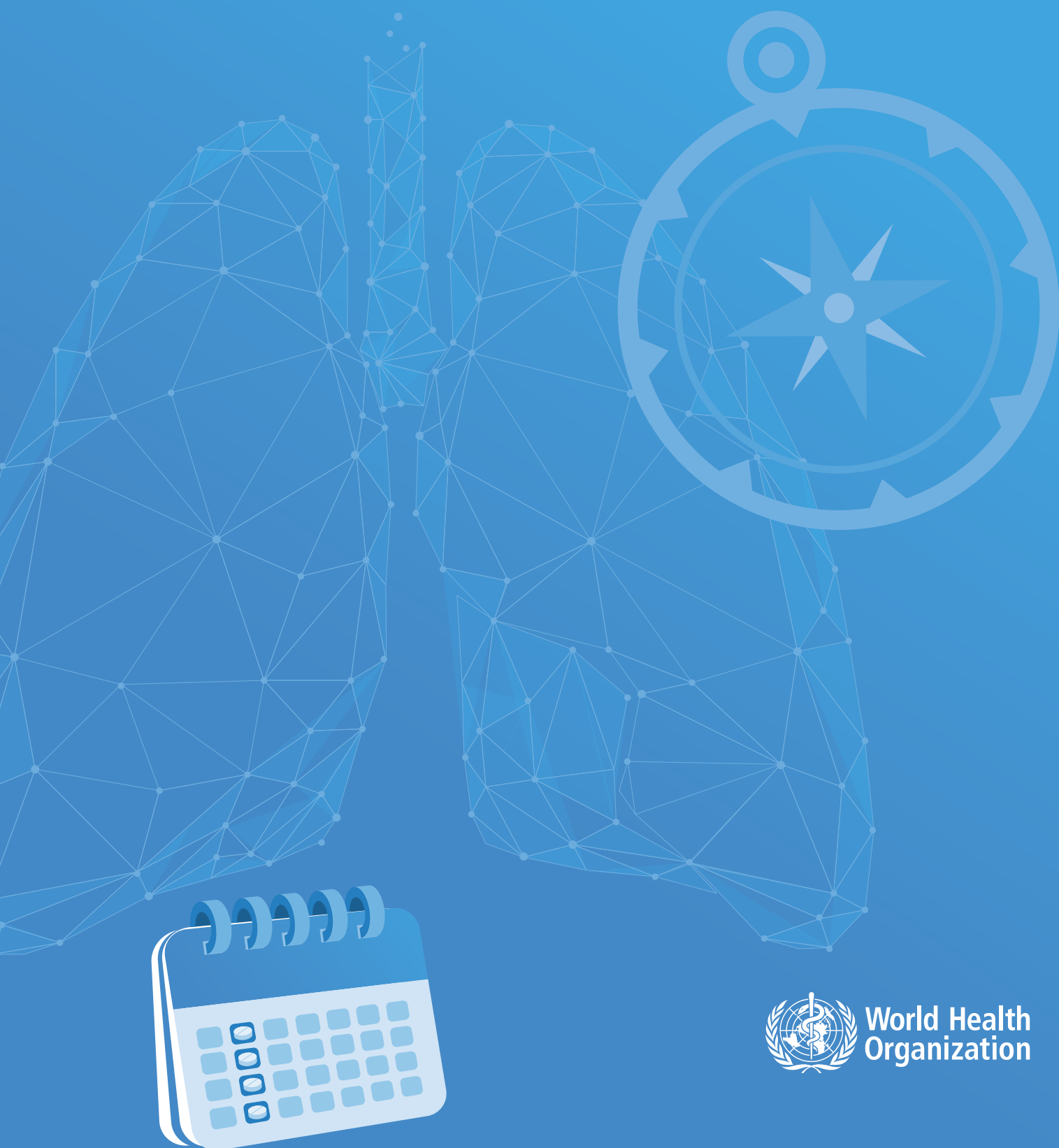


Guidance on evidence generation on new regimens for tuberculosis preventive treatment



World Health
Organization

Guidance on evidence generation on new regimens for tuberculosis preventive treatment

Guidance on evidence generation on new regimens for tuberculosis preventive treatment

ISBN 978-92-4-011718-1 (electronic version)

ISBN 978-92-4-011719-8 (print version)

© World Health Organization 2025

Some rights reserved. This work is available under the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 IGO licence (CC BY-NC-SA 3.0 IGO; <https://creativecommons.org/licenses/by-nc-sa/3.0/igo>).

Under the terms of this licence, you may copy, redistribute and adapt the work for non-commercial purposes, provided the work is appropriately cited, as indicated below. In any use of this work, there should be no suggestion that WHO endorses any specific organization, products or services. The use of the WHO logo is not permitted. If you adapt the work, then you must license your work under the same or equivalent Creative Commons licence. If you create a translation of this work, you should add the following disclaimer along with the suggested citation: "This translation was not created by the World Health Organization (WHO). WHO is not responsible for the content or accuracy of this translation. The original English edition shall be the binding and authentic edition".

Any mediation relating to disputes arising under the licence shall be conducted in accordance with the mediation rules of the World Intellectual Property Organization (<http://www.wipo.int/amc/en/mediation/rules/>).

Suggested citation. Guidance on evidence generation on new regimens for tuberculosis preventive treatment. Geneva: World Health Organization; 2025. Licence: [CC BY-NC-SA 3.0 IGO](#).

Cataloguing-in-Publication (CIP) data. CIP data are available at <https://iris.who.int/>.

Sales, rights and licensing. To purchase WHO publications, see <https://www.who.int/publications/book-orders>. To submit requests for commercial use and queries on rights and licensing, see <https://www.who.int/copyright>.

Third-party materials. If you wish to reuse material from this work that is attributed to a third party, such as tables, figures or images, it is your responsibility to determine whether permission is needed for that reuse and to obtain permission from the copyright holder. The risk of claims resulting from infringement of any third-party-owned component in the work rests solely with the user.

General disclaimers. The designations employed and the presentation of the material in this publication do not imply the expression of any opinion whatsoever on the part of WHO concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. Dotted and dashed lines on maps represent approximate border lines for which there may not yet be full agreement.

The mention of specific companies or of certain manufacturers' products does not imply that they are endorsed or recommended by WHO in preference to others of a similar nature that are not mentioned. Errors and omissions excepted, the names of proprietary products are distinguished by initial capital letters.

All reasonable precautions have been taken by WHO to verify the information contained in this publication. However, the published material is being distributed without warranty of any kind, either expressed or implied. The responsibility for the interpretation and use of the material lies with the reader. In no event shall WHO be liable for damages arising from its use.

Design and layout by Inis Communication

Contents

Acknowledgements	v
Abbreviations and acronyms	vii
Executive summary	ix
1. Introduction	1
1.1 Background	1
1.2 Purpose	2
1.3 Scope	3
1.4 Objectives	3
1.5 Audience	3
1.6 WHO guideline development and regulatory approval	4
1.7 Methodology for development of GEG document	8
2. WHO guideline development process and the GRADE approach	9
2.1 Development of WHO guidelines	9
2.2 Developing the scope and recommendation questions using the PICO format, evidence retrieval and synthesis	10
2.3 Determining values and DTs	10
2.4 Assessing the certainty of the evidence	11
2.5 Preparing evidence profiles and summary of findings tables	17
2.6 Making judgements across 12 EtD criteria	17
2.7 Developing recommendations	19
2.8 Managing potential conflicts of interest in guideline development	22
3. Guidance on evidence generation for TPT	23
3.1 Guidance on generating evidence on desirable and undesirable effects	23
3.2 Guidance on generating evidence on values, resources required, cost-effectiveness, equity, acceptability and feasibility	38
3.3 Role of qualitative research in evidence generation for TPT	42
4. Other WHO processes	45
4.1 WHO prequalification	45
4.2 Expert Review Panel	46
4.3 WHO model lists	47
4.4 WHO CSA procedure	47
4.5 WHO technical advisory groups	47
4.6 Development of WHO operational handbooks	47



References	48
Annex 1. Example of a GRADE evidence profile	54
References for Annex 1	54
Annex 2. GRADE evidence to decision framework: example of a summary of findings table	58
Annex 3. GRADE evidence to decision framework: other criteria used to develop recommendations	60
Annex 4. GRADE evidence to decision framework: example of summary judgements across the 12 evidence to decision criteria	70
Annex 5. Situations in which a strong recommendation coexists with evidence of low or very low certainty	72
References for Annex 5	75
Annex 6. Current best practice and options for tuberculosis preventive treatment trials	76
A6.1 Guidance on generating evidence on desirable and undesirable effects	76
A6.2 Guidance on generating evidence on values, resources required, cost-effectiveness, equity, acceptability and feasibility	96
References for Annex 6	101
Annex 7. Randomized controlled trials investigating tuberculosis preventive treatment: a systematic review	108
A7.1 Rationale	108
A7.2 Methods	108
A7.3 Results	108
References for Annex 7	118
Annex 8. The role of qualitative research in supporting WHO tuberculosis preventive treatment policy development: approaches and methods	122
A8.1 The role of qualitative research in supporting the development of WHO guidelines on tuberculosis preventive treatment	122
A8.2 Approaches and methods of qualitative and mixed-methods research for generating evidence	123
A8.3 How qualitative research can contribute to GRADE EtD frameworks	127
A8.4 How to generate high-certainty qualitative evidence to inform global TPT policy decision-making	129
References for Annex 8	134



Acknowledgements

This document has been prepared by the Prevention, Diagnosis, Treatment, Care and Innovation Unit in the Department for HIV, Tuberculosis, Hepatitis & Sexually Transmitted Infections (HTH) of the World Health Organization (WHO) in Geneva, Switzerland. The writing of this document was led by Avinash Kanchar (WHO/HTH) and Greg Fox and Tasnim Hasan (consultants, University of Sydney, Australia), with contributions from Dennis Falzon (WHO/HTH), under the guidance of Matteo Zignol (WHO/HTH) and the overall direction of Tereza Kasaeva, Director WHO/HTH. The content and structure of this document have been modelled on the WHO *Guidance on evidence generation on new regimens for tuberculosis treatment* released in 2024.

This document was developed through a dedicated Scientific Guidance on Evidence Generation Development Group (SGG), convened by WHO and other stakeholders, with input received via public consultation. The SGG was composed of professionals from a range of specialties who have extensive expertise and experience in public health policy, clinical trials and other studies, tuberculosis (TB) programme management, and the care and management of patients with TB: Helen Ayles (London School of Hygiene and Tropical Medicine, Zambia), Madhavi Bhargava (Yenepoya Medical College Hospital, India), Aldo Burua (National TB Programme, Uganda), Marco Cavaleri (European Medicines Agency, the Netherlands), Violet Chihota and Gavin Churchyard (The Aurum Institute, South Africa), David Dowdy (Johns Hopkins Bloomberg School of Public Health, United States of America [USA]), Mike Frick (Treatment Action Group, USA), Diana Gibb (Medical Research Council, United Kingdom of Great Britain and Northern Ireland [United Kingdom]), Amita Gupta (Johns Hopkins Bloomberg School of Public Health, USA), Yohhei Hamada (University College London, United Kingdom), Anneke Hesselning (Stellenbosch University Desmond Tutu TB Centre, South Africa), Alexander Kay (Baylor College of Medicine, Eswatini), Stephanie Law (McGill University, Canada), Christian Lienhardt (Institut de recherche pour le développement, France), Mqondisi Maphophe (South African Health Products Regulatory Authority [SAHPRA], South Africa), Alberto Matteelli (University of Brescia, Italy), Vidya Mave (Johns Hopkins Bloomberg School of Public Health, India), Dick Menzies (McGill University, Canada), Lindiwe Mvusi (National Department of Health, South Africa), Mahshid Nasehi (National TB Programme, Islamic Republic of Iran), Norbert Ndjeka (National Department of Health, South Africa), Khanyisile Nkuku (SAHPRA, South Africa), Nicole Salazar-Austin (Johns Hopkins Bloomberg School of Public Health, USA), Holger Schünemann (Humanitas University, Italy), Priya Shete (University of California San Francisco, USA), Tim Sterling (Vanderbilt University, USA), Yassen Tcholakov (WHO Civil Society Task Force, Canada), and Wenhong Zhang (Fudan University Shanghai, China). Puneet Dewan (Bill & Melinda Gates Foundation, USA) participated as an observer.

Valuable input and review was also provided by other colleagues from WHO/HTH, namely Annabel Baddeley, Annemieke Brands, Cecily Miller, Samuel Schumacher, Sabine Verkuijl and Kerri Viney, as well as Maria Mercedes Perez Gonzalez and Vaseeharan Sathiyamoorthy (WHO Science Division, Department of Research for Health), and Saskia den Boon (WHO Department of Immunization, Vaccines and Biologicals).



WHO acknowledges and is grateful for the time and support of all individuals who have significantly contributed to this document, both at the draft stage and during public consultation. External contributors signed the WHO declaration of interest forms, which were then reviewed by the Prevention, Diagnosis, Treatment, Care and Innovation Unit. The interests declared were assessed and deemed to be not significant.

The development of this guidance was made possible through funding from Unitaid and the European Respiratory Society.



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 event
API	active pharmaceutical ingredient
ART	antiretroviral therapy
CBI	community-based intervention
CCM	child contact management
CEA	cost-effectiveness analysis
CI	confidence interval
CSA	Coordinated Scientific Advice
CTCAE	Common Terminology Criteria for Adverse Events
DCE	discrete choice experiment
DCEA	distributional cost-effectiveness analysis
DST	drug-susceptibility testing
DT	decision threshold
EML	Essential Medicines List
EOI	expression of interest
ERP	Expert Review Panel
EtD	evidence to decision
FGD	focus group discussion
FPP	finished pharmaceutical product
FQ	fluoroquinolone
GDG	guideline development group
GEG	guidance on evidence generation
Global Fund	Global Fund to Fight AIDS, Tuberculosis and Malaria
GRADE	Grading of Recommendations Assessment, Development and Evaluation
GRC	Guidelines Review Committee
H	isoniazid
HIV	human immunodeficiency virus
HTH	Department for HIV, Tuberculosis, Hepatitis & Sexually Transmitted Infections
ICER	incremental cost-effectiveness ratio
IGRA	interferon-gamma release assay



IPD	individual participant data
ITT	intention-to-treat
Lfx	levofloxacin
MDR/RR-TB	multidrug-resistant and rifampicin-resistant tuberculosis
NRS	nonrandomized studies
NTP	national tuberculosis programme
P	rifapentine
PICO	population, intervention, comparator and outcome
PQ	prequalification
PQT/MED	Prequalification Team – Medicines
PRO	patient-reported outcome
QCL	quality control laboratory
QES	qualitative evidence synthesis
R	rifampicin
RCT	randomized controlled trial
ROBINS-I	Risk of Bias in Non-randomised Studies – of Interventions
SGG	Scientific GEG Development Group
SOC	standard of care
SRA	stringent regulatory authority
TB	tuberculosis
TBI	tuberculosis infection
TPP	target product profile
TPT	tuberculosis preventive treatment
TST	tuberculin skin test
WHO	World Health Organization
WLA	WHO listed authority
Z	pyrazinamide



Executive summary

The World Health Organization (WHO) has produced target product profiles (TPPs) to describe desired characteristics for new tuberculosis preventive treatment (TPT) regimens and to inform the plans of drug and regimen developers. However, the evidence available to WHO for use in guideline development has often presented challenges that have limited the strength of recommendations and their applicability to certain population groups, or have meant that the evidence is not suitable for policy development. These limitations have typically arisen in three main areas:

- the overall level of evidence is often of “low certainty” or “very low certainty”;
- supporting data may be incomplete, or inconsistently recorded or reported, which may hamper evidence synthesis and judgements about the certainty of evidence and guideline development; and
- the available evidence is often limited to the benefits and harms of a new regimen, with limited information on other Grading of Recommendations Assessment, Development and Evaluation (GRADE) decision criteria, such as feasibility, acceptability, equity, cost and cost-effectiveness.

The purpose of this guidance on evidence generation (GEG) document is to inform relevant stakeholders about key steps in the WHO guideline development process and the related evidence needs for novel regimens for TPT, including the planning, execution and reporting of studies. Although the TPPs describe what the minimal and optimal requirements are for various regimen characteristics, this document provides more detailed information on how the achievement of these requirements could or should be measured in clinical trials or other studies, from the perspective of the evidence needed to inform WHO guideline development.

In terms of scope, this document focuses mainly on providing guidance relating to generating evidence as part of or alongside clinical trials for drugs and regimens to treat tuberculosis infection. However, some of the principles outlined may also be applicable to other types of studies.

The objectives of this GEG document are to outline key steps in the WHO guideline development process and the GRADE approach; provide guidance on how evidence should be generated to optimally inform WHO guideline development; and provide a short outline of WHO processes, other than those related to guideline development, that may also be of relevance during the development and evaluation of new drugs, regimens or formulations.

The document outlines 22 key messages, summarized in Fig. A.





Fig. A. Key messages on providing GEG on new regimens for TB preventive treatment

- 1 Consider the WHO guideline development process during research design
- 2 Use inclusive selection criteria
- 3 Provide a precise specification of the intervention
- 4 Include an appropriate comparator arm in the trial
- 5 Use harmonized, disaggregated efficacy outcomes
- 6 Use harmonized, comprehensive safety outcomes
- 7 Characterize the tolerability of the TPT regimens
- 8 Characterize the acquisition of drug resistance during TPT
- 9 Measure the added value to health outcomes of shortening the regimen duration
- 10 Ensure sufficient follow-up of participants
- 11 Ensure sufficient sample size to achieve precise estimates for critical outcomes
- 12 Provide detailed analyses for each outcome
- 13 Share individual participant data
- 14 Analyse and report nonrandomized studies carefully
- 15 Formulate a strategy for extrapolation of evidence to excluded populations
- 16 Consider pharmacokinetic studies alongside trials evaluating novel drugs
- 17 Conduct research on values (i.e. the relative importance people place on health outcomes)
- 18 Gather evidence on the resources required to deliver the TPT regimen
- 19 Carry out cost-effectiveness analyses
- 20 Investigate the impact on health equity
- 21 Investigate the acceptability of the intervention
- 22 Investigate the feasibility of the intervention

TPT: tuberculosis preventive treatment;
WHO: World Health Organization.



1. Introduction

1.1 Background

About one quarter of the global population is estimated to have been infected with *Mycobacterium tuberculosis*, presenting a challenge to reducing the burden of morbidity and mortality from tuberculosis (TB) (1). 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 TB in the community. The World Health Organization (WHO) has published TPT guidelines that inform end users about the detection of TBI among target populations and the provision of TPT (2, 3), to support achieving the targets of the United Nations high-level meetings on TB and the End TB Strategy (4–6). In addition, WHO has produced target product profiles (TPPs) for TPT, which describe the optimal performance and operational characteristics of future TPT regimens (7, 8).

Evidence from research informs normative work of both WHO and regulatory bodies, with important differences in emphasis, processes and goals. Whereas regulators focus on the efficacy, safety and quality of a drug or regimen, WHO also works to determine the acceptability, feasibility, equity and resource implications of implementing a novel intervention within the contexts where it is commonly used.

The evidence available to WHO for development of guidelines on drugs and regimens often presents challenges that limit the strength of recommendations. The evidence may not apply to certain populations or may not be suitable for policy development. Limitations typically occur in three main areas. First, the evidence may give rise to recommendations that have “very low certainty”, such as when it emerges from nonrandomized studies (NRS), trials lacking a standard of care (SOC) arm as a comparator, or trials for which estimates are imprecise because of a small sample size. Second, supporting data may be incomplete, or inconsistently recorded or reported, hampering the ability of evidence reviewers to synthesize the evidence through meta-analyses (1, 9). Even when a meta-analysis can be performed, significant heterogeneity between studies, or the lack of common outcome definitions between studies, may make it difficult to interpret findings. Finally, studies presented to WHO may lack information about additional factors that are important to decision-making. In addition to evidence about efficacy and safety, the Grading of Recommendations Assessment, Development and Evaluation (GRADE) criteria 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 or efficacy or both.

This guidance on evidence generation (GEG) seeks to engage in dialogue with the stakeholders who generate evidence for the WHO guideline development process for TPT. This document

¹ Noninferiority trials are sometimes used when an established effective treatment exists, and the objective is to show that the new treatment is not unacceptably worse than the existing treatment. Demonstrating noninferiority is the statistical approach to show this. It is done in reference to a noninferiority margin that reflects a clinically meaningful difference that would be acceptable to patients and health care providers and that ensures the new treatment retains a significant proportion of the control treatment's efficacy.



serves as a companion to the WHO TPPs for TPT (7). It aims to maximize the level of certainty that guideline development groups (GDGs) can have in the evidence made available during the guideline development process. Evidence that engenders greater certainty will, in turn, result in more rapid adoption of new TPT regimens, in terms of both regulatory approval and implementation by the national TB and HIV programmes.

