World Health Organization

DRAFT for Public Consultation

Guidance on evidence generation on new tuberculosis preventive treatment regimens

Guidance on evidence generation on new tuberculosis preventive treatment regimens

ISBN XXXXX (electronic version)
ISBN XXXXX (print version)
Placeholder for disclaimer

Contents

Abbreviations and acronyms

1HP	1 month of daily rifapentine plus isoniazid
3HP	3 months of weekly rifapentine plus isoniazid
3HR	3 months of daily rifampicin plus isoniazid
4R	4 months of daily rifampicin monotherapy
6H	6 months of daily isoniazid monotherapy
9H	9 months of daily isoniazid monotherapy
AE	adverse events
CEA	cost-effective analysis
CSA	coordinated scientific advice
DR-TB	drug-resistant tuberculosis
DS-TB	drug-susceptible tuberculosis
DST	drug susceptibility testing
DT	decision threshold
ERP	Expert Review Panel
EtD	evidence to decision
FDA	United States Food and Drug Administration
FPP	finished pharmaceutical product
GDG	guideline development group
GEG	guidance on evidence generation
GRADE	Grading of Recommendations Assessment, Development and Evaluation
GRC	Guideline Review Committee
HIV	human immunodeficiency virus
HRZE	isoniazid, rifampicin, pyrazinamide and ethambutol
Lfx	levofloxacin
LTFU	lost to follow-up
MDR/RR-TB	multidrug-resistant/rifampicin resistant tuberculosis
Mfx	Moxifloxacin
NRS	nonrandomized studies
NTP	national tuberculosis program
PICO	population, intervention, comparator and outcome
PQ	prequalification
PRO	participant-reported outcome
QCL	quality control laboratory
ROBINS-I	risk of bias in nonrandomized studies – of interventions

SAE	severe adverse events
SOC	standard of care
SRA	stringent regulatory authority
TB	tuberculosis
TBI	tuberculosis infection
TPP	target product profiles
TPT	tuberculosis preventive treatment
WHO	World Health Organization
WLA	WHO listed authority

Table 1. Key messages on providing guidance for evidence generation on new TB preventive treatment regimens

#	Key message
1	Consider the requirements for the WHO guideline development process during research design
2	Be more inclusive in the selection of populations and settings
3	Include implementation considerations in the TPT study protocol
4	Include an appropriate comparator arm in the trial
5	Report on outcomes of importance for guideline development
6	Use harmonized definitions for safety outcomes and report them comprehensively
7	Characterise the acquisition of drug resistance
8	Ensure sufficient follow-up time
9	Characterise tolerability and acceptability of the TPT regimens
10	Report the effect of the shortening of TPT regimens on clinical and health system outcomes
11	Gather evidence within trials regarding the resources required to deliver the TPT regimen
12	Report cost-effectiveness
13	Ensure sufficient sample size to achieve precise estimates for critical outcomes
14	Consider the possibility of extrapolating study findings beyond the trial population
15	Share individual participant data
16	Investigate the impact on health equity
17	Evaluate the feasibility of implementation of TB preventive treatment
18	Data on the safety of novel regimens from large observational studies are important to identify infrequent but important adverse events
19	Trials evaluating novel drugs should be accompanied by pharmacokinetic studies in target populations where possible, to inform dosing recommendations.

1. Introduction

1.1 Background

Around one quarter of the global population is estimated to have been infected with *Mycobacterium tuberculosis* (1) presenting a challenge to the reduction of the burden of tuberculosis (TB) morbidity and mortality. TB preventive treatment (TPT) has been shown to reduce the rate of progression from TB infection (TBI) to TB disease, protecting affected individuals and preventing the spread of incident TB disease in the community. The World Health Organization (WHO) has published TPT guidance that informs end users about the detection of target populations and the provision of TPT regimens (2, 3). This guidance supports the achievement of the 2023 United Nations High-level meeting and End TB Strategy targets to achieve 90% coverage of the eligible populations with TPT. In addition, the WHO has produced target product profiles (TPP) for TPT (4, 5) which describe the optimal performance and operational characteristics of future TPT regimens.

Evidence from research informs normative work both of the WHO and regulatory bodies, with important differences in emphasis, processes, and goals. While regulators necessarily focus on the efficacy, safety, and quality of a drug or regimen, WHO additionally works to determine the acceptability, feasibility, equity, and resource implications of a novel intervention in the contexts in which it is commonly used (see Table 2).

The evidence available to the WHO on drugs and regimens for use in guideline development often presents challenges that limit the strength of WHO recommendations. Such evidence may not apply to certain populations or may not be suitable for policy development. Limitations have typically arisen in three main areas. First, the evidence may give rise to recommendations that have “very low certainty”, such as when the guidance relies upon nonrandomized studies, trials lacking a standard of care (SOC) arm as a comparator, or trials for which estimates are imprecise on account of a small sample size. These limitations may affect the strength of recommendations (1, 6). Second, supporting data may be incomplete, or inconsistently recorded or reported, hampering the ability of evidence reviewers to synthesise evidence by performing meta-analyses. This may lead to the strength of recommendations being downgraded. Significant heterogeneity between studies, or the lack of common outcome definitions between studies, may make it difficult to interpret findings even when a meta-analysis can be performed. Finally, studies presented to the WHO may lack information about factors that are important to decision-making. In addition to evidence about efficacy and safety, the Grading of Recommendations Assessment, Development and Evaluation (GRADE) decision criteria also include feasibility, acceptability, equity, cost, and cost-effectiveness. Information about these other dimensions is important for WHO decision-making and subsequent implementation following licensure. The inclusion of these additional criteria is of particular importance when evaluating novel regimens that are found to be “noninferior”[1] to existing options in terms of safety and efficacy.

[1] Non-inferiority trials are sometimes used when an established effective treatment exists, and the objective is to show that the new treatment is not unacceptably worse. Demonstrating non-inferiority is the statistical

This Guidance on Evidence Generation (GEG) seeks to engage in dialogue with the stakeholders who generate evidence for the WHO Guideline Development process to support the clarity of future WHO guidelines on TPT. This document serves as a companion to WHO's existing TPP for TPT (5). Its aim is to maximize the level of certainty that Guideline Development Groups (GDGs) can have in the evidence available during the guideline development process. Evidence that engenders greater certainty will, in turn, result in more rapid adoption of new regimens, both in terms of regulatory approval and adoption by national TB programmes (NTPs).

The TPP for TPT outlines the expectations for new regimens and informs the ambitions and plans of developers active in the field, describing the minimal and optimal requirements for various regimen characteristics (5). In contrast, this GEG document provides guidance regarding optimal approaches to the design, conduct, and evaluation of clinical trials and sub-studies generated from these trials, with a goal of improving the WHO guideline development process.

This guidance was developed for those who design and conduct research that has the potential to inform future WHO guidelines on TPT. It synthesises information that can be incorporated at each stage of the design and conduct of clinical trials, including sub-studies. It outlines a set of 19 key messages that will positively impact the development of WHO TPT guidelines, including trial design, population selection, the intervention and comparator regimens, outcome reporting, and data analysis. In addition to the outcomes of safety, efficacy, and tolerability, which have traditionally been the mainstay of TPT clinical trials, other important outcomes that may be of interest to GDGs may include the drug palatability, individual preferences, drug acceptability, feasibility of drug administration, cost effectiveness (from individual and/or heath system perspectives), drug-drug interaction, acquisition of drug resistance, quality of life and participant-reported outcomes. By generating evidence that addresses such a wide range of outcomes, issues of importance to individuals and TB programs can be adequately considered.

More trials of safer, shorter, and well-tolerated treatment options for TB infection are urgently needed. Given the substantial burden of TB infection globally, improved TPT options will be essential for its rapid uptake in high-risk populations. We hope that this guidance will assist the TB research community to develop high-quality evidence that will benefit affected communities around the world.

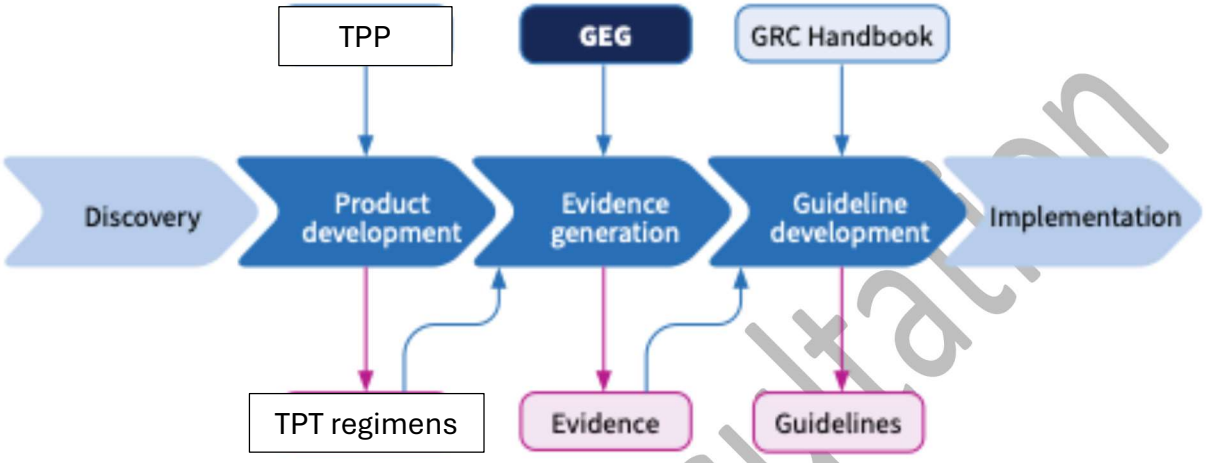
1.2 Purpose

This document aims to guide the design of clinical trials and other studies of TPT so that the evidence generated meets the requirements of the WHO guideline development process. This document addresses the planning, implementation, and reporting of clinical trials of TPT. Not only does it address efficacy and safety outcomes, but it also presents important additional research needs, including qualitative research, cost-effectiveness analyses, mathematical modelling, and other patient-related outcomes. It is intended that well-conducted clinical trials that adopt this guidance will contribute to

approach to demonstrate this and is done in reference to a non-inferiority margin that reflects a clinically meaningful difference that would be acceptable to patients and health care providers and that ensures that the new treatment retains a significant portion of the control treatment's efficacy.

strong recommendations from the WHO GDGs. Figure 1 shows the place of this guidance in the context of the WHO guideline development process

Figure 1: The role of GEG in relation to the TRPs and WHO guidelines within the discovery-to-implementation value chain



GEG: guidance on evidence generation; GRC: Guideline Review Committee; TPP: target product profile; TPT: tuberculosis preventive treatment; WHO: World Health Organization. The chevron process shows a simplified "discovery-to-implementation value chain"; the red boxes below each blue chevron show the outputs of some of the steps in this process, which then feed into the next step; the blue boxes above the chevrons show the guidance documents that inform steps in this process.

1.3 Scope

This Guidance on Evidence Generation focuses upon the evidence generated within, and alongside, phase 2, 3, and 4 clinical trials of TPT regimens. It describes important steps in the WHO guideline development process for TPT, including the GRADE processes for guideline development (Chapter 3). It is intended that alignment with the suggestions in this document may increase the strength, applicability, and uptake of the recommendations that emerge from a WHO guideline review process. This document complements a companion document for trials evaluating novel treatments for TB infection, published by WHO in 2024 (4). It, however, does not replace existing WHO normative documents such as the WHO handbook for guideline development (7), WHO guidance for best practice on clinical trials (8), and target product profiles (5), as well as guidance from the regulators. A TPT regimen may be composed of one or more drugs. This document focuses on trials and other studies of TPT regimens to determine whether a novel TPT regimen should be recommended by the WHO or not, including for populations of interest, such as people with HIV, household contacts of individuals with TB, people in prisons, or other at-risk populations. Mindful of the continuity between screening for TB and TPT under programmatic conditions, it would be important for studies to be able to assess the feasibility of integrating the two elements and their combined impact on TB incidence. This document does not cover the guidance on evidence generation on how eligible people should be evaluated to rule out TB disease and populations to be tested for TBI. Mindful of the continuity between screening for TB

infection and TPT under programmatic conditions, it would be important for studies to be able to assess the feasibility of integrating the two elements and their combined impact on TB incidence.

The guidance is not intended to stifle innovation, interfere with the conduct of clinical trials, or tell researchers how to do research. It presents methods that are currently used to evaluate TPT regimens. Evidence that is generated without reference to this advice is, of course, still eligible for review by WHO. There may be good reasons for researchers to diverge from the suggestions offered in this document; in which case, a rationale can be offered to the GDG. Also, this guidance does not replace national or international regulatory guidance for clinical trials.

1.4 Objectives

The objectives of this document are to:

- Outline the steps in the WHO guideline development process that are relevant to evidence generation for TPT regimens;
- Support the development of high-quality WHO guidelines for TPT by providing key messages about how research evidence should be generated;
- Describe other WHO processes that are relevant to the development and evaluation of new drugs, regimens, or formulations.

1.5 Audience

The target audience of this document is the research community that is generating new evidence about TPT that is relevant to future WHO guidelines. Key stakeholders to whom this guidance will be relevant include research funding agencies, academic researchers (including clinical trials investigators, health economists, qualitative researchers), drug developers, consumer groups contributing to clinical trials, and biostatisticians. It will also be informative for GDG members, since it describes important issues that are frequently the focus of discussion at WHO GDG meetings.

1.6 WHO guideline development and regulatory approval

There are similarities and differences between the evidence needed for regulatory approval and WHO recommendations. This document primarily pertains to the evidence needed for WHO policy development, while also covering some regulatory requirements. Table 2 summarizes the key differences in scope and approach between WHO recommendations and regulatory approval in relation to TPT. Regulators assess whether a drug or a regimen *can* be used, principally based on its efficacy and safety. In contrast, WHO guidelines are mainly concerned with the question of whether a new drug or regimen *should* be recommended, considering both clinical and programmatic aspects. The guidelines typically compare new regimens to existing options available to NTPs. Many drugs recommended as TPT are already recommended for the treatment of TB disease and registered for clinical use in countries. The inclusion of older drugs in TPT regimens thus rarely presents an impediment from a regulatory standpoint.

When novel drugs are being evaluated as a part of TPT regimens, clinical trials may serve both to address regulatory requirements, as well as WHO guideline development process to enable their use in resource-limited settings.

Table 2. Differences and similarities in scope and approach between WHO recommendations for TPT medicines and regimens, and regulatory approval

	WHO recommendation	Regulatory approval
Prerequisite for evaluation	Regulatory approval of the drugs within a regimen, although not necessarily for TPT indication, by at least one stringent regulatory authority (SRA) or WHO-listed authority (WLA) (9). Consideration is made on whether the drugs are being used to treat or prevent TB disease.	Submission to the applicable regulatory agency of a dossier; for example, common technical documents, availability of manufacturing and quality management processes, non-clinical and clinical data in accordance with common technical documents provisions and any additional data as required by national or regional legislation.
Goal	Main goal is to provide guidance on whether to use a specific drug or regimen and which regimen to prioritize, considering the balance of benefits (efficacy) and harms (toxicity and influence on quality of life) and additional factors (e.g., likelihood of TPT completion, cost). Choice of regimen also has a bearing on the likelihood that it can be used at scale and influence TB incidence at a population level. In TBI, the main focus for efficacy is on the overall comparative performance of ***regimens*** in relation to the incidence of TB disease to an established WHO-recommended SOC. Understanding the contribution of individual drugs to the safety and efficacy of a regimen is relevant, but is usually not critical for the recommendation of regimens per se. Instead the effect and safety of the regimen is considered.	Main goal is to ensure that the medicinal product meets the necessary standards of quality, safety and efficacy. A positive benefit–risk evaluation is a prerequisite for a regulatory approval. Historically, the process has focused on review and approval of single medicines but has also permitted approval of regimens comprising multiple drugs. Understanding the contribution of an individual drug to the safety and efficacy of a regimen is usually critical to their approval.
Meaning of a decision	A drug or regimen is recommended or suggested for use, or a recommendation is made against the use of a drug or regimen. A WHO recommendation may make the drug or regimen eligible for Global Fund grants and procurement via GDF, UN agencies, governments and other donors. The drug(s) involved may also be included in the Essential Medicines List once recommended by WHO and listed on the WHO prequalification Expression of Interest for manufacturers.	Approved products receive a marketing authorization and can be made commercially available within the country or countries of the respective national or regional regulatory agencies. Safety concerns that emerge in the post-marketing phase may lead regulatory agencies to recommend against the continued use of a drug and withdraw its marketing authorization.
Remit	Global (although national adaptation of global recommendations may be required, owing to	National or regional, although some national authorities may follow approvals

	WHO recommendation	Regulatory approval
	implementation considerations).	of SRAs or WLAs, or have abbreviated processes upon approval from an SRA or WLA. WHO Prequalification of a drug has a global remit.
Main criteria affecting decision-making	The main criteria in formulating recommendations are the so-called EtD criteria, which include: • certainty of evidence • values • desirable and undesirable effects • balance of effects • resources required and cost–effectiveness • equity • acceptability • feasibility	Criteria affecting decision-making by regulatory authorities include: • Preclinical pharmacology and toxicology • Dose selection • Clinical safety and efficacy • Manufacturing quality • Compliance with international and regional standards
Mechanism to ensure reliability and quality of evidence	Early discussion with WHO technical departments is encouraged, particularly where trials address important evidence gaps or are likely to change practice. Additionally, the Coordinated Scientific Advice (CSA) procedure is available for new drugs (See section 6.1). Procedure involves systematic and transparent review of evidence using the GRADE framework, including the use of evidence synthesis, evidence appraisal and management of conflicts of interest.	Regulatory agencies provide detailed guidance on requirements though the International Conference on Harmonisation, good practice frameworks and developer-specific consultations before clinical trials are performed. This typically includes trial design and the choice of primary and some secondary endpoints.
Evidence base for evaluation of benefits and harms	Systematic review or individual patient meta-analysis of all available trials, or NRS (pertinent single trials or individual patient data may also be reviewed).	Typically, one or two pivotal Phase 3 RCTs, supported by early phase clinical and preclinical data.
Approach to analysis and decision-making	Typically based on a summary of all available evidence with standard meta-analysis and meta-analysis of individual patient data from trials is provided to WHO. Such analyses should accord to an agreed statistical analysis plan. Guideline questions may differ from the hypotheses of trials included in the systematic review. Meeting or not meeting certain statistical criteria (e.g. $P<0.05$ or meeting pre-specified non-inferiority criteria) is not by itself relevant to decision-making. Rather, decision-making focuses on interpretation of effect sizes and confidence intervals of all critical outcome measures, considering the values placed on each outcome by different stakeholders (e.g. clinicians, patients, managers) and the certainty of evidence determined based on the GRADE framework.	Analysis of individual participant data provided to the regulatory agency by the sponsor of the trial or trials according to a predefined-agreed statistical analysis plan. Testing of a limited set of protocol-defined statistical hypotheses (often framed as superiority or non-inferiority) relating to the primary endpoint or endpoints against agreed levels of significance is central to decision-making.
Considerations after recommendation and approval	Identification of research gaps or requests to improve the strength of future recommendations. WHO's remit includes operational assistance and facilitation of implementation of recommended	Additional research may be required as a condition of full approval, post-marketing pharmacovigilance and (in some countries) population-specific research studies. Implementation of interventions with a

	WHO recommendation	Regulatory approval
	interventions. It also monitors the global uptake of the regimen and documents the resource requirements for large scale implementation (e.g. through country case studies). Guideline recommendations continually evolve based on reassessment of existing and novel regimens against each other. Changes in pharmaceutical presentation (e.g. alternate or coformulation) or dosing regimen are not usually subject to the guideline development process and are evaluated using other processes that are outside the scope of this document.	marketing approval is not a responsibility of the regulator. Once granted, there is usually no formal reassessment of marketing authorization against novel comparators. Health-economic analyses are generally not within scope. Changes in pharmaceutical presentation (e.g. alternate or coformulation) often require new approval. Changes in dosing regimen may be accommodated in an existing label.

EtD: evidence to decision; GDF: Global Drug Facility; Global Fund: Global Fund to Fight AIDS, Tuberculosis and Malaria; GRADE: Grading of Recommendations Assessment, Development and Evaluation; GTB: Global Programme on Tuberculosis & Lung Health; NRS: nonrandomized studies; RCT: randomized controlled trial; SOC: standard of care; SRA: stringent regulatory authority; TB: tuberculosis; UN: United Nations; WHO: World Health Organization; WLA: WHO listed authority.

2. Methodology for development of this guidance

2.1 Establishment of the Scientific GEG Development Group (SGG)

A Scientific GEG Development Group for TPT was formed, comprising leading trialists, scientists, public health officials, regulators, economists, social scientists, and experts involved in developing WHO policy recommendations. Civil society representatives, individuals with lived experience, and in-country end users were also included. The SGG supported the development of the GEG through participation in virtual meetings and review and input into multiple drafts of the document.

2.2 Development of the GEG document

The initial draft of the GEG document was developed by the WHO Global Programme on Tuberculosis & Lung Health secretariat and external consultants, based upon the approach followed for GEG on new regimens for tuberculosis treatment (10). The SGG reviewed the initial draft, providing detailed written feedback that was incorporated into an updated version. The feedback was also discussed at virtual meetings of the GEG.

An external review panel was established, including those with expertise in TPT clinical trials and guideline development. Staff from national TB programs were also consulted. They were asked to provide an independent peer review of the final draft version of GEG.

Following this, other stakeholders not represented on the SGG were engaged. These included funders, industry bodies via the International Federation of Pharmaceutical Manufacturers and Associations, and companies involved in the development of TPT drugs, with due diligence performed to identify potential conflicts of interest.

A final virtual consensus meeting (date) was convened to resolve any outstanding issues based on input from the external reviewers, funders, and industry.

3. WHO guideline development process and the GRADE approach

3.1 Development of WHO guidelines

One of the fundamental means through which WHO fulfils its technical leadership in health is review of evidence and development of normative products such as guidelines (Box 1); "WHO's legitimacy and technical authority lie in its rigorous adherence to the systematic use of evidence as the basis for all policies" (11).

Box 1. What is a WHO guideline?
A WHO guideline is any document developed by WHO that contains recommendations for clinical practice or public health policy. A recommendation tells the intended end users of the guideline what they can or should do in specific situations, individually or collectively, to achieve the best health outcomes possible. It offers a choice among different interventions or measures expected to have a positive impact on health and implications for the use of resources (7).

WHO uses the GRADE approach to assess the certainty of a body of evidence and to develop recommendations (6, 12-14). Key principles for the development of WHO guidelines include:

- explicit, inclusive, and transparent processes for developing recommendations (i.e., users can see how and why a recommendation was developed, by whom, and on what basis);
- use of processes and methods in each step of guideline development to minimize the risk of bias in the recommendations; and
- recommendations developed based on a systematic and comprehensive assessment of the balance of an intervention's potential health benefits and harms, and explicit consideration of other relevant factors.

The process for developing WHO guidelines is detailed in the WHO handbook for guideline development (7) which covers many activities beyond the assessment of available evidence (15).

This section of the document provides a brief overview of some of the critical steps that are particularly relevant in generating evidence that may be used by WHO during policy development, including:

- developing the scope and recommendation questions using the PICO (population, intervention, comparator and outcome) format for evidence retrieval and synthesis (Section 3.2);
- determining values and decision thresholds (Section 3.3);
- evaluating the certainty of the evidence (Section 3.4);
- preparing "evidence profiles" and "summary of findings" tables (Section 3.5);
- making judgements across 12 evidence-to-decision (EtD) criteria (Section 3.6); and
- developing recommendations (Section 3.7)

The software GRADEpro allows the information from the evidence profiles and the evidence-to-decision table to be managed and stored online (16). This is helpful for managing the discussions during the GDG meetings.

