* 2023; **388**: 873–87.
- Aber VR, Allen BW, Mitchison DA, Ayuma P, Edwards EA, Keyes AB. Quality control in tuberculosis bacteriology. 1. Laboratory studies on isolated positive cultures and the efficiency of direct smear examination. *Tubercle* 1980; **61**: 123–33.
- Mitchison DA, Keyes AB, Edwards EA, Ayuma P, Byfield SP, Nunn AJ. Quality control in tuberculosis bacteriology. 2. The origin of isolated positive cultures from the sputum of patients in four studies of short course chemotherapy in Africa. *Tubercle* 1980; **61**: 135–44.

- 31 Bryant JM, Harris SR, Parkhill J, et al. Whole-genome sequencing to establish relapse or re-infection with *Mycobacterium tuberculosis*: a retrospective observational study. *Lancet Respir Med* 2013; **1**: 786–92.
- 32 Nunn AJ, Phillips PPJ, Mitchison DA. Timing of relapse in short-course chemotherapy trials for tuberculosis. *Int J Tuberc Lung Dis* 2010; **14**: 241–42.
- 33 Naidoo K, Dookie N. Can the GeneXpert MTB/XDR deliver on the promise of expanded, near-patient tuberculosis drug-susceptibility testing? *Lancet Infect Dis* 2022; **22**: e121–27.
- 34 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; **3** (10 suppl 2): S231–79.
- 35 Phillips PPJ, Mendel CM, Nunn AJ, et al. A comparison of liquid and solid culture for determining relapse and durable cure in phase III TB trials for new regimens. *BMC Med* 2017; **15**: 207.
- 36 Diacon AH, Maritz JS, Venter A, van Helden PD, Dawson R, Donald PR. Time to liquid culture positivity can substitute for colony counting on agar plates in early bactericidal activity studies of antituberculosis agents. *Clin Microbiol Infect* 2012; **18**: 711–17.
- 37 Shenai S, Ronacher K, Malherbe S, et al. Bacterial loads measured by the Xpert MTB/RIF assay as markers of culture conversion and bacteriological cure in pulmonary TB. *PLoS One* 2016; **11**: e0160062.
- 38 Mtafya B, Sabi I, John J, et al. Systematic assessment of clinical and bacteriological markers for tuberculosis reveals discordance and inaccuracy of symptom-based diagnosis for treatment response monitoring. *Front Med (Lausanne)* 2022; **9**: 992451.
- 39 Koele SE, Phillips PPJ, Upton CM, et al. Early bactericidal activity studies for pulmonary tuberculosis: a systematic review of methodological aspects. *Int J Antimicrob Agents* 2023; **61**: 106775.
- 40 Timm J, Bateson A, Solanki P, et al. Baseline and acquired resistance to bedaquiline, linezolid and pretomanid, and impact on treatment outcomes in four tuberculosis clinical trials containing pretomanid. *PLoS Glob Public Health* 2023; **3**: e0002283.
- 41 Honeyborne I, McHugh TD, Phillips PPJ, et al. Molecular bacterial load assay, a culture-free biomarker for rapid and accurate quantification of sputum *Mycobacterium tuberculosis* bacillary load during treatment. *J Clin Microbiol* 2011; **49**: 3905–11.
- 42 Hellyer TJ, Desjardin LE, Hehman GL, Cave MD, Eisenach KD. Quantitative analysis of mRNA as a marker for viability of *Mycobacterium tuberculosis*. *J Clin Microbiol* 1999; **37**: 290–95.

© 2024 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).