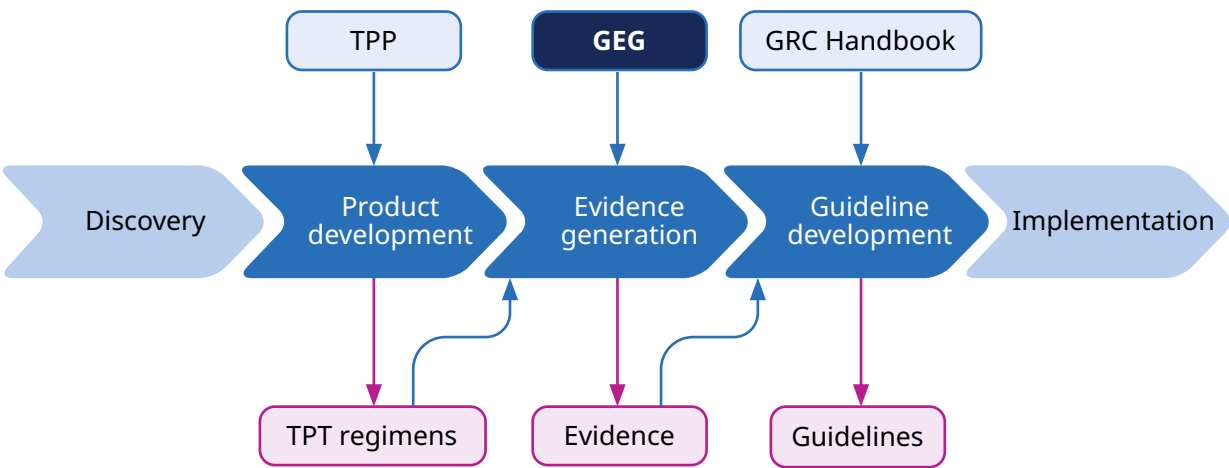
The TPPs for TPT outline the expectations for new regimens and inform the ambitions and plans of developers who are active in the field, describing the minimum and optimal requirements for various regimen characteristics (7). In contrast, this GEG presents optimal approaches to the design, conduct and evaluation of clinical trials and their substudies, with the intention of improving the evidence made available to WHO GDGs.

More trials of safer, shorter-duration and well-tolerated treatment options for TBI are urgently needed. Given the substantial burden of TBI globally, improved treatment options are essential for the rapid uptake of TPT in high-risk populations. This guidance was developed for those who design and conduct research that has the potential to inform future WHO TPT guidelines. It synthesizes information that can be incorporated at each stage of the design and conduct of clinical trials, including substudies. It outlines a set of 22 key messages that provide advice on trial design, population selection, intervention and comparator regimens, outcome reporting, data analysis and other crucial elements reviewed by GDGs. We hope that this guidance will assist the TB research community in developing high-quality evidence and 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. It addresses planning, implementation and reporting of clinical trials of TPT. Well-conducted clinical trials that adopt this guidance are expected to contribute to strong recommendations from GDGs and to enhance their applicability and uptake by implementers. Fig. 1.1 shows the place of GEG in the context of the WHO guideline development process.

Fig. 1.1. The role of GEG in the WHO discovery-to-implementation value chain^a



GEG: guidance on evidence generation; GRC: Guidelines Review Committee; TPP: target product profile; TPT: tuberculosis preventive treatment; WHO: World Health Organization.

^a The chevron process shows a simplified “discovery-to-implementation value chain”; each red box below a blue chevron shows the outputs of that step in the process, which then feeds into the next step; the blue boxes above the chevrons show the guidance documents that inform each step in this process.



1.3 Scope

This GEG focuses on 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 (see Chapter 2). This document complements a companion document for trials evaluating novel TB treatments, published by WHO in 2024 (9). However, it does not replace existing WHO normative documents, such as the *WHO handbook for guideline development*, the *WHO Guidance for best practices for clinical trials* or the TPPs (7, 10, 11), or guidance from regulators. It focuses on trials and other studies to determine whether a novel TPT regimen should be recommended or not recommended by WHO. It addresses not only the efficacy and safety outcomes but also important research needs, including qualitative research, cost-effectiveness analyses, mathematical modelling and patient-related outcomes. Mindful of the continuity between screening for TB disease and linkage to TPT under programmatic conditions, it is also important for studies to be able to assess the feasibility of integrating these two elements and to evaluate their combined impact on TB incidence in the target population. This document, however, does not include guidance on how to evaluate eligible people to rule out TB disease or on testing target populations for TBI.

This guidance is not intended to stifle innovation, impede the conduct of clinical trials or tell researchers how to undertake research. It presents methods that are currently used by WHO 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, including through harmonization of measurements; and
- 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 relevant to future WHO guidelines. Key stakeholders to whom this guidance will be relevant include research funding agencies, academic researchers (including clinical trial investigators, health economists and qualitative researchers), drug developers, consumer groups contributing to clinical trials, and biostatisticians. It will also be informative for prospective GDG members because 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 the evidence needed for WHO recommendations. This document primarily pertains to the evidence needed for WHO policy development, while also covering some regulatory requirements. Table 1.1 summarizes the key differences in scope and approach between the development of WHO recommendations and regulatory approval for TPT. Whereas regulators assess whether a drug or a regimen *can* be used, principally based on its efficacy and safety, WHO guidelines are mainly concerned with the question of whether a new drug or regimen *should* be recommended, considering both clinical and programmatic aspects. WHO guidelines typically compare new regimens with existing options available to national TB programmes (NTPs). Many drugs recommended for TPT are already recommended as part of regimens to treat TB disease and are registered for clinical use in many countries. The inclusion of such drugs in TPT regimens thus rarely presents an impediment from a regulatory standpoint. When novel drugs are being evaluated, clinical trials may serve to address both regulatory requirements and WHO guideline development and enable their use in resource-limited settings.

Table 1.1. Differences and similarities in scope and approach between WHO recommendations for and regulatory approval of TPT medicines and regimens

	WHO recommendation	Regulatory approval
Prerequisite for evaluation	Regulatory approval of the drugs within a regimen, although not necessarily for a TPT indication, by at least one SRA or WLA is required (12). Consideration is made as to whether the drugs are being used to treat or prevent TB disease.	Submission of a dossier to the applicable regulatory agency is required; for example, common technical documents, availability of manufacturing and quality management processes, nonclinical and clinical data in accordance with provisions of common technical documents, and any additional data as required by national or regional legislation.



	WHO recommendation	Regulatory approval
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), plus additional factors (e.g. likelihood of TPT completion, cost). Choice of regimen also takes into consideration the likelihood that it can be used at scale and influence TB incidence at a population level. For TBI, the main focus for efficacy is the overall comparative performance of regimens in relation to an established WHO-recommended SOC for reducing the incidence of TB disease. 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 a regimen per se. Instead, the effect and safety of the regimen are 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 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 its 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 in the WHO Prequalification Unit invitation to manufacturers to submit an expression of interest (13, 14).	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 postmarketing phase may lead regulatory agencies to recommend against the continued use of a drug and to withdraw its marketing authorization.



	WHO recommendation	Regulatory approval
Remit	WHO recommendations have a global remit (although national adaptation of global recommendations may be required, owing to implementation considerations).	Regulatory approval is national or regional, although some national authorities may follow approvals 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	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; and • feasibility.	Criteria affecting decision-making by regulatory authorities include: • preclinical pharmacology and toxicology; • dose selection; • clinical safety and efficacy; • manufacturing quality; and • 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 CSA procedure is available for new drugs (see Section 4.4). Assessment 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 through the International Council for 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 (15).
Evidence base for evaluation of benefits and harms	Systematic review or IPD meta-analysis of all available trials, or nonrandomized studies (pertinent single trials or IPD may also be reviewed), are required.	Typically, one or two pivotal Phase 3 RCTs, supported by early-phase clinical and preclinical data, are required.



	WHO recommendation	Regulatory approval
Approach to analysis and decision-making	Typically, a summary of all available evidence, with standard meta-analysis and meta-analysis of IPD from trials, is provided to WHO. Such analyses should conform 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 prespecified noninferiority criteria) is not by itself relevant to decision-making. Rather, decision-making focuses on the 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 by the GRADE framework.	Analysis of IPD is provided to the regulatory agency by the sponsor of the trial or trials, according to a predefined and agreed statistical analysis plan. Testing of a limited set of protocol-defined statistical hypotheses (often framed as superiority or noninferiority) relating to the primary endpoint or endpoints against agreed levels of significance is central to decision-making.
Considerations after recommendation and approval	Considerations include identification of research gaps or requests to improve the strength of future recommendations. WHO's remit includes operational assistance and facilitating implementation of recommended 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. alternative formulation 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.	Additional research may be required as a condition of full approval, postmarketing pharmacovigilance and (in some countries) population-specific research studies. Implementation of interventions with 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. alternative formulation or coformulation) often require new approval. Changes in dosing regimen may be accommodated in an existing label.

CSA: Coordinated Scientific Advice; 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; IPD: individual participant data; RCT: randomized controlled trial; SOC: standard of care; SRA: stringent regulatory authority; TB: tuberculosis; TBI: tuberculosis infection; TPT: tuberculosis preventive treatment; UN: United Nations; WHO: World Health Organization; WLA: WHO listed authority.



1.7 Methodology for development of GEG document

1.7.1 Establishment of the Scientific GEG Development Group

A Scientific GEG Development Group (SGG) 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 document through participation in a series of virtual meetings, as well as reviewing and providing input into multiple drafts of this document.

1.7.2 Development of the GEG document

The initial draft of the GEG document was developed by the WHO Department for HIV, Tuberculosis, Hepatitis & Sexually Transmitted Infections (HTH), Secretariat and external consultants, based on the approach followed for *Guidance on evidence generation on new regimens for tuberculosis treatment* (9). The SGG reviewed the initial draft and provided detailed written feedback, which was incorporated to develop an updated version. The feedback was also discussed at virtual meetings of the SGG. A systematic review of the literature on recent TPT trials was conducted to inform this process.

Once an advance draft was ready, the WHO HTH undertook a virtual public consultation by sharing the document widely using WHO *News Flash*, with a large outreach to the TB community. The consultation allowed stakeholders to provide written feedback on each chapter and all key messages separately. The members of the SGG also publicized the public consultation within their respective constituencies to garner more input.

The Secretariat and external consultants collectively reviewed the feedback from the public consultation and worked to finalize the GEG document.



2. WHO guideline development process and the GRADE approach

2.1 Development of WHO guidelines

One of the fundamental means by which WHO fulfils its technical leadership in health is review of evidence and development of normative products, such as guidelines (Box 2.1); “WHO’s legitimacy and technical authority lie in its rigorous adherence to the systematic use of evidence as the basis for all policies” (16).

Box 2.1. What is a WHO guideline?

A World Health Organization (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 (11).

WHO uses the GRADE approach to assess the certainty of a body of evidence and to develop recommendations (17–20). Common terms and concepts used in GRADE have been defined elsewhere (21). 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* (11), which covers many activities beyond the assessment of available evidence. This chapter provides a brief overview of some of the critical steps that are particularly relevant in the WHO guideline development process. These include:

- developing the scope and recommendation questions using the PICO (population, intervention, comparator and outcome) format for evidence retrieval and synthesis (see Section 2.2);
- determining values and decision thresholds (DTs) (see Section 2.3);
- evaluating the certainty of the evidence (see Section 2.4);
- preparing “evidence profiles” and “summary of findings” tables (see Section 2.5);



- making judgements across 12 evidence to decision (EtD) criteria (see Section 2.6); and
- developing recommendations (see Section 2.7).

WHO uses the GRADEpro software to facilitate deliberations of the GDG. GRADEpro allows the information from the evidence profiles and the EtD table to be managed and stored online (22). An example of how this approach has been applied in a GDG meeting on TPT is described in Annexes 1–4.

2.2 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 (23). 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* (11) and elsewhere – for example, in the *Cochrane handbook for systematic reviews of interventions* (24) – and is beyond the scope of this document (Box 2.2).

Box 2.2. 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 (11). It is intended mainly for WHO staff and covers the methods, processes and procedures for producing a document that meets WHO standards. The handbook is produced by WHO's Guidelines 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 best practices internationally.

2.3 Determining values and DTs

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 outcomes typically include adverse events (AEs), survival, TB incidence, TPT adherence and TPT completion. The GDG members are also required to rank all the possible outcomes from critical to not important, so that evidence reviewers focus on the most important outcomes and thus make the analyses more meaningful.

The GRADE EtD framework guides a GDG to assess and interpret the effect of an intervention and helps in making judgements about whether the desirable and undesirable health effects are absent, trivial, small, moderate or large (22, 23, 25–27). The framework allows GDG members to judge desirable and undesirable health effects, taking into account not only the size of their absolute effects but also the value of the respective outcomes from the



perspective of all stakeholders, including people affected by TB or TBI. Making judgements about whether or not a given magnitude of an intervention effect is significant is facilitated by agreeing upon DTs (i.e. quantitative reference values that can be used to classify the effect size of an intervention as absent, trivial, small, moderate or large) ahead of the assessment of the evidence (22). 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 (28). DTs were used by the GDG to determine the magnitude of effect during the development of the latest recommendations on drug-resistant tuberculosis (29). Empirical or otherwise defined DTs are used to:

- provide a reference for what is considered an absent, trivial, small, moderate or large effect, applicable to any health outcome;
- help increase the consistency and transparency of EtD judgements; and
- facilitate making these judgements and comparing them across multiple interventions.

2.4 Assessing the certainty of the evidence

Once the evidence has been retrieved and synthesized through a systematic review, a critical next step is assessing the certainty of the evidence (this was previously 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” (26–28). For qualitative evidence syntheses, the GRADE-CERQual (Confidence in the Evidence from Reviews of Qualitative Research) approach defines it as “the extent to which [a] review finding is a reasonable representation of the phenomenon of interest” (30). 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 are 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 2.1).

Table 2.1. Definitions of the four levels of certainty of evidence

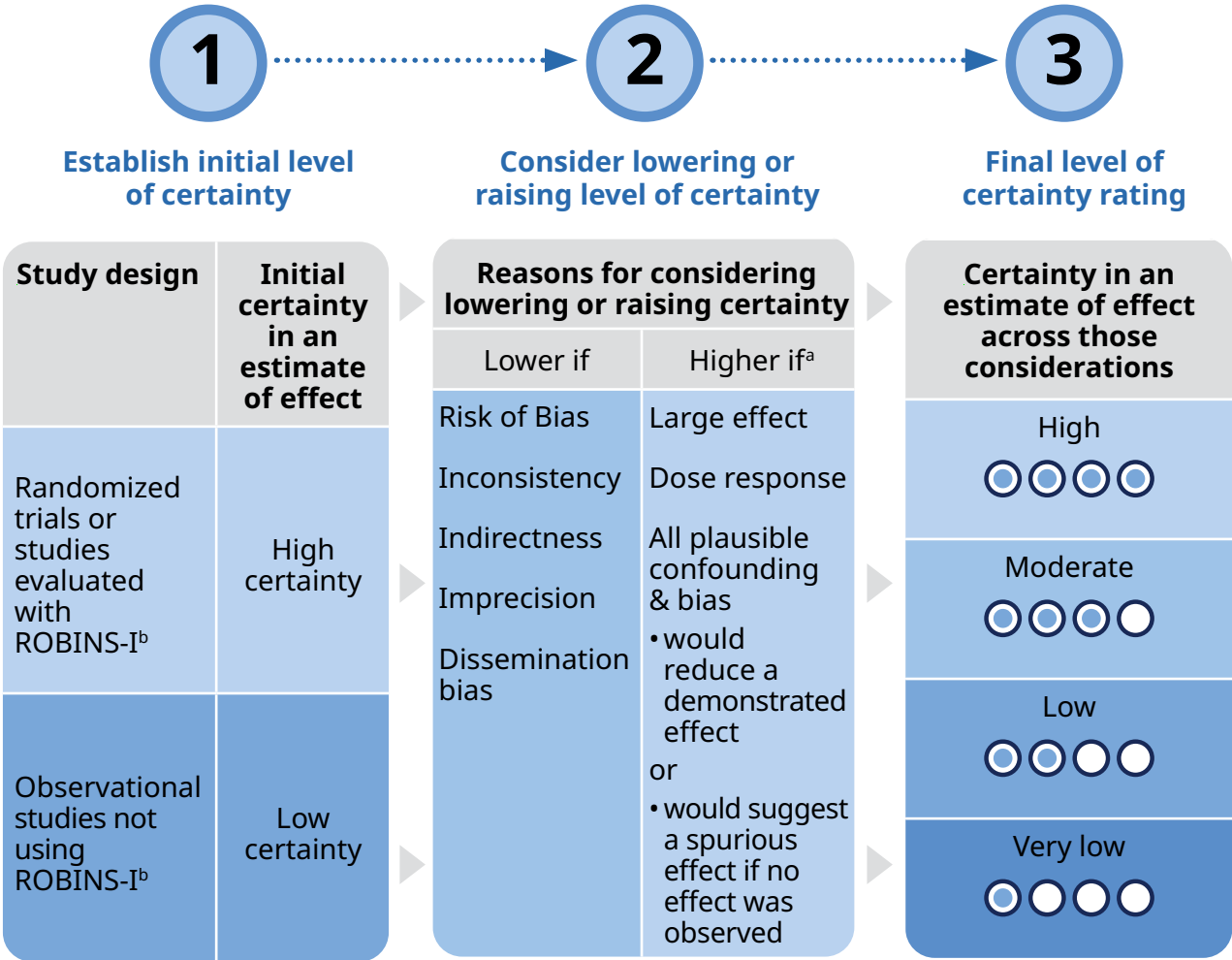
Certainty level	Definition or interpretation
High	We are very confident that the true effect lies close to that of the estimate of the effect or within a range or beyond a certain threshold
Moderate	We are moderately confident that the effect lies within a range or beyond a certain threshold: the true 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 (31).



A body of quantitative evidence based on randomized controlled trials (RCTs) is rated initially as being of high certainty, whereas evidence from NRS is rated as being of low certainty unless tools that allow for an assessment of NRS against randomized studies (e.g. ROBINS-I: Risk of Bias in Non-randomised Studies – of Interventions) find low risk of bias after assessment of confounding and selection domain (32). For both study types, the certainty of evidence starts with an initial rating. This rating can be downgraded based on five domains – risk of bias, inconsistency, indirectness, imprecision and dissemination bias (these are described in detail for quantitative evidence in Sections 2.4.1–2.4.5) – after an objective assessment. Conversely, the rating can be upgraded based on three domains: a large magnitude of effect, a dose–response relationship between intervention and outcome, and the presence of plausible residual confounding or bias that would likely increase the observed effect (if an effect is present) or strengthen confidence in a null finding (if no effect is observed). These upgrading criteria are rarely applied (Fig. 2.1). For a given body of quantitative evidence, the ratings are established separately for each outcome. Details on rating qualitative evidence are given in Section 2.4.6. The following sections give a brief outline of the certainty of evidence assessment; more detailed information can be found in the GRADE handbook (17).

Fig. 2.1. 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 Non-randomised Studies – of Interventions.

^a Upgrading criteria usually apply to observational studies only.

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

2.4.1 Risk of bias

This section outlines how judgements about the risk of bias are made; details are available elsewhere (17). Recognition of bias is essential in interpreting quantitative evidence, whether randomized or observational, during guideline development. GRADE prioritizes RCT evidence because of its lower risk of bias, and this is reflected in a higher certainty of evidence. In contrast, plausible sources of bias in observational datasets due to confounding or to residual confounding due to unknown variables, even after sophisticated statistical adjustment, may diminish confidence in the results and lead to a lower certainty of evidence (32). 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 an NTP, 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 (24).

For nonrandomized (observational) studies of interventions, the main criteria for assessment of bias depend on the study design and can be categorized as bias (33):

- 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 summarized across trials and studies for each outcome. Study limitations across the body of evidence for each outcome can be categorized 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 rating 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 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 rating for certainty of 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 assessing reasons to rate it up or down, or viewing the evidence as high certainty and assessing reasons to rate it down, the final certainty rating should be the same and should include a category of extremely serious risk of bias (32, 34). When ROBINS-I is used to rate conventional NRS of any design (e.g. cohort or case-control), often, the rating of risk of bias will be “high” after assessment of confounding and selection bias. 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 relationship or the direction of plausible confounding) or from the use of NRS with designs and analyses that have 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 this scenario.

2.4.2 Inconsistency

This section outlines how judgements about inconsistency are made; details are available elsewhere (23). 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 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.



2.4.3 Indirectness

This section outlines how judgements about indirectness are made; details are available elsewhere (23). 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; they relate to differences between the evidence identified and the recommendation question at hand, as defined in the PICO format (23, 34). Indirectness is present if:

- the population for which evidence is available differs in important ways from the population identified in the recommendation question;
- the intervention cannot be implemented with the same rigour in the intended settings of use as in the trials from which the data arose;
- no direct comparison of the intervention with the comparator of interest is available; or
- data are only available on intermediate and surrogate outcomes, because they do not provide direct evidence relating to 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.

2.4.4 Imprecision

This section outlines how judgements about imprecision are made; details are available elsewhere (23). The domain of imprecision evaluates the risk of random error within a body of evidence. Confidence intervals (CIs) around absolute estimates (e.g. risk differences or mean differences) are the main method for assessing imprecision. When these CIs 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 CIs, 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 CIs) surrounds the estimate of effect for a particular outcome.

For WHO guideline development, if the CI for the pooled estimate of effect crosses the threshold established for making one decision versus another (see Section 2.3), then the body of evidence is imprecise for that particular outcome, and the certainty of the evidence is lowered by one, two or three levels (35). Systematic review teams can use the 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 given 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 (35). Determining the acceptable threshold involves an explicit judgement.