An example of how this approach has been applied in a past GDG meeting on TPT is described in **Annex 1-3.**

3.2 Developing the scope and recommendation questions using the PICO format, evidence retrieval, and synthesis

One of the critical initial steps in the development of a guideline, often in response to significant changes in the available evidence, is the definition of the scope and formulation of questions in the PICO format, including the identification and selection of key outcome measures (17). Typically, systematic reviews for each PICO question are then commissioned through independent researchers. If a systematic review finds only a single study or trial providing pertinent evidence for the recommendation question, the evidence assessment will focus on that study or trial. Detailed guidance on the performance of systematic reviews is provided in the WHO handbook for guideline development (7) and elsewhere – for example, in the Cochrane handbook for systematic reviews of interventions (18) – and is beyond the scope of this document (Box 2).

Box 2. WHO Handbook for Guideline Development
The WHO Handbook for Guideline development provides step-by-step guidance on how to plan, develop and publish a WHO guideline (7). That document is intended mainly for WHO staff, and it covers the methods, processes and procedures for producing a document that meets WHO standards. The handbook is produced by WHO's Guideline Review Committee (GRC), which is an independent group that reviews all WHO guidelines during planning and before publication. The science underpinning evidence identification and synthesis and the translation of a body of evidence into recommendations continues to evolve; thus, the GRC also supports WHO staff by providing additional up-to-date guidance that reflects the latest methods and approaches in the peer-reviewed literature and the best practices internationally.

3.3 Determining values and decision thresholds

The GDG must make judgements about the size of the desirable and undesirable effects of interventions on health outcomes that are important to people who receive them. These typically include adverse events (AEs), survival, TB incidence, TPT adherence and completion. The GDG members are also required to rank the outcomes from critical to not important, so that evidence reviewers focus on the most important outcomes from among the whole range possible and thus make the analyses more meaningful.

The GRADE EtD framework guides a GDG to assess and interpret the effect of an intervention, making judgements about whether the desirable and undesirable health effects are absent, trivial, small, moderate or large (15-20). The framework allows GDG members to judge desirable and undesirable health effects of interventions, taking into account not only the size of their absolute effects but also the value of the respective outcomes from the perspectives of all stakeholders, including people affected by TB/TBI. Making judgements about whether a given magnitude of an intervention effect is significant or not are facilitated by agreeing upon decision thresholds (DTs) (i.e. quantitative reference values based on which one can classify the effect sizes of an intervention as trivial or no, small, moderate or large effects) ahead of the assessment of the evidence (15, 16). The GRADE working group has provided guidance for using generic, empirically derived, outcome-independent DTs as starting points for discussions with a decision-maker on what constitutes such DTs (21).

Empirical or otherwise defined DTs are used to:

- provide a reference for what is considered a trivial or no, small, moderate, or large effect, applicable to any health outcome;
- help to increase the consistency and transparency of EtD judgements; and
- facilitate making these judgements and comparing them across multiple interventions.

3.4 Assessing the certainty of the evidence

Once the evidence has been retrieved and synthesized through a systematic review, a critical next step is the assessment of the certainty of evidence (in the past this was also referred to as quality of evidence). In the context of quantitative evidence syntheses, the GRADE working group defines the certainty of the evidence as the "certainty that an estimate of association or effect is correct or, better, that a true effect lies on one side of a specified threshold or within a chosen range" (19-21). For qualitative evidence syntheses, the GRADE-CERQual (22) approach defines it as "the extent to which a review finding is a reasonable representation of the phenomenon of interest." In the context of guideline development, the certainty of the evidence reflects the confidence that the estimates of an effect, or the qualitative findings (e.g. themes and concepts), are adequate to support a particular decision or recommendation or within a range or beyond a certain threshold. GRADE and GRADE-CERQual categorize the certainty of the evidence as high, moderate, low or very low for each outcome (Table 3).

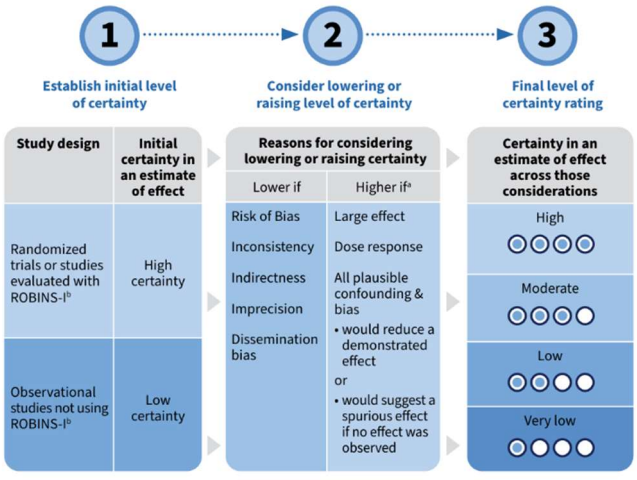
Table 3. Definitions of the four levels of certainty of quantitative evidence

Certainty level	Definition or interpretation
High	We are very confident that the true effect lies close to that of the estimate of the effect or within a range or beyond a certain threshold
Moderate	We are moderately confident that the effect lies within a range or beyond a certain threshold: the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low	Our confidence that the effect is within a range or beyond a certain threshold is limited: the true effect may be substantially different from the estimate of the effect
Very low	We have very little confidence that the effect lies within a certain range: the evidence is very uncertain

Source: Santesso et al, 2020 (23)

A body of quantitative evidence based on RCTs is rated initially as being of high certainty, whereas evidence from nonrandomized studies (NRS) is rated as being of low certainty unless tools are used that allow for an assessment of NRS against randomized studies (e.g. ROBINS-I: risk of bias in nonrandomized studies – of interventions finds low risk of bias after assessment of confounding and selection domain) (24). For both types of studies, five domains (described in detail below) can lower these initial ratings following objective assessment of the certainty of evidence and three domains can raise them (large effect, dose response relationship between intervention and outcome and plausible confounding and bias that may either reduce a demonstrated effect (making the true effect likely larger) or suggest a spurious effect if no effect was observed (increasing the confidence in the null finding), although these domains are rarely applicable (Figure 2). For a given body of quantitative evidence, the ratings are conducted separately for each outcome. The domains for rating down are described in more detail in subsections 3.4.1 to 3.4.5; details on rating qualitative evidence are found in subsection 5.1. The following sections provide a brief outline of the certainty of evidence assessment: more detailed information can be accessed in the Handbook (7).

Figure 2: GRADE's approach to rating the certainty of evidence for each outcome



GRADE: Grading of Recommendations Assessment, Development and Evaluation; ROBINS-I: risk of bias in nonrandomized studies – of interventions.

[a] Upgrading criteria usually apply to observational studies only.

[b] ROBINS-I is one of several risk of bias assessment tools.

3.4.1 Risk of bias in quantitative evidence

This section just outlines how judgements about the risk of bias are made; details are available elsewhere (6). Recognition of bias is essential in interpreting quantitative evidence during guideline development, whether randomized or observational. GRADE prioritizes randomized controlled trial evidence because of its lower risk of bias, and this is reflected in the higher certainty of evidence. In contrast, discussions around plausible sources of bias in observational datasets due to confounding, even after sophisticated statistical adjustment or residual confounding due to unknown variables, may diminish confidence in the results and lead to a lower certainty of evidence (24). It is important to maintain this hierarchy of evidence when conducting meta-analyses; however, there may be situations where observational data can strengthen and extend preliminary conclusions reached from RCTs. Furthermore, other factors that can increase the strength of recommendations, independent of trial design, relate to the ease of applicability of an intervention within the national TB program, such as the cost of an intervention.

For RCTs, several criteria are used to assess the risk of bias. The following characteristics are the distinguishing features of the studies that yield the best certainty of evidence:

- random sequence generation;
- concealment of treatment allocation to the treatment group;
- blinding of outcome assessors, including laboratory staff;
- blinding of participants and investigators, particularly if the outcomes were measured subjectively and thus may be subject to bias;
- reporting of data on all study participants, including attrition and exclusions from analysis; and
- complete reporting of all study outcomes that were specified *a priori* (25).

For nonrandomized (observational) studies of interventions, the main criteria for assessment of bias depend on the study design and can be categorised as bias (26):

- due to confounding;
- due to the selection of participants into the study;
- in the classification of interventions;
- due to deviations from intended interventions;
- due to missing outcome data;
- in the measurement of the outcome; and
- in the selection of reported results.

Once the risk of bias has been assessed for each individual trial or study, it is then summarised across trials and studies for each outcome. Study limitations across the body of evidence for each outcome can be categorised as follows:

- no serious limitations – most of the studies in the review meet all the minimum quality criteria for the particular study design;

- serious limitations – one of the minimum criteria for quality is not met by most of the studies in the review. This results in a lowering of the overall quality rating by one level (e.g. “high” becomes “moderate” for RCTs or “low” becomes “very low” for observational studies or NRS); and
- very serious limitations – the risk of bias may have a strong influence on the estimate of effect and study limitations are present in most of the studies contributing data on a given outcome in the review. This typically results in a lowering of the quality by two levels.

Other options that may be considered to judge the risk of bias include RoB 2 (Version 2 of the Cochrane risk-of-bias assessment tool for randomized trials); and the ROBINS-I tool for NRS. When using ROBINS-I for assessing risk of bias in NRS, the initial GRADE certainty in the evidence from a body of such studies would be high, given that assessment of selection bias and confounding is an integral part of ROBINS-I. This does not mean that GRADE sees randomization as the only secure way to guard against confounding bias. Thus, whether the starting point with a body of evidence from NRS is viewing the evidence as low certainty and assessment of reasons to rate it up or down, or viewing that evidence as high certainty and assess reasons to rate it down, the final certainty rating should be the same, and should include a category of extremely serious risk of bias (24, 27). When ROBINS-I is used to rate conventional NRS of any design (e.g. cohort or case–control), after assessment of confounding and selection bias, often, the rating of risk of bias will be “high”. Nevertheless, it is possible that a body of evidence from NRS, rated using ROBINS-I, will receive a final rating of high or moderate certainty of evidence. This could result from rating up (e.g. for large effect, dose response or the direction of plausible confounding) or from the use of NRS designs and analyses with greater protection against risk of bias (e.g. interrupted time series that would lead to rating down by only one level or not at all). However, to date, GRADE does not have convincing examples of that scenario.

3.4.2 Inconsistency

This section outlines how judgements about inconsistency are made; additional details are available elsewhere (17). Inconsistency is one of the four key GRADE domains used to assess the presence of systematic errors within a body of evidence. This domain evaluates whether there are systematic differences across the results of the studies included in the evidence synthesis. If only one study is available, clearly, there is no concern about inconsistency; also, if inconsistency can be explained by a small number of a priori subgroup hypotheses, GRADE users may choose to disaggregate the evidence based on these factors. However, if the inconsistency remains unexplained, the certainty of the evidence should be rated down. Statistical measures (e.g. I^2 and Cochran’s Q test) provide an initial assessment of inconsistency, but the final certainty judgement should be based on examining the effects of individual studies and their relationship with pre-established thresholds or ranges. The decision as to whether any inconsistency is important is guided by the magnitude of any differences in the direction and size of the effect observed in different studies, and whether any of these differences can be explained.

Inconsistency may arise from random variation or differences across studies, such as differences in study design or varying definitions of the population, interventions, comparator, and/or outcomes. To explore the sources of inconsistency, the GDG may have to review the study designs in detail or conduct sensitivity or subgroup analyses. The certainty of the evidence is rated down if an important inconsistency in study results remains after exploration of a priori hypotheses that might explain the heterogeneity.

3.4.3 Indirectness

This section outlines how judgments about indirectness are made(17). In GRADE, the term "directness" encompasses several characteristics of a study's results, which are often referred to as directness, generalizability, external validity, transferability, and applicability. Generally, four types of indirectness are differentiated in GRADE (27); they relate to differences between the evidence identified and the recommendation question at hand, as defined in the PICO format (17):

- Indirectness arising due to *differences in population* is present if the population for which evidence is available differs in important ways from the population identified in the recommendation question.
- Indirectness arising due to *differences in intervention* is present if the intervention cannot be implemented with the same rigour in the intended settings of use as in the trials from which the data arose.
- Indirectness arising due to *differences in comparator* is present when no direct comparison of the intervention with the comparator of interest is available.
- Indirectness arising due to *differences in outcome measures* is present if data are only available on intermediate and surrogate outcomes, because they do not provide direct evidence on the health outcomes that ultimately matter to individuals and populations.

Although most evidence is indirect to some degree, if indirectness is serious or very serious, it will cause the certainty of the body of evidence for a given outcome to be rated down by one or two levels. The combined effect of all four types of indirectness is considered when rating the certainty of evidence.

3.4.4 Imprecision

This section outlines how judgements about imprecision are made; details are available elsewhere (17). The domain of imprecision evaluates the risk of random error within a body of evidence. Confidence intervals around absolute estimates (e.g. risk differences or mean differences) are the main method for assessing imprecision. When these confidence intervals cross predetermined thresholds – ideally established before the analysis – the certainty in the evidence is lowered. Additionally, even when large effects seem precise based on their confidence intervals, it is important to evaluate whether the evidence is sufficiently robust. This assessment depends on the total number of participants and the events available to inform the body of evidence. If the effect estimates are based on a small number of participants or events, rating down due to imprecision may still be warranted. In general, results are

imprecise when studies include relatively few participants or few events, and thus, large uncertainty (i.e. wide confidence intervals) surrounds the estimate of effect for a particular outcome.

For WHO guideline development, if the confidence interval for the pooled estimate of effect crosses the thresholds established for making one decision versus another (Section 3.3), then the body of evidence is imprecise for the particular outcome in question, and the certainty of the evidence is lowered by one, two, or three levels (28). Systematic review teams can use the 95% confidence interval (95% CI) for the pooled estimate of effect as the primary criterion for judging the presence of imprecision.

In formulating a recommendation, all outcomes are considered, with attention to whether they are "critical" or "important (but not critical)" for decision-making. The decision to downgrade the certainty of the evidence for imprecision depends on the thresholds established as the basis for a decision or a recommendation, and on the trade-off between desirable and undesirable consequences (28). Determining the acceptable threshold involves an explicit judgement.

3.4.5 Dissemination bias

Dissemination bias may result in the systematic underestimating or overestimating of the underlying beneficial or harmful effect of an intervention or exposure, caused by the selective publication or similar limitations of studies based on the study results. Often, studies in which no effect is found are less likely to be published. Other forms of dissemination bias may include selective reporting of outcomes, delayed publication of negative results, grey literature bias, and language bias (preference for English language studies).

Public calls by WHO for data may help to mitigate the risk that unpublished studies remain unidentified even during systematic reviews. Searches of trial registries and the grey literature (i.e. information produced outside traditional publishing channels) can help to identify unpublished studies and thus reduce the risk of this bias. The inclusion of a broad range of experts on the GDG panel can also mitigate this risk. The risk of publication bias may be assessed using funnel plots and appropriate statistical tests (e.g., Egger's regression test). Such tests have limitations; however, the existence or non-existence of publication bias cannot be confirmed, it can only be suspected. When publication bias is suspected, the quality of the evidence should be downgraded by one level. In the context of the relative scarcity of contemporary trials of TPT regimens, publication bias has not been noted as a common problem to date (29, 30).

3.4.6 Assessing qualitative evidence

3.4.1–3.4.5 are primarily focused on the development of quantitative evidence, in accordance with the GRADE process. Qualitative research evidence can add value or complement quantitative evidence by providing an in-depth understanding of the question of why things are the way they are, rather than how much they are a certain way (e.g. why something is acceptable or feasible rather than to what degree people find something acceptable or feasible). Importantly, qualitative evidence also describes

participant-reported experiences that are not adequately reflected in quantitative outcomes, especially in the absence of validated participant-reported outcome measures (PROMs). GRADE CERQual is a transparent and structured approach for assessing how much confidence to place in qualitative review findings (i.e. "to assess the extent to which the review finding is a reasonable representation of the phenomenon of interest") (22). The review findings are the results of a qualitative evidence synthesis and can be presented in different formats (e.g. a theme, category, thematic framework, theory or contribution to theory) and at different levels (e.g. descriptive or aggregative and interpretive or narrow; for example, in relation to a specific health care setting or more broadly cutting across several different kinds of social care settings). At least two members of the review team will arrive at CERQual assessments for each review finding through discussion of four key components, with equal weight given to each component:

- methodological limitations of included studies;
- coherence of a review finding;
- adequacy of data; and
- relevance of included studies to the review question.

More detail on GRADE CERQual and qualitative research methods that can be used to generate evidence to support WHO guideline development on new TPT regimens is provided in Section 5.

3.5 Preparing evidence profiles and summary of findings tables

Evidence profiles are tables that display the ratings of the certainty of evidence together with summary effect estimates in a standardised format; summary of findings are tables that show abbreviated versions of the evidence profiles (31). These tables are a core element of the guideline development process. They represent the main format in which evidence is presented to the GDG members to support their judgements about the magnitude of desirable and undesirable effects.

3.6 Making judgements across 12 EtD criteria

Once the evidence has been retrieved, summarised and rated for certainty, WHO convenes a meeting of the GDG, where a summary of findings tables and other information are presented and discussed using a format of structured deliberation, under the guidance of a guideline methodologist. The outputs of the discussions are captured in a so-called EtD table (evidence to decision table), which shows how the factors that determine the direction and strength of a recommendation inform the process of developing the recommendation. These tables enhance the transparency of the process, focus the discussions of the GDG and permit recording of the judgements made about each factor and how each one contributed to the recommendation. Table 4 explains the 12 EtD criteria typically evaluated as part of the overall assessment of the evidence. Detailed guidance regarding the application of evidence to the EtD criteria and how GDG judgements are informed is provided in Section 4 of this document.

Table 4. Overview of the 12 EtD criteria typically evaluated as part of the overall assessment of the evidence

EtD criterion (GEG section)	Signalling questions
1. Problem	**Is the problem a priority?** Providing background on why the problem is a priority (not discussed in this document).
2. Desirable effects (3.6.2)	**How substantial are the desirable effects?** How large are the desirable effects of the intervention, considering the importance of the outcomes (i.e. how much they are valued), and the size of the effect (i.e. what is the likelihood of experiencing a benefit or how much of an improvement would individuals be likely to experience)? Here, the summary of findings table is displayed for the outcomes that favour the intervention.
3. Undesirable effects (3.6.2)	**How substantial are the undesirable effects?** How large are the undesirable effects of the intervention, considering the importance of the outcomes (i.e. how much they are valued) and the size of the effect? Here the summary of findings table is displayed for the outcomes that favour the comparator.
4. Certainty of evidence (3.4)	**What is the overall certainty of the evidence of effects?** How good an indication does the research provide of the likely effects across all critical outcomes (i.e. the likelihood that the effects will be sufficiently different from what the research found that it might affect a decision about the intervention)?
5. Values (3.3)	**Is there important uncertainty about or variability in how much people value the main outcomes?** How much do individuals value each of the main outcomes? Is uncertainty about how much they value each of the outcomes sufficiently large that it could lead to different decisions?
6. Balance of effects (3.6.2)	**Does the balance between desirable and undesirable effects favour the intervention or the comparison?** What is the balance between the desirable and undesirable effects, considering how much individuals value the main outcomes, how substantial the desirable and undesirable effects are, the certainty of those estimates, discount rates, risk aversion and risk seeking?
7. Resources required (4.9)	**How large are the resource requirements (costs)?** How large is the cost of the difference in resource use between the intervention and the comparator?
8. Certainty of evidence of required resources (3.4)	**What is the certainty of the evidence of resource requirements (costs)?**
9. Cost–effectiveness (4.9)	**Does the cost-effectiveness of the intervention favour the intervention or the comparison?**

	Is the intervention cost-effective, considering uncertainty about or variability in the costs, uncertainty about or variability in the net benefit, sensitivity analysis and its reliability, and applicability of the economic evaluation?
10. Equity (4.1.12)	**What would be the impact on health equity?** Are there plausible reasons for anticipating differences in the relative effectiveness of the interventions for disadvantaged subgroups or different baseline conditions across disadvantaged subgroups that affect the absolute effectiveness of that intervention or the importance of that problem?
11. Acceptability (4.7)	**Is the intervention acceptable to key stakeholders, in relation to the comparator?** Are key stakeholders likely not to accept the distribution of the benefits, harms, and costs, or are the costs or undesirable effects in the short term worth it to gain desirable effects (benefits) in the future? Are the stakeholders likely to disagree with the values attached to desirable or undesirable effects, or to not accept the intervention because of ethical concerns?
12. Feasibility (5.1, 4.1.3)	**Is the intervention feasible to implement in relation to the comparator?** Is it feasible to sustain the use of the intervention and to address potential barriers to using it?

EtD: evidence to decision; GEG: guidance on evidence generation.

3.6.1 Judgement of the magnitude of desirable and undesirable effects

During evidence assessment, outcomes are referred to as desirable and undesirable based not on their inherent nature (e.g. death is undesirable, lack of incident TB is desirable) but depending on whether the observed effects for a certain outcome favour the intervention or the comparator. Thus, outcomes for which effects favour the intervention will be listed as “desirable effects”, whereas those that favour the comparator will be listed as “undesirable effects” within the EtD tables. The GRADE EtD framework then classifies effect sizes for quantitative outcomes as trivial, small, moderate or large. This categorical determination is made based on a collective judgement by the GDG (see also Section 3.3). Judgements on the magnitude of desirable and undesirable effects are influenced by how guideline panels rate the effect sizes and the relative importance of prioritised outcomes.

The relative importance of the desirable and undesirable outcomes of TPT is not well understood. That is, the extent to which people treated for TB infection prioritise their reduced risk of developing TB disease (a ‘desirable outcome’) in comparison to their concerns about AEs (an ‘undesirable outcome’) has not been clearly identified for TPT, unlike for TB disease (32). Individuals’ priorities about whether to take TPT are particularly important because people with TBI are typically healthy, lack symptoms of TB disease, and most have a lifetime risk of developing incident TB of less than 10% (33). For this reason, their willingness to accept potential risks, harms and costs to receive the benefits of TPT might be considerably less than for people being treated for TB disease. Those who test negative for TBI and therefore have an even lower absolute risk of incident TB, may also consider the risks of treatment to outweigh the benefits. Hence, the frequency of both low-grade AEs and serious adverse events (SAEs) should be considered when assessing the undesirable effects of a TPT regimen.

3.6.2 Balance of desirable versus undesirable effects

The balance of effects reflects the risk–benefit ratio of an intervention (often referred to as the balance between "benefits and harms"), considers the overall certainty of the evidence and how the outcomes are valued by those receiving it. It is thus based on the combination of judgements on the four EtD criteria (desirable effects, undesirable effects, certainty of the evidence and values). This judgement about the balance of effects is a strong determinant of the direction and strength of the final recommendation for TPT, even after considering the other important GRADE criteria.

Having a well-informed judgement for both benefits and harms is an important consideration for GDGs, underscoring the need for reliable evidence from at least one Phase 3 trial.

3.7 Developing recommendations

Recommendations are developed based on the judgments made across the 12 EtD criteria, which may be displayed as a summary to serve as a basis for discussion. Typically, four main factors determine the direction and strength of a recommendation in public health:

- the certainty of the evidence (Section 3.4);
- values related to the health outcomes (Section 3.3);
- the balance of benefits and harms (Section 3.6.2); and
- resource implications (Section 4.9).

Five types of recommendations may be made:

- strong recommendation against the intervention;
- conditional recommendation against the intervention;
- conditional recommendation for either the intervention or the comparison;
- conditional recommendation for the intervention; and
- strong recommendation for the intervention.