2.4.5 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. Studies in which no effect is found are often less likely to be published. Other forms of dissemination bias include selective reporting of outcomes, delayed publication of negative results, grey literature bias and language bias (usually relating to a preference for English-language studies).

Public calls by WHO for data may help to mitigate the risk that unpublished studies remain unidentified 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 in the GDG can also mitigate this risk.

The risk of dissemination bias can be assessed using funnel plots and appropriate statistical tests (e.g. Egger's regression test). Such tests have limitations, however; the existence of dissemination bias can only be suspected, not confirmed. When dissemination 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, dissemination bias has not been noted as a common problem to date (36, 37).

2.4.6 Assessing qualitative evidence

Sections 2.4.1–2.4.5 are primarily focused on the development of quantitative evidence, following the GRADE process. Qualitative research evidence can add value to or complement quantitative evidence by providing an in-depth understanding 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 patient-reported outcome (PRO) measures. GRADE-CERQual is a transparent and structured approach to 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”) (30). 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 GRADE-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 3.3.



2.5 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 standardized format; summary of findings are tables that show abbreviated versions of the evidence profiles (see Annexes 1 and 2) (38). 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.

2.6 Making judgements across 12 EtD criteria

Once the evidence has been retrieved, summarized 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 an EtD 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 2.2 explains the 12 EtD criteria typically evaluated as part of the overall assessment of the evidence.

Table 2.2. 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? What is the background on why the problem is a priority (not discussed in this document)?
2. Desirable effects (2.6.2)	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 (2.6.2)	How substantial are the undesirable effects? How large are the undesirable effects of the intervention, considering the importance of the outcomes 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 (2.4)	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 they could affect a decision about the intervention)?



EtD criterion (GEG section)	Signalling questions
5. Values (2.3)	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 (2.6.2)	Does the balance between desirable and undesirable effects favour the intervention or the comparator? 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 (3.2.2)	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 (3.2.3)	What is the certainty of the evidence of resource requirements (costs)? Is the evidence on resource use coming from different sources than evidence of health outcomes? Does the trial situation reflect the circumstances that would be expected in clinical practice?
9. Cost-effectiveness (3.2.4)	Does the cost-effectiveness of the intervention favour the intervention or the comparator? 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 (3.2.5)	What would be the impact on health equity? Are there plausible reasons for anticipating differences in the relative effectiveness of the intervention for disadvantaged population subgroups, or different baseline conditions across disadvantaged subgroups that affect the absolute effectiveness of the intervention or the importance of the problem?
11. Acceptability (3.2.6)	Is the intervention acceptable to key stakeholders, in relation to the comparator? Are key stakeholders likely to not accept the distribution of the benefits, harms and costs, or are the costs or undesirable effects worthwhile in the short term 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 (3.2.7)	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.



2.6.1 Judgement of the magnitude of desirable and undesirable effects

During evidence assessment, outcomes are referred to as desirable or undesirable based not on their inherent nature (e.g. death is undesirable, lack of incident TB is desirable) but 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” in 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 2.3). Judgements on the magnitude of desirable and undesirable effects are influenced by how guideline panels rate the effect sizes and the relative importance of prioritized 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 TBI prioritize their reduced risk of developing TB disease (a “desirable outcome”) in comparison to their concerns about AEs (an “undesirable outcome”) has not been clearly characterized for TPT, unlike for TB disease (39). An individual’s priorities in relation to whether to take TPT are particularly important because people with TBI are typically healthy and lack symptoms of TB disease, and most have a lifetime risk of developing incident TB of less than 10% (40). For this reason, they may be considerably less willing than people being treated for TB disease to accept risks, potential harms and costs to receive the benefits of TPT. 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- and high-grade AEs should be considered when assessing the undesirable effects of a TPT regimen.

2.6.2 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). It considers the overall certainty of the evidence and how the outcomes are valued by those receiving the intervention. It is thus based on the combination of judgements on the four EtD criteria of 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, 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. For TPT, the balance between benefits and harms is particularly critical, given that an increased risk of progression to TB disease may not represent sufficient grounds to indicate systematic testing for TBI and commencement of TPT. Direct evidence of benefit from preventive medication is desirable before TPT is recommended. Establishing the balance between desirable and undesirable effects is typically the most time- and resource-intensive component of evidence generation, as it normally requires at least one Phase 3 trial.

2.7 Developing recommendations

Recommendations are developed based on the judgements made across the 12 EtD criteria, which may be displayed as a summary to serve as a basis for discussion (see Annexes 3 and 4). Typically, four main factors determine the direction and strength of a recommendation in public health (Table 2.3):

- the certainty of the evidence (see Section 2.4);



- values related to the health outcomes (see Section 2.3);
- the balance of benefits and harms (see Section 2.6.2); and
- resource implications (see Section 3.2.2).

Five types of recommendation may be made:

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

Table 2.3. Summary of key factors influencing the decision to give a strong or conditional recommendation (41)

Factor	A strong recommendation may be justified	A conditional recommendation should generally be expected
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 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 circumstances	The benefit of the intervention may not be justified in some circumstances

2.7.1 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. A strong recommendation implies that the recommendation can be adopted as policy directly by most WHO 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 (Table 2.4) (42).



Table 2.4. Implications of the strength of a recommendation for different stakeholders (11)

Perspective	Strong recommendation	Conditional recommendation
Patients	• Most individuals in this situation would accept the recommended course of action, and only a small proportion would not. • Individuals are unlikely to require aid in making decisions consistent with their values and preferences.	• The majority of individuals in this situation would accept the suggested course of action, but many would not.
Clinicians	• Most individuals should receive the intervention. • Adherence to this recommendation, according to the guidelines, could be used as a quality criterion or performance indicator.	• Different choices will be appropriate for different patients, and patients should be assisted to arrive at a management decision consistent with their values and preferences. • Decision aids may be useful in helping individuals make decisions consistent with their values and preferences.
Policy-makers	• The recommendation can be adopted as policy in most situations.	• Policy-making will require substantial debate and the involvement of various stakeholders.

The WHO GDG 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 (43). 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.

2.7.2 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. A conditional recommendation implies that substantial debate may be required before the policy is adopted by WHO Member States, clinicians may need to discuss different management options with each patient, the majority of patients would want the recommended course of action but many may not, and additional research would be likely to strengthen and possibly alter the recommendation (Table 2.4) (42).

2.7.3 Conditional recommendations for either the intervention or the comparator

Guideline users benefit from clear recommendations. A conditional recommendation for either the intervention or the comparator 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, as this will not provide guidance and will often be meaningless. Furthermore, a conditional recommendation for either the intervention or the comparator 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.

2.8 Managing potential conflicts of interest in guideline development

WHO has defined procedures that require GDG members, evidence reviewers and other experts involved in the development of its guidelines to declare potential financial or intellectual conflicts of interest. WHO assesses these declarations to determine if any represent a significant conflict of interest that needs to be mitigated by part or complete recusal of an individual from an expert panel. The declarations of interests are reported in the guidelines document. In the interests of transparency, it is also useful to understand the provenance of funding for the generation of evidence that is assessed for a guideline (e.g. what involvement the pharmaceutical industry may have had in the sponsorship or support of a trial). An initiative to promote transparent documentation of clinical trial costs, prospective and open sharing of protocols, registering of clinical trials, and timely sharing of results and data has been recently launched (44).



3. Guidance on evidence generation for TPT

3.1 Guidance on generating evidence on desirable and undesirable effects

The balance of desirable and undesirable effects (often referred to as the balance between benefits and harms) is often the most important determinant of a strong or conditional recommendation in the GRADE framework, alongside the certainty of the evidence. It is typically the most time and resource-intensive component of evidence generation because it normally requires at least one pivotal Phase 3 trial. This chapter describes key messages from WHO regarding major study protocol elements and related considerations that may help to increase the strength of recommendations.

Although most of these messages are specifically relevant to researchers developing clinical trial protocols, many apply to non-trial research studies. Suggestions for additional substudies and considerations relating to trial design and study protocols are also described. These additional substudies may inform implementation considerations for WHO recommendations within NTPs. Annex 6 offers methodological guidance on current best practice and options for generating robust evidence on novel TPT regimens. It serves as a complement to the key messages on evidence generation for TPT in this Chapter, expanding on operational and technical considerations to optimize the design, conduct and reporting of clinical trials and associated studies. We also make references to case studies from trials that have been completed, or are in progress, as illustrative examples in Annex 6. In addition, Annex 7 presents the findings of a systematic review evaluating the design characteristics of recent TPT trials that have informed WHO evidence-based guidelines.



3.1.1 Research design



Key message #1

Consider the WHO guideline development process during research design

Downgrading of evidence because of imprecision has been by far the most frequent hurdle to achieving moderate or high certainty of evidence based on single trials; therefore, finding ways to conduct more or larger high-quality trials is critical. When designing trials, researchers should consider how the data generated by their research may be presented to a future WHO-convened GDG and how their findings might best inform global policy recommendations. Early during the protocol development, researchers should consider engaging with WHO and national policy-makers to identify relevant policy gaps and align their research design with stakeholder priorities, in accordance with the TPPs for novel TPT regimens (7). Study endpoints can then be defined with future policy recommendations in mind. Early interactions with WHO technical teams can also enable the design of TPT trials to be aligned with the priorities identified in the TPPs for TPT (7, 8).

Why this is important: Trial design needs to support the specific aims of the trial (45). However, if researchers aim to change clinical practice and increase the uptake of TPT globally, they should also consider how their evidence will inform global policy. Researchers should consider the design factors that often limit the potential impact of research by contributing to low-certainty evidence or recommendations that are otherwise limited in scope. Understanding WHO's guideline development process and its related data needs can help avert common issues that might lower the certainty of evidence or the strength of recommendations.

See Section A6.1.1 for more background on research design



3.1.2 Study population, selection criteria and setting



Key message #2

Use inclusive selection criteria

Clinical trials of TPT should aim to enrol populations from the settings where subsequent policy guidelines will be applied. Researchers are encouraged to conduct trials in a wide range of epidemiological settings and to include individuals at high risk of TB disease from diverse geographical regions (e.g. rural and urban areas, or countries with varying disease burdens). In efficacy trials, the participants should reflect the demographic, geographical and epidemiological diversity of the target population for the study regimens. Although many TPT trials include people with HIV, other key subpopulations are less often represented. A systematic review of recent Phase 3 and 4 clinical trials of TPT (see Annex 7) found that, of 46 studies, only one study included pregnant women, and one included older people (aged 50–69 years), with eight studies explicitly including individuals with diabetes, seven including those with hepatitis C infection, and two including people receiving renal transplantation.

High priorities for inclusion may be specific populations among the target groups (e.g. infants, children, pregnant and breastfeeding women, older people) and other populations with important comorbid conditions (e.g. silicosis, diabetes, undernourishment in people of any age or cardiovascular diseases). Additionally, in high transmission settings, trials involving the re-treatment of individuals previously treated for TBI may be valuable. Exclusion criteria should be clearly stated and justified.

Why this is important: The WHO GDG examines the entire body of evidence when formulating a recommendation. To maximize the generalizability and utility of recommendations, it is imperative to ensure that evidence from diverse, clinically relevant populations and epidemiological settings is available to the GDG. Excluding certain populations can mean that such groups are unable to benefit from the innovation under study or are only able to access the novel regimen after a significant delay. An inadvertent consequence of clinical trials omitting key populations is that decisions about using TPT in those populations must be made in the absence of relevant evidence. As a result, clinicians and individuals treated with TPT may bear the risks of using treatment beyond the scope of available guidance.

See Section A6.1.2 for more background on study population, selection criteria and setting



3.1.3 Specification of the intervention



Key message #3

Provide a precise specification of the intervention

TPT regimens evaluated in clinical trials comprise multiple components, each of which must be described in detail. Components of the intervention include the regimen itself (i.e. drugs, doses and durations), the process of screening to exclude TB disease before TPT, the tests for TBI, the frequency of clinic visits, the methods of adherence monitoring, AE management, provision of companion drugs (e.g. antiretrovirals) and drug-susceptibility testing of the index case. Careful consideration should be given to the feasibility of implementing each component of the intervention by NTPs in high-burden countries. These elements should also be detailed for the comparator regimens (see Section 3.1.4).

Why this is important: Understanding precisely how an intervention was delivered in a trial is essential to its interpretation. Clearly describing the trial methods allows GDGs to assess the feasibility of delivering the same intervention under real-world programmatic conditions. Furthermore, it helps determine the potential significance of deviations from the trial's methods on clinical outcomes in a programmatic setting (e.g. in terms of how treatment support is delivered). Providing comprehensive evidence on key elements in the delivery of the intervention helps GDGs to reach informed judgements for each of the EtD criteria, determine the certainty of evidence and clearly articulate implementation considerations. Precise details of how trials were implemented serve to directly inform the implementation guidance that accompanies GDG recommendations.

See Section A6.1.3 for more background on the specification of the intervention



3.1.4 Comparator regimens



Key message #4

Include an appropriate comparator arm in the trial

WHO guideline questions, often framed in the PICO format, are inherently comparative, yet the use of a placebo or no treatment as a comparator is often unethical. Therefore, it is important to include a relevant comparator arm alongside the intervention regimen. Participants in this arm should receive SOC regimens consistent with current WHO recommendations. Random allocation to the intervention or comparator regimen would increase the certainty of the evidence that is generated. For trials aiming to shorten the duration of treatment, the comparator regimen has often been a longer WHO-approved regimen (e.g. 9 months of daily isoniazid monotherapy). If no internal comparator is included in the trial, it is important to be sufficiently clear about inclusion and exclusion criteria and other decision-making to permit the selection of relevant nonconcurrent comparators from other studies. However, reliance on an external cohort as the principal comparator group may result in downgrading of the certainty of evidence due to indirectness, selection bias or unmeasured confounding.

Why this is important: Benchmarking new regimens against the current SOC is most likely to enable changes and updates to WHO recommendations. The outcomes of trials that lack an internal comparator group are difficult to interpret and do not lend themselves to strong recommendations. Comparisons to external populations and comparator groups that are not enrolled concurrently are subject to biases, leading to a lowering in the certainty of evidence. Head-to-head comparisons between different shorter regimens (e.g. 3 months of weekly rifapentine plus isoniazid versus 1 month of daily rifapentine plus isoniazid) may enable stronger recommendations than indirect comparisons using network meta-analysis.

See Section A6.1.4 for more background on comparator regimens



3.1.5 Efficacy outcomes



Key message #5

Use harmonized, disaggregated efficacy outcomes

The preferred primary outcome of TPT trials for treated individuals is survival without incident TB after an adequate follow-up period. A systematic review of published TPT trials showed that incident TB disease is the primary outcome most often assessed (56%) in clinical trials (see Annex 7), whereas many trials report other primary and secondary efficacy outcomes. Comparing trials that report different outcomes can be difficult. The trial outcomes should be clearly defined in protocols, including definitions used for incident pulmonary, extrapulmonary and paediatric TB disease. Outcome descriptors should be used to enable a common understanding by GDG members and those affected by an intervention. It may be desirable to develop and validate composite outcomes that integrate both efficacy and harms, to reflect real-world relevance (see Box A6.4 in Annex 6). When composite endpoints are used, they should also be transparently disaggregated for reporting.

Why this is important: The primary goal of TPT is to prevent incident TB disease, so study outcomes should measure this. Intermediate outcomes, or other markers that have not been formally validated as surrogate endpoints, are not typically considered for WHO guideline development. The efficacy outcomes typically rated as critical by GDG members, for the purposes of a WHO recommendation, are those that are important to programmes or patients (e.g. TB incidence, TPT tolerability; a more complete listing is provided in Box A6.3 in Annex 6). However, there may be important differences in the way in which outcomes are defined between trials. In these circumstances, novel outcome definitions should be mapped against established definitions and transparently reported. Disaggregating composite endpoints allows the GDG to make judgements about the magnitude of effect of each component, because the consequences or significance of each component outcome may differ (e.g. a composite maternal–infant outcome (46)).

See Section A6.1.5 for more background on efficacy outcomes



3.1.6 Safety outcomes



Key message #6

Use harmonized, comprehensive safety outcomes

Information on AEs should be reported, particularly the proportions of people who experienced one or more of AEs of Grade 3 or higher, AEs of special interest (see **Table A6.2** in Annex 6), and AEs of Grade 1 or 2, as each of these types of AE influences TPT adherence and completion. A systematic review of recent Phase 3 and 4 clinical trials of TPT (see Annex 7), which was conducted to inform this GEG, found that not all studies reported AEs. Of the 46 studies included, nine (20%) reported AEs as the primary outcome and 36 (78%) reported them as either a primary or a secondary outcome. A recognized instrument to measure toxicity should be used, such as the Common Terminology Criteria for Adverse Events or the grading criteria of the Council for International Organizations of Medical Sciences or the Division of AIDS (47–49). The systematic review (see Annex 7) found that just over 50% of studies used such an instrument. Reporting of events should cover a fixed period that is equal to the treatment duration of the longest regimen under evaluation (whether intervention or control) plus an appropriate washout period. Data on treatment interruptions or permanent discontinuation resulting from AEs, the long-term impact of AEs, and the resolution of such events should be included. Secondary analyses considering attribution of AEs to treatment, as determined by a blinded panel, can add value, as can analyses of the rate (rather than the proportion) of AEs.

Why this is important: Most people offered TPT are otherwise healthy and will likely be unwilling to accept substantial risks of harm from taking TPT. Consistent and comprehensive reporting of safety outcomes is critical for evaluating the performance of TPT regimens. Consistent measurement of safety outcomes between intervention and control regimens is recommended in most circumstances. Documenting AEs is important to assess treatment acceptability and to balance the risks of novel regimens against their potential benefits. Potential long-term impacts of AEs are also important; AEs that reduce treatment adherence will affect their effectiveness.

See Section A6.1.6 for more background on safety outcomes



3.1.7 Tolerability outcomes



Key message #7

Characterize the tolerability of the TPT regimens

The United States Food and Drug Administration defines drug tolerability as the degree to which overt adverse effects can be tolerated by patients (50). Measures of tolerability should be reported and include, as a minimum, detailed reporting on interruption and discontinuation of a regimen or individual drug and reporting on all AEs of Grades 1–2. Additional qualitative studies to capture any challenges with tolerability would be valuable. If available, a validated PRO measure of tolerability should be used (51). However, the systematic review of recent Phase 3 and 4 clinical trials on TPT found that only a few of the 46 included studies reported PROs through substudies (see Annex 7). In pragmatic trials, differences in tolerability between compared regimens may affect the effectiveness outcome, owing to nonadherence to a regimen that is poorly tolerated. Hence, evaluating effectiveness could be an additional or alternative way to assess differences in tolerability.

Why this is important: The tolerability profile of a given drug or regimen is important to patients. It is just as important for GDGs to assess tolerability as efficacy and safety because tolerability largely determines adherence to TPT, which in turn determines TPT success or failure. Tolerability is a key characteristic considered by GDGs because it has a direct impact on treatment uptake and risk of treatment discontinuation. It is particularly important to capture measures of tolerability for participants in closely managed trials, where the level of support provided to help participants adhere to treatment is well above that which can be provided in routine care settings in high-burden countries.

See Section A6.1.7 for more background on tolerability outcomes



3.1.8 Acquisition of drug resistance



Key message #8

Characterize the acquisition of drug resistance during TPT

M. tuberculosis, as well as the complex community of bacteria comprising the human microbiome, may develop resistance to component antibiotics as a result of TPT. Some recommended TPT regimens also include a broad-spectrum antibiotic, such as levofloxacin (Lfx). Where possible, the TB drug resistance profile should be measured and reported for specimens collected from the patient who is the likely source of TB and from a contact at the time they develop incident TB. Where the two isolates are genotypically matched, paired comparisons of resistance between the two can demonstrate acquired resistance. Whole genome sequencing can help verify the clonality of the strains and identify the appearance of new mutations in *M. tuberculosis* induced by exposure to a drug in the TPT regimen. Changes to the human microbiome as a result of TPT can be complex and may include changes to bacterial diversity, as well as the acquisition of resistance. Efforts should be made to obtain an antibiogram on the *M. tuberculosis* strain in a trial participant who develops TB after taking TPT, for comparison with the source patient. Reporting results from the monitoring of drug resistance acquisition during trials enhances messaging about the risks associated with TPT.

Why this is important: Even though the evidence is scarce, the concern that TPT may cause drug-resistant TB is prevalent among health care workers (52), and this is a known barrier to TPT implementation. This risk should be minimized when TB is carefully excluded before initiating TPT. However, the acquisition of drug resistance while on TPT is infrequently reported in studies. Having data on the risk of *M. tuberculosis* and other bacterial flora acquiring drug resistance from TPT will enable a better comparison with other established risk factors and more trustworthy assessment of the benefit-to-harm balance of TPT.

See Section A6.1.8 for more background on acquisition of drug resistance



3.1.9 Treatment duration



Key message #9

Measure the added value to health outcomes of shortening the regimen duration

Where everything else is equal, shortening TPT duration is advantageous to both the individual and the health system. Shorter TPT can improve convenience and minimize the duration of AEs and drug–drug interactions (53). It can also save health systems costs for treatment, adherence support and monitoring, although novel regimens may be expensive initially. Nonetheless, people on treatment may not necessarily value a shorter regimen (e.g. if it is less tolerable). Outcomes that are typically included in the GRADE evidence profile (see Annex 1) do not provide a direct, quantified indication of the effect of regimen duration on key outcomes. As a result, GDGs commonly draw inferences from the overall rate of AEs and treatment adherence and completion, all of which are likely to be adverse with longer TPT duration (54). Better metrics are therefore needed to express the cumulative effect of treatment shortening, including pill burden and indirect costs.

Why this is important: Over the past 50 years, the duration of TPT regimens has reduced from 9–12 months to 1–4 months, based largely on evidence from noninferiority trials of efficacy and safety (54). It is typically expected that a shorter treatment duration would enhance adherence and thus improve the effectiveness of treatment. However, the potential benefits of shorter TPT duration for patients are not captured by traditional measures of efficacy and safety in clinical trials. If a trial shows noninferior efficacy and safety for a regimen that is shorter than the SOC, but no apparent benefit of duration shortening on efficacy or safety outcomes is directly observed in programmatic settings, the magnitude of the added value is uncertain (55). It then becomes challenging to weigh the putative benefits against potential downsides of a shorter TPT regimen (e.g. increased cost). Therefore, clear and explicit quantification of the benefits of TPT shortening is preferable. More pragmatic trials or more direct ways to capture patients' experiences (e.g. via PROs) could overcome this challenge (56).

See Section A6.1.9 for more background on treatment duration



3.1.10 Duration of follow-up



Key message #10

Ensure sufficient follow-up of participants

Participants in both intervention and comparator arms should be followed up for an equivalent period; this is typically 24 months after TPT initiation or at least 12 months after completion of TPT with the longest regimen under evaluation. In high transmission settings, extended follow-up may capture the increased likelihood of reinfection after randomization 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 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 of having the outcome of interest (57). Retention of participants is important to avoid missing or misclassifying incident TB after TPT, which is an uncommon event and may not be equally distributed between those who do and do not attend follow-up.

Why this is important: Post-treatment outcomes usually determine differences in the efficacy of regimens. Shorter periods of post-treatment monitoring and lower retention of participants in follow-up are likely to decrease confidence in the evidence of efficacy, particularly in noninferiority comparisons and when duration of the novel TPT regimen is shorter than that for SOC. For TPT in particular, long-term follow-up data help assess the durability of protection provided by different regimens. Sufficient follow-up time is thus required to determine if TPT has been effective and has programmatic value. Extended follow-up can show differences in efficacy that may not be apparent in shorter-term assessments. As these events are expected to be rare, capturing them as completely as possible is critically important. The approach of assessing outcomes after a fixed duration from randomization aligns with recommendations from regulatory authorities and is typically conservative, favouring longer control regimens (58).

See Section A6.1.10 for more background on duration of follow-up



3.1.11 Sample size and statistical issues



Key message #11

Ensure sufficient sample size to achieve precise estimates for critical outcomes

Meeting certain statistical criteria (e.g. statistical significance or noninferiority) is not sufficient for decision-making in WHO guideline development. Instead, the magnitude of effects (see Section 2.6.1), how different outcomes are valued (see Section 2.3) and the precision of estimates (see Section 2.4.4) are what matter during the assessment of benefits and harms. When determining the sample size of a trial, it is important to consider the desired level of precision for all critical outcome measures (see Sections 3.1.5–3.1.7), not just for the primary (composite) outcome on which the trial hypothesis may have been based. It is useful to aim for a level of precision that avoids downgrading the certainty of evidence by more than one level.

Why this is important: It is sometimes assumed that meeting the protocol-defined criteria, such as statistical significance or noninferiority, equates to having sufficient evidence to support strong recommendations for an intervention. This is not generally true because, for example:

- the GDG considers evidence beyond the benefits and harms;
- statistical criteria are typically protocol-defined for only some of the outcomes (e.g. they may be limited to efficacy outcomes and not 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 (59).

Although any clinical trial must justify its chosen sample size on its own terms, it often does not provide sufficient evidence for a strong recommendation (although it is possible). This is because of the desirability of reproducing findings in different settings, potential limitations in sample size and power, risk of biases such as selection bias, and a lack of generalizability to all subpopulations. The traditional regulatory “two-trial rule” has not been applied in TPT policy-making in the way it has been in 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, in the absence of trials, high-quality observational studies, may be needed to strengthen GDG recommendations.

See Section A6.1.11 for more background on sample size and statistical issues



3.1.12 Analysis and reporting



Key message #12

Provide detailed analyses for each outcome

Trial findings should be reported according to relevant Consolidated Standards of Reporting Trials (CONSORT) guidelines, and the full trial protocol, statistical analysis plan and trial report should be published. Analyses should be provided for all the outcome measures listed in Annex 6 (see Sections A6.1.15 and A6.1.16), with estimates provided within each arm, and risk ratios and risk differences between arms (including 95% CIs). Subgroup analyses should be provided, stratified by variables that are known to be important predictors of treatment outcome. Analyses should be based on a harmonized approach of defining outcomes, analysis populations and estimands. Sensitivity analyses for possible alternative definitions appropriate to the trial question should be conducted and presented. Additional details on each outcome should be reported (e.g. cause of death, separate analyses for failure, and available information on those lost to follow-up).

Why this is important: Complete and transparent reporting is a prerequisite for a full and adequate understanding and assessment of available evidence. Publication of full trial protocols and statistical analysis plans (with confidential information redacted as necessary) allows a complete understanding of methods. Analyses of disaggregated outcome measures are the basis for evaluating the balance of benefits and harms during WHO guideline development, even if the number of events for certain outcomes is small. Stratified analyses (e.g. by age, HIV status, country, body mass index, measures of bacteriological burden or extent of lung damage) are important to confirm that regimens are safe and effective for key population subgroups, although it is best to be cautious and avoid overinterpreting subgroup analyses, owing to problems with smaller sample sizes (60–63). Definitions of outcomes, analysis populations and estimands vary across trials (61–63), which affects the ability to compare or combine estimates across different trials. Sensitivity analyses should be carried out because they support a better understanding of the extent to which results depend on critical assumptions or analytical approaches.

See Section A6.1.12 for more background on analysis and reporting



3.1.13 Data sharing



Key message #13

Share individual participant data

De-identified individual participant data (IPD) should be made widely available. This should preferably be done using established data repositories where data can be found and obtained through secure and standardized processes, such as the global individual patient data platform for TB treatment supported by WHO (64). Creation of a similar platform for TPT could provide benefits to researchers, such as enabling a more granular approach to analysis and adjustment and, in the case of observational data from programmatic settings, an understanding of feasibility and barriers for TPT interventions in the field.

Why this is important: Making de-identified IPD publicly available facilitates IPD meta-analyses and other research to gain further insight into 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 (65). Public data repositories can facilitate data sharing, dissemination and access to enable further research using existing data (66, 67).

See Section A6.1.13 for more background on data sharing

3.1.14 Nonrandomized studies



Key message #14

Analyse and report nonrandomized studies carefully

Where there are gaps in randomized evidence, available evidence from large longitudinal observational cohorts with high-quality data collection methods may be of value. Analyses should be conducted using modern epidemiological approaches that explicitly aim to address causal questions, to minimize bias, and assumptions should be transparently outlined. The results should be carefully interpreted in the context of other relevant existing direct or indirect randomized evidence.

Why this is important: Some aspects of TPT may not have been evaluated in randomized trials and therefore rely on NRS. This is the case for drug safety, given that AE monitoring in trials is generally insufficient to detect rare events of significance. It is therefore useful to focus operational research on surveillance of unexpected safety concerns under programmatic conditions. GDGs may consider observational data from cohort studies where trials are not conducted because of a lack of clinical equipoise or because 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 (68). When interpreting nonrandomized data, potential sources of bias need to be identified, considered and accounted for in statistical analyses (15, 69–71).

See Section A6.1.14 for more background on NRS



3.1.15 Extrapolation to excluded populations



Key message #15

Formulate a strategy for extrapolation of evidence to excluded populations

Inclusivity and direct evidence of the effectiveness of TPT for populations of interest are generally preferable. However, extrapolation of trial results to excluded populations can be justifiable and may allow recommendations to be extended to such populations. Additional information that could allow a GDG to extrapolate findings beyond the study population of the trial would be helpful (e.g. knowledge about the use of a drug in children or people with comorbidities). The possibility of extrapolation should be considered carefully at an early stage when planning a trial, to include any ancillary studies that may be needed.

Why this is important: Direct evidence of the effectiveness of TPT in a given population avoids lowering of the certainty of evidence for indirectness. Although it is not always possible to have all population subgroups included in a study, having a means to extrapolate results to the wider population who could benefit can enable NTPs to apply TPT regimens to the most eligible populations. Guidance from regulatory authorities suggests that evidence of efficacy from trials in adults may be extrapolated to children; only studies to determine the appropriate dose, safety and tolerability among children may be needed (72).

See Section A6.1.15 for more background on extrapolation to excluded populations

3.1.16 Pharmacokinetic studies



Key message #16

Consider pharmacokinetic studies alongside trials evaluating novel drugs

Drug dosing recommendations for TPT require evidence from pharmacokinetic (PK) studies conducted in the target populations for whom the treatment will be recommended. Ideally, PK data on novel TPT regimens should be generated within the study population of trials that evaluate the efficacy of the drugs. Adequately powered PK studies are also recommended when data are lacking among specific populations, such as older people, children, pregnant and lactating women, and people with HIV who are naïve to or already on antiretroviral therapy. Such studies can also facilitate understanding of concerns related to drug–drug interactions, particularly when using a rifamycin-based regimen. However, a systematic review of recent Phase 3 and 4 clinical trials on TPT (see Annex 7) found that only four of the 46 studies reported PK data for the study drugs.



Why this is important: PK studies provide information about the appropriateness of drug dosing for TPT regimens by weight band, informing implementation guidance.

See Section A6.1.16 for more background on PK studies

3.2 Guidance on generating evidence on values, resources required, cost-effectiveness, equity, acceptability and feasibility

Deliberations on the remaining EtD criteria (values, resources required, cost-effectiveness, equity, acceptability, and feasibility) should be informed by research evidence on these criteria because they are all pertinent to the direction, strength and wording of WHO recommendations. Having evidence available on these criteria is particularly important if the evidence on desirable and undesirable effects suggests that, on balance, neither the intervention nor the comparator would be favoured (i.e. they appear to be equivalent in terms of benefits and harms), as may be the case in successful noninferiority trials. In such situations, it may be possible to determine which regimen should be prioritized based on careful assessment of values, resources required, cost-effectiveness, equity, acceptability and feasibility.

Often, there is little or no evidence available to inform judgements about these EtD criteria, which can be challenging for the GDG. Trialists should therefore consider collecting evidence on these criteria alongside trials; even general, overarching findings may be useful or broadly applicable to the guideline development process.

3.2.1 Values



Key message #17

Conduct research on values (i.e. the relative importance people place on health outcomes)

Values affect the weighting of desirable and undesirable effects, thereby potentially modifying a recommendation based on the balance of effects derived from clinical trial evidence (see Sections 2.3 and 2.6.1). Better data from quantitative and qualitative studies about how different stakeholders (especially people receiving TPT, but also health care providers and policy-makers) value different outcomes would make it easier to incorporate this information more explicitly into GDG processes. Having more direct evidence on the relative importance that people with TBI place on the different health outcomes considered by GDGs would be valuable. The strength of a recommendation can also be influenced when there is uncertainty about how those affected by an intervention value its outcomes. Various types of studies can be employed to this end, embedded in or run independently of trials, and the results should be synthesized in a systematic review (73–75).

Why this is important: Judgements about the balance of effects depend not only on the magnitude of the effects but also on the importance of outcomes. For example, a small



positive effect on an outcome that is of the highest importance to stakeholders may outweigh a larger negative effect that is seen as less important.

See Section A6.2.1 for more background on values

3.2.2 Resources required



Key message #18

Gather evidence on the resources required to deliver the TPT regimen

Data should be collected on total TPT costs to both the health system and patients. A systematic review of recent Phase 3 and 4 clinical trials on TPT (see Annex 7), which was conducted to inform this GEG, found that only two of the 46 included studies (4%) reported health care costs associated with TPT interventions. Health system costs include not only the drugs but also other components of TPT provision (e.g. staff salaries during outpatient visits and testing for TBI). Patient costs, including out-of-pocket expenses, travel, time lost, income disruption and costs associated with managing side-effects, are important determinants of the acceptability and completion of TPT. It can be useful to gather evidence on the resources required to deliver both the intervention and comparator arms directly within a trial. This can help to quantify the financial burden to patients and the budget impact for implementers. Costs typically vary significantly between settings. Therefore, where possible, costs should be measured across multiple settings so that the findings are broadly representative and can inform global recommendations.

Why this is important: The resources required to implement a TPT regimen may affect the strength of a recommendation. If costs are high, the strength of the recommendation may be downgraded to “conditional” (see **Table 2.3**). The cost of an intervention significantly affects its adoption, even if there is a WHO recommendation for its use. Many NTPs struggle to implement expensive TPT regimens on a large scale, even if those regimens are considered “cost-effective”. Regimens with novel drugs are often more expensive than the SOC regimen. It is important that any evidence of potential savings with new regimens, such as by reducing the number of clinic visits or avoiding the need for medication to treat adverse drug reactions, is presented to the GDG (76).

See Section A6.2.2 for more background on resources required

3.2.3 Certainty of evidence of required resources

The certainty of evidence concerning required resources is appraised using the same criteria as for health outcomes (see Section 2.4). Evidence of actual resource use is generally preferable to indirect estimates of the costs of those resources because the certainty that a TPT regimen is associated with high or low resource requirements will have a bearing on the strength of the recommendation and trustworthiness of resource data (77).

See Section A6.2.2 for more background on certainty of evidence of required resources



3.2.4 Cost-effectiveness



Key message #19

Carry out cost-effectiveness analyses

Cost-effectiveness analyses (CEAs) combine estimates of costs associated with resource use (e.g. staff time, consultation time, diagnostic tests and medicines) with estimates of health effects (e.g. reductions in the incidence of TB), providing additional value to support policy-making. Uncertainties and variability in health effects, health system costs, resources used and willingness-to-pay thresholds all need to be carefully considered. Sensitivity analyses can evaluate the impact of varying key costs and clinical parameters on the cost-effectiveness of a new TPT regimen.

Why this is important: CEAs can inform GDG decisions by evaluating the incremental costs of a new TPT regimen (against the SOC) per unit of health improvement. In situations where the costs of an effective novel intervention that can improve outcomes exceed those of the SOC, a formal CEA can have an important influence on the strength of the recommendation. Alongside considerations of resource use (see Section 3.2.2), a CEA will influence the prioritization of an intervention (e.g. a new TPT regimen) by a programme with constrained financial resources.

See Section A6.2.3 for more background on cost-effectiveness

3.2.5 Equity



Key message #20

Investigate the impact on health equity

The following steps are proposed for a trial to evaluate the impact of a TPT intervention on health equity.

- Identify populations who may experience inequities as a result of a TPT intervention and determine their baseline risks for outcomes of interest.
- Collect demographic, socioeconomic and clinical data to assess the representation of vulnerable groups.
- Expand eligibility criteria to enrol traditionally underrepresented groups (e.g. children, pregnant women).
- Identify real-world barriers, preferences and delivery strategies for implementers (e.g. supply chain gaps, staff attitudes) and individuals (e.g. stigma, language, entitlements, transport).
- Analyse and report indicators reflecting equity, such as the differential impact of the intervention by sex or socioeconomic status.

Results from such analyses can help GDGs assess potential health inequities associated with the introduction of a TPT intervention, including how the intervention may affect these inequities or introduce new ones (78, 79).



Why this is important: TBI disproportionately affects certain high-risk populations. Identifying inequities in access to TPT helps identify potential barriers to TPT provision in these populations and guides the development of strategies to improve availability and uptake across all subgroups. Some TPT interventions, such as those that consume more resources, may increase the benefit for some groups at the expense of others. WHO guidelines should support measures that enable fair provision of care according to the needs of all populations. Guidelines can play a crucial role in promoting health equity for populations who are vulnerable or marginalized by explicitly considering how recommendations may affect them. For example, less complex interventions that could be widely implemented and made accessible to all populations, including remote, underserved or other marginalized groups at high risk of TB, are typically more likely to increase equity. Other approaches that may improve equity include streamlining care pathways, improving drug-safety monitoring for those at risk of toxicity and developing strategies to strengthen drug supply chains or drug storage.

See Section A6.2.4 for more background on equity

3.2.6 Acceptability



Key message #21

Investigate the acceptability of the intervention

Acceptability is a multifaceted concept that has been defined as “the extent to which people delivering or receiving a healthcare intervention consider it to be appropriate, based on anticipated or experienced cognitive and emotional responses to the intervention” (80). Characteristics of TPT interventions that often affect acceptability include the health outcomes themselves, uncertainties about efficacy, the duration of the regimen, pill burden, the possibility of coformulation to reduce pill burden, dosage form (e.g. dispersible formulations for children) and route of administration, dosing frequency, palatability, adherence monitoring requirements, costs, drug-safety monitoring procedures, drug-drug interactions and broader forms of treatment support needs (e.g. community, institutional and household supports). Surveys of key stakeholders can provide helpful direct or indirect evidence about the acceptability of a novel intervention. Such surveys can describe the intervention and key components and can ask direct, comparative questions about their acceptability compared with the current SOC in the same way as this question is posed in the EtD framework during guideline development (see Annex 3).

Why this is important: The characteristics of a TPT intervention affect its acceptability to patients, policy-makers and providers. Where an intervention is not acceptable to patients, they will refuse to use it or will not adhere to it in the intended manner; if it is not acceptable to policy-makers, it will not be taken up by NTPs and other providers of TPT; and if it is not acceptable to health care providers, they will hesitate to use it if they have alternative options.

See Section A6.2.5 for more background on acceptability



3.2.7 Feasibility



Key message #22

Investigate the feasibility of the intervention

The feasibility of an intervention refers to the likelihood that it can be carried out appropriately in a particular context. Considerations regarding feasibility are often based on barriers to implementation (81). Characteristics of TPT interventions that often affect feasibility include drug availability, drug procurement and supply chains, infrastructural requirements (e.g. for interferon-gamma release assay testing, health worker support), complexity of patient pathways, drug-drug interactions, drug-safety monitoring, adherence monitoring, dosing frequency, the number of component drugs, drug stability and shelf life, and the possibility of fixed-dose combinations and child-friendly formulations. Potential barriers to implementation, from the perspective of patients, providers and policy-makers, should be measured directly through stakeholder surveys, qualitative research or other methods. Ideally, evidence should also be generated on how these barriers can be overcome.

Why this is important: The feasibility of implementing a novel TPT regimen can have important implications for its uptake by countries. Logistical considerations relating to implementing a novel TPT regimen will determine whether barriers to implementation can be overcome in most or all settings. Evidence related to feasibility and how potential challenges can be dealt with plays an important part in the guideline development process, as well as in the formulation of operational guidance.

See Section A6.2.6 for more background on feasibility

3.3 Role of qualitative research in evidence generation for TPT

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 data may provide additional evidence to answer questions posed by GDGs. In particular, qualitative research may contribute to evaluating the EtD criteria of values, cost-effectiveness, 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;
- no quantitative research informing certain EtD criteria of interest;
- a requirement to understand what participants think about certain aspects of treatment, such as formulation, adherence monitoring, drug palatability or treatment monitoring; or
- a need to understand implementation barriers or feasibility issues from the perspectives of other stakeholders, including health care workers and NTP managers.



Qualitative evidence concerning an intervention or context that is similar to the one of interest (e.g. patient opinions about the acceptability of similar regimens) can be provided to the GDG as indirect evidence. Case study 1 illustrates the value of qualitative research involving policy-makers, clinicians and patients in forming recommendations for Lfx to prevent drug-resistant TB (Box 3.1).

Box 3.1. Case study 1

A World Health Organization (WHO) guideline development group (GDG) assessed the use of levofloxacin (Lfx) to treat tuberculosis (TB) infection among people at high risk of multidrug-resistant and rifampicin-resistant tuberculosis (MDR/RR-TB) (2). A qualitative survey was conducted in 30 high-burden MDR/RR-TB countries to assess the feasibility of the programmatic use of Lfx for TB preventive treatment (TPT) among contacts of MDR/RR-TB cases (82). This qualitative survey aimed to evaluate the feasibility, affordability, participant acceptance and impact on equity of Lfx.

The study found that national TB programme (NTP) managers considered Lfx to be affordable, and half of the respondents believed programmatic implementation could be achieved without additional resources. Many managers felt this intervention would promote equity. However, about one quarter were concerned that implementing TPT for MDR/RR-TB might divert resources from existing services. Stakeholders, including doctors and NTP managers, also raised concerns about potential increases in Lfx resistance and possible strain on funding and resources from scaling up.

The qualitative survey highlighted important factors influencing household contacts' decisions about accepting TPT. A small group of individuals made firm decisions based on personal beliefs about preventive medicine, without weighing the specific benefits and risks of TPT. In addition, trust in the health care system, medical research and health professionals strongly influenced perceptions of TPT's risks and benefits; these factors are largely beyond the control of trial designers and regimen developers.

Overall, the survey results, presented to the WHO GDG, reflected generally positive views about treatment. These findings contributed to the strong recommendation issued in favour of the regimen.