Table 5: Summary of key factors influencing the decision to give strong or conditional recommendations

	A strong recommendation may be justified if:	In general, we should expect a conditional recommendation when:
Overall confidence in effect estimates	There is high or moderate confidence in effect estimates (or in special circumstances when the confidence is low or very low) AND	There is low or very-low confidence in effect estimates OR
Balance between benefits and harms	The benefits clearly outweigh the harms or vice versa AND	The balance between benefits and harms is close OR
Uncertainty and variability in patients' values and preferences	All or almost all fully informed patients will make the same choice. AND	There is variability or uncertainty in what fully informed patients may choose OR

Resource considerations (optional)	The benefit of the intervention is clearly justified (or not) in all or almost all the circumstances	The benefit of the intervention may not be justified in some circumstances

3.7.1 Strong recommendations

When we can be very certain about the balance of effects (i.e. the desirable consequences clearly outweigh the undesirable consequences or vice versa, and the certainty is high or at least moderate), and other EtD criteria support this, WHO may issue a strong recommendation in favour of or against an intervention. The implications of strong recommendations are that the recommendation can be adopted as policy directly by most Member States, most clinicians would follow it, most patients would want the recommended course of action, and additional research is unlikely to alter the recommendation (34). Currently, the WHO GDGs has issued a few strong recommendations on TPT. Many of its conditional recommendations resulted from limitations in the available evidence base (2). The weak evidence base and resulting conditional nature of many recommendations influence the wider uptake and implementation of WHO guidelines (35). A few paradigmatic situations where strong recommendations may be made despite the evidence being of low or very low certainty are outlined in Annex 5.

3.7.2 Conditional recommendations

When we are uncertain about the balance of effects or where the balance may depend on circumstances specific to an individual or context (e.g. based on judgements on other EtD criteria), WHO will typically issue a conditional recommendation. The implication of conditional recommendations are that substantial debate may be required before the policy is adopted by Member States; clinicians may need to discuss different management options with each patient; most patients may want the recommended course of action, although some or even many may not; and additional research would be likely to strengthen and possibly alter the recommendation (34).

3.7.3 Conditional recommendation for either the intervention or the comparison

Guideline users benefit from clear recommendations. A conditional recommendation for either the intervention or the comparison should be reserved for rare situations when two alternative intervention options appear to have equivalent net desirable consequences across the EtD criteria after careful evaluation. This option should not be chosen if an intervention is compared with current practice or no intervention – this will not provide guidance and will often be meaningless. Furthermore, a conditional recommendation for either the intervention or the comparison may be based on a comparator that has a strong recommendation as a basis (e.g. if it was previously compared with no intervention); logically, this suggests that the new intervention would also be strongly recommended if compared with no intervention.

In summary, the balance of desirable and undesirable effects (often referred to as the balance between "benefits and harms") is an important determinant of the strength of a recommendation in the GRADE

framework for TPT. This is the most time- and resource-intensive component of evidence generation, compared with other forms of evidence generation, since it normally requires at least one Phase 3 trial.

4. Guidance on evidence generation for TPT

This section describes key messages from WHO to consider across major study protocol elements and related considerations that may help to increase the strength of recommendations. They are intended to help ensure that clinical trials are conducted in a manner likely to result in a strong recommendation from a Guideline Development Group (GDG). These messages address general concerns, as well as considerations relating to the trial population, intervention, comparator, outcomes, generalisability, and statistical issues relevant to trial design. Although most of these messages specifically target researchers developing clinical trial protocols, many apply to non-trial research as well. Suggestions for additional sub-studies or aspects to consider when designing trial and study protocols are also described here. These additional studies may facilitate the implementation of WHO recommendations within NTPs.

4.1 Research design

Key message #1: Consider the requirements for the WHO guideline development process during research design

When designing trials, researchers should consider how the data generated can be presented to the WHO-convened GDG and inform global policy recommendations. Early during the protocol development, researchers should consider engaging with the WHO and policymakers within countries to identify relevant policy gaps. Study endpoints can then be defined in such a way that they are relevant to future policy recommendations. Early interactions with WHO technical teams can also enable the trial design to be aligned with the priorities identified in the TPP for TPT (4, 5).

Why this is important?

Trial design needs to be chosen to support the specific aims of the trial (36). Understanding the WHO's guidelines development process and its related data needs can help overcome common reasons that lower the certainty in evidence or the strength of recommendations. Downgrading of evidence due to imprecision has been by far the greatest hurdle to achieving moderate or high certainty evidence based on single trials; therefore, finding ways to conduct more or larger high-quality trials is critical.

4.2 Study Populations

Key message #2: Be more inclusive in the selection of populations and settings

In efficacy trials, the trial participants should reflect the demographic, geographic, and epidemiological diversity of the target population for the study regimens while preserving trial internal validity. While many trials for TPT include people with HIV, other key subpopulations are less frequently represented (Annex 6). High-priority populations for inclusion may be those with important co-morbid conditions (e.g., silicosis, diabetes, or cardiovascular diseases or abnormalities) and other specific populations (e.g., infants, children, pregnant and breastfeeding women, older people, and undernourished people of any age). Although several existing TPT trials include children, fewer focus on elderly populations (Annex 6).

Additionally, in high-transmission settings, trials involving the re-treatment of individuals previously treated for TB infection may also be valuable.

Clinical trials of TPT should aim to enrol populations from the settings where the subsequent policy guidelines will be applied. Researchers are encouraged to conduct trials in a wide range of epidemiological settings. This should include individuals at high risk of TB disease from diverse geographic regions (e.g. rural and urban areas, or countries with varying disease burdens). It is recognised that the cost of conducting multi-country studies may exceed available resources. In such instances, researchers may consider including a diverse range of sub-national regions to enhance the broader applicability of the findings and strengthen subsequent recommendations.

Exclusion criteria should be thoroughly justified and focused on identifying individuals whose participation would expose them to undue risk compared to potential benefits (e.g. due to insufficient dosing information or specific medical history). Children or pregnant women, in particular, should not be excluded unless there is a strong medical or scientific justification for doing so (34).

Why is this important?

WHO Guideline Development Group (GDG) examines the entire body of evidence when formulating a recommendation. To maximize the generalizability and utility of these recommendations, it is imperative to ensure the inclusion of diverse, clinically relevant populations and epidemiological settings. Exclusion of key cohorts, such as individuals living with HIV or contacts of patients with multidrug-resistant/rifampicin resistant (MDR/RR)-TB, would inherently constrain the scope and applicability of the GDG's recommendations. Excluding certain population groups can result in such groups never benefiting from the innovation under study or benefiting only with significant delay; it can also mean that clinicians and individuals bear the risk of giving treatment beyond the scope of available evidence.

The populations and settings selected significantly influence the assessment of evidence applicability and directness during the GDG meetings. Availability of consistent TPT outcomes across different populations and settings provides greater certainty regarding a regimen's effectiveness and tolerability. Hence, recruiting participants from a range of populations and settings will likely provide stronger evidence than a single trial conducted in one setting. Inclusive eligibility criteria increase the relevance of research findings to target populations and settings and reduce the risk that the certainty of evidence is downgraded due to indirectness when WHO recommendations are developed (Case study #1). Including countries with a high TB burden can facilitate the extrapolation of recommendations to all regions and countries and later accelerate the uptake by countries, given that the findings will be directly applicable to those settings. This approach thus promotes equity in access to novel TPT regimens.

Case study #1

In 2020, the WHO GDG evaluated evidence regarding the 1-month daily rifapentine and isoniazid (1HP) regimen for TPT in people with HIV, from the BRIEF TB study (37), a randomized, open-label,

Phase 3, non-inferiority trial. This study compared the efficacy and safety of 1HP against 9 months of isoniazid, with the primary endpoint being the diagnosis of incident TB disease or death.

When this study was reviewed by the GDG, the certainty of the evidence was downgraded due to indirectness for several reasons: the study exclusively enrolled people with HIV, rather than all individuals at risk of TB; TB infection was not confirmed in approximately 80% of participants; and the comparator was the 9-month isoniazid regimen (9H) instead of the more widely used 6-month regimen (6H).

Although the trial did not include an untreated control group, the GDG determined that the benefits observed in the 1HP arm among participants who were TST or IGRA positive would likely extend to other at-risk populations. Consequently, the recommendation was expanded to include HIV-negative individuals over 13 years of age. In this recommendation, the GDG stated that "if the findings can be replicated by other studies, the confidence in the estimates would increase." Based on the above considerations, the overall evidence was downgraded to moderate certainty, and the regimen was conditionally recommended.

4.3 Intervention regimens

Key message #3: Include implementation considerations in the TPT study protocol

When developing recommendations regarding TPT regimens, GDGs consider not only the drug composition and dosing but also crucial implementation factors. These include the methods of drug delivery and monitoring, as well as elements influencing feasibility and effectiveness. Specifically, the regimen's formulation, the process of screening for TB disease prior to enrolment, the tests for TB infection used, the frequency of clinic visits, and the monitoring methods all play a significant role. Therefore, when designing trials, researchers should prioritize interventions that are feasible to implement at scale within a programmatic setting (38).

The screening methods used should be those widely available in resource-limited settings to ensure trial findings are applicable in such contexts. Radiology and WHO-approved molecular rapid diagnostic tests are recommended for screening and can also detect asymptomatic TB (Case study #2) (39). Regimens and formulations likely to be readily accessible to National Tuberculosis Programs (NTPs) outside the trial setting should be prioritized. For children, paediatric formulations (typically dispersible, palatable, and scored) should be preferred over adult formulations. Adherence to the WHO harmonized weight bands for dosing is desirable (3).

Factors such as the methods of adherence monitoring, AE management, use of drugs in combination with companion drugs (such as antiretrovirals), drug susceptibility test (DST) results of the index case, maximum allowed treatment duration, and the approach to management of treatment interruption will influence trial outcomes and are relevant to GDG discussions. While clinical trials are needed to evaluate optimal approaches for managing treatment interruptions, this specific area may be best examined

within larger cohorts. Phase 4 trials, conducted after a regimen's efficacy has been demonstrated, are particularly well-suited for this type of investigation. Such methods should be described in the trial protocol and manuals of procedures and be available to GDGs for both intervention and comparator regimens.

Why this is important:
Providing comprehensive evidence on key elements in the delivery of the intervention is critical for GDGs to reach informed judgements across EtD criteria, the certainty of evidence, and implementation considerations. Understanding precisely how an intervention was delivered in a trial is essential. It allows GDGs to assess the feasibility of delivering that same intervention under real-world programmatic conditions. Furthermore, it helps determine the potential impact on observed outcomes if there are deviations from the trial's conditions (for example, in terms of treatment support). This detailed information also serves to directly inform the implementation guidance that accompanies GDG recommendations.

Case study #2
The TB-YOUTH study investigated short-course TB preventive treatment (TPT) among students who were close contacts of TB patients. In the initial screening phase, approximately 10,000 students who were contacts were tested, and 8% of them were found to be IGRA-positive. Further screening for TB disease among these IGRA-positive individuals revealed 14 individuals with symptomatic TB and 33 asymptomatic TB. Of the asymptomatic individuals, 10 were bacteriologically confirmed. These results highlight the critical importance of thoroughly excluding TB disease, including asymptomatic forms, before starting TPT in clinical trials (40).

4.4 Comparator regimens

Key message #4: Include an appropriate comparator arm in the trial

WHO guideline questions, often framed as PICO questions, are inherently comparative. Therefore, it is crucial to include a relevant, randomly enrolled comparator arm. Participants in this arm should receive standard-of-care (SOC) regimens consistent with WHO recommendations current at the time of the guideline review. For trials aiming to shorten the duration of treatment, the comparator regimens have often been a longer WHO-approved standard regimen. Where no SOC is recommended, a placebo may be used (Case study #3). If no internal comparator regimen is included and historical comparators are used in a trial, the strength of recommendations will be downgraded due to indirectness, selection bias, or unmeasured confounding. It is important to be sufficiently clear about inclusion and exclusion criteria and other decision-making to permit the selection of relevant non-concurrent comparators from other studies. Analyses based on external, non-concurrent controls typically result in very low certainty evidence.

Given that the time between trial design and reporting can span many years, WHO recommendations may change during this period. As a result, researchers may change the comparator regimen during the

trial. In such cases, stratified analyses should be presented to compare the effectiveness and tolerability of different regimens to help interpret the outcomes.

Case study #3
The selection of a comparator depends directly on the trial's aim. A placebo-controlled design is optimal when the goal is to prove efficacy. The PROTID trial compares three months of weekly isoniazid and rifapentine against a placebo in individuals with diabetes (41). Given the current lack of specific TPT recommendations for this patient group, using a placebo as a control is acceptable. This design is expected to yield critical efficacy data for this at-risk population.

Another example is the BALANCE trial, an open-label, randomized controlled trial that evaluates the effectiveness, safety, and benefit-risk profile of 1HP compared to placebo in people with diabetes (42). Its objectives include assessing the feasibility, acceptability, and overall benefit of TPT in this population, alongside any proportionate harm. Individuals with diabetes often manage multiple medications for their condition and other comorbidities. Consequently, the additional pill burden from 1HP could decrease acceptability and negatively affect adherence, potentially impacting even antidiabetic treatment adherence.

Why is this important?

Benchmarking new regimens against existing regimens will enable changes to existing recommendations to be evaluated and incorporated into updated WHO guidelines. An appropriate comparator arm allows accurate assessment of the efficacy of the new TPT regimen with reference to existing recommended regimens. It also allows a comparison of the safety profile of the two alternative regimens, which is particularly important for TPT trials, given that tolerability is an important priority for otherwise healthy people who take these drugs. The outcomes of trials that lack an internal comparator (control) group are difficult to interpret and do not lend themselves to strong recommendations. Comparisons to external populations, and not concurrently enrolled comparator groups, are subject to selection bias and confounding, leading to very low certainty in the evidence, making it impossible to reach any clear conclusions with confidence. Head-to-head comparisons between different shorter regimens (e.g., 3HP vs 1HP) may enable stronger recommendations than indirect comparison using network meta-analysis.

4.5 Outcomes

Key message #5: Report on outcomes of importance for guideline development

Although incident TB is the primary outcome assessed in most clinical trials of TPT, existing trials report varied primary and secondary efficacy and safety outcomes, making the comparison difficult (Annex 6). The preferred primary outcome of TPT trials for treated individuals is survival without incident TB after an adequate follow-up period (Key message #8). Trial outcomes should be clearly defined in protocols, including definitions for incident extrapulmonary and paediatric TB. Outcomes that the GDG has previously rated as important or critical in developing recommendations are presented in Table 6, below. For the incidence of TB disease, it is important to differentiate between microbiologically and

clinically diagnosed TB in the results. Annex 6 presents the outcome measures used in clinical trials of TPT published since 2000.

Table 6: Outcomes rated by Guideline Development Groups as important or critical for TPT regimens

1	Incidence of TB disease (in all forms)
2	Incidence of TB disease (microbiologically confirmed)
3	Incidence of TB disease among participants who test positive at entry on an antigen-based skin test, such as TST/TBST or IGRA
4	Mortality (all cause) during TPT
5	Mortality (related to drug) during TPT
6	Serious adverse events (related)
7	Adverse events (grades 3-5)
8	Adverse events (related grades 3-5)
9	TPT completion (ever)*
10	Incidence of TB disease (microbiologically confirmed) in people with HIV
11	Incidence of TB disease (all forms) in people with HIV
12	Incidence of TB disease among participants who are ART-naïve at entry
13	Incidence of TB disease among participants who are on ART at entry
14	Adverse events (grades 3-5) in people with HIV
15	Adverse events (related grades 3-5) in people with HIV
16	Adverse events of special interest (e.g., grade 3-5 hepatotoxicity, tendinopathy, or QT prolongation)
17	Adherence to TPT (proportion of doses taken)
18	Changes in the gastrointestinal or respiratory microbiome
19	Acquired drug resistance

**Treatment completion may be defined according to the proportion of treatment completed within an assigned time (such as at least 80% of doses completed within 150% of the minimum treatment period).*

***Pre-specified definitions may be used to define possible or probable incident clinical TB. Expert clinical panels, blinded to treatment group, or local clinicians may apply standard definitions in order to assign outcomes.*

In some trials, two or more outcomes are presented as a composite primary outcome, without evaluating each component separately. However, the use of composite outcomes alone can mask the effect of two component outcomes that have effects in different directions. The use of composite outcomes can result in downgrading due to indirectness. Disaggregating composite endpoints is essential, given that judgements about the magnitude of effects are made by evaluating each outcome separately. That is, the consequences or severity of some outcomes may be more important to patients and clinicians than others.

Why is this important?

Trials that report on outcomes considered critical or important by a GDGs are most likely to result in a WHO recommendation. The primary goal of TPT is to prevent incident TB disease, so study outcomes should directly measure this. Such standardized reporting can facilitate network meta-analyses and allow comparison between TPT regimens that have not been evaluated directly, notwithstanding concerns about the indirectness of evidence. Reporting data on treatment adherence is important to allow for ITT and per-protocol analysis to determine the likely level of compliance with the use of the

novel TPT regimen. Mortality and safety outcomes similarly reflect the direct impact of the TPT regimen on patient well-being and the overall risk-benefit profile of the intervention. Moreover, granular, outcome-specific analyses are essential to fully elucidate the efficacy, safety, and tolerability of TPT regimens, providing robust evidence for consideration of the GDG.

Key message #6: Use harmonized definitions for safety outcomes and report them comprehensively

Data on the incidence of AEs should be reported for all individuals who commence treatment, in an ITT analysis. The proportion of people starting treatment who experienced one or more of the following should be reported: SAE, AEs of Grade 3 or higher, grade 1 or 2 AEs (often under-reported in trials but associated with high discontinuation of TPT), treatment limiting AEs, and AEs of special interest, such as hepatotoxicity and neuropathy with isoniazid. It is important to harmonize the timing of AE assessment for unbiased comparisons of safety outcomes, and the data should be collected at regular intervals (e.g. monthly). Trials investigating new medicinal products must comply with the ICH standard (International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use) and report all grade AEs, including grade 1 and 2, as they may affect tolerability and acceptability. A recognized toxicity instrument should be used for reporting AEs; for example, Common Terminology Criteria for Adverse Events (CTCAE) (43), Council for International Organizations of Medical Sciences, or Division of Microbiology and Infectious Diseases (44, 45).

Secondary analyses considering attribution of AEs to treatment as determined by a blinded panel can add value in some circumstances, as can analyses on the rate (rather than the proportion) of AEs. AEs related to drug-drug interactions and AEs of special interest (such as hepatotoxicity) should also be reported. Patient reported outcomes can also provide GDGs with valuable insights into the priorities of individuals taking TPT. These may include an assessment of quality of life, taste of the medication, or measures of specific symptoms or forms of intolerance (see Key message #9). Hepatotoxicity is a particularly important complication of most established TPT regimens.

Why is this important?

Safety outcomes should be reported using standardised definitions, so that valid comparisons can be made between different trials and treatment regimens, and consistently interpreted. AEs that affect adherence indirectly affect efficacy. Understanding AEs is important in understanding the acceptability and balancing the risks of a novel regimen against potential benefits. Understanding the long-term impact is also important for individual acceptability. Most people taking TPT are otherwise healthy and likely unwilling to accept the substantial risks due to TPT.

Key message #7: Characterise the acquisition of drug resistance

Development of resistance in TB strains and human microbiome to component drugs during TPT should be measured, given the serious implications for future TPT options. Measuring acquired resistance due to TPT requires comparison of isolates grown from a known source case and those from their contacts who received TPT, using genomic testing to compare mutations. In trials of TPT for household or close contacts of people with TB disease, DST results of the isolate in the index case should be systematically reported. Where this is not feasible, the isolates from index patients may be stored for later analysis should the contacts develop incident TB. In children with incident TB, the collection of isolates for DST may be difficult. Overall, the number of people on TPT who develop resistant strains will be limited even in large efficacy trials. A meta-analysis of cases may best estimate the potential of a TPT regimen to generate resistance. Sensitive TB screening algorithms (including symptoms, chest X-ray) should be used to exclude prevalent TB disease before starting TPT, to reduce the likelihood of treating TB disease with insufficient regimens. Nested studies that evaluate the role of TPT regimens upon the human microbiome should be considered for regimens that include broad-spectrum antibiotics, given the potential adverse impacts upon the diversity of gastrointestinal and respiratory tract flora and their acquisition of resistance to TPT.

Why is this important?

Understanding the risk of acquisition of drug resistance is essential for assessing the long-term effectiveness of TPT regimens and the potential impact of its widespread use on the future burden of drug-resistant TB and other bacterial resistance. Even though the evidence is scarce, the perception that TPT may cause drug-resistant TB is prevalent among healthcare workers (46) and this is a known barrier to TPT implementation. Having data that shows that TPT does not contribute to drug resistance may facilitate the uptake of TPT. Moreover, the acquisition of drug resistance while on TPT is infrequently reported. The TB strain of the index patients and the concurrent disease discovered among household contacts or the incident TB following TPT are at times not identical, especially in high-prevalence settings, so the acquisition of drug resistance from the source case cannot be inferred.

4.6 Duration of follow-up

Key message #8: Ensure sufficient follow-up time

Participants in both intervention and comparator arms should be followed for an equivalent period, typically for 24 months after treatment initiation or at least 12 months after completion of TPT in the longest regimen under evaluation. In high-transmission settings, extended follow-up may capture the increased likelihood of reinfection post-randomisation in both intervention and comparator arms. The time point for analyses of efficacy outcomes should be based on this fixed duration (i.e., equal total follow-up time from time of randomization between intervention and control) to avoid immortal time bias (a form of bias that arises when not all individuals have a similar chance to have the outcome of interest) (47). Retention of participants is important to avoid misclassification of incident TB post-TPT,

which is an uncommon event and may not be equally distributed between those who do and do not attend follow-up.

Why is this important?

Sufficient follow-up time is required to determine if TPT has been effective. Incident TB disease is most likely to occur in the first one to two years after infection with *M. tuberculosis* (48). Long-term follow-up data helps assess the durability of protection provided by different TPT regimens. Extended follow-up can reveal differences in efficacy that may not be apparent in shorter-term assessments. For treatment-shortening trials, assessing outcomes in all groups after a fixed duration from randomization aligns with recommendations from regulatory authorities for trials of TB disease (49) and is typically conservative, favouring longer comparator regimens. The duration of follow-up needs to balance the detection of incident TB against the potential for re-infection with a new strain of TB post-randomisation, particularly in high-transmission settings. TB that occurs after the completion of TPT comprises a substantial proportion of the incident cases of TB that occur post-randomisation. Since these events are relatively rare, capturing as high a proportion as possible is also important to avoid biasing trial results towards the null. Shorter periods of post-treatment monitoring and lower retention of participants during follow-up may reduce the number of events, reducing the power to detect a difference (or lack thereof) between groups.

4.7 Regimen tolerability and acceptability

Key message #9: Characterise tolerability and acceptability of the TPT regimens

Tolerability and acceptability are important constructs to assess when balancing the potential harms and benefits of the TPT regimens, as they can influence the likelihood that an individual completes treatment (38). Acceptability describes the overall satisfaction with the treatment, including factors like impact on quality of life, ease of use, and individual preference, and is a multidimensional construct encompassing cognitive, emotional, and practical responses that must be assessed across stakeholders, including policymakers, providers, and patients (50), (51).

Tolerability can be measured during trials using commonly reported participant reported outcomes (PROs). PROs have been defined as measures of a person's health condition that come directly from the individual, without interpretation by a clinician or anyone else. These could be an important complement to bacteriological and clinical outcomes during the guideline development process. Standardised participant questionnaires can be developed based upon the likely tolerability of the regimens being evaluated, preferably based upon standardised tools (such as the PRO-CTCAE tool, which measures participant experience with AEs whilst on treatment) (52). Any such tool should be demonstrated to reflect the person's experience with TPT that is valid, reliable, and responsive (i.e., has longitudinal validity).

Acceptability can be measured directly through stakeholder surveys, qualitative research, or other methods; it may also be reflected indirectly through adherence and discontinuation rates. During protocol development, Community Advisory Boards can provide input into participant perception of these factors, which impact acceptability. Acceptability of treatment incorporates usability, receptivity, and integration within the local context (53). When developing a TPT intervention, it is critical to think about what is likely to be acceptable to key stakeholders, and acceptability across key stakeholder groups should be evaluated directly. Quantitative studies can provide information on the acceptability (e.g., the percentage of a group finding the intervention acceptable), whereas qualitative studies can provide insights into why a particular intervention may be more or less acceptable and under what circumstances (see Section 5).