Annex 8 provides more details on the role of qualitative research in supporting WHO policy development, as well as approaches and methods of qualitative and mixed-methods research.



4. Other WHO processes

4.1 WHO prequalification

Problems with the quality of medicines and their supply led to the creation of the WHO medicines prequalification (PQ) process in 2001. The PQ process aims to ensure that key health products (diagnostics, medicines, vaccines, immunization-related equipment and devices, and vector control products) for high-burden diseases meet global standards of quality, safety and efficacy or performance. The aim is to optimize the use of health resources and improve health outcomes.

At WHO, the Prequalification Team – Medicines (PQT/MED) ensures that finished pharmaceutical products (FPPs) and active pharmaceutical ingredients (APIs) are safe, effective and meet internationally accepted stringent quality standards. This section briefly describes PQT/MED activities and provides references for further information.

For TPT, multiple medicines and formulations are available and recommended by WHO, including isoniazid, rifampicin and rifapentine, as well as Lfx for contacts of multidrug-resistant and rifampicin-resistant TB. Child-friendly and fixed-dose combination formulations are also prequalified to support treatment in various populations. The WHO PQT/MED plays a key role in ensuring the quality of these medicines by prequalifying both APIs and FPPs for TPT. The PQ process offers two pathways: either full assessment of a generic product or an abridged assessment for innovator or generic products already approved by a stringent regulatory authority (hereafter referred to as a WHO listed authority [WLA]). The primary mandate of PQT/MED in the area of TPT is to prequalify generic medicines or alternative formulations that have been recommended in WHO guidelines, thereby expanding access to high-quality, safe and effective TPT globally.

Key PQ activities for medicines are:

- assessment of product dossiers (for FPPs or master files and for APIs);
- inspection of manufacturing and clinical sites; and
- 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 and elsewhere (64, 72, 83).

An essential prerequisite for PQ of a medicine is an indication that the class of medical product and the specific product (dosage strength, formulation and administration route) are of public health value, as judged by the relevant WHO technical department (i.e. the disease programme). This assessment is usually based on a recommendation for use in WHO guideline documents that have undergone quality oversight through the WHO Guidelines Review Committee or inclusion in the Essential Medicines List (EML). For a product to be eligible for submission to PQT/MED, it must be included in an invitation for expression of interest (EOI). The EOIs are publicly available on the PQT/MED website; they are living documents that are revised by the corresponding WHO technical department as necessary.



Pre-submission meetings are mandatory for new applicants to PQT/MED; specific details of requirements and expectations are provided in these meetings or in other advisory type meetings. PQT/MED is ready to respond to questions regarding the development of products that reflect unmet medical needs in relation to WHO's priorities.

For generic medicines, efficacy and safety will generally be documented by a bioequivalence study (or a biowaiver in certain circumstances). In most cases, therefore, clinical efficacy studies will not be needed for PQ of generic medicines. PQT/MED covers a range of disease areas; hence, there is minimal specific PQT/MED-oriented guidance on clinical studies for specific diseases or product classes. 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 Specifications for Pharmaceutical Preparations. 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.

PQT/MED organizes annual quality assessment training of regulators and annual quality workshops for manufacturers. It provides advice within the framework of its guidelines, including review of bioequivalence protocols, but owing to conflict of interest, PQT/MED cannot provide technical assistance to manufacturers (there is a separate unit within WHO that offers technical assistance).

For a prequalified product, the assessment and inspection reports can be shared with national regulatory authorities to facilitate national registration via the so-called Collaborative Registration Procedure. WHO also prequalifies quality control laboratories (QCLs) – specifically QCLs that carry out chemical and microbiological testing of medicines – but this is an activity outside of the remit of PQT/MED.

Manufacturers of medicines are encouraged to contact PQT/MED early for guidance if a submission is planned.

- PQT/MED is available to provide technical and scientific advice to manufacturers and clinical research organizations (84).
- Pre-submission meetings are mandatory for all new applicants; the goal is to help the applicant resolve critical issues before submission (85). A pre-submission meeting allows PQT/ MED to have an overview of the product and to ensure that the applicant is on the right track, provide general guidance on how to proceed and provide guidance on identified issues that should be dealt with before submission. At the same time, the meeting is an opportunity for the applicant to introduce and discuss the intended dossier, raise questions and gain valuable feedback, and address issues before submission.

4.2 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 risks and potential benefits associated with the use of FPPs and makes recommendations to the Global Fund to Fight AIDS, Tuberculosis and Malaria (Global Fund) 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 EOIs. 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 (86), where further information can be found. Each individual invitation details the specific documents to include in a submission.

4.3 WHO model lists

The WHO Model List of Essential Medicines and Model List of Essential Medicines for Children are updated and published every 2 years. These lists are intended as a guide for countries or regional authorities to adopt or adapt into national EMLs, in accordance with local priorities and treatment guidelines for the development and updating of national EMLs. Selection of a limited number of 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 patients. Information for applicants preparing a submission for meetings of the WHO Expert Committee on Selection and Use of Essential Medicines is available from the WHO website (87).

4.4 WHO CSA procedure

The CSA procedure is used for developers working on a well-defined specific product that will meet WHO conditions for scientific advice. Consequently, if the developer of a single product would benefit from an early-stage interaction with WHO to ask specific questions or receive feedback on their product development plan, the WHO CSA procedure is applicable. The CSA occurs based on a submission to WHO from a product developer. It follows an open submission process, with a clearly defined single entry point that is the WHO Science Division's Research for Health Department. WHO's advice to the product developer will be coordinated between one or more relevant technical departments and the PQT/MED through a formal and harmonized process. Further information about the scope and process is available on the WHO website (88).

4.5 WHO technical advisory groups

In addition to GDGs, WHO 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. An example is the Technical Advisory Group on dosing of TB medicines for adults and children, which revised TPT weight-band dosing (for the 3-month weekly rifapentine plus isoniazid regimen, based on PK data for children aged below 2 years) and made recommendations for dosing of Lfx in children, ahead of the 2024 update of the WHO TPT guidelines and operational handbook (89).

4.6 Development of WHO operational handbooks

When a GDG makes significant changes to recommendations for TPT, WHO may revise the operational handbook to accompany the new guidance. The handbook addresses issues of practical importance to NTPs, including dosing considerations, approaches to monitoring for toxicity and AEs of special interest (3). It also includes other implementation considerations, based on expert opinion and operational research. This operational guidance is often informed by the design of the trials on which the new guidance is based. The operational handbook may address issues that are raised as priorities by the GDG 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:e1002152 (<https://doi.org/10.1371/journal.pmed.1002152>).
2. WHO consolidated guidelines on tuberculosis: module 1: prevention: tuberculosis preventive treatment, 2nd ed. Geneva: World Health Organization; 2024 (<https://iris.who.int/handle/10665/378536>). License: CC BY-NC-SA 3.0 IGO.
3. WHO operational handbook on tuberculosis: module 1: prevention: tuberculosis preventive treatment, 2nd ed. Geneva: World Health Organization; 2024 (<https://iris.who.int/handle/10665/378535>). License: CC BY-NC-SA 3.0 IGO.
4. The end TB strategy. Geneva: World Health Organization; 2015 (<https://iris.who.int/handle/10665/331326>).
5. Political declaration of the high-level meeting of the General Assembly on the fight against tuberculosis : resolution / adopted by the General Assembly (A/RES/73/3). New York: United Nations; 2018 (<http://digitallibrary.un.org/record/1649568>).
6. Political declaration of the high-level meeting on the fight against tuberculosis: resolution / adopted by the General Assembly (A/RES/78/5). New York: United Nations; 2023 (<https://digitallibrary.un.org/record/4025280>).
7. Target product profiles for tuberculosis preventive treatment. Geneva: World Health Organization; 2020 (<https://iris.who.int/handle/10665/333947>). License: CC BY-NC-SA 3.0 IGO.
8. 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;26:302-9 (<https://doi.org/10.5588/ijtld.21.0667>).
9. Guidance on evidence generation on new regimens for tuberculosis treatment. Geneva: World Health Organization; 2024 (<https://iris.who.int/handle/10665/379830>). License: CC BY-NC-SA 3.0 IGO.
10. Guidance for best practices for clinical trials. Geneva: World Health Organization; 2024 (<https://iris.who.int/handle/10665/378782>). License: CC BY-NC-SA 3.0 IGO.
11. WHO handbook for guideline development, 2nd ed. Geneva: World Health Organization; 2014 (<https://iris.who.int/handle/10665/145714>).
12. WHO Listed Authorities [website]. World Health Organization; 2024 (<https://www.who.int/news-room/questions-and-answers/item/who-listed-authorities>).
13. WHO electronic Essential Medicines List (eEML). World Health Organization; 2023 (<https://list.essentialmeds.org/>). License: CC BY 3.0 IGO.
14. 22nd invitation to manufacturers of antituberculosis medicines to submit an expression of interest (EOI) for product evaluation to the WHO Prequalification Unit [website]. World Health Organization; 2024 (<https://extranet.who.int/prequal/news/22nd-invitation-manufacturers-antituberculosis-medicines-submit-expression-interest-eoi>).
15. ICH harmonised guideline: guideline for good clinical practice E6(R3). International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use; 2025 ([https://database.ich.org/sites/default/files/ICH_E6\(R3\)_Step4_FinalGuideline_2025_0106.pdf](https://database.ich.org/sites/default/files/ICH_E6(R3)_Step4_FinalGuideline_2025_0106.pdf)).



16. Twelfth General Programme of Work 2014–2019: not merely the absence of disease. Geneva: World Health Organization; 2014 (<https://iris.who.int/handle/10665/112792>).
17. Schunemann H, Brozek J, Guyatt G, Oxman A, editors. GRADE handbook for grading quality of evidence and strength of recommendations. The GRADE Working Group; 2013 (<https://guidelinedevelopment.org/handbook>).
18. Global Programme on Evidence for Health Policy. Guidelines for WHO guidelines. Geneva: World Health Organization; 2003 (<https://iris.who.int/handle/10665/68925>).
19. 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;328:1490 (<https://doi.org/10.1136/bmj.328.7454.1490>).
20. Schunemann HJ, Best D, Vist G, Oxman AD, Group GW. Letters, numbers, symbols and words: how to communicate grades of evidence and recommendations. CMAJ. 2003;169:677-80 (<https://www.ncbi.nlm.nih.gov/pubmed/14517128>).
21. Chapter 9. Glossary of terms and concepts. In: Schunemann H, Brozek J, Guyatt G, Oxman A, editors. GRADE handbook for grading quality of evidence and strength of recommendations. The GRADE Working Group; 2013 (<https://gdt.gradeapro.org/app/handbook/handbook.html#h.buaodtl66dyx>).
22. GRADEpro GDT: GRADEpro Guideline Development Tool [website]. McMaster University and Evidence Prime; 2025 (<https://www.gradeapro.org/>).
23. Neumann I, Mustafa R, Morgan R, Zhang Y, Dahm P, Langendam M et al. Developing questions about interventions. In: Neumann I, Schünemann H, editors. The GRADE Book version 1.0. The GRADE Working Group; 2024 (<https://book.gradeapro.org/guideline/questions-about-interventions-diagnostic-test-prognosis-and-exposures>).
24. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ et al., editors. Cochrane handbook for systematic reviews of interventions version 6.5. Cochrane; 2024 (<http://www.cochrane.org/handbook>).
25. 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 (<https://doi.org/10.1136/bmj.i2089>).
26. 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 (<https://doi.org/10.1016/j.jclinepi.2017.05.006>).
27. Schunemann 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 (<https://doi.org/10.1016/j.jclinepi.2016.03.018>).
28. Balshem H, Helfand M, Schunemann HJ, Oxman AD, Kunz R, Brozek J et al. GRADE guidelines: 3. Rating the quality of evidence. J Clin Epidemiol. 2011;64:401-6 (<https://doi.org/10.1016/j.jclinepi.2010.07.015>).
29. WHO consolidated guidelines on tuberculosis: module 4: treatment and care. Geneva: World Health Organization; 2025 (<https://iris.who.int/handle/10665/380799>). License: CC BY-NC-SA 3.0 IGO.
30. GRADE-CERQual [website]. The GRADE Working Group; 2018 (<https://www.cerqual.org/>).
31. 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 (<https://doi.org/10.1016/j.jclinepi.2019.10.014>).



32. Schunemann 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 (<https://doi.org/10.1016/j.jclinepi.2018.01.012>).
33. Sterne JA, Hernan MA, Reeves BC, Savovic 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 (<https://doi.org/10.1136/bmj.i4919>).
34. 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;64:1303-10 (<https://doi.org/10.1016/j.jclinepi.2011.04.014>).
35. Schunemann 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 (<https://doi.org/10.1016/j.jclinepi.2022.07.015>).
36. Sabella-Jimenez V, Hoyos Mendez Y, Benjumea-Bedoya D, Estupinan-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:e001551 (<https://doi.org/10.1136/bmjpo-2022-001551>).
37. Balcells ME, Thomas SL, Godfrey-Faussett P, Grant AD. Isoniazid preventive therapy and risk for resistant tuberculosis. *Emerg Infect Dis.* 2006;12:744-51 (<https://doi.org/10.3201/eid1205.050681>).
38. Schünemann HJ, Higgins JP, Vist GE, Glasziou P, Akl EA, Skoetz N et al. Chapter 14: Completing 'Summary of findings' tables and grading the certainty of the evidence. In: Higgins JPT, Thomas J, Chandler J, Cumpston M et al., editors. *Cochrane handbook for systematic reviews of interventions version 6.5.* Cochrane; 2024 (<https://www.cochrane.org/authors/handbooks-and-manuals/handbook/current/chapter-14>).
39. 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:e0002804 (<https://doi.org/10.1371/journal.pgph.0002804>).
40. 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;162:1851-4 (<https://doi.org/10.1164/ajrccm.162.5.2004154>).
41. Section 6.2 Factors determining direction and strength of recommendations. In: Schunemann H, Brozek J, Guyatt G, Oxman A, editors. *GRADE handbook for grading quality of evidence and strength of recommendations.* The GRADE Working Group; 2013 (<https://gdt.grade.org/app/handbook/handbook.html#h.1yd7iwhn8pxp>).
42. Schunemann 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;2:3198-225 (<https://doi.org/10.1182/bloodadvances.2018022954>).
43. 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;68:703-7 (<https://doi.org/10.1016/j.jclinepi.2014.11.006>).
44. Transparency CORE: clinical trial cost reporting toolkit [website]. Médecins Sans Frontières; 2024 (<https://msfaccess.org/transparency-core-clinical-trial-cost-reporting-toolkit>).
45. Report of the technical consultation on innovative clinical trial designs for evaluating new TB preventive treatments, virtual meeting, 15-17 September 2021. Geneva: World Health Organization; 2022 (<https://iris.who.int/handle/10665/353393>). License: CC BY-NC-SA 3.0 IGO.



46. Montepiedra G, Kim S, Weinberg A, Theron G, Sterling TR, LaCourse SM et al. Using a composite maternal-infant outcome measure in tuberculosis-prevention studies among pregnant women. *Clin Infect Dis*. 2021;73:e587-e93 (<https://doi.org/10.1093/cid/ciaa1674>).
47. CTCAE and adverse event reporting [website]. National Cancer Institute; 2025 (<https://dctd.cancer.gov/research/ctep-trials/for-sites/adverse-events>).
48. CIOMS Working Group VI on the Management of Safety Information from Clinical Trials, Council for International Organizations of Medical Sciences. Management of safety information from clinical trials: report of CIOMS Working Group VI. Geneva: Council for International Organizations of Medical Sciences; 2005 (<https://kohahq.searo.who.int/bib/29403>).
49. Division of AIDS (DAIDS) table for grading the severity of adult and pediatric adverse events, version 2.1. Bethesda, Md: Division of AIDS, National Institute of Allergy and Infectious Diseases; 2017 (<https://rsc.niaid.nih.gov/sites/default/files/daidsgradingcorrectedv21.pdf>).
50. Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research. E9 statistical principles for clinical trials: guidance for industry. Rockville: Food and Drug Administration; 1998 (<https://www.fda.gov/media/71336/download>).
51. Section 7.4. Indirect evidence and impact on patient-important outcomes. In: Schunemann H, Brozek J, Guyatt G, Oxman A, editors. GRADE handbook for grading quality of evidence and strength of recommendations. The GRADE Working Group; 2013 (<https://gdt.gradeapro.org/app/handbook/handbook.html#h.t09m6hxd5iz>).
52. Chandra DK, Moll AP, Altice FL, Didomizio E, Andrews L, Shenoi SV. Structural barriers to implementing recommended tuberculosis preventive treatment in primary care clinics in rural South Africa. *Glob Public Health*. 2022;17:555-68 (<https://doi.org/10.1080/17441692.2021.1892793>).
53. 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;19:1039-44, i-v (<https://doi.org/10.5588/ijtld.14.0829>).
54. Salazar-Austin N, Dowdy DW, Chaisson RE, Golub JE. Seventy years of tuberculosis prevention: efficacy, effectiveness, toxicity, durability, and duration. *Am J Epidemiol*. 2019;188:2078-85 (<https://doi.org/10.1093/aje/kwz172>).
55. Kendall EA, Durovni B, Martinson NA, Cavalcante S, Masonoke K, Saraceni V et al. Adherence to tuberculosis preventive therapy measured by urine metabolite testing among people with HIV. *AIDS*. 2020;34:63-71 (<https://doi.org/10.1097/qad.0000000000002380>).
56. du Cros P, Greig J, Alffenaar JC, Cross GB, Cousins C, Berry C et al. Standards for clinical trials for treating TB. *Int J Tuberc Lung Dis*. 2023;27:885-98 (<https://doi.org/10.5588/ijtld.23.0341>).
57. Levesque 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 (<https://doi.org/10.1136/bmj.b5087>).
58. Pulmonary tuberculosis: developing drugs for treatment. Guidance for industry. Rockville: US Food and Drug Administration; 2022 (<http://www.fda.gov/files/drugs/published/16856dft-Pulmonary-Tuberculosis.pdf>).
59. Thomas LE, Pencina MJ. Do not over (P) value your research article. *JAMA Cardiol*. 2016;1:1055 (<https://doi.org/10.1001/jamacardio.2016.3827>).
60. Schandelmaier S, Briel M, Varadhan R, Schmid CH, Devasenapathy N, Hayward RA et al. Development of the Instrument to assess the Credibility of Effect Modification Analyses (ICEMAN) in randomized controlled trials and meta-analyses. *CMAJ*. 2020;192:E901-E6 (<https://doi.org/10.1503/cmaj.200077>).



61. Hills NK, Lyimo J, Nahid P, Savic RM, Lienhardt C, Phillips PPJ. A systematic review of endpoint definitions in late phase pulmonary tuberculosis therapeutic trials. *Trials*. 2021;22:515 (<https://doi.org/10.1186/s13063-021-05388-1>).
62. Pham TM, Tweed CD, Carpenter JR, Kahan BC, Nunn AJ, Crook AM et al. Rethinking intercurrent events in defining estimands for tuberculosis trials. *Clin Trials*. 2022;19:522-33 (<https://doi.org/10.1177/17407745221103853>).
63. Weir IR, Dufault SM, Phillips PPJ. Estimands for clinical endpoints in tuberculosis treatment randomized controlled trials: a retrospective application in a completed trial. *Trials*. 2024;25:180 (<https://doi.org/10.1186/s13063-024-07999-w>).
64. Global individual participant data platform for tuberculosis treatment [website]. World Health Organization; 2023 (<https://www.who.int/news/item/31-01-2023-launch-of-global-individual-patient-data-platform-for-tuberculosis-treatment>).
65. WHO data principles [website]. Geneva: World Health Organization; 2025 (<https://www.who.int/data/principles>).
66. TB-Platform for Aggregation of Clinical TB Studies (TB-PACTS) [website]. Tucson: Critical Path Institute; 2025 (<https://c-path.org/tools-platforms/tb-pacts/>).
67. Goodall RL, Fabiane SM, Hakiman A, Crook AM, Mirzayev F, Schumacher S et al. A publicly accessible global data repository - the WHO TB-IPD platform. *IJTL Open*. 2024;1:151-3 (<https://doi.org/10.5588/ijtldopen.24.0131>).
68. Tahan TT, Rossoni AMO, Bedim Dos Santos M, da Silveira JTP, de Oliveira SP, Rodrigues CO. Tuberculosis preventive treatment in children and adolescents: an observational study of secondary data. *J Pediatr (Rio J)*. 2023;99:399-405 (<https://doi.org/10.1016/j.jped.2023.02.002>).
69. Sterne JAC, Hernán MA, McAleenan A, Reeves BC, Higgins JPT. Chapter 25: Assessing risk of bias in a non-randomized study. In: Higgins JPT, Thomas J, Chandler J, Cumpston M et al., editors. *Cochrane handbook for systematic reviews of interventions version 6.5*. Cochrane; 2024 (<https://www.cochrane.org/authors/handbooks-and-manuals/handbook/current/chapter-25>).
70. Park SY. Reducing bias in observational studies. *Korean J Anesthesiol*. 2020;73:269-70 (<https://doi.org/10.4097/kja.20348>).
71. Sterrantino AF. Observational studies: practical tips for avoiding common statistical pitfalls. *Lancet Reg Health Southeast Asia*. 2024;25:100415 (<https://doi.org/10.1016/j.lansea.2024.100415>).
72. 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:e043954 (<https://doi.org/10.1136/bmjopen-2020-043954>).
73. Lewin S, Glenton C, Lawrie TA, Downe S, Finlayson KW, Rosenbaum S et al. Qualitative evidence synthesis (QES) for guidelines: Paper 2 – using qualitative evidence synthesis findings to inform evidence-to-decision frameworks and recommendations. *Health Res Policy Syst*. 2019;17:75 (<https://doi.org/10.1186/s12961-019-0468-4>).
74. Selva A, Sola I, Zhang Y, Pardo-Hernandez H, Haynes RB, Martinez Garcia L et al. Development and use of a content search strategy for retrieving studies on patients' views and preferences. *Health Qual Life Outcomes*. 2017;15:126 (<https://doi.org/10.1186/s12955-017-0698-5>).
75. Karam SG, Zhang Y, Pardo-Hernandez H, Siebert U, Koopman L, Noyes J et al. ROBVALU: a tool for assessing risk of bias in studies about people's values, utilities, or importance of health outcomes. *BMJ*. 2024;385:e079890 (<https://doi.org/10.1136/bmj-2024-079890>).
76. Collaborative Group for the Meta-Analysis of Individual Patient Data in MDR-TB treatment–2017, Ahmad N, Ahuja SD, Akkerman OW, Alffenaar JC, Anderson LF et al. Treatment correlates of successful outcomes in pulmonary multidrug-resistant tuberculosis: an individual patient data meta-analysis. *Lancet*. 2018;392:821-34 ([https://doi.org/10.1016/s0140-6736\(18\)31644-1](https://doi.org/10.1016/s0140-6736(18)31644-1)).



77. Brunetti M, Shemilt I, Pregno S, Vale L, Oxman AD, Lord J et al. GRADE guidelines: 10. Considering resource use and rating the quality of economic evidence. *J Clin Epidemiol*. 2013;66:140-50 (<https://doi.org/10.1016/j.jclinepi.2012.04.012>).
78. Dewidar O, Pardo JP, Welch V, Hazlewood GS, Darzi AJ, Barnabe C et al. Operationalizing the GRADE-equity criterion to inform guideline recommendations: application to a medical cannabis guideline. *J Clin Epidemiol*. 2024;165:111185 (<https://doi.org/10.1016/j.jclinepi.2023.10.001>).
79. Welch VA, Akl EA, Guyatt G, Pottie K, Eslava-Schmalbach J, Ansari MT et al. GRADE equity guidelines 1: considering health equity in GRADE guideline development: introduction and rationale. *J Clin Epidemiol*. 2017;90:59-67 (<https://doi.org/10.1016/j.jclinepi.2017.01.014>).
80. 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;17:88 (<https://doi.org/10.1186/s12913-017-2031-8>).
81. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ*. 2010;340:c332 (<https://doi.org/10.1136/bmj.c332>).
82. A5.4 A survey to explore the programmatic feasibility of levofloxacin (Lfx) TPT for MDR-TB contacts. In: WHO consolidated guidelines on tuberculosis: module 1: prevention: tuberculosis preventive treatment, 2nd ed. Geneva: World Health Organization; 2024 (<https://tbksp.who.int/en/node/2748>).
83. 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*. London: European Medicines Agency; 2017 (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).
84. Prequalification of Medical Products. Technical advice [website]. World Health Organization; 2023 (<https://extranet.who.int/prequal/medicines/technical-advice>).
85. Prequalification of Medical Products. Pre-submission meetings [website]. World Health Organization; 2023 (<https://extranet.who.int/prequal/content/pre-submission-meetings>).
86. Expert Review Panel [website]. Global Fund to Fight AIDS, Tuberculosis and Malaria; 2025 (<https://www.theglobalfund.org/en/sourcing-management/quality-assurance/expert-review-panel/>).
87. Instructions for applicants preparing a submission for the 2025 meeting of the WHO expert committee on selection and use of essential medicines. Geneva: World Health Organization; 2024 (<https://iris.who.int/handle/10665/376713>). License: CC BY-NC-SA 3.0 IGO.
88. Coordinated Scientific Advice for health product R&D [website]. World Health Organization; 2023 (<https://www.who.int/our-work/science-division/research-for-health/who-coordinated-scientific-advice-for-health-product-r-d>).
89. Technical Advisory Group on dosing of TB medicines for adults and children [website]. World Health Organization; 2025 (<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 Grading of Recommendations Assessment, Development and Evaluation (GRADE) system is an internationally recognized approach for rating the quality of evidence and grading the strength of recommendations in health care (1). GRADE evidence profiles display the ratings of the certainty of evidence, along with effect estimates, in standardized, transparent and informative tables, including the following key information for each critical outcome:

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

Table A1.1 gives an example of an evidence profile for the PICO (population, intervention, comparator and outcome) question, “Should 6 months of levofloxacin versus another regimen or no tuberculosis preventive treatment be used for people who are household contacts of someone with multidrug-resistant and rifampicin-resistant tuberculosis?”. The table’s contents summarize the findings from a meta-analysis of individual participant data from two clinical trials presented to the guideline development group (GDG). The table presents quantitative estimates for each outcome of importance to the GDG. Common terms and concepts used in GRADE have been defined elsewhere (2).



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

Question: Should 6 months of Lfx versus another regimen or no TPT be used for people who are household contacts of someone with MDR/RR-TB?

			Certainty assessment				No. of patients		Effect			
No. of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	6 months of Lfx	Other regimen or no TPT	Relative risk (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%)	0.38 (0.17–0.86)	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%)	1.26 (0.34–4.68)	1 more per 1000 (from 2 fewer to 10 more)	⊕⊕○○ Low	Critical
AEs of Grade 3 or above at least possibly related to study drug (TB-CHAMP, participants aged below 18 years; follow-up: 6 months plus 21 days)												
1	Randomized trial	Not serious	Not serious	Not serious	Serious ^b	None	4/452 (0.9%)	8/469 (1.7%)	0.53 (0.16–1.70)	8 fewer per 1000 (from 14 fewer to 12 more)	⊕⊕⊕○ Moderate	Critical
AEs of Grade 3 or above at least possibly related to study drug (V-QUIN, 97% of participants aged over 14 years; follow-up: 6 months plus 30 days)												
1	Randomized trial	Not serious	Not serious	Not serious	Not serious	None	10/960 (1.0%)	2/962 (0.2%)	5.26 (1.16–23.95)	9 more per 1000 (from 0 fewer to 48 more)	⊕⊕⊕⊕ High	Critical





Certainty assessment							No. of patients		Effect		Certainty	Importance
No. of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	6 months of Lfx	Other regimen or no TPT	Relative risk (95% CI)	Absolute (95% CI)		
AEs 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%)	6.32 (3.43–11.63)	45 more per 1000 (from 20 to 89 more)	⊕⊕⊕⊕ High	Critical
Treatment completion (opposite of discontinuation)												
2	Randomized trials	Not serious	Not serious	Not serious	Not serious	None	1078/1476 (73.0%)	1233/1487 (82.9%)	0.88 (0.85–0.92)	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%)	0.88 (0.85–0.91) ^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	Additional resistance to Lfx or other antimicrobials was not detected in any of eight strains from index–incident pairs in the V-QUIN trial that were tested with whole genome sequencing ^d				⊕○○○ Very low	Important
Emergence of additional fluoroquinolone resistance in microbiome other than TB (e.g. gut flora) not measured												
–	–	–	–	–	–	–	–				–	Important

AE: adverse event; CI: confidence interval; Lfx: levofloxacin; MDR/RR-TB: multidrug-resistant and rifampicin-resistant tuberculosis; TB: tuberculosis; TB-CHAMP: Tuberculosis Child Multidrug-Resistant Preventive Therapy trial; TPT: TB preventive treatment; V-QUIN: Levofloxacin for the Prevention of Multidrug-Resistant Tuberculosis in Vietnam trial.

- ^a We rated down two levels because the CI includes appreciable harm and appreciable benefit (RR: 1.26; 95% CI: 0.34–4.68).
- ^b We rated down one level because the CI includes appreciable harm and some benefit (RR: 0.53; 95% CI: 0.16–1.70).
- ^c Treatment completion in the Lfx arm was 86% in TB-CHAMP (placebo arm: 86%) and 70% in V-QUIN (placebo arm: 85%), with RRs of 1.00 (95% CI: 0.95–1.06) and 0.83 (95% CI: 0.79–0.87), respectively.
- ^d We rated down one level for risk of bias. The results are not from a randomized comparison. In V-QUIN, of the 43 people with presumed TB after randomization, 17 had laboratory-confirmed incident TB, in four of whom an isolate could not be recovered. Whole genome sequencing with drug-susceptibility test results were available for 8/13 people with confirmed TB. Of these, six were in the placebo arm and two in the Lfx arm. None had acquired resistance to Lfx. In TB-CHAMP, 14 people in the placebo arm and seven in the Lfx arm developed TB, of whom seven and three, respectively, had confirmed TB. No results for Lfx susceptibility were available for the strains isolated.
- ^e We rated down one level for indirectness. Data were 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 for Annex 1

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;33:176-82 (<https://doi.org/10.1017/s0266462317000447>).
2. Chapter 9. Glossary of terms and concepts. In: Schunemann H, Brozek J, Guyatt G, Oxman A, editors. *GRADE handbook for grading quality of evidence and strength of recommendations*. The GRADE Working Group; 2013 (<https://gdt.gradeapro.org/app/handbook/handbook.html#h.buaodtl66dyx>).



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

The Grading of Recommendations Assessment, Development and Evaluation (GRADE) evidence to decision (EtD) framework is a systematic and transparent approach to translating evidence into recommendations. It provides a structure for assessing benefits, harms, certainty of evidence and other key decision criteria, and it ensures that recommendations are informed by the best available evidence and aligned with stakeholder values and contextual factors. The GRADE EtD criteria organize population, intervention, comparator and outcome (PICO)-derived evidence into a comprehensive decision-making process by evaluating key factors, such as problem relevance, clinical benefits and harms, certainty of evidence, and patient and societal values. They are also used to 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 health care settings.

The summary of findings table presents a concise overview of the key evidence considered in the development of a particular recommendation in the EtD framework. It 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, to support an informed and balanced decision-making process. Table A2.1 gives an example of a summary of findings table from an EtD process that led to a recommendation on the use of 6 months of levofloxacin as tuberculosis preventive treatment.

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

Question: Should 6 months of Lfx versus another regimen or no TPT be used for people who are household contacts of someone with MDR/RR-TB?

Patients: people in contact with someone with MDR/RR-TB

Setting: outpatient

Intervention: 6 months of Lfx

Comparison: placebo



Outcomes	No. of participants (studies)	Certainty of the evidence (GRADE)	Anticipated absolute effects		
			Relative effect (95% CI) ^a	Risk with other regimen or no TPT	Risk with 6 months of Lfx
TB 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 1000	5 per 1000 (2 to 12)
Treatment completion (opposite of discontinuation)	2963 (2 RCTs)	⊕⊕⊕⊕ High	RR 0.88 (0.85 to 0.92)	829 per 1000	730 per 1000 (705 to 763)
Treatment completion (80% or more of doses taken by 6 months)	2928 (2 RCTs)	⊕⊕⊕⊕ High	RR 0.88 (0.85 to 0.91) ^b	850 per 1000	748 per 1000 (723 to 774)
Death (any cause)	2963 (2 RCTs)	⊕⊕○○○ Low ^c	RR 1.26 (0.34 to 4.68)	3 per 1000	3 per 1000 (1 to 13)
AEs of Grade 3 or above at least possibly related to study drug (TB-CHAMP, participants aged below 18 years; follow-up: 6 months plus 21 days)	921 (1 RCT)	⊕⊕⊕○ Moderate ^d	RR 0.53 (0.16 to 1.70)	17 per 1000	9 per 1000 (3 to 29)
AEs of Grade 3 or above at least possibly related to study drug (V-QUIN, 97% of participants aged over 14 years; follow-up: 6 months plus 30 days)	1922 (1 RCT)	⊕⊕⊕⊕ High	RR 5.26 (1.16 to 23.95)	2 per 1000	11 per 1000 (2 to 50)
AEs 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 1000	53 per 1000 (29 to 98)
Emergence of additional fluoroquinolone resistance in TB strains	8 (2 RCTs) ^e	⊕○○○○ Very low ^{f,g,h}		Additional resistance to Lfx or other antimicrobials was not detected in any of eight strains from index–incident pairs in the V-QUIN trial that were tested with whole genome sequencing	

AE: adverse event; CI: confidence interval; GRADE: Grading of Recommendations Assessment, Development and Evaluation; Lfx: levofloxacin; MDR/RR-TB: multidrug-resistant and rifampicin-resistant tuberculosis; RCT: randomized control trial; RR: relative risk; TB: tuberculosis; TB-CHAMP: Tuberculosis Child Multidrug-Resistant Preventive Therapy trial; TPT: TB preventive treatment; V-QUIN: Levofloxacin for the Prevention of Multidrug-Resistant Tuberculosis in Vietnam trial.

^a The risk in the intervention group (and its 95% CI) is based on the assumed risk in the comparator group and the relative effect of the intervention (and its 95% CI).

^b Treatment completion in the Lfx arm was 86% in TB-CHAMP (placebo arm: 86%) and 70% in V-QUIN (placebo arm: 85%), with RRs of 1.00 (95% CI: 0.95–1.06) and 0.83 (95% CI: 0.79–0.87), respectively.

^c We rated down two levels because the CI includes appreciable harm and appreciable benefit (RR: 1.26; 95% CI: 0.34–4.68).

^d We rated down one level because the CI includes appreciable harm and some benefit (RR: 0.53; 95% CI: 0.16–1.70).

^e Out of 17 laboratory-confirmed incident TB strains.

^f We rated down one level for risk of bias. The results are not from a randomized comparison. In V-QUIN, of the 43 people with presumed TB after randomization, 17 had laboratory-confirmed incident TB, in four of whom an isolate could not be recovered. Results were only available for 8/13 people with confirmed TB. Of these, six were in the placebo arm and two were in the Lfx arm. In TB-CHAMP, 14 people in the placebo arm and seven in the Lfx arm developed TB, of whom seven and three, respectively, had confirmed TB. No results for Lfx susceptibility were available for the strains isolated.

^g We rated down one level for indirectness. Data were only available for V-QUIN; all strains were from individuals aged over 15 years.

^h 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 Grading of Recommendations Assessment, Development and Evaluation (GRADE) evidence to decision (EtD) framework contains a total of 12 criteria for the overall assessment of evidence. In addition to the summary of findings table (see Annex 2), the EtD process includes information on other characteristics that are critical to the formulation of recommendations. Table A3.1 gives an example of how these criteria were used during the development of recommendations for 6 months of levofloxacin as tuberculosis preventive treatment for household contacts of people with multidrug-resistant and rifampicin-resistant tuberculosis.

For this table, a meta-analysis of individual participant data from two clinical trials, Levofloxacin for the Prevention of Multidrug-Resistant Tuberculosis in Vietnam (V-QUIN) and Tuberculosis Child Multidrug-Resistant Preventive Therapy (TB-CHAMP), was evaluated by the guideline development group. Data from additional substudies of cost-effectiveness and pharmacokinetics, completed by both the V-QUIN and TB-CHAMP study teams, were used to inform the respective criteria of the EtD framework. Acceptability and feasibility were assessed through separate systematic reviews of evidence and surveys of both national tuberculosis programmes and affected populations. The V-QUIN study also provided data regarding the effects of levofloxacin on the gut and nasal microbiome.



Table A3.1. EtD criteria used to answer the question: “Should 6 months of Lfx versus another regimen or no TPT be used for people who are HHCs of someone 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 TB 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 and 160 000 deaths due to MDR/RR-TB 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 TPT 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, such as TPT, are crucial were as follows. • The 2020 TPT guidelines included a recommendation for TPT for contacts of people with MDR/RR-TB that was conditional and based on evidence of very low certainty. The recommendation was not specific to any regimen, and its implementation since it was first published in 2017 has been poor. • Now that trial-based evidence for a defined treatment regimen has become available, it is more important to review the new evidence to assess the efficacy of the 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 studies that assessed TB prevention (reduction in incidence) with FQ (alone or in combination with other TB drugs) and one study that assessed TB prevention with isoniazid. All four were observational studies with substantial risk of bias, notably selection bias. The three studies involving FQ did not detect any reduction in TB incidence with FQ use, compared with no TPT. The results from the systematic review and from the individual participant data for isoniazid could not be summarized in the GRADE table.	Key considerations expressed by GDG members when making a judgement of moderate for desirable effects were as follows. • The efficacy of Lfx in the trials was similar to that observed in other studies of TPT, although uncertainty was expressed regarding the durability of effect. • The risk of MDR/RR-TB in an exposed person 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/RR-TB in different contexts. Any intervention that can reduce this risk would be welcome. • It was observed that the two outcomes presented here – TB incidence and TPT completion – are going in opposite directions, making judgement difficult, as the judgements for incidence and treatment completion may be different. • 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 based on the pooled data because separation by age (adults and children) would reduce precision and lower the certainty 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
☐ Large ☐ Moderate ☐ Small ☐ Trivial ☒ Varies ☐ Don't know	A systematic review of studies published between June 2016 and September 2023 identified five studies that assessed AEs with FQ (alone or in combination with other TB drugs). All were observational studies with substantial risk of bias, notably selection bias. Detection, judgement of severity, and attribution were not blinded, potentially leading to ascertainment bias. FQ monotherapy (i.e. Lfx, ofloxacin or moxifloxacin alone) was seen to be generally safe in three studies, with some mild or moderate drug-related AEs in children, but no Grade 3 or 4 AEs or serious AEs. 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 with the studies of 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. In a small study among inmate contacts (Bedini et al., 2016), FQ with pyrazinamide was found to have very low tolerability (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. Although the two trials collected data on the emergence of additional FQ resistance to TB strains and other flora, results of DST or whole genome sequencing were incomplete at the time of the GDG meeting. Only one outcome from eight TB strains tested in V-QUIN (two of which were from the Lfx arm) was included in the summary of findings table, which showed no additional resistance acquired.	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 AEs between children (trivial) and adults (moderate), with very good tolerance in children and much lower tolerability with increasing age, which has likely contributed to lower adherence to TPT among 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.



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 was judged to be high for TB incidence, treatment completion and AEs of Grade 3 or above that were at least possibly associated with the study drug in adults; moderate for AEs of Grade 3 or above that were at least possibly associated with the study drug in children; and low for death (all critical outcomes). Certainty was considered very low for the emergence of additional FQ resistance in TB strains and was not estimable for the emergence of additional FQ resistance in the 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 AEs (all studies were observational). The low incidence of Grade 3–4 AEs and the low occurrence of discontinuation of FQ TPT due to AEs in adults and children in observational studies is consistent with evidence from the trials.	Key considerations expressed by GDG members when making a judgement that the certainty of the evidence of effects was moderate 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 FQ as TPT for MDR/RR-TB in the foreseeable future (the PHOENix MDR-TB trial is using 26 weeks of delamanid and is expected to be completed at the end of 2026). • However, uncertainties were expressed about the serious or very serious imprecision in the AEs and the fact that there are only two trials. It was highlighted that there may be difficulties in standardizing 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 FQs 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 (involving two published studies on acceptability of starting TPT for MDR/RR-TB, two published studies on willingness to take hypothetical TPT for MDR/RR-TB, one published study on acceptability of a novel child-friendly Lfx formulation, and one published explorative qualitative study) suggested that overall acceptability of TPT for MDR/RR-TB to prevent incident TB disease was high. However, based on the qualitative acceptability study (among 36 HHCs in five countries), there is an indication of possibly important uncertainty or variability. Although the sample size was relatively small, this study included people with a wide range of TB and MDR/RR-TB knowledge and experience, as well as very different socioeconomic and cultural backgrounds, and found meaningful differences in TPT acceptability. For example, although most people valued a lowered risk of developing MDR/RR-TB, some refused to accept any risk of serious AEs due to TPT, which overrode any value they placed on avoiding MDR/RR-TB. The study suggests that, where there is an absence of trained HCWs or researchers taking the time to explain TPT, the value people place on prevention is quite low, their understanding of the severity or risk of MDR/RR-TB seems very low, and the value placed on their present health is very high by contrast. Arguably, this is a very important variability in values that could affect real-world uptake of TPT for MDR/RR-TB.	Key considerations expressed by GDG members when making a judgement that there was probably no important uncertainty or variability in values were as follows. • Values are likely to depend on how much people being offered FQ TPT are well informed about the efficacy and downsides of TPT and the seriousness of MDR/RR-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 TPT adherence. These may be overcome with education but are still important. • Acceptance for people who started TPT was quite high, and higher than is seen with comparable interventions under programmatic settings. However, the evidence reviewed was from small samples so may not be generalizable.



Balance of effects

Does the balance between desirable and undesirable effects favour the intervention or the comparator?