Why is this important?

Participant acceptance and tolerability are vital to the successful implementation of a TPT program at a national level. A positive participant experience can improve uptake of the regimen for current and future participants through word of mouth and shared experiences. Where an intervention is not acceptable to policymakers it is unlikely to be adopted by national programs; if it is not acceptable to health care providers, they may hesitate to use the intervention if they have alternative choices; and if it is not acceptable to individuals they will refuse to take it or will not adhere to it in the way it is intended to be used. It is particularly valuable to collect data using measures of tolerability and acceptability during trials, as this enables GDGs to incorporate the views of individuals receiving therapy into their decision-making. This is important in phase 3 trials in which the level of support provided to individuals to stay on treatment is frequently beyond the support that can be provided in most high burden countries (i.e. high efficacy of a regimen that can only be tolerated with substantial investment in adherence support may not be reproducible under programmatic conditions, reducing the likelihood of benefit for those who stop treatment early). Tolerability is a priority for individuals on TPT who are usually healthy and less willing to accept even low-grade symptoms than individuals with TB disease (5), (54). Poor tolerability can lead to treatment interruptions or discontinuation, reducing the regimen's effectiveness.

4.8 Treatment duration

Key message #10: Report the effect of the shortening of TPT regimens on clinical and health system outcomes

Where everything else is equal, shortening treatment duration is advantageous to both the individual and the health service. Outcomes that are typically included in the GRADE Evidence Profile (Annex 1) do not provide a direct, quantified indication of the effect of regimen duration. As a result, GDGs commonly infer this from the increased risk of AEs and lower rates of treatment adherence and completion, which are associated with longer treatment (55). This is not ideal. Trials that compare shorter treatments to longer SOC regimens could report the effect of the new shorter regimen upon specific outcomes of importance to participants, such as effectiveness, tolerability, or other PROs.

Why is this important?

Over the last fifty years, the duration of the commonly used TPT regimens has reduced from 9-12 months to 1-4 months, based upon non-inferiority trials of efficacy and safety (55). However, there has been limited research evidence regarding the relative importance that people with TB infection attach to treatment duration, in comparison to the safety and tolerability of these novel regimens. Shorter TPT regimens can improve convenience and minimize the time over which AE and drug-drug interactions can occur (56). Nonetheless, participants may not necessarily value a shorter regimen, for example, if it is less tolerable. Shorter treatments can also benefit health systems through reduced costs for treatment and monitoring, although some novel regimens may be more expensive initially. Given that individuals may have differing priorities, having a range of TPT options may be an advantage. Better metrics are needed to express the cumulative treatment burden, including the pill burden and indirect costs. Inclusion of such metrics in the GRADE evidence profile would allow the GDG to deliberate on the benefits of shortening a treatment early on in the EtD process, alongside other discussions on desirable and undesirable effects. In the absence of direct evidence on the benefit of duration-shortening, this discussion typically occurs later, when issues of cost, acceptability, and feasibility are considered.

4.9 Cost considerations

Key message #11: Gather evidence within trials regarding the resources required to deliver the TPT regimen

Health economic outcomes comprise an important component of the EtD framework. Health system costs that may differ between the treatment arms not only include the cost of drugs, but also other aspects of management (e.g. hospitalization costs, adherence support, outpatient visits, routine monitoring). Where feasible, cost data should be collected during trials from health system and/or participant perspectives for both intervention and comparator arms. Costs often vary substantially between countries; therefore, for multi-country studies, costing should be carried out across multiple settings to allow for broad representativeness. The true costs of scaling up TPT regimens, however, may differ substantially from costs within a trial, given that routine clinical encounters may be shorter and fewer scheduled visits required. For this reason, programmatic studies that measure costs can complement the in-trial analyses of phase 3 studies. The effects of varying cost parameters, such as changes in the costs of the investigational product, should be presented where possible. Costs from the health system perspective should be presented for a range of settings.

Why is this important?

Understanding resource requirements is essential for translating trial results into practice. It helps predict the feasibility of implementing TPT regimens in various healthcare settings and allows healthcare systems to better prepare for the adoption of new TPT regimens. The resources required to deliver TPT may affect the strength of recommendations. Some recommendations may be made conditional on the availability of sufficient resources within national programs. Regardless of WHO recommendations, the cost of interventions significantly impacts their adoption. Many NTPs may struggle to implement more expensive treatment regimens on a large scale, even if those regimens are “cost-effective”. Regimens

with novel drugs often incur higher costs than the SOC regimen. Therefore, it is crucial to provide evidence showing how a regimen with more expensive drugs can reduce overall costs for participants and remain affordable for programmes by leading to savings in non-drug expenses (57).

Key message #12: Report cost-effectiveness

Cost–effectiveness analyses (CEA) combine estimates of costs associated with resource use (for example, staff time, consultation time, laboratory equipment, medicines) with estimates of health effects, providing additional value to support decision-making. This approach is particularly important to consider if an intervention improves health outcomes but costs more than the current SOC. Uncertainties and variability in health effects, health system costs, resources used, and willingness to pay thresholds all need to be carefully considered. CEA should use parameters from a range of high and low resource settings if the study is conducted in multiple settings, to make findings more generalizable.

Why is this important?

CEA can inform GDG decisions by evaluating the incremental costs of a new TPT regimen (against an SOC) per incremental health improvement (Case study #4). In situations where the costs of an effective novel intervention exceed those of the SOC, formal CEA can have an important influence on the strength of the recommendation. CEA provides valuable insights for policymakers and helps determine which TPT regimens offer the best value for money, guiding decisions on which treatments to implement or prioritize. This information is crucial for making evidence-based decisions in resource-limited settings where budget constraints are significant.

Case study #4

A systematic review of studies published between June 2016 and September 2023 (2) identified one high-quality cost-effectiveness (CE) study of TB prevention with fluoroquinolone (FQ) for MDR contacts. This study found that CE was highest when implementing levofloxacin (Lfx)/moxifloxacin (Mfx) for children <5 and children <15 with HIV (ICER, US$738 per DALY) but it averted fewer total deaths and years of life lost than providing Lfx/Mfx for all children <15 (870 deaths averted compared to 1240 respectively). The CE of Lfx/Mfx decreased in countries with higher FQ resistance, with a greater number of contacts under the age of 15 years needing to be treated per TB episode averted. This analysis was updated with the efficacy estimates from the TB CHAMP and V-QUIN trials, and results were found to be similar. A sub-study of V-QUIN estimated that for every 1000 adult MDR contacts provided Lfx as TPT, compared to monitoring only, would result in: (i)A total health system cost saving of US$2,091, and a total health gain of 40.96 QALYs; (ii) Lfx TPT would also result in prevention of 0.56 MDR-TB cases. The WHO GDG judged that these data favour the intervention.

4.10 Statistical issues and sample size

Key message #13: Ensure sufficient sample size to achieve precise estimates for critical outcomes

Standard TPT regimens (e.g. 6/9 H) are highly effective. Demonstrating superiority of a new regimen is often impractical given the low event rates (progression to TB disease) and hence the requirement of a large sample size. TPT trials, therefore, frequently employ a non-inferiority design (58, 59) evaluating whether a new treatment is not worse than an existing treatment by a pre-specified margin (60). Meeting certain statistical criteria (e.g., statistical significance or non-inferiority) alone is not relevant to the decision making for WHO guideline development. Guidance regarding the statistical issues relating to non-inferiority studies has been published previously (61, 62).

One of the key considerations should be the choice of Non-Inferiority Margin that defines what is considered an "acceptable" loss of efficacy. This choice must be clinically justified and statistically sound, balancing the risk of adopting a less effective regimen against the potential benefits, also taking into account the effect size of the reference regimen, public health implications of small differences in efficacy, and added benefits (e.g. better adherence, safety). Secondly, the choice of participants (e.g. positive test for TB infection) and random allocation with strict allocation concealment are essential to avoid selection bias and false declaration of non-inferiority. Stratified randomization (HIV status, age, geographic locations by TB burden), combined with blinding of outcome assessors, will enhance the certainty of evidence. Both ITT and per-protocol analyses are typically required to confirm non-inferiority.

When determining the sample size of a trial, it is important to consider levels of desired precision for each important outcome measure, not just the primary outcome on which the trial hypothesis may have been based. It is beneficial to aim for a level of precision that avoids downgrading of the certainty of evidence by more than one level. Clinical trials of TPT have consistently observed a low incidence of TB among people with TB infection. Consequently, very large sample sizes are required for phase 3 trials of TPT where the primary outcome is incident TB.

Why is this important?
A sufficient sample size is essential to detect clinically meaningful differences between treatments with adequate statistical power, obtain precise estimates of treatment effects, as indicated by narrow confidence intervals, and reduce the risk of false-negative findings, which could lead to rejecting potentially effective treatments. In non-inferiority trials, precision is particularly important because the goal is to demonstrate that a new treatment is not worse than the SOC.

It is sometimes assumed that meeting protocol-defined criteria, such as “statistical significance” or “non-inferiority,” equates to having sufficient evidence to support strong recommendations for an intervention. This is not generally true; for example, because:

- the GDG considers evidence beyond the benefits and harms;
- statistical criteria are typically protocol-defined for only some of the outcomes; for example, they may be limited to efficacy outcomes but may not be reported for safety outcomes or the acquisition of drug resistance; and

- using a single p-value or any other arbitrary measures or boundaries to dichotomize estimates of effects as being either “present” or “absent” is generally discouraged (63).

A single trial often does not provide sufficient evidence required for a strong recommendation (although it is possible), on account of the need for reproducibility, potential limitations in sample size and power, risk of biases such as selection bias, and a lack of generalizability to all sub-populations. The traditional regulatory “two-trial rule” has not been applied in TB in the way it has in some other therapeutic areas, as conducting multiple trials can be infeasible and other relevant research (such as evidence of safety and tolerability in other populations) may provide sufficient supporting data for a GDG to make a recommendation. Thus, if the only available trial is not large enough to adequately address most or all critical outcomes, additional data, preferably from additional randomized trials or high-quality observational cohorts (in the absence of trials), may need to be generated to strengthen GDG recommendations.

Key message #14: **Consider the possibility of extrapolating study findings beyond the trial population**

Inclusivity and direct evidence of the effectiveness of TPT for populations of interest are generally preferable. However, extrapolation of trial results to some excluded populations can be justifiable and may allow recommendations to be extended to such populations. This possibility should be considered carefully at an early stage when planning a trial, to include any ancillary studies that may be needed. However, if certain populations are excluded from a trial and thus no direct evidence exists, it is sometimes possible and desirable to use indirect evidence to extrapolate the findings. Guidance from regulatory authorities suggests, for example, that evidence of efficacy from trials in adults may be extrapolated to children; thus, only studies to determine the appropriate dose and safety, and tolerability among children with the recommended dose may be needed (64).

Mathematical modelling can provide insights into the population level impacts of an intervention that has been shown to be effective in the context of a clinical trial. After the trial data is reported, trialists may consider undertaking modelling to extrapolate findings to the different target populations. Although modelling has limitations, such as the extent to which assumptions in the model align with the true parameters in the target population, when accompanied by trial data, they may provide useful information for a GDG to determine the effect of the intervention on health equity

Why this is important

Excluding certain population groups in clinical trials can result in such groups never benefiting from the innovation under study or benefiting only with significant delay; it can also mean that clinicians and individuals bear the risk in extrapolating beyond the evidence. Whilst it is not always possible to have all population sub-groups in a study, having a means to extrapolate into the wider population can enable NTPs to apply treatment regimens to the most eligible populations.

4.11 Data sharing

Key message #15: Share individual participant data

De-identified individual participant data, which respects participant confidentiality, should be made widely available following the trial, preferably through established data repositories where data can be found and obtained through secure and standardized processes, e.g. global individual patient data (IPD) platform for tuberculosis treatment supported by WHO (65). Creation of a similar setup for TPT could provide benefits to researchers and help assessment of the programmatic feasibility of TPT interventions. Sharing the statistical code used to calculate the primary outcomes of the trials also increases transparency of the analytic process.

Having publicly available data can also facilitate alignment of outcome definitions and harmonised data collection in the future. (Case study #5).

Why this is important:

Making de-identified individual participant data publicly available provides the possibility for individual participant meta-analyses and other research to gain further insight and understanding of TBI. Providing open and equitable but secure access to trial data offers the greatest opportunity for learning and is in the spirit of open data (66). The use of public data repositories can facilitate data sharing, dissemination, and access for further research on existing data (61, 67). Sharing individual participant data from TPT trials can facilitate the comparison and combination of data from different studies and facilitate meta-analyses. Combined analyses can increase the generalisability of the study findings and increase the directness of the evidence. It may also enable verification of the conclusions and improve data validity and support testing of new hypotheses using existing data, maximizing the value of research investments. This may also enhance the ability of GDGs to identify subgroups that may benefit most from TPT.

Case study #5

The effectiveness of Lfx as TPT for contacts of individuals with MDR/RR-TB was evaluated in two independent clinical trials in Vietnam (VQUIN) and South Africa (TB CHAMP). The VQUIN trial (68) enrolled primarily adults, with some adolescents and children, while the TB-CHAMP trial (69) investigated Lfx treatment predominantly in children and adolescents. Collaboration between investigators during trial design ensured participants of all ages were collectively recruited between the two trials. Researchers from both trials collaborated before unblinding results to design a combined analysis strategy. Combining data from these two trials from two geographical locations, through a pre-planned individual patient data meta-analysis, increased the generalisability of the study finding. Bayesian analyses enabled direct estimates of the probability that Lfx reduced the incidence of TB. The separate trial findings, as well as an individual participant data meta-analysis combining trial findings were presented to the WHO GDG. A systematic review was also commissioned by WHO. These combined analyses contributed to the strong recommendation issued by WHO, with moderate certainty of the estimate of effects.

4.12 Health equity

Key message #16: Investigate the impact on health equity

WHO guidelines should support the equitable care and meet the needs of all populations, particularly the vulnerable. Equity in relation to recommendations for TPT may be evaluated according to the following steps:

- identify populations who may experience inequities as a result of the TPT intervention,
- determine the baseline risk for prioritized outcomes in these populations,
- evaluate whether these populations have been included in the trial,
- conduct analyses for these populations, if possible,
- identify barriers to the implementation of interventions within populations experiencing inequities.

Equipped with these analyses, GDGs can understand current health inequities and evaluate how the introduction of the TPT intervention may affect existing health inequities or introduce new inequities (70, 71). The ability of a TPT regimen to address health inequities may also be influenced by acceptability (see also Key message #9).

To ensure that trials benefit populations in the countries where they are held, it is recommended that clinical trialists promote access to study drugs beyond the life of the trial (72).

Why is this important?

Guidelines can play a crucial role in promoting health equity for populations who are vulnerable or marginalized by explicitly considering how recommendations affect them. This requires explicit consideration of whether and how the introduction of a novel intervention may improve or worsen existing health inequities or lead to new ones. For example, less complex interventions that could be implemented widely and made accessible to all populations (including high-risk, remote, underserved, or other marginalised groups) are typically more likely to increase equity. However, such effects may differ between population groups (e.g. some interventions may increase equity for some populations but decrease it for others). Other approaches that may improve equity include changes to care pathways, drug-safety monitoring for those at risk for toxicity, and strategies to strengthen drug supply chains or drug storage.

TB infection disproportionately affects at-risk populations. Understanding how TPT regimens impact different groups can help reduce these disparities. Identifying concerns about equity helps identify potential barriers to access for certain populations and guides the development of strategies to improve availability and uptake of TPT across all subgroups. Quantitative data about the effect of an intervention upon equity can help policymakers make informed decisions to implement TPT and direct resource allocation to reach those most in need.

4.13 Feasibility

Key message #17: Evaluate the feasibility of implementation of TB preventive treatment

The feasibility of an intervention refers to the likelihood that it can be carried out appropriately in a particular context. Potential barriers to the implementation of the intervention should be assessed directly through stakeholder surveys, qualitative research, or other methods, among participants, providers, and policymakers. These can be incorporated within the trial or performed separately. Ideally, evidence should be generated on how those barriers can be addressed. Elements that deserve consideration include the requirements for drug availability, drug-drug interactions, drug-safety monitoring, adherence monitoring, dosing frequency, the number of component drugs, drug procurement and supply chains, drug stability and shelf life, and the possibility of fixed dose combinations and child-friendly formulations.

Why is this important?

The feasibility of a novel TPT regimen can have important implications for the strength of recommendations and their uptake. Considerations regarding feasibility are often based upon deliberations about barriers to implementation (73). Logistical considerations relating to the implementation of a novel TPT regimen will determine whether those barriers can be overcome in most or all settings. Evidence related to the feasibility of addressing these potential challenges can be important in the GDG process.

4.14 Non-randomized studies

Key message #18: Data on the safety of novel regimens from large observational studies are important to identify infrequent but important adverse events

While Phase 3 trials of TPT typically enrol between 1,000 and 4,000 participants, this sample size may not be sufficiently large to detect rare but clinically important AEs. Additional data from cohort studies beyond a clinical trial may be of value, such as large longitudinal cohorts. Analyses of safety in such prospective cohort studies should be conducted using approaches that explicitly aim to address questions related to causality, address potential biases and confounders, and identify assumptions about how these factors relate to study outcomes transparently (Case study #6). For example, the use of Directed Acyclic Graphs provides a visual depiction of the relationship between potential confounders that can make the clinical assumptions underlying a statistical model more explicit (74). Such results should be carefully interpreted in the context of evidence of safety concerns from clinical trials and observational data.

Why this is important:
Some aspects of treatment with TPT may not have been evaluated in randomised trials and therefore rely on non-randomised studies, for example, unexpected safety concerns. Observational data from cohort studies may be considered by GDGs where trials are not conducted due to a lack of clinical equipoise, or where ethical considerations preclude conducting a trial in a certain population or setting. These studies also shed light on the practicalities and feasibility of implementation under programmatic conditions. When interpreting non-randomized data, potential sources of bias need to be identified, considered, and accounted for in statistical analyses (75-77).

Case study #6
For the WHO TPT guideline update in 2020 (78), a systematic review was conducted assessing adverse pregnancy outcomes for those using TPT (79). While one RCT showed an increased risk of adverse pregnancy outcomes in women taking TPT during pregnancy, three observational studies demonstrated otherwise, and meta-analysis confirmed this finding in the observational studies. As a result, the GDG was able to conclude that there was insufficient evidence to change the current recommendation on the use of TPT among women who are pregnant and that deferral of TPT would result in more women being vulnerable to developing TB disease. However, the GDG emphasized that further research is required to strengthen the evidence and recommendations in this sub-population.

4.15 Pharmacokinetic studies

Key message #19: Trials evaluating novel drugs should be accompanied by pharmacokinetic studies in target populations where possible, to inform dosing recommendations.

Adequately powered studies of the PK of novel TPT regimens are recommended when PK data are lacking among specific populations such as the elderly and children, pregnant and lactating women, and people with HIV (naïve/on ART). Ideally, these data should be generated in the study population of trials that evaluate the efficacy of the drugs. PK studies can also facilitate the understanding of concerns related to drug-drug interactions, particularly when using a rifamycin-based regimen.

PK studies may be considered to evaluate how physiological changes in pregnancy affect dosing and consequently efficacy/safety, along with human lactation studies. In children, PK studies can be used to determine optimal safety or palatability of a novel TPT regimen (Case study #7). Trials also offer the opportunity to assist in the development and evaluation of appropriate formulations in children.

Why this important:
Drug dosing recommendations for TPT require evidence from PK studies conducted in the target populations for whom the treatment will be recommended. Such studies can provide information regarding the appropriateness of drug dosing, informing guidelines, and also accompanying

implementation guidance. Separate studies to determine the appropriate dose (based on PK studies) and safety (and acceptability) in children may also be needed (64).

Case study #7
The ASTEROID study will evaluate six weeks of daily rifapentine as TPT compared to 12-16 weeks of rifampicin. The BREACH study will evaluate one month of bedaquiline with Lfx. Both studies included children during enrolment and intend to complete PK studies in children to increase the understanding of the drug dosing and efficacy in this sub-population. Hence, even if there are small numbers of children enrolled, the associated PK studies will allow a greater depth of understanding of safety within this sub-population and enable WHO to develop dosing guidance.

5. Role of qualitative research in TB preventive treatment

5.1 The role of qualitative research in supporting WHO policy development

Participant preferences play an important role in decisions about whether treatment should be given for TBI. Participants may conclude that uncommon but serious risks, or common low-grade toxicity, may not be acceptable when weighed against their risk of developing TB disease. For this reason, nested studies evaluating participant values and preferences can provide important information for GDGs.

Qualitative evidence may provide additional data to questions posed by GDGs. In particular, qualitative research may contribute to the evaluation of the EtD framework on the subjects of values, cost, equity, acceptability, and feasibility. Qualitative evidence can be particularly relevant if:

• there is a desire to understand the participants' perceptions of the balance between risks and benefits

• there is no quantitative research informing certain EtD criteria of interest; or

• there is a requirement to understand what participants think about certain aspects of treatment, such as the formulation, adherence monitoring, drug palatability, or treatment monitoring;

• there is a need to understand implementation barriers or feasibility issues from the perspectives of other stakeholders, including health care workers or national programme managers.

Qualitative evidence that concerns a related intervention or context to the one of interest can be provided to the GDG as indirect evidence (Case study #8).

Case study #8
A GDG assessed the use of Lfx to treat TB infection among people at high risk of MDR/RR-TB. A qualitative survey was conducted in 30 high burden MDR-TB countries to assess the feasibility of the programmatic use of Lfx for TPT among contacts of MDR-TB cases. This qualitative survey aimed to evaluate the feasibility, affordability, participant acceptance, and impact of Lfx upon equity.

The study found that NTP managers thought Lfx was affordable, and 50% of respondents thought that programmatic implementation could be achieved without additional resources. NTP managers thought that this intervention would increase equity. However, approximately 25% thought implementation of TPT for MDR/RR-TB prevention would divert resources away from other existing services. Concerns raised by stakeholders, including doctors and national TB program managers, included a possible increase in Lfx resistance and the effect of scale-up on funding and resources. The qualitative component of this survey also revealed key considerations of household contacts when deciding whether to accept TPT. Notably, there was a small subset of individuals who did not weigh the benefits and harms of TPT per se, but rather, made immutable decisions on taking TPT based on their personal values around preventive medicine. Further, it noted that the trust in the health care system, in medical research, in doctors/nurses, influences the perceived harms/benefits of TPT, which is often outside the control of 'trialists' or regimen designs.

Overall, findings from this survey were presented to the WHO GDG, which demonstrated a favourable participant view of treatment. This informed the strong recommendation given for this regimen.

5.2 Approaches and methods of qualitative and mixed-methods research

Social science approaches can provide a framework or a lens through which to make sense of data. Typically, qualitative research draws on social and behavioural theories (of which there are hundreds). A theory can be defined as a set of analytical principles or statements designed to structure our observation, understanding, and explanation of the world (80).

Theories are useful because they help to:

- organize and clarify findings, and connect them to research questions;
- move research from description (i.e., what the data say) to interpretation (i.e., what the data mean);
- generate new theoretical notions (e.g., about people's behaviours or experiences);
- reconnect the data to the research question or spur new questions; and
- improve the transferability or applicability of the data.

6. Other WHO processes

6.1 WHO prequalification

Problems with the quality of medicines and their supply led to the creation of WHO medicines prequalification in 2001. Medicines prequalification activities are:

- assessment of product dossiers (for finished pharmaceutical products (FPPs) or master files (for active pharmaceutical ingredients (APIs) responding to an expression of interest (EOI)
- inspection of manufacturing and clinical sites
- organization of quality control testing of products.

This information, in conjunction with other procurement criteria, is used by the United Nations and other procurement agencies to make purchasing decisions regarding medicines. Further information can be found on the WHO website *(64-66)*.

The standards used to evaluate APIs and FPPs, and their manufacturing sites, are based on the principles and practices agreed by the world's leading regulatory agencies and as adopted by the WHO Expert Committee on Specification for Pharmaceutical Preparations.

At WHO, the Prequalification Team – Medicines (PQT/MED) ensures that finished pharmaceutical products (FPPs) and active pharmaceutical ingredients (APIs) are safe and effective, and meet internationally accepted, stringent-quality standards. PQT/MED also provides advice regarding the development of new products or already invited products intended for additional indications or uses. This advice is given via WHO's Coordinated Scientific Advice (CSA) procedure, jointly with the corresponding disease programme.

WHO also prequalifies quality control laboratories (QCLs), specifically those QCLs that carry out chemical and microbiological testing of medicines.

Other medicines prequalification activities include:

- training (for manufacturers, regulators, and QCLs)
- provision of technical assistance (for manufacturers and QCLs)
- implementation of the collaborative procedure for registration.

6.2 WHO essential medicines list

The WHO Model List of Essential Medicines and Model List of Essential Medicines for Children are updated and published every two years, intended as a guide for countries or regional authorities to adopt or adapt in accordance with local priorities and treatment guidelines for the development and updating of national essential medicines lists. Selection of a limited number of essential medicines as essential, taking into consideration national disease burden and clinical need, can lead to improved access through streamlined procurement and distribution of quality-assured medicines, support more rational or appropriate prescribing and use and lower costs for both health care systems and for

participants. Currently isoniazid, rifampicin, rifapentine and levofloxacin are on the essential medicines list (81).

6.3 Expert Review Panel

The Expert Review Panel (ERP) procedure is a transitional process for much-needed medicines that are undergoing PQ or that are not yet authorized by a WLA. The ERP is a group of independent experts that reviews the potential risks and benefits associated with the use of finished pharmaceutical products and makes recommendations to the Global Fund to Fight AIDS, Tuberculosis and Malaria on their use. WHO oversees the selection of the experts and hosts the panel.

Manufacturers are invited to submit their products for ERP evaluation upon publication of a request for EOI. Invitations are published as either round calls, which occur each semester and have a submission deadline, or ad hoc with no specified deadline. Invitations are published on the website of the Global Fund, where further information can be found. Each individual invitation details the specific documents to include in a submission.

6.4 WHO Technical Advisory Groups

In addition to GDGs, WHO also appoints technical advisory groups to provide expertise on key technical matters that are not usually included in sufficient detail in guidelines and that are important for implementation. One example was the revision of the TPT regimen weight-band dosing by the Technical Advisory Group on the dosing of TB medicines for adults and children (for the 3HP regimen, based upon PK data for the age group under 2 years) and recommendations for dosing of Lfx in children, ahead of the 2024 update of the WHO TPT guidelines and operational handbook (82).

6.5 Development of WHO Operational Handbooks

When significant changes in recommendations for TPT are made by a GDG, WHO may revise the Operational Handbook to accompany the new guidance. The Handbook addresses issues of practical importance to TB programmes, including dosing considerations, approaches to monitor for toxicity and AEs of special interest. It also includes other implementation considerations, based upon expert opinion and operational research. This operational guidance is often informed by the design of the trials upon which the new guidance is based. The Operational Handbook may address issues that are raised as priorities by GDGs during the EtD review process.

References

1. Houben RM, Dodd PJ. The Global Burden of Latent Tuberculosis Infection: A Re-estimation Using Mathematical Modelling. PLoS Med. 2016;13(10):e1002152.
2. World Health Organisation. WHO consolidated guidelines on tuberculosis. Module 1: prevention – tuberculosis preventive treatment, second edition. Geneva; 2024.
3. World Health Organisation. WHO operational handbook on tuberculosis. Module 1: prevention - tuberculosis preventive treatment, second edition. Geneva; 2024.
4. Den Boon S, Lienhardt C, Zignol M, Schwartzman K, Arinaminpathy N, Campbell JR, et al. WHO target product profiles for TB preventive treatment. Int J Tuberc Lung Dis. 2022;26(4):302-9.
5. World Health Organization. Target product profiles for tuberculosis preventive treatment. 2020.
6. Schünemann H BJ, Oxman A. GRADE handbook for grading quality of evidence and strength of recommendations. The GRADE Working Group; 2013.
7. World Health Organization. WHO handbook for guideline development Geneva, Switzerland 2014 [Available from: https://www.who.int/publications/i/item/9789241548960 (Accessed 10 February 2025).
8. World Health Organization. Guidance for best practices for clinical trials. Geneva: World Health Organization, Licence: CC BY-NC-SA 3.0 IGO; 2024.
9. World Health Organisation. WHO listed authorities 2024 [Available from: https://www.who.int/news-room/questions-and-answers/item/who-listed-authorities.
10. World Health Organisation. Guidance on evidence generation on new regimens for tuberculosis treatment. 2024.
11. World Health Organisation. Twelfth General Programme of Work 2014–2019: not merely the absence of disease. 2017.
12. World Health Organization. Guidelines for WHO guidelines. 2003.
13. Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S, et al. Grading quality of evidence and strength of recommendations. Bmj. 2004;328(7454):1490.
14. Schünemann HJ, Best D, Vist G, Oxman AD. Letters, numbers, symbols and words: how to communicate grades of evidence and recommendations. Cmaj. 2003;169(7):677-80.
15. World Health Organisation. Understanding and using WHO guidelines on tuberculosis 2022 [Available from: https://openwho.org/.
16. Gradepro GDT. Gradepro 2021 [26 June 2026]. Available from: https://www.gradepro.org/.
17. Neumann I S-PB, Meerpohl J, Dahm P, Brennan S, Alonso P et al. Making answerable questions. Neumann I & Schünemann H (eds), The GRADE Book version 10 (updated September
2024): The GRADE Working Group; 20242024.
18. Alonso-Coello P, Oxman AD, Moberg J, Brignardello-Petersen R, Akl EA, Davoli M, et al. GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 2: Clinical practice guidelines. Bmj. 2016;353:i2089.
19. Hultcrantz M, Rind D, Akl EA, Treweek S, Mustafa RA, Iorio A, et al. The GRADE Working Group clarifies the construct of certainty of evidence. J Clin Epidemiol. 2017;87:4-13.

20. Schünemann HJ. Interpreting GRADE's levels of certainty or quality of the evidence: GRADE for statisticians, considering review information size or less emphasis on imprecision? J Clin Epidemiol. 2016;75:6-15.
21. Balshem H, Helfand M, Schünemann HJ, Oxman AD, Kunz R, Brozek J, et al. GRADE guidelines: 3. Rating the quality of evidence. J Clin Epidemiol. 2011;64(4):401-6.
22. GRADE CERQual Project Group. GRADE-CERQual: confidence in the evidence from reviews of qualitative research 2024 [Available from: https://www.cerqual.org/.
23. Santesso N, Glenton C, Dahm P, Garner P, Akl EA, Alper B, et al. GRADE guidelines 26: informative statements to communicate the findings of systematic reviews of interventions. J Clin Epidemiol. 2020;119:126-35.
24. Schünemann HJ, Cuello C, Akl EA, Mustafa RA, Meerpohl JJ, Thayer K, et al. GRADE guidelines: 18. How ROBINS-I and other tools to assess risk of bias in nonrandomized studies should be used to rate the certainty of a body of evidence. J Clin Epidemiol. 2019;111:105-14.
25. Higgins JPT AD, Sterne JAC (editors). Chapter 8: Assessing risk of bias in included studies. In: Higgins JPT CR, Chandler J,, MS C, editors. Cochrane Handbook for Systematic Reviews of Interventions
version 520: Cochrane; 2017.
26. Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. Bmj. 2016;355:i4919.
27. Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, et al. GRADE guidelines: 8. Rating the quality of evidence--indirectness. J Clin Epidemiol. 2011;64(12):1303-10.
28. Schünemann HJ, Neumann I, Hultcrantz M, Brignardello-Petersen R, Zeng L, Murad MH, et al. GRADE guidance 35: update on rating imprecision for assessing contextualized certainty of evidence and making decisions. J Clin Epidemiol. 2022;150:225-42.
29. Sabella-Jiménez V, Hoyos Mendez Y, Benjumea-Bedoya D, Estupiñán-Bohorquez AF, Acosta-Reyes J, Florez ID. Effectiveness and safety of available preventive tuberculosis treatment regimens for children and adolescents: protocol for a systematic review and network meta-analysis. BMJ Paediatr Open. 2022;6(1).
30. Balcells ME, Thomas SL, Godfrey-Faussett P, Grant AD. Isoniazid preventive therapy and risk for resistant tuberculosis. Emerg Infect Dis. 2006;12(5):744-51.
31. Holger J Schünemann JPH, Gunn E Vist, Paul Glasziou, Elie A Akl, Nicole Skoetz, Gordon H Guyatt; on behalf of the Cochrane GRADEing Methods Group (formerly Applicability and Recommendations Methods Group) and the Cochrane Statistical Methods Group. Chapter 14: Completing 'Summary of findings' tables and grading the certainty of the evidence. In: Julian Higgins JT, editor. Cochrane Handbook for Systematic Reviews of Interventions2024.
32. Kamchedzera W, Quaife M, Msukwa-Panje W, Burke RM, Macpherson L, Kumwenda M, et al. Treatment preferences among people at risk of developing tuberculosis: A discrete choice experiment. PLOS Glob Public Health. 2024;4(7):e0002804.
33. Marks GB, Bai J, Simpson SE, Sullivan EA, Stewart GJ. Incidence of tuberculosis among a cohort of tuberculin-positive refugees in Australia: reappraising the estimates of risk. Am J Respir Crit Care Med. 2000;162(5):1851-4.

34. Schünemann HJ, Cushman M, Burnett AE, Kahn SR, Beyer-Westendorf J, Spencer FA, et al. American Society of Hematology 2018 guidelines for management of venous thromboembolism: prophylaxis for hospitalized and nonhospitalized medical patients. Blood Adv. 2018;2(22):3198-225.
35. Nasser SM, Cooke G, Kranzer K, Norris SL, Olliaro P, Ford N. Strength of recommendations in WHO guidelines using GRADE was associated with uptake in national policy. J Clin Epidemiol. 2015;68(6):703-7.
36. World Health Organisation. Report of the technical consultation on innovative clinical trial designs for evaluating new TB preventive treatments Geneva, Switzerland; 2021.
37. Swindells S, Ramchandani R, Gupta A, Benson CA, Leon-Cruz J, Mwelase N, et al. One Month of Rifapentine plus Isoniazid to Prevent HIV-Related Tuberculosis. N Engl J Med. 2019;380(11):1001-11.
38. Wademan DT, Palmer M, Purchase S, van der Zalm MM, Osman M, Garcia-Prats AJ, et al. Toward a conceptual framework of the acceptability of tuberculosis treatment in children using a theory generative approach. PLOS Glob Public Health. 2022;2(12):e0001267.
39. Teo AKJ, MacLean EL, Fox GJ. Subclinical tuberculosis: a meta-analysis of prevalence and scoping review of definitions, prevalence and clinical characteristics. Eur Respir Rev. 2024;33(172).
40. Huashan Hospital. TB YOUTH - TB sYstemic Management Using One-month, Ultra-short TPT Regimen for scHool Contacts (TB-YOUTH) 2025 [26 June, 2025]. Available from: https://clinicaltrials.gov/study/NCT06022146.
41. Ntinginya N. Preventive treatment of latent tuberculosis infection in people with diabetes mellitus (PROTID) 2023 [Available from: https://clinicaltrials.gov/study/NCT04600167?term=NCT04600167&rank=1.
42. Rangaka MX. The effectiveness, safety, and benefit risk balance of one month of daily rifapentine with isoniazid added to diabetes standard of care, compared to diabetes standard of care alone, to prevent TB in people with diabetes 2024 [Available from: https://www.isrctn.com/ISRCTN37955921.
43. National Institutes of Health. Common Terminology Criteria for Adverse Events (CTCAE) v5.0. 2017.
44. National Institute of Allergy and Infectious Diseases. Safety Reporting and Pharmacovigilence 2025 [Available from: https://www.niaid.nih.gov/research/dmid-safety-reporting-pharmacovigilance.
45. Council for International Organizations of Medical Sciences. Management of Safety Information from Clinical Trials: Report of CIOMS Working Group VI. . Geneva, Switzerland; 2005.
46. Chandra DK, Anthony P. Moll, Frederick L. Altice, Elizabeth Didomizio, Laurie Andrews, and Sheela V. Shenoi. . Structural Barriers to Implementing Recommended Tuberculosis Preventive Treatment in Primary Care Clinics in Rural South Africa. Yale School of Medicine, Department of Medicine; 2020.
47. Lévesque LE, Hanley JA, Kezouh A, Suissa S. Problem of immortal time bias in cohort studies: example using statins for preventing progression of diabetes. Bmj. 2010;340:b5087.

48. Velen K, Shingde RV, Ho J, Fox GJ. The effectiveness of contact investigation among contacts of tuberculosis patients: a systematic review and meta-analysis. Eur Respir J. 2021;58(6).
49. U.S. Department of Health and Human Services. Guidance for Industry Pulmonary Tuberculosis: Developing Drugs for Treatment. 2013.
50. Sekhon M, Cartwright M, Francis JJ. Acceptability of healthcare interventions: an overview of reviews and development of a theoretical framework. BMC Health Serv Res. 2017;17(1):88.
51. European Medicines Agency. ICH E9 statistical principles for clinical trials - Scientific guideline 2025 [11th April, 2025]. Available from: https://www.ema.europa.eu/en/ich-e9-statistical-principles-clinical-trials-scientific-guideline.
52. United States National Cancer Institute. NCI Patient Reported Outcomes version of the Common Terminology Criteria for Adverse Events 2025 [Available from: https://healthcaredelivery.cancer.gov/pro-ctcae/background.html (Accessed 18 Feb 2025).
53. Wademan DT, Palmer M, Purchase S, van der Zalm MM, Osman M, Garcia-Prats AJ, et al. Toward a conceptual framework of the acceptability of tuberculosis treatment in children using a theory generative approach. PLOS Global Public Health. 2022;2(12):e0001267.
54. US Department of Health and Human Services FaDA. Guidance for Industry. E9 Statistical Principles for Clinical Trials. . 1998.
55. Salazar-Austin N, Dowdy DW, Chaisson RE, Golub JE. Seventy Years of Tuberculosis Prevention: Efficacy, Effectiveness, Toxicity, Durability, and Duration. Am J Epidemiol. 2019;188(12):2078-85.
56. Bliven-Sizemore EE, Sterling TR, Shang N, Benator D, Schwartzman K, Reves R, et al. Three months of weekly rifapentine plus isoniazid is less hepatotoxic than nine months of daily isoniazid for LTBI. Int J Tuberc Lung Dis. 2015;19(9):1039-44, i-v.
57. Ahmad N, Ahuja SD, Akkerman OW, Alffenaar JC, Anderson LF, Baghaei P, et al. Treatment correlates of successful outcomes in pulmonary multidrug-resistant tuberculosis: an individual patient data meta-analysis. Lancet. 2018;392(10150):821-34.
58. Dorman SE, Nahid P, Kurbatova EV, Phillips PPJ, Bryant K, Dooley KE, et al. Four-Month Rifapentine Regimens with or without Moxifloxacin for Tuberculosis. N Engl J Med. 2021;384(18):1705-18.
59. Menzies D, Adjobimey M, Ruslami R, Trajman A, Sow O, Kim H, et al. Four Months of Rifampin or Nine Months of Isoniazid for Latent Tuberculosis in Adults. N Engl J Med. 2018;379(5):440-53.
60. Mauri L, D'Agostino Ralph B. Challenges in the Design and Interpretation of Noninferiority Trials. N Engl J Med.377(14):1357-67.
61. Food and Drug Administration. Non-inferiority clinical trials to establish effectiveness guidance for industry Silver Spring, MD2016 [Available from: https://www.fda.gov/media/78504/download.
62. The European Agency for the Evaluation of Medicinal Products Evaluation of Medicines for Human Use. Committee for proprietary medicinal products. 2000.
63. Thomas LE, Pencina MJ. Do Not Over (P) Value Your Research Article. JAMA Cardiology. 2016;1(9):1055-.

64. Stringer B, Lowton K, James N, Nyang'wa BT. Capturing patient-reported and quality of life outcomes with use of shorter regimens for drug-resistant tuberculosis: mixed-methods substudy protocol, TB PRACTECAL-PRO. BMJ Open. 2021;11(9):e043954.
65. World Health Organisation. Global individual participant data platform for tuberculosis treatment 2023 [Available from: https://www.who.int/news/item/31-01-2023-launch-of-global-individual-patient-data-platform-for-tuberculosis-treatment#:~:text=WHO%20is%20supporting%20the%20publicly%20accessible%20global%20individual,modalities%20for%20tuberculosis%20%28TB%29%20and%20stimulate%20TB%20research.
66. 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 [Available from: https://www.ema.europa.eu/en/documents/scientific-guideline/addendum-guideline-evaluation-medicinal-products-indicated-treatment-bacterial-infections-address-clinical-development-new-agents-treat-pulmonary-disease-due-mycobacterium-tubercul_en.pdf.
67. Odem-Davis K, Fleming TR. A simulation study evaluating bio-creep risk in serial non-inferiority clinical trials for preservation of effect. Stat Biopharm Res. 2015;7(1):12-24.
68. Fox GJ, Nhung NV, Cam Binh N, Hoa NB, Garden FL, Benedetti A, et al. Levofloxacin for the Prevention of Multidrug-Resistant Tuberculosis in Vietnam. N Engl J Med. 2024;391(24):2304-14.
69. Hesseling AC, Purchase SE, Martinson NA, Fairlie L, Schaaf HS, Brigden J, et al. Levofloxacin Preventive Treatment in Children Exposed to MDR Tuberculosis. N Engl J Med. 2024;391(24):2315-26.
70. Trevisi L, Hernán MA, Mitnick CD, Khan U, Seung KJ, Rich ML, et al. Effectiveness of Bedaquiline Use beyond Six Months in Patients with Multidrug-Resistant Tuberculosis. Am J Respir Crit Care Med. 2023;207(11):1525-32.
71. Rodriguez CA, Lodi S, Horsburgh CR, Mitnick CD, Bastard M, Huerga H, et al. Comparative effectiveness of adding delamanid to a multidrug-resistant tuberculosis regimen comprised of three drugs likely to be effective. PLOS Glob Public Health. 2023;3(4):e0000818.
72. World Medical Association. Declaration of Helsinki - ethical principles for medical research involving human participants: Helsinki, Finland; 2024 [Available from: https://www.wma.net/policies-post/wma-declaration-of-helsinki/#:~:text=In%20advance%20of%20a%20clinical,reasonably%20safe%20in%20the%20trial.).
73. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. Bmj. 2010;340:c332.
74. Tennant PWG, Murray EJ, Arnold KF, Berrie L, Fox MP, Gadd SC, et al. Use of directed acyclic graphs (DAGs) to identify confounders in applied health research: review and recommendations. Int J Epidemiol. 2021;50(2):620-32.
75. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use. Good clinical practice (GCP): E6(R3). Geneva2023 [Available from: https://database.ich.org/sites/default/files/ICH_E6%28R3%29_DraftGuideline_2023_0519.pdf.
76. European Union. Risk proportionate approaches in clinical trials: Recommendations of the expert group on clinical trials for the implementation of Regulation (EU) No 536/2014 on

clinical trials on medicinal products for human use. Brussels2017 [Available from: https://health.ec.europa.eu/system/files/2017-08/2017_04_25_risk_proportionate_approaches_in_ct_0.pdf.
77. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use. A selective approach to safety data collection in specific late-stage pre-approvalor post-approval clinical trials: E19 Geneva2022 [Available from: https://database.ich.org/sites/default/files/ICH_E19_Guideline_Step4_2022_0826_0.pdf.
78. World Health Organisation. WHO consolidated guidelines on tuberculosis: module 1: prevention: tuberculosis preventive treatment. Geneva, Switzerland; 2020.
79. Hamada Y, Figueroa C, Martín-Sánchez M, Falzon D, Kanchar A. The safety of isoniazid tuberculosis preventive treatment in pregnant and postpartum women: systematic review and meta-analysis. Eur Respir J. 2020;55(3).
80. Nilsen P. Making sense of implementation theories, models and frameworks. Implement Sci. 2015;10:53.
81. World Health Organisation. Model list of essential medicines 2025 [Available from: https://list.essentialmeds.org/.
82. World Health Organisation. Technical Advisory Group on dosing of TB medicines for adults and children 2025 [26 June, 2025]. Available from: https://www.who.int/groups/technical-advisory-group-on-dosing-of-tb-medicines-for-adults-and-children.

Annex 1. Example of a GRADE evidence profile

The GRADE (Grading of Recommendations Assessment, Development and Evaluation) system is an internationally recognized approach for rating the quality of evidence and grading the strength of recommendations in health care (1). The GRADE evidence profiles summarize evidence in a succinct, transparent, and informative summary of findings tables, including the following key information for each critical outcome:

- The type and number of studies included (e.g. randomized trials, observational studies)
- An assessment of the quality of evidence for each outcome, based on explicit criteria
- The magnitude of the effect (both relative and absolute)
- Reasons for downgrading or upgrading the quality of evidence

The following table shows an example of an evidence profile table for the PICO question "Should 6 months of levofloxacin vs. other regimen or no TPT be used for people in contact with MDR/RR-TB?". The contents summarize the findings from an individual participant data meta-analysis of two clinical trials presented to the Guideline Development Group (GDG). Table A1.1 presents quantitative estimates for each outcome of importance to the GDG.