Judgement	Research evidence	Additional considerations
☐ Favours the comparator ☐ Probably favours the comparator ☐ Does not favour either the intervention or the comparator ☒ Probably favours the intervention ☐ Favours the intervention ☐ Varies ☐ Don't know	The reduction in incidence of MDR/RR-TB by 60% in adults and children with the intervention of Lfx is offset only by mild Grade 1–2 AEs. Both desirable and undesirable effect estimates were derived from two RCTs that were judged to be of high quality overall. The ascertainment of these outcomes was also free of bias, and there was sufficient precision to be reasonably certain of these effects. The estimates of low rates of Grade 1–2 AEs and very low rates of Grade 3–4 AEs are supported by observational studies included in the systematic review, although it was not possible to estimate a pooled rate of mild or severe AEs in the review because of the heterogeneity of interventions reported and the definitions of AEs used.	Key considerations expressed by GDG members when making a judgement that the balance of effects probably favours the intervention were as follows. • It was noted that, based on the evidence presented to the GDG, the benefits outweigh the risks, especially in children. To a large extent, AEs were mild and self-limiting. • Although not critical for this assessment, the emergence of other resistance is important, and there was uncertainty about how this 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 individuals in the long term are unknown. • It was highlighted that the use of FQ as TPT for MDR/RR-TB should be considered an appropriate use of antimicrobial agents (inappropriate use is more likely to generate avoidable resistance). • It was noted that the effects of wide-scale use of Lfx in a population are 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 NTP managers of 30 high-burden MDR/RR-TB countries, of whom 18 (60%) responded, seven respondents stated that the cost of additional resource requirements may be a barrier to implementation, with some specifically mentioning the concurrent need for DST, 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 Lfx is much more expensive than the adult formulation (a tenfold difference per milligram at current GDF prices: about US\$ 0.12/100 mg tablet versus US\$ 0.03/250 mg tablet; (https://www.stoptb.org/sites/default/files/documents/2025.09.02_gdf_medicines_catalog.pdf))	Key considerations expressed by GDG members when making a judgement that the resources required would involve moderate costs were as follows. • The cost of Lfx, a generic medicine in widespread 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, DST, monitoring for AEs, capacity-building to improve the skills of HCWs, 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 people with MDR/RR-TB. • It was also noted that, overall, the burden of MDR/RR-TB is relatively low compared with drug-susceptible TB. • NTPs not procuring TPT through the GDF mechanism may face a higher price for a product of guaranteed quality, as well as differences in costs if the 750 mg formulation is used instead of the 250 mg formulation. However, this variation in the exact per-patient budget impact may not have had a major influence on 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	Certainty of the evidence of resource requirements was based on a self-administered questionnaire of NTP managers with a completion rate of 60%. Of 18 respondents, seven stated that the cost of additional resource requirements may be a barrier to implementation, with some specifically mentioning the concurrent need for DST, screening, monitoring and follow-up in the programme, as well as the already limited human resources and budgets within programmes. Pricing of the GDF medications is standardized for all eligible countries.	A key consideration expressed by GDG members when making a judgement that the certainty of evidence of required resources was low was as follows. • Only one survey was reviewed, and there was no evidence on costs for implementation.



Cost-effectiveness

Does the cost-effectiveness of the intervention favour the intervention or the comparator?

Judgement	Research evidence	Additional considerations
☐ Favours the comparator ☐ Probably favours the comparator ☐ Does not favour either the intervention or the comparator ☐ Probably favours the intervention ☒ Favours the intervention ☐ Varies ☐ No included studies	Cost-effectiveness was based on a single self-administered questionnaire of NTP managers with a completion rate of 60%. Pricing of the GDF medications is standardized for all eligible countries.	Key considerations expressed by GDG members when making a judgement that cost-effectiveness favours the intervention were as follows. • Cost-effectiveness favours the intervention because it saves money rather than generating costs. • Although the paediatric formulation is more expensive, the cost-effectiveness analysis still finds it cost-effective, given the high risks of progression to TB disease from infection. • The cost-effectiveness subanalysis is based on a setting where the risk of progressing to TB disease is 2.5%; in areas where the risk is lower, the analysis may not have the same findings.

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/RR-TB countries, of whom 18 (60%) responded, overall equity was expected to increase for contacts through the avoidance of TB disease. However, six NTP managers mentioned that certain remote areas may not have an adequate supply of Lfx. Additionally, 11 NTP managers mentioned increased out-of-pocket spending for contacts, with two commenting on the need for health insurance to cover TPT, to ensure equity. Importantly, interviews with contacts themselves in the qualitative acceptability study (36 HHCs in five countries) suggested that those with little income, unstable or no employment, or little or no social support will likely not be able to agree to and complete a 6-month TPT regimen that would require at least monthly check-ups and may cause mild side-effects (especially at the beginning of treatment) that could affect their daily activities and responsibilities. Also, caregivers for the MDR/RR-TB index patients or other HHCs are unlikely to be able to start or accept TPT unless they have access to improved socioeconomic support systems. Hence, findings from this qualitative study suggest that equity may be reduced by the introduction of TPT for contacts of people with MDR/RR-TB, unless this is accompanied by improved social and financial support.	Key considerations expressed by GDG members when making a judgement that equity would probably be increased were as follows. • Some people might benefit from Lfx more than others. From a drug perspective, equity is increased 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 MDR/RR-TB who are generally marginalized and 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 costs 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 for the drug itself and other health care system components. In situations where the health system covers the expenditure for Lfx, another consideration is the opportunity cost of investing in Lfx as TPT for MDR/RR-TB. Will the cost of treatment be deducted from another important programmatic component, such as TPT for non-MDR/RR-TB or 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 of TPT when offered, willingness to take a hypothetical TPT regimen for MDR/RR-TB, and acceptability (ability and willingness to use TPT as directed) of TPT with FQ. A sixth qualitative study was conducted in South Africa as a substudy 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 among caregivers and 70% among adults to take TPT for MDR/RR-TB, and one study indicated high levels of acceptability among caregivers administering a novel dispersible child-friendly formulation of Lfx to their children. The published qualitative study also found overall high acceptability of Lfx among caregivers of children, but with some pragmatic difficulties in relation to the financial and care burden of providing TPT to the children, especially for caregivers undergoing treatment for TB disease themselves (which was a motivator for accepting treatment but limited their 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 contacts of people with MDR/RR-TB in five countries (Georgia, India, Indonesia, South Africa and Viet Nam) concluded that TPT for MDR/RR-TB was acceptable and of high social value among participants in all five 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 substudy conducted by the V-QUIN investigators examined acceptability among a randomly selected sample of 240 participants in V-QUIN (about equal numbers took placebo and Lfx). They found no major differences in ratings of medication taste, size and frequency of TPT 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 among those who did not complete the study drug. Only a minority of participants would take the treatment again or would recommend it to others.	Key considerations expressed by GDG members when making a judgement of probably yes for acceptability were as follows. • In the survey of NTP managers, many stated that they would accept the recommendation only if it were strong. • The 6-month duration of treatment may be a challenge, although this is the same as the minimum duration of treatment with isoniazid, which 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 regimen 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 AEs in adults in particular was highlighted. • Providing clear information on benefits and risks and a supporting environment for caregivers and beneficiaries is likely to improve acceptability. People's perceptions of the effectiveness and value of TPT are important.



A prospective quantitative substudy among all participants in the TB-CHAMP trial examined acceptability at every treatment phase visit and found that children disliked the taste of Lfx more than placebo, but the children in both arms adapted to the taste over the course of treatment. Caregivers found it more difficult to administer Lfx than placebo, but more than 95% of caregivers reported no difficulty in giving Lfx. Overall, the investigators concluded that acceptability was reasonable but noted an association between poor acceptability and poor adherence.

In addition, a semistructured interview was conducted to evaluate the caregiver experience of administering a novel child-friendly Lfx formulation among 10 child-caregiver dyads in TB-CHAMP. There was relatively high overall acceptability. One major motivator was the caregivers' own experiences with MDR/RR-TB illness and treatment. Pragmatic difficulties were expressed in relation to the financial and care burden on the household due to TPT. Challenges were exacerbated for caregivers who were receiving treatment for their own MDR/RR-TB disease, limiting their capacity to care for children. Caregivers who received greater social support reported better capability for themselves and their children in adhering 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/RR-TB countries, of whom 18 (60%) responded, eight countries were ready to implement Lfx programme-wide if WHO made a strong recommendation (in addition to the six that had already implemented it). A conditional recommendation made implementation less likely for seven NTPs. All managers anticipated that drug storage, transportation and distribution were sustainable. However, the need for additional resources (DST, monitoring and follow-up) was raised as a concern or barrier to implementation by seven of 18 NTP managers.	Key considerations expressed by GDG members when making a judgement of yes for feasibility were as follows. • There is already a WHO recommendation on the use of TPT for MDR/RR-TB, which has been implemented to some degree despite being conditional, with Lfx being one of the options proposed. • Feasibility will depend on additional resources being available to implement the intervention properly, such as DST of the presumed index patient, testing for TB infection (in TB-CHAMP, a positive test result for infection was not required in most individuals; in V-QUIN, adults could participate if they were TST-positive, plus a small number with HIV or malnutrition who were TST-negative) and chest X-ray (done for participants in both trials). • Lfx is widely available as a generic drug in both adult and paediatric formulations.

AE: adverse event; BPaL(M): bedaquiline, pretomanid and linezolid (plus moxifloxacin); DST: drug-susceptibility testing; EMB: ethambutol; EtD: evidence to decision; FQ: fluoroquinolone; ETH: ethionamide; GDF: Global Drug Facility; GDG: guideline development group; GRADE: Grading of Recommendations Assessment, Development and Evaluation; HCW: health care worker; HHC: household contact; HIV: human immunodeficiency virus; Lfx: levofloxacin; MDR/RR-TB: multidrug-resistant and rifampicin-resistant tuberculosis; NTP: national tuberculosis programme; PHOENIX MDR-TB: Protecting Households on Exposure to Newly Diagnosed Index Multidrug-Resistant Tuberculosis Patients; RCT: randomized controlled trial; TB: tuberculosis; TB-CHAMP: Tuberculosis Child Multidrug-Resistant Preventive Therapy trial; TPT: tuberculosis preventive treatment; TST: tuberculin skin test; V-QUIN: Levofloxacin for the Prevention of Multidrug-Resistant Tuberculosis in Vietnam trial; WHO: World Health Organization.



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

Table A4.1 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 offers a structured overview of the GDG's deliberations and presents all judgements in a succinct form that enables their holistic assessment. This synthesis ensures transparency and consistency in the decision-making process, ultimately guiding the GDG's formulation of robust, evidence-based recommendations on the PICO (population, intervention, comparator and outcome) question for clinical practice. The example shown here is for the judgements made for the comparison of the 6-month levofloxacin regimen versus no treatment.



Table A4.1. Example of summary judgements across the 12 EtD criteria

ETD CRITERION	JUDGEMENT						
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	Favours the comparator	Probably favours the comparator	Does not favour either the intervention or the comparator	Probably favours the intervention	Favours 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	Favours the comparator	Probably favours the comparator	Does not favour either the intervention or the comparator	Probably favours the intervention	Favours 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

EtD: evidence to decision.



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

In guideline development, the strength of a recommendation typically reflects the certainty of the underlying evidence. However, there are exceptional circumstances where guideline development groups (GDGs) may issue strong recommendations despite the evidence being of low or very low certainty. 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.

In well-defined clinical contexts – such as life-threatening situations, situations where there is clear potential for catastrophic harm, or situations where alternative options are equally uncertain but one is less harmful or costly – GDGs 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 interests of patients, particularly in urgent or high-stakes clinical scenarios.

Table A5.1 lists paradigmatic situations in which a discordant strong recommendation may be indicated, despite low-certainty or very-low-certainty evidence.





Table A5.1. Paradigmatic situations where strong recommendations are made with evidence of low or very low certainty

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/RR-TB, the BPaLM regimen should be used (1).
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 cost 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 (2).
Potentially equivalent options, one clearly less risky or costly than the other	Low or very low	High or moderate	Both options show similar, though uncertain, benefits, but one is certainly less harmful or expensive than the other	A high value is placed on avoiding harm	High incremental cost or use of resources relative to benefits may justify recommending the comparator, if less harmful	Strong recommendation in favour of the less harmful or costly comparator	For management of postpartum haemorrhage, oxytocin should be preferred over ergometrine alone, a fixed-dose combination of ergometrine and oxytocin, carbetocin and prostaglandins (3).



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					
High confidence in benefits being similar, but one option potentially more risky or costly than the other	High or moderate	Low or very low	Established that alternative management strategies afford similar benefits, but one may be more harmful than the other (low certainty)	A high value is placed on avoiding harm	High incremental cost or use of resources 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 pregnant or planning conception, the American College of Chest Physicians' guidelines recommended against the use of certain anticoagulants (4). High-confidence estimates suggest similar effects of different anticoagulants, but indirect evidence (low confidence in effect estimates) suggests potential harm to the unborn infant with oral direct thrombin inhibitors (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 use of resources 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 presumed 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 (5).

BPaLM: bedaquiline, pretomanid, linezolid and moxifloxacin; HIV: human immunodeficiency virus; MDR/RR-TB: multidrug-resistant and rifampicin-resistant tuberculosis; TB: tuberculosis.

Source: World Health Organization, 2014 (6); adapted from Andrews et al. (2013) (7) and Alexander et al. (2014) (8).

References for Annex 5

1. WHO consolidated guidelines on tuberculosis: module 4: treatment and care. Geneva: World Health Organization; 2025 (<https://iris.who.int/handle/10665/380799>). License: CC BY-NC-SA 3.0 IGO.
2. 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;95:2536-59 (<https://doi.org/10.1210/jc.2009-2354>).
3. WHO guidelines for the management of postpartum haemorrhage and retained placenta. Geneva: World Health Organization; 2009 (<https://iris.who.int/handle/10665/44171>).
4. 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:e691S-e736S (<https://doi.org/10.1378/chest.11-2300>).
5. Rapid advice: treatment of tuberculosis in children. Geneva: World Health Organization; 2010 (<https://iris.who.int/handle/10665/44444>).
6. WHO handbook for guideline development, 2nd ed. Geneva: World Health Organization; 2014 (<https://iris.who.int/handle/10665/145714>).
7. Andrews JC, Schunemann 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;66:726-35 (<https://doi.org/10.1016/j.jclinepi.2013.02.003>).
8. 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;67:629-34 (<https://doi.org/10.1016/j.jclinepi.2013.09.020>).



Annex 6. Current best practice and options for tuberculosis preventive treatment trials

This annex offers methodological guidance on current best practice and options for generating robust evidence on novel tuberculosis (TB) preventive treatment (TPT) regimens. It serves as a complement to the key messages on evidence generation for TPT (see Chapter 3), expanding on operational and technical considerations to optimize the design, conduct and reporting of clinical trials and associated studies. These practices support the generation of high-quality evidence that is more likely to inform evidence-based, inclusive and actionable World Health Organization (WHO) guidelines.

A6.1 Guidance on generating evidence on desirable and undesirable effects

A6.1.1 Research design

Clinical trial design is an important factor that guideline development groups (GDGs) consider when determining the strength of their recommendations. A factor that often contributes to downgrading of the certainty of evidence during guideline development is imprecision. Imprecise estimates of treatment effects may result from an insufficient sample size. Further to the need for both larger studies and early discussion with WHO (see Section 3.1.1), there are important additional methodological details relating to clinical trial design. TPT trials should be strategically designed to support their specific aims and with consideration of how the generated evidence will inform global and national policies. Most TPT trials employ a standard two-arm, parallel-group, randomized controlled trial (RCT) design. To strengthen a trial's integrity and minimize bias, double blinding, where both the outcome assessors and the participants are unaware of group allocation, should be considered (1). When trial arms have different treatment durations or numbers of pills, blinding can be maintained by adding a matching placebo to the shorter regimen, or matching placebo pills can be added to equalize pill burden. However, this approach may have cost limitations and could complicate the assessment of adherence.

RCTs, although methodologically robust, typically assess only one new regimen against a single control. Few trials of TPT incorporate more agile methods, such as adaptive, Bayesian or platform designs, despite their successful application in fields such as HIV prevention and oncology (2, 3). This limits the rapid incorporation of new TPT regimens or delivery strategies. Multi-arm multistage and platform trials can efficiently compare multiple regimens in a single overarching protocol and may reduce costs, timelines and sample sizes for new candidate regimens (4). Adaptive approaches (e.g. using prespecified interim analyses to adjust sample size, stopping poorly performing arms early, or revising inclusion criteria to increase recruitment) could enable faster responses to interim findings, such



as slower-than-expected accrual or lower-than-anticipated event rates (5). Noninferiority designs are frequently employed in TPT trials. These trials often face significant issues with the study's power, when the event rate (i.e. progression from TB infection [TBI] to disease) is lower than anticipated, making it difficult to detect meaningful differences between interventions. Choice of the noninferiority margin (i.e. the largest acceptable loss of efficacy) is a critical decision in these trials. This choice should be grounded in clinical judgement and reflect what is considered "not unacceptably worse" for both patients and public health. Both the absolute and relative noninferiority margins should be carefully selected during trial design. The choice will substantially affect the sample size, the precision of the trial outcome and the subsequent interpretation of results (6). Adaptive approaches, such as revising the sample size during interim analyses based on observed event rates, can help maintain statistical power and relevance as a trial progresses and conditions change.

[Return to Section 3.1.1 in the main guidance on evidence generation \(GEG\) document](#)

A6.1.2 Study population, selection criteria and setting

To improve the external validity and generalizability of trial findings, study populations should reflect the demographic, geographical and epidemiological diversity of the intended target populations. Inclusive eligibility criteria enhance the relevance of research findings to target populations, reducing the risk of indirectness when WHO recommendations are developed (Box A6.1).

Box A6.1. Case study 2

In 2020, the World Health Organization guideline development group (GDG) evaluated evidence on the 1-month daily rifapentine and isoniazid (1HP) regimen for tuberculosis (TB) preventive treatment of people with HIV, from the Brief Rifapentine–Isoniazid Evaluation for Tuberculosis Prevention (BRIEF TB) study (7), a randomized, open-label, Phase 3, noninferiority trial. This study compared the efficacy and safety of 1HP against 9 months of treatment with isoniazid, with the primary endpoint being the diagnosis of incident TB disease or death. When the GDG reviewed this study, 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 about 80% of participants; and the comparator was the 9-month isoniazid regimen instead of the more widely used 6-month regimen.

Although the trial did not include an untreated control group, the GDG determined that the benefits observed in the 1HP arm among participants who tested positive by tuberculin skin test or interferon-gamma release assay would likely extend to other at-risk populations. Consequently, the recommendation was expanded to include HIV-negative individuals aged over 13 years. 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.

Proactive enrolment of high-priority target populations is important, including infants, children, pregnant and breastfeeding women, older adults, undernourished individuals and those with significant comorbidities, such as silicosis or diabetes. In many previous trials, these populations have been excluded (see Annex 7). For subgroups that



constitute a small proportion of the target population but have key risk factors (such as HIV status, nutritional status, age, pregnancy, substance use, exposure to multidrug-resistant and rifampicin-resistant TB [MDR/RR-TB], incarcerated people, migrants and Indigenous populations), stratified or targeted enrolment strategies are advisable. This approach facilitates adequate subgroup representation, enabling stratified analyses that can increase certainty about a regimen's effectiveness and tolerability, enhance generalizability and inform health equity considerations. Inclusion of participants taking concomitant medicines (e.g. antiretrovirals, oral contraceptives, opioids) is encouraged, unless there is a known risk of clinically significant drug–drug interactions or insufficient safety data (8).

Including high TB burden countries facilitates the extrapolation of recommendations to other countries and regions, promotes uptake of the regimen and fosters equity in access to novel TPT regimens. Studies with a broad geographical span are particularly helpful (e.g. the Protecting Households on Exposure to Newly Diagnosed Index Multidrug-Resistant Tuberculosis Patients [PHOENIx MDR-TB] trial, which is evaluating delamanid for prevention of MDR/RR-TB in 31 locations around the world) (9). It is recognized that the cost of conducting multicountry studies may exceed the available resources. In such instances, researchers can consider including a diverse range of subnational regions to enhance the broader applicability of the findings and strengthen subsequent recommendations. Overall, recruiting participants from a range of populations and settings is likely to be more generalizable, potentially leading to stronger recommendations than would be possible from a single trial conducted in one setting.

Exclusion criteria should be thoroughly justified and focused on identifying individuals whose participation would expose them to undue risk compared with potential benefits (e.g. due to insufficient dosing information or concerns about specific comorbidities). Children or pregnant women should not be excluded without a strong medical or scientific justification. Exclusion of key cohorts, such as people with HIV or pregnant and breastfeeding women, would inherently constrain the scope and applicability of the GDG's recommendations.

A WHO expert consultation identified key issues affecting the representativeness of enrolled populations in TPT trials. These issues, and suggested solutions, are shown in Table A6.1 (4).