Table A1.1 Example of an evidence profile table, displaying five prioritized outcome measures with assessments of the certainty of evidence, together with event rates and effect estimates for each outcome

Certainty assessment							No of patients		Effect			
No of studies	**Study design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**6 months Lfx**	**Other regimen or no TPT**	**Relative (95% CI)**	**Absolute (95% CI)**	**Certainty**	**Importance**
TB incidence (bacteriologically confirmed or clinically defined TB, TB-related death at 54 weeks)												
2	Randomized trials	not serious	not serious	not serious	not serious	none	8/1474 (0.5%)	21/1483 (1.4%)	RR 0.38 (0.17 to 0.86)	9 fewer per 1000 (from 12 to 2 fewer)	High	Critical
Death (any cause)												
2	Randomized trials	not serious	not serious	not serious	very serious[a]	none	5/1476 (0.3%)	4/1487 (0.3%)	RR 1.26 (0.34 to 4.68)	1 more per 1000 (from 2 fewer to 10 more)	Low	Critical

Certainty assessment							No of patients		Effect			
No of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	6 months Lfx	Other regimen or no TPT	Relative (95% CI)	Absolute (95% CI)	Certainty	Importance
Adverse events (follow-up: 6 months plus 21 days; Grade 3 or above at least possibly related to study drug (TB CHAMP; under 18y))												
1	Randomized trials	not serious	not serious	not serious	serious[b]	none	4/452 (0.9%)	8/469 (1.7%)	RR 0.53 (0.16 to 1.70)	8 fewer per 1000 (from 14 fewer to 12 more)	Moderate	Critical
Adverse events (follow-up: 6 months plus 30 days; Grade 3 or above at least possibly related to study drug (V-QUIN; 97% of participants >14y))												
1	Randomized trials	not serious	not serious	not serious	not serious	none	10/960 (1.0%)	2/962 (0.2%)	RR 5.26 (1.16 to 23.95)	9 more per 1000 (from 0 fewer to 48 more)	High	Critical
Adverse events of any grade leading to treatment discontinuation (follow-up: 6 months plus 21 or 30 days)												
2	Randomized trials	not serious	not serious	not serious	not serious	none	77/1412 (5.5%)	12/1431 (0.8%)	RR 6.32 (3.43 to 11.63)	45 more per 1000 (from 20 to 89 more)	High	Critical
Treatment completion (opposite of discontinuation)												
2	Randomized trial	not serious	not serious	not serious	not serious	none	1078/1476 (73.0%)	1233/1487 (82.9%)	RR 0.88 (0.85 to 0.92)	100 fewer per 1000 (from 124 to 66 fewer)	High	Critical
Treatment completion (80% or more of doses taken by 6 months)												
2	Randomized trials	not serious	not serious	not serious	not serious	none	1092/1460 (74.8%)	1248/1468 (85.0%)	RR 0.88 (0.85 to 0.91)[c]	102 fewer per 1000 (from 128 to 77 fewer)	High	Critical
Emergence of additional fluoroquinolone resistance in TB strains												
2	Randomized trials	serious[d]	not serious	serious[e]	serious[f]	none	In none of 8 strains from index-incident pairs in the V-QUIN trial that were tested with whole genome sequencing was additional resistance to levofloxacin or other antimicrobials detected[d]				Very low	Important
Emergence of additional fluoroquinolone resistance in microbiome other than TB (e.g. gut flora) not measured												
-	-	-	-	-	-	-	-				-	Important

CI: confidence intervals; Lfx: levofloxacin; RR: relative risk; TB: tuberculosis; TPT: TB preventive treatment

a We rated down two levels because the confidence intervals include appreciable harm and appreciable benefit: RR 1.26 (0.34 to 4.68)

b We rated down one level because the confidence intervals include appreciable harm and some benefit. RR 0.53 (0.16 to 1.70)

c Treatment completion in the levofloxacin arm was 86% in TB CHAMP (placebo arm: 86%) and 70% in V-QUIN (placebo arm: 85%) – RRs 1.00 [95% CI 0.95 to 1.06] and 0.83 [0.79 to 0.87] respectively

d We rated down one level for risk of bias. The results are not from a randomized comparison. In the V-QUIN Trial, of the 43 persons with suspected TB post-randomization, 17 had a laboratory-confirmed incident TB, in 4 of whom an isolate could not be recovered. Whole Genome Sequencing with drug susceptibility results were available for 8/13 with confirmed TB. Of these, 6 were in the placebo group and 2 from the Lfx arm. None had acquired resistance to levofloxacin. In TB CHAMP, 14 individuals in the placebo arm and 7 in the Lfx arm developed TB, of which 7 and 3, respectively, with confirmed TB. No results for levofloxacin susceptibility were available for the strains isolated.
e We rated down one level for indirectness. Data was only available for V-QUIN; all strains were from individuals aged over 15 years.
f We rated down one level for imprecision due to the small number of samples and zero events

References

1. Parmelli E, Amato L, Oxman AD, Alonso-Coello P, Brunetti M, Moberg J, et al. Grade Evidence to Decision (EtD) framework for coverage decisions. Int J Technol Assess Health Care. 2017;33(2):176-82.

Annex 2. GRADE evidence to decision framework: Example of a summary of findings table

The GRADE Evidence to Decision (EtD) framework is a systematic and transparent approach for translating evidence into recommendations. It provides a structure for assessment of benefits, harms, certainty of evidence, and other key decision criteria and ensures that recommendations are informed by the best available evidence and aligned with stakeholder values and contextual factors. The GRADE EtD criteria organize PICO-derived evidence into a comprehensive decision-making process by evaluating key factors like problem relevance, clinical benefits and harms, certainty of evidence, and patient/societal values. They also assess economic impact, equity, stakeholder acceptability, and practical feasibility. This structured approach combines biomedical, economic, and social considerations to produce clear, actionable recommendations that are both evidence-based and relevant to real-world healthcare settings.

The Summary of Findings table presents a concise overview of the key evidence considered in the development of a particular recommendation within the EtD framework. This table highlights the critical outcomes, the certainty of the evidence, and the magnitude of effects for each outcome. It is intended to provide decision-makers and stakeholders with a transparent and accessible summary of the most relevant findings, supporting an informed and balanced decision-making process. Table A2.1 provides an example from an EtD that led to a recommendation on the use of 6 months of levofloxacin as TPT.

Table A2.1. Example of a summary of findings table, displaying five prioritized outcome measures with overall rating of the certainty of evidence, together with the baseline event rates and effect estimates for each outcome

A 6-month regimen of levofloxacin compared to no therapy in household contacts of MDR/RR-TB					
Patient: people in contact with MDR/RR-TB **Setting:** outpatient **Intervention**: 6 months of levofloxacin **Comparison**: placebo					
				Anticipated absolute effects	
Outcomes	No of participants (studies) Follow-up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	**Risk with the nil treatment (comparator)**	**Risk difference with a 6 month levofloxacin**
Tuberculosis incidence (bacteriologically confirmed or clinically defined TB, TB-related death at 54 weeks)	2957 (2 RCTs)	⊕⊕⊕⊕ High	**RR 0.38** (0.17 to 0.86)	14 per 1,000	**5 per 1,000** (2 to 12)
Treatment completion (opposite of discontinuation)	2963 (2 RCTs)	⊕⊕⊕⊕ High	**RR 0.88** (0.85 to 0.92)	829 per 1,000	**730 per 1,000** (705 to 763)

				Anticipated absolute effects	
Outcomes	No of participants (studies) Follow-up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	**Risk with the nil treatment (comparator)**	**Risk difference with a 6 month levofloxacin**
Treatment completion (80% or more of doses taken by 6 months)	2928 (2 RCTs)	⊕⊕⊕⊕ High	**RR 0.88** (0.85 to 0.91)[a]	850 per 1,000	**748 per 1,000** (723 to 774)
Death (any cause	2963 (2 RCTs)	⊕⊕◯◯ Low[b]	**RR 1.26** (0.34 to 4.68)	3 per 1,000	**3 per 1,000** (1 to 13)
Adverse Events (AE) Grade 3 or above at least possibly related to study drug (TB CHAMP; under 18y) follow-up: 6 months plus 21 days	921 (1 RCT)	⊕⊕⊕◯ Moderate[c]	**RR 0.53** (0.16 to 1.70)	17 per 1,000	**9 per 1,000** (3 to 29)
Adverse events: Grade 3 or above at least possibly related to study drug (V-QUIN; 97% of participants >14y) follow-up: 6 months plus 30 days	1922 (1 RCT)	⊕⊕⊕⊕ High	**RR 5.26** (1.16 to 23.95)	2 per 1,000	**11 per 1,000** (2 to 50)
Adverse events of any grade leading to treatment discontinuation, follow-up: 6 months plus 21 or 30 days	2843 (2 RCTs)	⊕⊕⊕⊕ High	**RR 6.32** (3.43 to 11.63)	8 per 1,000	**53 per 1,000** (29 to 98)
Emergence of additional fluoroquinolone resistance in TB strains	8 (2 RCTs)[e]	⊕◯◯◯ Very low[d,f,g]		In none of 8 strains from index-incident pairs in the V-QUIN trial that were tested with whole genome sequencing was additional resistance to levofloxacin or other antimicrobials detected	
*The risk in the intervention group (and its 95% confidence interval) is based on the assumed risk in the comparison group and the relative effect of the intervention (and its 95% CI).					

CI: confidence intervals; Lfx: levofloxacin; RCTs: randomized control trial; RR: risk ratio; TB: tuberculosis; TPT: TB preventive treatment

Explanations

a. Treatment completion in the levofloxacin arm was 86% in TB CHAMP (placebo arm: 86%) and 70% in the V-QUIN trial (placebo arm: 85%) - RRs 1.00 [95% CI 0.95 to 1.06] and 0.83 [0.79 to 0.87] respectively
b. We rated down two levels because the confidence intervals include appreciable harm and appreciable benefit: RR 1.26 (0.34 to 4.68)
c. We rated down one level because the confidence intervals include appreciable harm and some benefit. RR 0.53 (0.16 to 1.70)
d. We rated down one level for risk of bias. The results are not from a randomized comparison. In V-QUIN, of the 43 persons with suspected TB post-randomization, 17 had a laboratory-confirmed incident TB, in 4 of whom an isolate could not be recovered. Results were only available for 8/13. Of these 6 were in the placebo group and 2 from the Lfx arm. In TB CHAMP, 14 individuals in the placebo arm and 7 in the Lfx arm developed TB, of which 7 and 3 respectively with confirmed TB. No results for levofloxacin susceptibility were available for the strains isolated.
e. out of 17 laboratory-confirmed incident TB strains
f. We rated down one level for indirectness. Data was only available for V-QUIN; all strains were from individuals aged over 15 years.
g. We rated down one level for imprecision due to the small number of samples and zero events.

Annex 3. GRADE evidence to decision framework: other criteria used to develop recommendations

The evidence to decision (EtD) framework contains a total of 12 criteria for overall assessment of the evidence. In addition to the summary of findings table (Annex 2), the EtD includes information on other characteristics that are critical to the formulation of recommendations. Table A3.1 demonstrates an example of how these criteria were used during the development of recommendations for 6Lfx as TPT for household contacts of MDR/RR-TB.

For this table, an individual participant meta-analysis combining data from two clinical trials, V-QUIN and TB CHAMP, was evaluated by the Guideline Development Group. Data from additional sub-studies of cost-effectiveness and pharmacokinetics, completed by both the V-QUIN and TB CHAMP study teams, were used to inform the respective domains of the EtD framework. Acceptability and feasibility were assessed through separate systematic reviews of evidence and surveys of both national programmes and affected populations. The V-QUIN study also provided data regarding the effects of levofloxacin upon the gut and nasal microbiome.

Table A3.1: EtD criteria used to answer the question: "Should 6 months of levofloxacin vs other regimen or no TPT be used for people in contact with MDR/RR-TB?"

Problem Is the problem a priority?		
Judgement	**Research evidence**	**Additional considerations**
○ No ○ Probably no ○ Probably yes ● Yes ○ Varies ○ Don't know	Drug-resistant tuberculosis is one of the most prominent causes of morbidity and mortality from an antimicrobial resistant organism. Globally, there were an estimated 410,000 incident cases of MDR/RR-TB in 2022. An estimated 160,000 deaths due to MDR/RR-TB occurred in 2022. With recent advances in therapeutics and increased global access to more effective medication, treatment success has improved over time. However, it remains lower than for rifampicin-susceptible TB (63% for people starting treatment in 2021). People with MDR/RR-TB may infect other individuals. It is thus important to take all measures possible to lower the risk of secondary cases of MDR/RR-TB. This includes the use of appropriate TB preventive treatment with regimens of proven effectiveness.	Key considerations expressed by GDG members when deciding that MDR/RR-TB is a priority problem and that measures to prevent it, like TPT, were crucial were as follows: The 2020 TPT guidelines included a recommendation for TPT of contacts of MDR/RR-TB that is conditional and based on evidence of very low certainty. The recommendation was not specific to any regimen, and its implementation since first published in 2017 has been poor. Now that trial-based evidence for a defined treatment regimen has become available, it became more important to review the new evidence to assess the efficacy of this new regimen in preventing this formidable public health problem.
Desirable Effects How substantial are the desirable anticipated effects?		
Judgement	**Research evidence**	**Additional considerations**

○ Trivial ○ Small ● Moderate ○ Large ○ Varies ○ Don't know	See Annex 2 for GRADE Summary of Findings table A systematic review of studies published between June 2016 and September 2023 identified three observational studies that assessed TB prevention (reduction in incidence) with FQ (alone or in combination with other TB drugs), and one assessed prevention of TB with isoniazid. All four were observational studies with substantial risk of bias, notably selection bias. The three studies with FQ did not detect any reduction in TB incidence with FQ use, compared to no TPT. The results from the systematic review and from the isoniazid IPD could not be summarized in the GRADE table.	Key considerations expressed by GDG members when making a judgement of MODERATE desirable effects were as follows: The efficacy of levofloxacin (Lfx) in the trials was similar to what was observed in other studies of TPT, although uncertainty was expressed regarding the durability of effect. The risk for MDR-TB in a person exposed and the seriousness of the disease, with its high lethality, more complicated treatment and likelihood to relapse unless properly treated, are important considerations, regardless of the background risk of MDR-TB in different contexts. Any intervention that can reduce this risk would be welcome. There is an observation that the two outcomes presented here – TB incidence and TPT completion – are going in opposite directions, making it difficult to judge, as the judgements for incidence may be different than for treatment completion. It was noted that the "number needed to treat" to prevent one incident case of TB was different in V-QUIN (193 [98-5495]) and TB CHAMP (56 [30-389]). The decision was made on the pooled data because separation by adults and children would reduce precision and lower the quality of evidence. This will be developed further in the Subgroup considerations.
Undesirable Effects **How substantial are the undesirable anticipated effects?**		
Judgement	**Research evidence**	**Additional considerations**
○ Small ○ Moderate ○ Large ○ Trivial ● Varies ○ Don't know	A systematic review of studies published between June 2016 and September 2023 identified five observational studies that assessed adverse events with FQ (alone or in combination with other TB drugs). All were observational studies with substantial risk of bias, notably selection bias. Detection, judgment of severity, and attribution were not blinded, potentially leading to ascertainment bias. FQ monotherapy (i.e. Lfx, Ofx, or Mfx alone) was observed in three studies to be generally safe, with some mild or moderate drug-related AEs in children, but no grade 3 or 4 AE or serious AE. In a study evaluating FQ with a companion drug (ETH/EMB), the regimen had a higher observed rate of grade 1 or 2 drug-related AEs compared to the studies with FQ monotherapy (ETH+FQ had a significantly higher AE rate than EMB), but no serious AEs were reported, and AEs were not associated with treatment discontinuation. FQ with PZA was found to have very low tolerability in a small study among inmate contacts by Bedini et al 2016 (7/12 contacts discontinued treatment due to AEs). The GDG scored the two outcomes on the emergence of additional resistance as IMPORTANT rather than CRITICAL. While the two trials collected data on the emergence of additional fluoroquinolone resistance to TB strains and other flora, results of drug-susceptibility testing or whole genome sequencing were	Key considerations expressed by GDG members when making a judgement of VARIES for undesirable effects were as follows: There was an important difference in the risk of adverse events between children (trivial) and adults (moderate), with very good tolerance in children and much less tolerability with increasing age, that has likely contributed to lower adherence to TPT in adults. Some forms of toxicity should not be discounted, given that the regimen would be rolled out for use in programmatic settings. The results on the emergence of resistance were inconclusive, although these were not CRITICAL outcomes.

	incomplete at the time of the GDG meeting. Only one outcome from 8 TB strains tested (2 of which from the Lfx arm) in the V-QUIN trial was included in the evidence summary table, which showed no additional resistance acquired.	
Certainty of evidence **What is the overall certainty of the evidence of effects?**		
Judgement	**Research evidence**	**Additional considerations**
○ Very low ○ Low ● Moderate ○ High ○ No included studies	Certainty is judged to be HIGH for TB incidence, treatment completion, adverse events GRADE 3 or above at least possibly associated with study drug in adults, MODERATE for adverse events GRADE 3 or above at least possibly associated with study drug in children, and LOW for death (all CRITICAL outcomes). It was considered VERY LOW for the emergence of additional fluoroquinolone resistance in TB strains and was not estimable for the emergence of additional fluoroquinolone resistance in microbiome other than TB (e.g., gut flora) (both IMPORTANT outcomes). Evidence from other studies identified by the systematic review was considered of very low certainty for efficacy and low certainty for adverse events (all studies were observational). The low incidence of Grade 3-4 adverse events, as well as the low occurrence of discontinuation of FQ TPT due to adverse events, in adults and children, from observational studies, is consistent with evidence from the trials.	Key considerations expressed by GDG members when making a judgement of MODERATE certainty of the evidence of effects were as follows: The two trials were well conducted, large and independently showed very similar estimates of reduction in TB incidence in two different settings with populations of different characteristics. It was acknowledged that we are unlikely to get such high-quality evidence from trials of fluoroquinolone as a TPT for MDR-TB in the foreseeable future (PHOENIX trial is using 26 weeks of delamanid and is expected to be completed at the end of 2026) However, uncertainties were expressed given the serious or very serious imprecision on the adverse events and the fact that there are only two trials. It was highlighted that there may be difficulties to standardize some of the endpoints between the two trials. Effects from pooled estimates were felt to be less robust. The evidence for the emergence of additional resistance to fluoroquinolones was considered uncertain.
Values **Is there important uncertainty about or variability in how much people value the main outcomes?**		
Judgement	**Research evidence**	**Additional considerations**
○ Important uncertainty or variability ○ Possibly important uncertainty or variability ● Probably no important uncertainty or variability ○ No important uncertainty or variability	Evidence from the systematic review (2 published studies on acceptance to start MDR TPT, 2 published studies on willingness to take hypothetical MDR TPT, 1 published study on acceptability of a novel child friendly Lfx formulation, and 1 published explorative qualitative study included in the systematic review) suggested that OVERALL acceptability of MDR TPT to prevent incident TB disease was high. However, based on the qualitative acceptability study (among 36 HHCs from 5 countries), there is an indication of possibly important uncertainty or variability. Although the sample size was still relatively small, this study included people with a wide range of TB and MDR knowledge and experience, as well as with very different socioeconomic and cultural backgrounds, found meaningful differences in TPT acceptability. For example, although most people valued a lowered risk of developing MDR-TB, some refused to accept any risk of serious adverse events due to TPT, which overrode any value they placed in avoiding MDR-TB. The study suggests that in the case where there is an absence of trained HCWs or researchers recruiting them, and taking the time to explain TPT to them, the value for prevention is quite low, the understanding of the severity/risk of MDR also seems very low,	Key considerations expressed by GDG members when making a judgement of PROBABLY NO IMPORTANT UNCERTAINTY OR VARIABILITY in values were as follows: The values are likely to depend on how much people being offered fluoroquinolone TPT are well informed about the efficacy and downsides of TPT, and the seriousness of MDR-TB. In all situations, safety is paramount, particularly for a person who is not ill. There were some financial, emotional, and psychological factors that played into adherence. They may be overcome with education, but still important. Acceptance for people who started TPT was quite high and more than is seen with comparable interventions under programmatic settings. However, the evidence reviewed is from small samples so maybe not generalizable.

	and the value in one's present health is very high by contrast. Arguably, this is a very important variability in values that could really affect real-world uptake of MDR TPT	
Balance of effects Does the balance between desirable and undesirable effects favor the intervention or the comparison?		
Judgement	**Research evidence**	**Additional considerations**
○ Favors the comparison ○ Probably favors the comparison ○ Does not favor either the intervention or the comparison ● Probably favors the intervention ○ Favors the intervention ○ Varies ○ Don't know	The reduction of MDR-TB incidence with the intervention of Lfx by 60% in adults and children is offset only by mild Grade 1-2 AEs. Both desirable and undesirable effect estimates are derived from two RCTs that are judged to be of high quality overall, and the ascertainment of these outcomes was also free of bias, and there was sufficient precision that we can be reasonably certain of these effects. The estimates of low rates of Grade 1-2 adverse events and very low rates of Grade 3-4 adverse events are supported by observational studies found in the systematic review, although it was not possible to estimate a pooled rate of mild or severe adverse events in the review due to heterogeneity of interventions reported, and definitions of adverse events used.	Key considerations expressed by GDG members when making a judgement of PROBABLY FAVOURS THE INTERVENTION for the balance of effects were as follows: It is noted that, based on the evidence presented to the GDG, the benefits outweigh the risks, especially in children. To a large extent, the adverse events were mild and self-limiting. Although not critical for this assessment, the emergence of other resistance is important, and there is uncertainty about how it could reduce the potential benefit from the intervention. The evidence reviewed was incomplete, and the implications of the effects reported for the overall population and for the individual in the long term are unknown. It was highlighted that the use of fluoroquinolone as a TPT for MDR-TB should be considered as an appropriate use of antimicrobial agents, unlike inappropriate use, which is more likely to generate avoidable resistance. It is noted that the effects of using levofloxacin at a wide scale in a population is unknown.
Resources required **How large are the resource requirements (costs)**?		
Judgement	**Research evidence**	**Additional considerations**
○ Large costs ● Moderate costs ○ Negligible costs and savings ○ Moderate savings ○ Large savings ○ Varies ○ Don't know	Based on a self-administered questionnaire survey among national TB programme (NTP) managers of 30 high-burden MDR-TB countries, of whom 18 (60%) responded, 7 of 18 respondents stated that the cost of additional resource requirements may be a barrier to implementation, with some mentioning specifically the concurrent need for drug-susceptibility testing, screening, monitoring, and follow-up in the programme as well as the already limited human resources and budgets within programmes. The paediatric dispersible formulation of levofloxacin is much more expensive than the adult formulation (a tenfold difference per mg at current GDF prices - approx. US$0.12/100mg tablet vs. US$0.03/250mg tablet respectively; https://www.stoptb.org/sites/default/files/gdf_medicines_catalog_5.pdf).	Key considerations expressed by GDG members when making a judgement of MODERATE COSTS for the resources required were as follows: The cost of levofloxacin, a generic medication in wide use, is relatively low when compared with other TPT or no TPT. However, the health system costs to deliver the overall intervention may entail additional investments in programmatic components that are weak, such as screening and identifying contacts, drug-susceptibility testing, monitoring for adverse events, capacity building to improve the skills of healthcare workers, engaging communities, increasing treatment literacy, and providing social support. There is no reason to consider that these costs will be excessive. Investments may generate gains in the long term and the need for additional expenditure should not stop programmes from doing what is necessary to prevent and care for MDR-TB.

		It was also noted that overall, the burden of MDR-TB patients is relatively low compared with drug-susceptible TB. Those not procuring through the Global Drug Facility mechanism may face a higher price for a product of guaranteed quality, as well as differences in costs if the 750mg formulation is used instead of the 250mg. However, this variation in the exact per-patient budget impact may not have had a major influence in the NTP survey responses.
Certainty of evidence of required resources **What is the certainty of the evidence of resource requirements (costs)?**		
Judgement	**Research evidence**	**Additional considerations**
○ Very low ● Low ○ Moderate ○ High ○ No included studies	A single self-administered questionnaire and completion rate of only 60%. The pricing of the Global Drug Facility medications is standardized for all countries eligible.	Key considerations expressed by GDG members when making a judgement of LOW for the certainty of evidence of required resources were that there was only one survey reviewed, and that there was no evidence on costs for implementation.
Cost effectiveness **Does the cost-effectiveness of the intervention favor the intervention or the comparison?**		
Judgement	**Research evidence**	**Additional considerations**
○ Very low ● Low ○ Moderate ○ High ○ No included studies	A single self-administered questionnaire and a completion rate of only 60%. The pricing of the Global Drug Facility medications is standardized for all countries eligible.	Key considerations expressed by GDG members when making a judgement of LOW for the certainty of evidence of required resources were that there was only one survey reviewed and that there was no evidence on costs for implementation.
Equity What would be the impact on health equity?		
Judgement	**Research evidence**	**Additional considerations**
○ Reduced ○ Probably reduced ○ Probably no impact ● Probably increased ○ Increased ○ Varies ○ Don't know	Based on a self-administered survey questionnaire among NTP managers of 30 high-burden MDR-TB countries, of whom 18 (60%) responded, overall equity was expected to increase, from the perspective of the managers, for contacts through the avoidance of TB disease incidence. However, 6 NTPs mentioned that certain remote areas may not have an adequate supply of Lfx. Additionally, 11 NTPs mentioned increased out-of-pocket spending for contacts, with 2 stating the need for health insurance to cover TPT to ensure equity. Importantly, interviews with contacts themselves in the qualitative acceptability study (36 HHCs from 5 countries) suggested that those with little income, unstable or no employment, little or no social support, will likely NOT be able to accept and complete a 6-month TPT regimen that will require at least monthly check-ups, and maybe some mild side effects, especially at the beginning of treatment that could impact their daily activities and responsibilities. Also, caregivers for the MDR index patients or other contacts within the household are unlikely to be able to start/accept TPT as well, unless they have access to improved socioeconomic support systems. Hence, findings from this qualitative study suggest that equity may be reduced by the	Key considerations expressed by GDG members when making a judgement of PROBABLY INCREASED for equity were as follows: Some people might benefit more from levofloxacin than others. From a drug perspective, there is more equity because we can prevent TB in more people, given the efficacy of the drug. Equity may increase if services are provided to contacts at high risk of drug-resistant TB and who are generally marginalised and who have difficulty accessing services. From a model of care perspective, equity is more likely in situations where drug costs are covered by the public health system. Otherwise, the intervention might shift cost to the affected person and lead to out of pocket payments that can reduce equity. So, it is important to think about improving models of care to protect the person needing the drug from incurring costs

	introduction of TPT for MDR contacts, unless this is accompanied by improved social and financial support.	from the drug and other healthcare system components. In situations where the health system covers the expenditure for levofloxacin, another consideration is the opportunity cost of investing in levofloxacin as a TPT for MDR-TB. Will the cost of treatment be deducted from another important programmatic component, like TPT for non-MDR-TB or the treatment of people with MDR/RR-TB?
Acceptability Is the intervention acceptable to key stakeholders?		
Judgement	**Research evidence**	**Additional considerations**
○ No ○ Probably no ● Probably yes ○ Yes ○ Varies ○ Don't know	A systematic review of studies published between June 2016 and September 2023 identified five observational quantitative studies that assessed acceptance in starting TPT when offered, willingness to take a hypothetical MDR TPT regimen, and acceptability (ability and willingness to use TPT as directed) of TPT with Fluoroquinolone, and a sixth qualitative study conducted in South Africa as a sub-study of the TB CHAMP trial. Two studies indicated an 80% acceptance rate among caregivers, for their children to be started on TPT, and among adolescents and adult contacts. Two studies indicated 90% willingness by caregivers and 70% among adults to take TPT for MDR-TB, and one study indicated high levels of acceptability by caregivers administering a novel dispersible child-friendly formulation of Lfx to their children. The published qualitative study found an overall high acceptability of Lfx among caregivers of children as well but found that there were some pragmatic difficulties around the financial and care burden of providing TPT to their children, especially for caregivers undergoing treatment for TB disease themselves (which was a motivator for accepting treatment but limited capacity to care for children). Greater social support led to greater capability to ensure adherence to treatment for both caregivers and children. A qualitative study conducted among 36 MDR-TB contacts in 5 countries (Georgia, India, Indonesia, South Africa, and Viet Nam) concluded that: TPT for MDR was acceptable and of high social value among participants in all 5 settings. The most acceptable TPT regimen would have a high degree of effectiveness in preventing MDR/RR-TB, no risk of side effects that are permanent or that could interfere with daily activities, few pills and a short duration, low socioeconomic cost, and minimal clinical follow-up visits. A retrospective quantitative sub-study conducted by the V-QUIN investigators examined acceptability among a randomly selected sample of 240 participants in the V-QUIN trial (about equal numbers took placebo, and Lfx). They found no major differences in ratings of medication taste, size, and frequency of preventive treatment between arms. Of all participants, less than 20% rated the duration ideal, and almost one third rated the duration as too long. Acceptability was somewhat worse in those who did NOT complete study drug. Only a minority of participants would take the treatment again or would recommend to others. A prospective quantitative sub-study among all participants in the TB CHAMP trial examined acceptability on every treatment phase visit and found that the taste of levofloxacin was disliked by children more than placebo, but the children in both arms adapted to the taste over the course of treatment. Caregivers found it more	Key considerations expressed by GDG members when making a judgement of PROBABLY YES for acceptability were as follows: In the survey of national TB programme managers many stated that they would accept the recommendation only if it is strong. The 6-month duration of treatment may be a challenge although this is the same as the minimum duration of isoniazid that is still one of the most widely used TPT regimens worldwide. Six months has also been the duration of standard treatment for drug-susceptible TB and for the new BPaL(M) regimen for MDR/RR-TB. However, a shorter TPT would be preferred in future. Other factors such as cost, administration issues and the taste of medication were also mentioned as challenges. The high frequency of adverse events in adults in particular was highlighted. Providing clear information on benefits and risk and a supporting environment to caregivers and beneficiaries is likely to improve acceptability: people's perceptions of the effectiveness and value of TPT are important.

	difficult to administer levofloxacin than placebo, but overall, more than 95% of caregivers reported NO difficulty in giving levofloxacin. Overall, the investigators concluded that acceptability was reasonable but noted an association between poor acceptability and poor adherence. In addition, a semi-structured interview was conducted to evaluate the caregiver experience of administering novel child-friendly levofloxacin formulation in 10 child/caregiver dyads on the side of TB CHAMP. There was a relatively high overall acceptability. One major motivator was the caregivers' own experiences with MDR-TB illness and treatment. Pragmatic difficulties were expressed around financial and care burden on the household due to TPT. Challenges were exacerbated for caregivers who were on treatment for their own MDR-TB disease, limiting their capacity to care for their children. Caregivers who received greater social support reported better capability for them and their children to adhere to treatment.	
Feasibility Is the intervention feasible to implement?		
Judgement	**Research evidence**	**Additional considerations**
○ No ○ Probably no ○ Probably yes ● Yes ○ Varies ○ Don't know	Based on a self-administered survey questionnaire among NTP managers of 30 high-burden MDR-TB countries, of whom 18 (60%) responded, in the case of a strong WHO recommendation, an additional 8 countries (apart from the 6 that were already implementing 6 Lfx) were ready to implement Lfx programme-wide. A conditional recommendation made it less likely for 7 NTPs. All managers anticipated that drug storage, transportation, and distribution was sustainable. However, the need for additional resources (DST, monitoring, and follow-up) were raised as concerns/barriers to implementation by 7 of 18 managers.	Key considerations expressed by GDG members when making a judgment of YES for feasibility were as follows: There is already a WHO recommendation for the use of TPT in MDR-TB, which has been implemented to some degree despite it being conditional, with levofloxacin being one of the options proposed. Feasibility will depend upon additional resources being available to implement the intervention properly, such as drug-susceptibility testing of the presumed source case and testing for TB infection (in the TB-CHAMP trial a positive tests for infection was not required in most individuals; in the V-QUIN trial adults could participate if TST positive, with a small number who were TST negative with HIV or malnutrition) and chest X-ray (done for participants in both trials) Levofloxacin is widely available as a generic drug in both adult and paediatric formulations,

Annex 4. GRADE evidence to decision framework: Example of summary judgements across the 12 EtD criteria

Table A4 gives an example of summary judgements across the 12 Evidence to Decision (EtD) criteria typically evaluated as part of the overall assessment of the evidence. This summary is generated at the conclusion of the guideline development group (GDG) discussions and serves as a key resource to inform final recommendations. This summary table presents all judgements in a succinct form that enables a holistic assessment of the judgements and provides guidance for the GDG to make recommendations around the PICO question. The example reproduced here is for the judgements made for the comparison of the 6-month levofloxacin regimen vs nil treatment.

Table. A4 Example of summary judgements across the 12 EtD criteria

Problem	No	Probably no	Probably yes	**Yes**		Varies	Don't know
Desirable effects	Trivial	Small	**Moderate**	Large		Varies	Don't know
Undesirable effects	Large	Moderate	Small	Trivial		**Varies**	Don't know
Certainty of evidence	Very low	Low	**Moderate**	High			No included studies
Values	Important uncertainty or variability	Possibly important uncertainty or variability	**Probably no important uncertainty or variability**	No important uncertainty or variability			
Balance of effects	Favors the comparison	Probably favors the comparison	Does not favor either the intervention or the comparison	**Probably favors the intervention**	Favors the intervention	Varies	Don't know
Resources required	Large costs	**Moderate costs**	Negligible costs and savings	Moderate savings	Large savings	Varies	Don't know
Certainty of evidence of required resources	Very low	**Low**	Moderate	High			No included studies
Cost effectiveness	Favors the comparison	Probably favors the comparison	Does not favor either the intervention or the comparison	Probably favors the intervention	**Favors the intervention**	Varies	No included studies

Equity	Reduced	Probably reduced	Probably no impact	**Probably increased**	Increased	Varies	Don't know
Acceptability	No	Probably no	**Probably yes**	Yes		Varies	Don't know
Feasibility	No	Probably no	Probably yes	**Yes**		Varies	Don't know

In summary, the above table offers a structured overview of the GDG's deliberations and final judgements for each EtD criterion. This synthesis ensures transparency and consistency in the decision-making process, ultimately guiding the formulation of robust, evidence-based recommendations for clinical practice.

Annex 5. Situations in which a ‘strong’ recommendation coexists with ‘low’ or ‘very low’ certainty evidence

In guideline development, the strength of a recommendation typically reflects the certainty of the underlying evidence. However, there are exceptional circumstances where guideline panels may issue strong recommendations despite low or very low certainty evidence. These scenarios arise when the potential benefits, risks, or contextual factors compel a decisive course of action, even in the absence of robust evidence. This annex outlines the situations in which such “discordant” recommendations may be justified.

Table A5 lists paradigmatic situations in which a “discordant” strong recommendation may be indicated, despite low or very low certainty evidence.

Table A5: Paradigmatic situations

Situation	**Confidence in effect estimates (evidence quality)**		**Benefits versus harms**	**Value judgements and preferences**	**Resource considerations**	**Type of recommendation**	**Example of a discordant recommendation**
	Benefits	**Harms**					
Life threatening situation	Low or very low	Immaterial (very low to high)	Intervention may save lives in a life-threatening situation. Adverse events not prohibitive.	A very high value is placed on an uncertain but potentially life-preserving benefit.	Small incremental cost or use of resources relative to benefits justifies the intervention.	Strong recommendation in favour of the intervention.	In the treatment of patients with MDR-TB, the BPaLM regimen should be used (2).
Uncertain benefit, certain harm	Low or very low	High or moderate	Possible but uncertain benefit. Substantial established harm.	A much higher value is placed on the harmful effects, which are certain, than on the benefits, which are uncertain.	Possible high incremental costs or use of resources in the face of uncertain benefits may dictate the need for a recommendation against the intervention.	Strong recommendation against the intervention (or in favour of a less harmful or costly comparator).	We recommend against screening for androgen deficiency in the general population (3).
Potentially equivalent options, one	Low or very low	High or moderate	Both alternatives show similar – though uncertain, benefits, but	A high value is placed on avoiding harm.	High incremental cost (or resource use) relative to	Strong recommendation in favour of the	For management of post partum haemorrhage, oxytocin should be preferred over ergometrine alone, a fixed-dose

clearly less risky or costly than the other			one is certainly less harmful or expensive than the other.		benefits may justify recommending the comparator, if less harmful.	less harmful or costly comparator.	combination of ergometrine and oxytocin, carbetocin and prostaglandins (4).
High confidence in benefits being similar, but one option potentially more risky or costly than the other	High or moderate	Low or very low	Have established that alternative management strategies afford similar benefits, but one of them may be more harmful than the other (low certainty).	A high value is placed on avoiding harm.	High incremental cost (or resource use) of one intervention may justify recommending the comparator, if less harmful.	Strong recommendation against the potentially more harmful or costly comparator.	In women requiring anticoagulation and planning conception or in pregnancy, the American College of Chest Physicians' guidelines recommended *against* the use of certain anticoagulants (5). For example, high confidence estimates suggest similar effects of different anticoagulants. However, indirect evidence (low confidence in effect estimates) suggests potential harm to the unborn infant with oral direct thrombin (e.g. dabigatran) and factor Xa inhibitors (e.g. rivaroxaban, apixaban).
Potential catastrophic harm	Immaterial (very low to high)	Low or very low	Intervention potentially quite harmful, while its benefit varies in magnitude.	A high value is placed on avoiding greater harm.	High incremental cost (or resource use) of the potentially more harmful intervention may further justify recommending the less harmful comparator.	Strong recommendation against the intervention (or in favour of the less harmful or less expensive comparator).	Children with suspected or confirmed pulmonary TB or TB peripheral lymphadenitis living in settings with high HIV prevalence (or with confirmed HIV infection) should *not* be treated with intermittent regimens (6).

BPaLM Bedaquiline, Pretomanid, Linezolid, Moxifloxacin; HIV: human immunodeficiency virus; MDR-TB: multidrug-resistant TB; TB: tuberculosis.

Source: WHO, 2014 *(7)*; adapted from Andrews et al. (2013) *(8)* and Alexander et al. (2014) *(9)*.

In conclusion, the presence of low or very low certainty evidence does not always preclude the formulation of strong recommendations. Well-defined clinical contexts—such as life-threatening conditions, clear potential for catastrophic harm, or situations where alternative options are equally uncertain but one is less harmful or costly – guideline groups may issue strong recommendations by carefully considering and transparently justifying these decisions. In such situations, the rationale for discordant recommendations should be clearly communicated. This approach balances the need for practical guidance with the imperative to act in the best interest of patients, particularly in urgent or high-stakes clinical scenarios.

References

1. Parmelli E, Amato L, Oxman AD, Alonso-Coello P, Brunetti M, Moberg J, et al. Grade Evidence to Decision (EtD) framework for coverage decisions. Int J Technol Assess Health Care. 2017;33(2):176-82.
2. World Health Organisation. WHO consolidated guidelines on tuberculosis. Module 4: Treatment. Drug-resistant tuberculosis treatment. Geneva, Switzerland; 2022.
3. Bhasin S, Cunningham GR, Hayes FJ, Matsumoto AM, Snyder PJ, Swerdloff RS, et al. Testosterone therapy in men with androgen deficiency syndromes: an Endocrine Society clinical practice guideline. J Clin Endocrinol Metab. 2010;95(6):2536-59.
4. World Health Organisation. WHO guidelines for the management of postpartum haemorrhage and retained placenta. . Geneva: World Health Organization; 2009.
5. Bates SM, Greer IA, Middeldorp S, Veenstra DL, Prabulos AM, Vandvik PO. VTE, thrombophilia, antithrombotic therapy, and pregnancy: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. Chest. 2012;141(2 Suppl):e691S-e736S.
6. World Health Organisation. Rapid advice: treatment of tuberculosis in children. Geneva: World Health Organisation; 2010.
7. World Health Organisation. WHO handbook for guideline development. 2025.
8. Andrews JC, Schünemann HJ, Oxman AD, Pottie K, Meerpohl JJ, Coello PA, et al. GRADE guidelines: 15. Going from evidence to recommendation-determinants of a recommendation's direction and strength. J Clin Epidemiol. 2013;66(7):726-35.
9. Alexander PE, Bero L, Montori VM, Brito JP, Stoltzfus R, Djulbegovic B, et al. World Health Organization recommendations are often strong based on low confidence in effect estimates. J Clin Epidemiol. 2014;67(6):629-34.

Annex 6. Randomised control trials investigating tuberculosis preventative therapy: a systematic review

Rationale

Published clinical trials evaluating the effectiveness of tuberculosis preventive therapy TPT report a range of primary and secondary trial outcomes. This systematic review aimed to summarise the outcomes reported by recent phase 3 and 4 clinical trials of TPT.

Methods

This systematic review was performed in accordance with the preferred reporting items for systematic reviews (PRISMA) (1). Electronic searches were conducted in PubMed, Cochrane Library, and EMBASE databases for English language articles published from 1 January 2000 until 30 March 2025.

Inclusion criteria included: i) The study population comprised people taking TPT, regardless of confirmed TB infection or presumed drug resistance of the source case; ii) study design included randomised clinical trials (phase 3 or 4), including publications reporting secondary outcomes of clinical trials and published protocol; iii) English language studies.

Exclusion criteria included: i) trials describing treatment of tuberculosis disease only, ii) trials where protocols or registries lacked clear definitions of study outcome, iii) Phase 1 or 2 trials; iv) trials of TB vaccines.

Results

The search strategy identified 6149 studies for abstract and title review. Of 94 articles identified for full-text review, 46 were included in the systematic review. The features of the 46 included randomised trials are presented in Table A6.

Populations included

Geographically, 14 (30.4%) of the trials were conducted in Africa and an equal percentage in Asia, five (10.9%) in Europe, and seven across multiple regions (15.2%). Regarding specific population, 25 studies (54.3%) enrolled people living with HIV, with 16 (34.8%) exclusively including this group. Children were included in 12 studies (26.1%), and 7 (15.2%) of these focused solely on children. Only one study (2.2%) included pregnant women, and another (2.2%) reported outcomes in an elderly population (50-69 years) in rural China. Other comorbidities studied included diabetes (8 studies, 17.4%), hepatitis C virus infection (7 studies, 15.2%), and renal transplantation (2 studies, 4.3%).

Regimens used

Of the 46 included studies, 14 (30.4%) evaluated use of isoniazid of different durations. Other interventions included rifampicin and pyrazinamide in (3 studies, 6.5%), isoniazid and rifampicin (1 study, 2.2%), rifampicin alone (7 studies, 15.2%), and various combinations of weekly or

twice weekly isoniazid with rifapentine 1 to 3 months (11 studies, 23.9%). Six studies (13.0%) investigated multiple TPT regimens. Two studies (4.3%) specifically investigated 6 months of levofloxacin therapy in contacts of MDR/RR-TB.

Outcomes assessed

The most frequently assessed primary outcome was the incidence of microbiologically or clinically confirmed TB, reported in 26 (56.5%) studies. Overall, 31 studies (69.6%) assessed TB incidence as a primary or secondary outcome. Adverse events (AEs) were the primary outcome in 9 studies (19.6%) and were assessed as either a primary or secondary outcome in 36 studies (78.3%). A standardized method for AE reporting, such as CTCAE or DAIDS tables, was used in 24 studies (52.2%). Adverse events of special interest (AESI) were assessed in 33 studies (71.7%), with hepatotoxicity being the most common (29 studies, 63.0%). Serious adverse events (SAEs) were assessed in 28 studies (60.9%), and treatment discontinuation due to AEs was assessed in 10 studies (21.7%).

Treatment completion was the primary outcome in 7 studies (15.2%) and was reported as a primary or secondary outcome in 29 studies (63.0%) overall. However, methods for outlining treatment completion varied considerably among the 19 studies (41.3%) that reported them. Treatment adherence was assessed in 28 studies (60.9%). Only two studies (4.3%) reported costs associated with TPT interventions, and no studies reported patient-reported outcomes. Only four studies (8.7%) investigated the pharmacokinetics of the drugs used.

Limitations of included studies

The 46 studies included in this systematic review provide valuable insight into the information already available about TPT regimens from clinical trials. However, there are several limitations of the studies, and future studies should consider these to address research gaps. The studies included have limited reporting on adverse events of special interest and severity. Several studies either did not classify AEs by grade (1-5) or did not report on Grade >3 SAEs. This makes comparisons between studies difficult. Furthermore, included adverse events are often generalized (e.g., "Hepatotoxicity"), and there is limited information on non-hepatotoxic adverse events. Secondly, the studies reported inconsistent adherence measurement and reporting. Some trials report the use of pill count, Directly Observed Therapy, and some report measurement of biomarkers such as urine metabolites. There was also variability in the definition of treatment completion in terms of the portion of pills completed as well as the time taken to complete therapy. Such variations also make comparisons difficult, limiting applicability in real-life scenarios.

For the included studies, there was underrepresentation of specific subpopulations and comorbidities, in particular pregnant women, young children, the elderly, and those with comorbidities. There was significant heterogeneity in the assessment of prevalent and incident

TB. The specific diagnostic tests (phenotypic versus genotypic), follow-up duration, and diagnostic algorithms vary between trials. The true incidence of TB between trials during the follow-up period is therefore difficult to ascertain. There was also sparse data on long-term outcomes beyond two years of follow-up. Finally, several trials compare newer short-course regimens (e.g., 3HP, 1HP, 4R) to traditional isoniazid monotherapy (6H or 9H), but few head-to-head comparisons between regimens exist. Overall, the heterogeneity between studies makes direct comparisons and generalisability difficult. Harmonisation of study methods can improve the evidence available for the WHO guideline development process and increase the strength of recommendations.

Table A6: Demographic features of tuberculosis preventative therapy trials published since the year 2000

Author (ordered by year of publication)	Setting(s)	Regimens (intervention vs control)	Inclusion criteria	Age range (years)	%children (< 15)	% pregnant women	% PLWH	Included subpopulations	Total sample size (n)	Male sex (%)	Primary outcome measured	Secondary outcomes measured	Duration of follow up (from treatment initiation)
Fitzgerald et al 2000 (4)	Haiti	12H vs placebo	≥18 years, post primary TB disease	≥18	0	0	61	NA	233	50.2	Recurrence of TB disease	NA	24 months
Gordin et al 2000 (5)	US, Mexico, Haiti, Brazil	2RZ vs 12H	≥13 years PLWH, with positive TST	16-70	0	0	100	NA	1583	71.5	TB incidence	AE, mortality, probable TB	Mean 37 months (range NA)
Johnson et al 2001 (6)	Uganda	3HR/3HRZ vs 6H	PLWH, with positive TST	18-50	0	0	100	NA	2736	NA	TB incidence	Mortality, acquired drug resistance	2 years
Quigley et al 2001 (7)	Zambia	6H, 3RZ vs placebo	≥15 years PLWH	≥15	0	0	100	NA	1053	NA	TB incidence	Mortality	Mean 3 years (up to several years)
Leung et al 2003 (8)	Hong Kong	2RZ vs 6H	People with silicosis and positive TST	mean age 61.6 (in 2RZ arm)	0	0	0	Silicosis	76	98.7	Hepatotoxicity of any grade	NA	Up to ten years
Rivero et al 2003 (9)	Spain	6H, 3RH, 2RZ vs nil	PLWH with negative TST	18-65	0	0	100	NA	319	72	TB incidence	AE, treatment completion, mortality	2 years
Agarwal et al 2004 (10)	India	12H vs no treatment	People with ESRF renal transplant	15 to 58	0	0	0	Renal transplant	85	92.6	TB incidence	Hepatotoxicity grade ≥ 3	30 months
Menzies et al 2004 (11)	Canada	4R vs 9H	Adults with positive TST	≥18	0	0	0	NA	116	56	Percentage of TPT doses taken	AE causing discontinuation, treatment costs	Treatment duration
Tortajada et al 2005 (12)	Spain	2RZ vs 6H	Contacts of confirmed TB cases with positive TST	≥1	17.3 (≤19 years)	0	0	NA	352	50.9	Safety and tolerability	AE	Treatment duration

Author (ordered by year of publication)	Setting(s)	Regimens (intervention vs control)	Inclusion criteria	Age range (years)	%children (< 15)	% pregnant women	% PLWH	Included subpopulations	Total sample size (n)	Male sex (%)	Primary outcome measured	Secondary outcomes measured	Duration of follow up (from treatment initiation)
Vikrant et al 2005 (13)	India	12H vs no treatment	≥14 years on renal replacement therapy	16-53 for 12H; 14-62 for control arm	0	0	0	ESRF, HBV, HCV	109	84.4	TB incidence	Hepatotoxicity grade ≥ 3	Up to 3 years
Naqvi et al 2006 (14)	Pakistan	9H vs no treatment	Renal allograft recipients	NA	0	0	0	Renal transplant, HCV	480	NA	TB incidence	NA	2 years
Zar et al 2007 (16)	South Africa	H vs placebo	≥8 weeks old neonates with HIV who are contacts of TB cases	≥8 weeks	100	0	100	NA	263	56	TB incidence	Grade 3/4 AE, mortality	Median 5.7 months [IQR 2.0-9.7 months]
Mohammed et al 2007 (17)	South Africa	12H vs placebo	PLHW with negative TST	≥18	0	0	100	NA	118	49.2	TB incidence	AE, adherence, death, change in CD4	2 years
Spyridis et al 2007 (18)	Greece	3HR, 4HR vs 9H	≤15 years with TBI	0-15	100	0	0	NA	926	51.4	Adherence, treatment completion	AE, TB incidence	3 years
Menzies et al 2008 (19)	Saudi Arabia, Brazil, Canada	4R vs 9H	Adults with TBI	≥18	0	0	1.5	NA	847	52.7	AE causing discontinuation	Treatment completion	Treatment duration
Madhi et al 2011 (20)	South Africa, Botswana	H vs placebo	HIV exposed infants 91 to 120 days	91 to 120 days	100	0	40.4	NA	1352	48	TB incidence	Mortality	96-108 weeks
Martinson et al 2011 (21)	South Africa	3HP, 3HR, 6 years H vs 6H	PLWH with positive TST	IQR 26.4-34.7	0	0	100	NA	1148	16.7	TB incidence	AE causing discontinuation, adherence	Mean 4 years (up to 6 years)
Samandari et al 2011 (22)	Botswana	36H vs 6H + placebo	≥18 years PLWH	IQR 28-39	0	0	100	NA	1995	28	TB incidence	Mortality	36 months
Sterling et al 2011 (23)	United States, Canda, Brazil, Spain	3HP vs 9H	≥2 years, close contacts of TB cases with positive TST	IQR 25-46	0	0	2.6	HBV, HCV	7731	54.5	TB incidence	AE causing discontinuation, grade 3-5 AE, mortality, acquired	33 months

Author (ordered by year of publication)	Setting(s)	Regimens (intervention vs control)	Inclusion criteria	Age range (years)	%children (< 15)	% pregnant women	% PLWH	Included subpopulations	Total sample size (n)	Male sex (%)	Primary outcome measured	Secondary outcomes measured	Duration of follow up (from treatment initiation)
												drug resistance, treatment completion	
Chan et al 2012 (24)	Taiwan	4R vs 6H	HIV negative prison inmates	≥18	0	0	0	HCV, diabetes	373	100	AE causing discontinuation	Hepatotoxicity of any grade, other causes of treatment cessation	Treatment duration
Swaminathan et al 2012 (25)	India	6H vs 36H	≥18 years PLWH	≥18	0	0	100	NA	712	37	TB incidence	AE, mortality	36 months
White et al 2012 (26)	United States	4R vs 9H	Prison inmates with TBI	NA	0	0	0	HCV	364	93	Toxicity, treatment completion	AE, adherence	Treatment duration
Jimenez et al 2013 (27)	Spain	3HR vs 6H	Immigrants with TBI	12-40	0	0	0	NA	590	67.8	Adherence, AE	TB incidence at 5 years	5 years
Gray et al 2014 (28)	South Africa	H vs placebo	≥8 weeks infants with HIV	IQR 17-63 months	100	0	100	NA	167	49.7	TB disease, mortality	AE, hospital admissions, adherence	34 months
Kim et al 2015 (29)	South Korea	9H vs no treatment	IGRA positive renal pancreas transplant recipients	≥18	0	0	0	Diabetes, hypertension	263	65	TB incidence	Mortality, transplant rejection	Mean 21 months (IQR 1.1 - 2.5 years)
Rangaka et al 2015 (30)	South Africa	12H vs placebo	≥18 years PLWH on ART	IQR 30-40	0	0	100	NA	1329	24.9	TB incidence,	AE, mortality, loss to follow up	Median 2.5 years (up to 3.7 years)
Temprano et al 2015 (31)	Ivory Coast	6H + ART vs ART	≥18 years PLWH with CD4 <800 cells/mm^3	IQR 29-42	0	0	100	NA	2056	21.5	Mortality, AIDs defining illness, non-AIDs defining cancer or invasive	Grade 3/4 AE	30 months

Author (ordered by year of publication)	Setting(s)	Regimens (intervention vs control)	Inclusion criteria	Age range (years)	%children (< 15)	% pregnant women	% PLWH	Included subpopulations	Total sample size (n)	Male sex (%)	Primary outcome measured	Secondary outcomes measured	Duration of follow up (from treatment initiation)
											bacterial disease		
Villarino et al 2015 (32)	United States, Canada, Brazil, Spain, Hong Kong	3HP vs 9H	2-17 years with TBI	2-17	25 (≤12 years)	0	0.5	NA	905	50.9	AE causing discontinuation	AE of any grade, mortality, TB incidence at 33 months	33 months
Denholm et al 2016 (33)	Australia	3HP vs 9H	≥18 years with TBI	18-77	0	0	0	NA	80	48.8	Health system costs	Treatment completion	Treatment duration
Gao et al 2018 (34)	China	3HP, $2H_2P_2$ vs no treatment	Rural residents aged 50-69 years	50-69	0	0	0	Diabetes	3738	54.9	TB incidence	Grade 3-5 AE, treatment completion, mortality, AE causing discontinuation	2 years
Diallo et al 2018 (35)	Australia, Benin, Brazil, Canada, Ghana, Guinea, Indonesia, Saudi Arabia, South Korea	4R vs 9H	Children with TBI	0-17	100 (≤ 17 years)	0	0	NA	844	49.7	TB incidence	Grade 3-5 AE, treatment completion	16 months
Menzies et al 2018 (36)	Australia, Benin, Brazil, Canada, Ghana, Guinea, Indonesia, Saudi Arabia, South Korea	4R vs 9H	Adults with TBI	18-90	0	0	4	Immunosuppression	6012	40.9	TB incidence	Grade 3-5 AE, treatment completion	28 months
Sun et al 2018 (37)	Taiwan	3HP vs 9H	≥12 years, close contacts of TB cases with positive TST	≥12	0	0	0	NA	263	57.8	Treatment completion	AE	2 years
Gupta et al 2019 (38)	Botswana, Haiti, India, South Africa, Tanzania, Uganda,	6H during pregnancy vs 6H post-partum	≥18 years pregnant women living with HIV	IQR 24-33	0	100	100	NA	956	0	Pregnancy related AE	Grade 3-5 AE, mortality, TB incidence	48 weeks

Author (ordered by year of publication)	Setting(s)	Regimens (intervention vs control)	Inclusion criteria	Age range (years)	%children (< 15)	% pregnant women	% PLWH	Included subpopulations	Total sample size (n)	Male sex (%)	Primary outcome measured	Secondary outcomes measured	Duration of follow up (from treatment initiation)
	Zimbabwe, Thailand												
Swindells et al 2019 (39)	Ten countries	1HP vs 9H	≥1 years living with HIV	IQR 28-43	0	0	100	NA	3000	46	TB incidence, mortality	AE	Median 3.3 years (range NA)
Churchyard et al 2021 (40)	South Africa, Ethiopia, Mozambique	3HP vs 6H	PLWH on ART	IQR 35-49	0.3 (≤ 18 years)	0	100	NA	4014	30.4(in 3HP)	Treatment completion	Grade 3-5 AE, TB incidence, mortality, discontinuation	24 months
LaCourse et al 2021 (41)	Kenya	12H vs no treatment	HIV exposed infants	6-10 weeks	100	0	0	NA	298	52.7	TB infection incidence	Grade 3-5 AE, mortality	12 months
Ruan et al 2021 (42)	China	3HP vs no treatment	≥18 years with silicosis and HIV negative	18-65	0	0	0	NA	513	100	TB incidence	AE, treatment completion, mortality, acquired drug resistance	37 months
Surey et al 2021 (43)	United Kingdom	3HP vs 3RH	16-65 years with TBI	16-65	0	0	0	Diabetes, immunosuppression, alcohol, smoker	52	50	Treatment completion	AE	Treatment duration
Chaisson et al 2023 (44)	Uganda	3HP + ART vs ART	ART naïve PLWH with CD4 <350 cells/mm^3	IQR 25-37	0	0	100	NA	453	36.6	CD4 count pre and post treatment	NA	2 years
Tamez Torres et al 2023 (45)	Mexico	3R vs 6H	≥18 years with diabetes and positive TST	IQR 50-62	0	0	0	Diabetes, alcohol, smoker, renal disease	131	41.2	AE causing discontinuation	AE of special interest, tolerability, adherence, treatment completion	Treatment duration
Zhang et al 2023 (46)	China	6 weeks H_2P_2	18-75 years with TBI	18-75	0	0	0	HBV, diabetes,	677	58.1	TB incidence	AE, treatment	24 months

Author (ordered by year of publication)	Setting(s)	Regimens (intervention vs control)	Inclusion criteria	Age range (years)	%children (< 15)	% pregnant women	% PLWH	Included subpopulations	Total sample size (n)	Male sex (%)	Primary outcome measured	Secondary outcomes measured	Duration of follow up (from treatment initiation)
								alcohol, smoker				completion, AE causing discontinuation, mortality	
Chen et al 2024 (47)	China	$3H_2P_2$ vs 6H	Close contacts of confirmed TB cases	5-64	12.5	0	0	NA	2434	41.2	Treatment completion	NA	Treatment duration
Fox et al 2024 (48)	Vietnam	6Lfx vs placebo	Household contacts of MDR-TB cases, of any age	IQR 28-52	2.90	0	0.4	Diabetes, renal disease, HBV, HCV, HIV, lung disease	2041	36	TB incidence	Grade 3-5 AE, mortality, acquired drug resistance	30 months
Hesseling et al 2024 (49)	South Africa	6Lfx vs placebo	Children who are contacts of MDR-TB cases	IQR 1.3 - 4.2	100	0	2.1	NA	922	49.2	TB incidence	Death, grade 3-5 AE	72 weeks
Huang et al 2024 (50)	Taiwan	1HP vs 3HP	≥12 years, close contacts of TB cases	≥12	0	0	0	Diabetes, HBV, HCV	490	49.2	Drug reactions	Treatment completion	Treatment duration

AE adverse events, AIDS acquired immunodeficiency syndrome, ART antiretroviral therapy, ESRF end stage renal failure, H isoniazid, HP isoniazid and rifapentine weekly, H_2P_2 twice weekly isoniazid and rifapentine, HBV hepatitis B virus, HCV hepatitis C virus, HIV human immunodeficiency virus, IGRA interferon gamma release assay, IQR interquartile range, MDR-TB multi-drug resistant tuberculosis, NA not applicable, P rifapentine, PLWH people living with HIV, R rifampicin TB tuberculosis, Lfx levofloxacin, TB tuberculosis , TBI tuberculosis infection, TPT tuberculosis preventative therapy, TST tuberculin skin test, Z pyrazinamide

References

1. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Bmj. 2021;372:n71.
2. National Cancer Institute Division of cancer treatment and diagnosis. Common Terminology Criteria for Adverse Events (CTCAE) 2025 [Available from: https://ctep.cancer.gov/protocoldevelopment/electronic_applications/ctc.htm.
3. National Institute of Health. DAIDS Adverse Experience Reporting System (DAERS) 2025 [Available from: https://rsc.niaid.nih.gov/sites/default/files/DAERS-User-Instructional-Guide.pdf.
4. Fitzgerald DW, Desvarieux M, Severe P, Joseph P, Johnson WD, Jr., Pape JW. Effect of post-treatment isoniazid on prevention of recurrent tuberculosis in HIV-1-infected individuals: a randomised trial. Lancet. 2000;356(9240):1470-4.
5. Gordin F, Chaisson RE, Matts JP, Miller C, de Lourdes Garcia M, Hafner R, et al. Rifampin and pyrazinamide vs isoniazid for prevention of tuberculosis in HIV-infected persons: an international randomized trial. Terry Beirn Community Programs for Clinical Research on AIDS, the Adult AIDS Clinical Trials Group, the Pan American Health Organization, and the Centers for Disease Control and Prevention Study Group. Jama. 2000;283(11):1445-50.
6. Johnson JL, Okwera A, Hom DL, Mayanja H, Mutuluuza Kityo C, Nsubuga P, et al. Duration of efficacy of treatment of latent tuberculosis infection in HIV-infected adults. Aids. 2001;15(16):2137-47.
7. Quigley MA, Mwinga A, Hosp M, Lisse I, Fuchs D, Porter JDH, et al. Long-term effect of preventive therapy for tuberculosis in a cohort of HIV-infected Zambian adults. Aids. 2001;15(2):215-22.
8. Leung CC, Law WS, Chang KC, Tam CM, Yew WW, Chan CK, et al. Initial Experience on Rifampin and Pyrazinamide vs Isoniazid in the Treatment of Latent Tuberculosis Infection among Patients with Silicosis in Hong Kong. Chest. 2003;124(6):2112-8.
9. Rivero A, López-Cortés L, Castillo R, Verdejo J, García MA, Martínez-Marcos FJ, et al. [Randomized clinical trial investigating three chemoprophylaxis regimens for latent tuberculosis infection in HIV-infected patients]. Enferm Infecc Microbiol Clin. 2007;25(5):305-10.
10. Agarwal SK, Gupta S, Dash SC, Bhowmik D, Tiwari SC. Prospective randomised trial of isoniazid prophylaxis in renal transplant recipient. Int Urol Nephrol. 2004;36(3):425-31.
11. Menzies D, Dion MJ, Rabinovitch B, Mannix S, Brassard P, Schwartzman K. Treatment completion and costs of a randomized trial of rifampin for 4 months versus isoniazid for 9 months. American Journal of Respiratory and Critical Care Medicine. 2004;170(4):445-9.
12. Tortajada C, Martinez-Lacasa J, Sanchez F, Jimenez-Fuentes A, De Souza ML, Garcia JF, et al. Is the combination of pyrazinamide plus rifampicin safe for treating latent tuberculosis infection in persons not infected by the human immunodeficiency virus? International Journal of Tuberculosis and Lung Disease. 2005;9(3):276-81.

13. Vikrant S, Agarwal SK, Gupta S, Bhowmik D, Tiwari SC, Dash SC, et al. Prospective randomized control trial of isoniazid chemoprophylaxis during renal replacement therapy. Transplant infectious disease. 2005;7(3-4):99-108.
14. Naqvi R, Akhtar S, Noor H, Saeed T, Bhatti S, Sheikh R, et al. Efficacy of isoniazid prophylaxis in renal allograft recipients. Transplantation proceedings. 2006;38(7):2057-8.
15. Schechter M, Zajdenverg R, Falco G, Barnes GL, Faulhaber JC, Coberly JS, et al. Weekly rifapentine/isoniazid or daily rifampin/pyrazinamide for latent tuberculosis in household contacts. Am J Respir Crit Care Med. 2006;173(8):922-6.
16. Zar HJ, Cotton MF, Strauss S, Karpakis J, Hussey G, Schaaf HS, et al. Effect of isoniazid prophylaxis on mortality and incidence of tuberculosis in children with HIV: randomised controlled trial. Bmj. 2007;334(7585):136.
17. Mohammed A, Myer L, Ehrlich R, Wood R, Cilliers F, Maartens G. Randomised controlled trial of isoniazid preventive therapy in South African adults with advanced HIV disease. International journal of tuberculosis and lung disease. 2007;11(10):1114-20.
18. Spyridis NP, Spyridis PG, Gelesme A, Sypsa V, Valianatou M, Metsou F, et al. The effectiveness of a 9-month regimen of isoniazid alone versus 3- and 4-month regimens of isoniazid plus rifampin for treatment of latent tuberculosis infection in children: Results of an 11-year randomized study. Clinical Infectious Diseases. 2007;45(6):715-22.
19. Menzies D, Long R, Trajman A, Dion MJ, Yang J, Al Jahdali H, et al. Adverse events with 4 months of rifampin therapy or 9 months of isoniazid therapy for latent tuberculosis infection: A randomized trial. Annals of Internal Medicine. 2008;149(10):689-97.
20. Madhi SA, Nachman S, Violari A, Kim S, Cotton MF, Bobat R, et al. Primary isoniazid prophylaxis against tuberculosis in HIV-exposed children. New England Journal of Medicine. 2011;365(1):21-31.
21. Martinson NA, Barnes GL, Moulton LH, Msandiwa R, Hausler H, Ram M, et al. New regimens to prevent tuberculosis in adults with HIV infection. New England Journal of Medicine. 2011;365(1):11-20.
22. Samandari T, Agizew TB, Nyirenda S, Tedla Z, Sibanda T, Shang N, et al. 6-month versus 36-month isoniazid preventive treatment for tuberculosis in adults with HIV infection in Botswana: a randomised, double-blind, placebo-controlled trial. Lancet. 2011;377(9777):1588-98.
23. Sterling TR, Villarino ME, Borisov AS, Shang N, Gordin F, Bliven-Sizemore E, et al. Three months of rifapentine and isoniazid for latent tuberculosis infection. N Engl J Med. 2011;365(23):2155-66.
24. Chan PC, Yang CH, Chang LY, Wang KF, Lu BY, Lu CY, et al. Latent tuberculosis infection treatment for prison inmates: a randomised controlled trial. International journal of tuberculosis and lung disease. 2012;16(5):633-8.
25. Swaminathan S, Menon PA, Gopalan N, Perumal V, Santhanakrishnan RK, Ramachandran R, et al. Efficacy of a six-month versus a 36-month regimen for prevention of tuberculosis in HIV-infected persons in India: a randomized clinical trial. PLoS One. 2012;7(12):e47400.
26. White MC, Tulsky JP, Lee JR, Chen L, Goldenson J, Spetz J, et al. Isoniazid vs. rifampin for latent tuberculosis infection in jail inmates: toxicity and adherence. J Correct Health Care. 2012;18(2):131-42.

27. Jimenez-Fuentes MA, De Souza-Galvao ML, Mila Auge C, Solsona Peiro J, Altet-Gomez MN. Rifampicin plus isoniazid for the prevention of tuberculosis in an immigrant population. International Journal of Tuberculosis and Lung Disease. 2013;17(3):326-32.
28. Gray DM, Workman LJ, Lombard CJ, Jennings T, Innes S, Grobbelaar CJ, et al. Isoniazid preventive therapy in HIV-infected children on antiretroviral therapy: a pilot study. International journal of tuberculosis and lung disease. 2014;18(3):322-7.
29. Kim SH, Lee SO, Park IA, Kim SM, Park SJ, Yun SC, et al. Isoniazid treatment to prevent TB in kidney and pancreas transplant recipients based on an interferon-γ-releasing assay: an exploratory randomized controlled trial. J Antimicrob Chemother. 2015;70(5):1567-72.
30. Rangaka MX, Wilkinson RJ, Boulle A, Glynn JR, Fielding K, van Cutsem G, et al. Isoniazid plus antiretroviral therapy to prevent tuberculosis: a randomised double-blind, placebo-controlled trial. Lancet. 2014;384(9944):682-90.
31. Danel C, Moh R, Gabillard D, Badje A, Le Carrou J, Ouassa T, et al. A Trial of Early Antiretrovirals and Isoniazid Preventive Therapy in Africa. N Engl J Med. 2015;373(9):808-22.
32. Villarino ME, Scott NA, Weis SE, Weiner M, Conde MB, Jones B, et al. Treatment for preventing tuberculosis in children and adolescents: a randomized clinical trial of a 3-month, 12-dose regimen of a combination of rifapentine and isoniazid. JAMA pediatrics. 2015;169(3):247-55.
33. Denholm JT, McBryde ES, Eisen D, Street A, Matchett E, Chen C, et al. SIRCLE: a randomised controlled cost comparison of self-administered short-course isoniazid and rifapentine for cost-effective latent tuberculosis eradication. Intern Med J. 2017;47(12):1433-6.
34. Gao L, Zhang H, Xin H, Liu J, Pan S, Li X, et al. Short-course regimens of rifapentine plus isoniazid to treat latent tuberculosis infection in older Chinese patients: a randomised controlled study. The european respiratory journal. 2018;52(6).
35. Diallo T, Adjobimey M, Ruslami R, Trajman A, Sow O, Obeng Baah J, et al. Safety and Side Effects of Rifampin versus Isoniazid in Children. N Engl J Med. 2018;379(5):454-63.
36. Menzies D, Adjobimey M, Ruslami R, Trajman A, Sow O, Kim H, et al. Four Months of Rifampin or Nine Months of Isoniazid for Latent Tuberculosis in Adults. New England journal of medicine. 2018;379(5):440-53.
37. Sun HY, Huang YW, Huang WC, Chang LY, Chan PC, Chuang YC, et al. Twelve-dose weekly rifapentine plus isoniazid for latent tuberculosis infection: A multicentre randomised controlled trial in Taiwan. Tuberculosis. 2018;111:121-6.
38. Gupta A, Montepiedra G, Aaron L, Theron G, McCarthy K, Bradford S, et al. Isoniazid Preventive Therapy in HIV-Infected Pregnant and Postpartum Women. New England journal of medicine. 2019;381(14):1333-46.
39. Swindells S, Ramchandani R, Gupta A, Benson CA, Leon-Cruz J, Mwelase N, et al. One Month of Rifapentine plus Isoniazid to Prevent HIV-Related Tuberculosis. N Engl J Med. 2019;380(11):1001-11.
40. Churchyard G, Cárdenas V, Chihota V, Mngadi K, Sebe M, Brumskine W, et al. Annual Tuberculosis Preventive Therapy for Persons With HIV Infection : a Randomized Trial. Annals of internal medicine. 2021;174(10):1367-76.

41. LaCourse SM, Richardson BA, Kinuthia J, Warr AJ, Maleche-Obimbo E, Matemo D, et al. A Randomized Controlled Trial of Isoniazid to Prevent Mycobacterium tuberculosis Infection in Kenyan Human Immunodeficiency Virus-Exposed Uninfected Infants. Clinical infectious diseases. 2021;73(2):e337-e44.
42. Ruan QL, Huang XT, Yang QL, Liu XF, Wu J, Pan KC, et al. Efficacy and safety of weekly rifapentine and isoniazid for tuberculosis prevention in Chinese silicosis patients: a randomized controlled trial. Clinical microbiology and infection. 2021;27(4):576-82.
43. Surey J, Stagg HR, Yates TA, Lipman M, White PJ, Charlett A, et al. An open label, randomised controlled trial of rifapentine versus rifampicin based short course regimens for the treatment of latent tuberculosis in England: the HALT LTBI pilot study. BMC infectious diseases. 2021;21(1):90.
44. Chaisson LH, Semitala FC, Nangobi F, Steinmetz S, Marquez C, Armstrong DT, et al. Viral suppression among adults with HIV receiving routine dolutegravir-based antiretroviral therapy and 3 months weekly isoniazid-rifapentine. Aids. 2023;37(7):1097-101.
45. Tamez-Torres KM, Mongua-Rodríguez N, Ferreyra-Reyes L, Torres-Gonzalez P, Delgado-Sánchez G, Martínez-Hernández M, et al. Safety and Tolerability of Six Months of Isoniazid Plus Pyridoxine or Three Months of Rifampicin for Tuberculosis among Subjects with Diabetes Mellitus: A Randomized Trial. Microorganisms. 2023;11(8).
46. Zhang H, Xin H, Du Y, Cao X, Pan S, Liu J, et al. Tuberculosis preventive treatment among individuals with inactive tuberculosis suggested by untreated radiographic abnormalities: a community-based randomized controlled trial. Emerg Microbes Infect. 2023;12(1):e2169195.
47. Chen H, Zhang H, Cheng J, Sun D, Wang Q, Wu C, et al. Adherence to preventive treatment for latent tuberculosis infection in close contacts of pulmonary tuberculosis patients: A cluster-randomized controlled trial in China. International Journal of Infectious Diseases. 2024;147(no pagination).
48. Fox GJ, Nhung NV, Cam Binh N, Hoa NB, Garden FL, Benedetti A, et al. Levofloxacin for the Prevention of Multidrug-Resistant Tuberculosis in Vietnam. N Engl J Med. 2024;391(24):2304-14.
49. Hesseling AC, Purchase SE, Martinson NA, Fairlie L, Schaaf HS, Brigden J, et al. Levofloxacin Preventive Treatment in Children Exposed to MDR Tuberculosis. N Engl J Med. 2024;391(24):2315-26.
50. Huang HL, Lee MR, Lee CH, Cheng MH, Lu PL, Sheu CC, et al. One-month daily and three-month weekly rifapentine plus isoniazid are comparable in completion rate and safety for latent tuberculosis infection in non-HIV Population: a randomized controlled trial. Clinical Microbiology and Infection. 2024;30(11):1410-7.