Table A6.1. Key challenges in study population, selection criteria and setting for TPT trial design, and suggested solutions (4)

	Key challenge	Suggested solutions
Study population	Underrepresentation of high-risk and special subgroups (e.g. children, people with diabetes, pregnant women)	• Use basket and stratified trial designs to ensure subgroup inclusion and enable subgroup analysis
	Low enrolment in small high-risk groups, leading to data gaps	• Use data borrowing from similar subgroups or external trials • Use adaptive recruitment strategies
	Patient-level barriers (e.g. beliefs, stigma, out-of-pocket costs) affecting participation	• Design person-centred, pragmatic trials that address barriers, preferences and care costs
Selection criteria	Overly narrow or restrictive eligibility, excluding key populations	• Broaden eligibility • Deliberately include diverse subgroups
	Imperfect TB infection tests (TST and IGRA); TST has reduced specificity among BCG-vaccinated individuals, and neither test can predict progression to TB disease. This hinders accurate identification of optimal TPT target populations and enrolment of lower-risk or uninfected participants, and increases the risk of false noninferiority conclusions for a regimen	• Combine both epidemiological and clinical criteria during enrolment • Invest in prognostic biomarker research
	Inconsistent global definitions of “high-risk” and “high-burden” populations	• Harmonize definitions across trials • Use multisetting and multinational approach
Setting	Heterogeneity across trial settings (e.g. urban/rural, different TB burdens) reduces generalizability	• Conduct trials across varied countries and settings • Use multisite enrolment
	Contextual barriers (e.g. delivery model, health system factors)	• Use pragmatic and implementation science methods to assess delivery strategies in real-world care

BCG: Bacillus Calmette–Guérin; IGRA: interferon-gamma release assay; TB: tuberculosis; TPT: tuberculosis preventive treatment; TST: tuberculin skin test.



It is important to note the trade-off between inclusivity and statistical validity, particularly when designing noninferiority trials for TPT (10). Efforts to be inclusive should not violate the criteria that determine risk of TB in the target study participants. Enrolling participants with low TB risk can dilute the event rate and create a situation where noninferiority is “easy” to show, regardless of the true value of the intervention. This undermines the ability of the trial to detect a difference (if one exists) and the constancy assumption (i.e. the expectation that the control regimen will perform as in prior trials). If a trial enrolls a population at a lower risk of developing TB, the constancy assumption is violated, meaning the findings may not be valid or generalizable. Measures of TB incidence in such “diluted”, low-risk populations will be misleadingly low, so differences between novel regimens and established comparators in noninferiority studies may be small, misinforming the GDG. To overcome this statistical concern, trialists may need to prioritize the enrolment of individuals who are truly at risk of developing TB based on their history, clinical status and TBI test results, especially when the efficacy of TPT has already been demonstrated in such groups. Where possible, analyses should be stratified by risk of developing TB (i.e. high versus low risk), so the GDG can interpret the findings in the appropriate context. Providing these measures would enhance both the certainty of evidence and the generalizability of findings.

Beyond increasing the inclusion of underrepresented populations as trial participants, it is also important to ensure that study teams and study advisory boards are inclusive of different stakeholders. The inclusion of people with lived experience of TB or TPT is a valuable complement to the technical expertise of health care workers and other clinical experts (11). Overall, trial participant and community engagement should be centrally placed in the trial planning and implementation phases, to maximize the likelihood that research meets public needs and maintains their trust. It is also important to share the findings of research studies with the participants.

[Return to Section 3.1.2 in the main GEG document](#)

A6.1.3 Specification of the intervention

Further to the need for a detailed description of different components of the intervention (see Section 3.1.3), this section highlights additional background information. WHO encourages trialists to evaluate the safety, efficacy and feasibility of regimens that can be implemented in resource-limited settings. The following additional regimen characteristics may be considered in the choice of intervention: high efficacy, safety, good tolerability, low cost, minimal drug interactions, and availability in formulations suitable for all target populations (12). Preference should be given to regimens and formulations that are likely to be readily accessible to national TB programmes (NTPs) outside the trial setting (except in trials studying new medications), regimens with low pill burden or available fixed-dose combinations, or long-acting formulations. For children, paediatric formulations (typically dispersible, palatable and scored) should be preferred over adult formulations. The dosing of intervention drugs should align with WHO harmonized weight bands (13).

Diagnostic tools to detect TBI and rule out TB are becoming increasingly available in resource-limited settings, although barriers to access persist. It is therefore important for trial findings to be applicable to the setting where the intervention studied is most needed. Chest radiography and WHO-approved molecular rapid diagnostic tests are recommended for screening to exclude TB, including asymptomatic TB, and avoid the inadvertent administration of TPT to those with unrecognized TB disease (14). The diagnosis of TBI should be made by following WHO guidelines (15).



Other implementation elements that influence trial outcomes and are also relevant to GDG discussions include adverse event (AE) management, drug–drug interactions (e.g. with antiretroviral drugs), drug-susceptibility testing (DST), methods of adherence monitoring, maximum allowed duration of TPT when interrupted, and the approaches to management of treatment interruption (including patient education and awareness or community engagement strategies). These should be described in the trial protocol and manuals of procedures and be made available to GDGs for both intervention and comparator regimens. Detection and appropriate management of AEs are ethically important in Phase 2 and 3 trials. Provider decisions to stop TPT may be more frequent in a trial than in programmes; on the other hand, patient decisions to stop TPT are variably reported in trials. Without collecting data on the reason why study participants discontinue treatment, factors affecting adherence – such as tolerability, perceptions, stigma or access to treatment – may be incompletely understood. As a result, GDGs may conclude that rates of treatment interruption observed in a trial may not be directly generalizable to programmatic settings. They may also recommend that additional patient support is required to overcome potentially underrecognized barriers to adherence.

[Return to Section 3.1.3 in the main GEG document](#)

A6.1.4 Comparator regimens

Direct head-to-head comparisons of different TPT regimens are preferred over indirect comparisons using network meta-analysis. Trials should include a concurrently and randomly enrolled comparator arm receiving a standard of care (SOC) TPT regimen, consistent with current WHO recommendations. An appropriate comparator arm allows accurate assessment of the efficacy of the new TPT regimen. 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 a critical priority for participants who are otherwise healthy. If no internal comparator is included, or historical or external nonconcurrent controls are used, the strength of recommendations will be downgraded due to indirectness, selection bias or unmeasured confounding (8, 16, 17). Clear documentation of inclusion and exclusion criteria is crucial if nonconcurrent comparators are used.

The selection of a comparator depends directly on the aim of the trial. Use of a placebo comparator may be justified in (increasingly rare) situations where there is no available or recommended SOC regimen for the target population. This was previously the case for TPT for contacts of patients with MDR/RR-TB or other high-risk groups for whom no proven comparator existed (12). A placebo-controlled design is optimal when the goal is to prove efficacy in a population for which there is uncertainty about the balance of potential benefits and risks of TPT – that is, where there is equipoise. For example, the Levofloxacin for the Prevention of Multidrug-Resistant Tuberculosis in Vietnam (V-QUIN) trial used a placebo control due to lack of an established SOC regimen at the time (18). Similarly, diabetes increases the risk of TB, but there has been equipoise as to whether TPT is more beneficial or harmful in people with diabetes. Therefore, a placebo-controlled trial was ethically justified to address this clinical uncertainty (Box A6.2). Selecting an appropriate comparator allows the efficacy and safety of new regimens to be assessed relative to current best practice, and to optimize the guidance developed for individuals who may benefit from TPT.



Box A6.2. Case study 3

There is uncertainty about the balance of benefits and harms of giving tuberculosis (TB) preventive treatment (TPT) to people with diabetes, even if they are at higher risk of TB. There is thus no recommendation for systematic testing for TB infection and providing TPT for people with diabetes. Placebo-controlled TPT trials are underway in people with diabetes. The Preventive Treatment of Latent Tuberculosis Infection in People with Diabetes Mellitus (PROTID) trial, which began recruitment in 2022, is comparing the use of 3 months of weekly isoniazid and rifapentine against a placebo (19). This design is expected to yield critical efficacy data. Similarly, the BALANCE trial, which started in 2025, is an open-label, randomized controlled trial evaluating the effectiveness, safety and risk-benefit profile of 1 month of daily rifapentine with isoniazid compared with placebo for people with diabetes (20), with objectives also including assessment of feasibility and acceptability.

The time between trial design, implementation and reporting can span many years, and WHO recommendations may change during this period. Researchers may have difficulty in altering the comparator regimen midway through a trial, should the SOC change due to new evidence emerging before the trial is completed. If comparator regimens do change during the trial, stratified analyses should present the effectiveness and tolerability of the intervention in comparison to each comparator regimen to help interpret the outcomes (21).

Network meta-analysis may be useful to indirectly compare regimens that have not undergone head-to-head studies. However, assessment of the certainty of evidence from these analyses needs to be done with caution (22).

[Return to Section 3.1.4 in the main GEG document](#)

A6.1.5 Efficacy outcomes

Most recent TPT trials include progression to TB disease as the primary endpoint. However, this endpoint alone may not fully capture the spectrum of outcomes of importance to patients, communities and decision-makers, particularly in low-incidence settings or in studies with short follow-up durations. The heterogeneity of TB risk between populations (e.g. based on HIV status, TBI test results, age and region) is often inadequately described when trials are reported. Limited stratification and insufficient power to detect subgroup-specific effects restrict a GDG's ability to assess the generalizability of the findings. Trial designs should allow for more granular stratification and prespecified subgroup analyses to inform the policy for use of TPT in diverse populations (4). However, it is acknowledged that trials may not be sufficiently powered to measure a significant effect in each subgroup.

Standardized outcome definitions and reporting are essential to enable valid network meta-analyses and to permit comparisons between TPT regimens that have not been evaluated directly, notwithstanding concerns about the indirectness of evidence. Reported outcomes should be clearly defined and would preferably include those rated as "important" or "critical" by the respective WHO GDG (Box A6.3).



Box A6.3. Examples of outcomes rated by guideline development groups as important or critical for TPT regimens in the past

- Incidence of TB disease (all forms^a and microbiologically confirmed, including drug-resistant TB); TB among participants who test positive at study entry on an antigen-based skin test, such as TST/TBST or IGRA
- Incidence of TB disease for people with HIV (all forms, microbiologically confirmed); TB among participants who are ART-naïve or receiving ART at entry
- Mortality during TPT (all-cause and related; i.e. possibly, probably or definitely caused by or associated with the treatment)
- AEs of Grades 3–5 and AEs of Grades 3–5 that are assessed as being related to TPT
- AEs of special interest (e.g. Grade 3–5 hepatotoxicity, tendinopathy or QT prolongation)
- TPT discontinuation due to AEs
- TPT completion (ever, or within a specified period)^b
- Adherence to TPT (proportion of doses taken)
- Changes to the gastrointestinal or respiratory tract microbiome
- Acquired drug resistance of *Mycobacterium tuberculosis*, where comparison to a known or presumed source case is possible

AE: adverse event; ART: antiretroviral therapy; HIV: human immunodeficiency virus; IGRA: interferon-gamma release assay; TB: tuberculosis; TBST: tuberculosis antigen-based skin test; TPT: tuberculosis preventive treatment; TST: tuberculin skin test.

^a Prespecified definitions of possible or probable incident clinical TB may be used. Expert clinical panels blinded to treatment group or local clinicians may apply standard definitions to assign outcomes.

^b Treatment completion may be defined according to the proportion of treatment completed within an assigned time frame (e.g. at least 80% of doses completed within 150% of the minimum treatment period).

Table A6.2 outlines the relevance of priority outcomes for different aspects of the WHO guideline development process.



Table A6.2. Examples of priority outcomes for WHO TPT guideline development and their relevance to the EtD framework

Outcome example	Relevance to EtD framework
Incidence of TB (confirmed or clinical)	Informs the balance of desirable and undesirable effects of TPT
Mortality (all-cause or TB-specific)	Informs the balance of desirable and undesirable effects of TPT
Grade 3–5 AEs	Informs the balance of desirable and undesirable effects of TPT
Grade 1–2 AEs and discontinuation due to side-effects (tolerability)	Informs the balance of desirable and undesirable effects and acceptability of TPT
Acquisition of TB drug resistance during or after TPT	Informs the undesirable effects of TPT
Completion of and adherence to TPT (e.g. proportion of participants completing $\geq 80\%$ of doses)	Informs the acceptability and feasibility of TPT
Patient-reported acceptability or satisfaction	Informs the acceptability of TPT
Distributional cost-effectiveness by age and social status	Informs the equity and feasibility of TPT (consideration of subgroups)
Cost per TB case or DALY averted	Informs the cost-effectiveness (incremental cost) of TPT

AE: adverse event; DALY: disability-adjusted life-year; EtD: evidence to decision; TB: tuberculosis; TBST: tuberculosis antigen-based skin test; TPT: tuberculosis preventive treatment; WHO: World Health Organization.

Efficacy (i.e. prevention of TB disease) and safety (i.e. AEs) are typically reported as separate endpoints. This places the burden of synthesizing the risk–benefit profile on GDGs, rather than the researchers reporting the trial. To aid GDGs in assessing how patients and clinicians balance the risks and benefits of TPT in real-world practice, composite outcomes that integrate both efficacy and harms can be developed and validated (Box A6.4). Where possible, outcomes should incorporate the priorities of patients and other stakeholders, as their views are often underrepresented in measures of efficacy (4).



Box A6.4. Case study 4

In the TB Ante versus Postpartum Prevention with Isoniazid in HIV Seropositive mothers and their Exposed infants (TB APPRISE) trial, a composite maternal–infant outcome was constructed based on maternal tuberculosis (TB), infant TB, severe adverse pregnancy outcomes and other serious adverse events (AEs) in either the mother or infant. Panels of stakeholders assigned scores to these combined outcomes to reflect their relative importance, enabling a more unified and patient-centred analysis of trial results (23).

In the Phase 3 Brief Rifapentine–Isoniazid Evaluation for TB Prevention (BRIEF TB) trial, a composite outcome (of whether participants developed TB, the severity of AEs and death) was defined and scored to simultaneously assess both efficacy and safety (7). The categories included no TB and only mild or moderate (Grade ≤ 2) AEs, no TB with Grade 3 or 4 AEs, development of TB with various degrees of AEs, and death. The scores reflected clinical desirability, with “no TB and Grade ≤ 2 AEs” considered the best outcome and “death” the worst. This composite was used to provide an integrated risk–benefit assessment, beyond reporting efficacy and safety separately.

When a composite outcome is employed, it is essential to disaggregate and report its individual components separately to the GDG (see Sections 3.1.5 **and** 3.1.6). This will avoid the evidence being downgraded due to indirectness, given that aggregated reporting may obscure differential effects and lead to misinterpretation, particularly if the components differ in clinical importance or frequency.

Treatment adherence data should be consistently collected and reported to enable intention-to-treat (ITT) and per-protocol analyses, providing insights into real-world effectiveness. Mortality and safety outcomes must also be captured, as they directly inform the overall risk–benefit assessment (24). To enhance methodological rigour and policy relevance, trialists should consider incorporating outcomes that matter to the patient (using tools such as discrete choice experiments or direct patient engagement) and adopting trial designs that accommodate rare events and population heterogeneity. Outcomes such as quality of life, treatment acceptability and long-term health impacts should also be considered, given their importance for TPT uptake and adherence (25).

[Return to Section 3.1.5 in the main GEG document](#)

A6.1.6 Safety outcomes

Systematic reporting of TPT safety outcomes is important for GDGs, as it allows WHO recommendations to be based on a clear understanding of benefits and risks of an intervention across diverse populations. For a comprehensive safety assessment, data on AEs should be reported for all individuals who received at least one dose of TPT, constituting the safety population for analysis. This must include the proportion of participants experiencing AEs of Grade 3 or higher (both overall and treatment-related AEs, in the general adult population and in specific groups, such as children and people with HIV), and, critically, Grade 1 or 2 AEs. Additionally, AEs of special interest – such as hepatotoxicity and neuropathy with isoniazid, tendinopathy or QT prolongation (Box A6.3) – and AEs related to drug–drug interactions should be reported. Hepatotoxicity is a particularly important complication of most established TPT regimens. Another key safety outcome that trials frequently report



is all-cause and drug-related mortality during TPT. Differences in susceptibility to AEs based on age, sex, comorbidities (especially HIV and diabetes) and concurrent medications are often not adequately considered during trial design. Many trials lack sufficient subgroup analyses or adequate power for these comparisons. Special populations, such as pregnant women, people with diabetes and children, may experience different safety profiles, warranting tailored safety endpoints and analytical approaches (4, 26).

There is inconsistency in the grading, definition, monitoring and reporting of AEs across TPT trials, which hinders comparison and meta-analysis. Trials investigating new medicinal products must therefore comply with the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use standard and report AEs of all grades (27). To ensure consistency and enable unbiased comparisons between studies, a recognized toxicity instrument, such as the Common Terminology Criteria for Adverse Events (CTCAE) or grading criteria of the Council for International Organizations of Medical Sciences or the Division of AIDS, can be used (28–30). Safety outcomes reported using standardized definitions support valid comparisons between different TPT trials and consistent interpretation of results.

The timing of AE assessments should be harmonized for unbiased comparisons of safety outcomes, and the data should be collected at regular intervals (e.g. fortnightly or monthly). Secondary analyses considering attribution of AEs to treatment, as determined by a blinded panel, can add value, as can analyses of the rate (rather than the proportion) of AEs.

As TB and AEs from TPT are relatively rare, and trials often have limited sample size and duration, some uncommon but serious AEs may not occur during an initial Phase 2 or Phase 3 trial (31, 32). Trials often do not follow participants for long enough to capture delayed toxicity, which is crucial for new or long-acting regimens or in vulnerable groups (e.g. children, pregnant women and people with comorbidities). To identify rare but important safety outcomes, data from larger observational studies or pragmatic Phase 4 trials would be helpful for WHO guideline development.

Most safety endpoints in TPT trials focus on severe AEs or laboratory findings, neglecting other outcomes that matter to patients, such as quality of life, tolerability, treatment burden and mild-to-moderate AEs that may affect adherence to and acceptability of TPT. There is a recognized need to solicit patient, community and stakeholder input when defining safety outcomes, to ensure relevance and improve treatment uptake (4). As TPT research increasingly shifts towards pragmatic (real-world, implementation-focused) and decentralized trial designs, there may be less direct onsite supervision, more variation between study settings and delegation of certain trial functions to nonspecialist staff. In such trials, safety monitoring frameworks need to be adapted through, for example, use of innovative digital tools, remote monitoring, standardized protocols or increased participant education, so that participant safety and data integrity are not compromised (33).

[Return to Section 3.1.6 in the main GEG document](#)

A6.1.7 Tolerability outcomes

Tolerability is an important consideration for individuals taking TPT, who are usually healthy and less willing to accept even low-grade symptoms compared with individuals with TB disease (34). Poor tolerability can lead to treatment interruption or discontinuation, reducing



the regimen's effectiveness. As tolerability outcomes influence TPT completion, they are crucial for assessing the balance between the harms and benefits of TPT (35).

It is desirable to collect data on tolerability during TPT trials, as this enables GDGs to incorporate the views of individuals receiving TPT into their decision-making. A high-efficacy regimen that is poorly tolerated may not be feasible to implement under programmatic conditions (36). This is important in Phase 3 trials, in which the level of support provided to individuals to adhere to treatment is often beyond that which can be provided during routine care in most high-burden countries.

Tolerability can be measured during trials using standardized patient-reported outcome (PRO) questionnaires. 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 can be an important complement to bacteriological and clinical outcomes during the guideline development process (37, 38). Standardized participant questionnaires can be developed, preferably using standardized tools (e.g. the PRO-CTCAE tool, which measures participant experiences with AEs while receiving treatment), based on the likely tolerability of the regimens being evaluated (39). Any such tool should be shown to be valid, reliable and responsive (i.e. has longitudinal validity) in reflecting the person's experience with TPT.

[Return to Section 3.1.7 in the main GEG document](#)

A6.1.8 Acquisition of drug resistance

TPT is indicated for people who are expected to have a low *Mycobacterium tuberculosis* bacillary load in their body and thus a lower likelihood of selection of a mutant strain. However, when inadvertently given to individuals with unrecognized TB, TPT is unlikely to cure TB disease (due to its higher bacterial load than TBI) and may thus favour the emergence of resistant strains. TPT trials should therefore implement an effective rule-out algorithm and use sensitive tools (e.g. chest X-ray) to exclude prevalent TB disease, including asymptomatic TB, among the target populations before starting TPT. Although poor TPT adherence could also lead to the emergence of resistance (40), the risk of acquired resistance with TPT is uncertain, particularly for regimens including newer TB drugs (e.g. bedaquiline, delamanid). The emergence of drug-resistant TB is especially consequential in vulnerable populations, such as children, people with HIV and pregnant women, where microbiological confirmation of disease may be more difficult and the implications of treating drug-resistant disease more serious.

TPT trials need protocols to monitor the emergence of drug resistance during and after TPT. Systematic DST is required for participants who develop incident TB disease, as well as the presumed source patient. Interim analyses may help to quickly identify higher-than-expected rates of acquired resistance among TPT trial participants, potentially enabling the trial design to be modified (4). However, even in large clinical trials of TPT, the absolute number of individuals who develop incident TB is small. For this reason, the number of people taking TPT who develop resistance will be limited, and data from larger prospective studies may be required to obtain precise estimates of acquired drug resistance. Extended follow-up of trial participants may increase the number of individuals identified with incident TB, potentially increasing the power to detect acquired resistance. A meta-analysis of cases may best estimate the potential of a TPT regimen to generate resistance (41).

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. The development of drug



resistance during TPT can be measured with genomic testing by comparing mutations in isolates from contacts with those from the person with incident TB while receiving TPT. Where it is not feasible to carry this out immediately, the isolates from index patients may be stored for later analysis. Collection of isolates in children may be more difficult. The strain isolated from index patients and household or close contacts may be different, especially in high-prevalence settings; hence, acquisition of drug resistance cannot be inferred from the presence of drug resistance alone (42, 43).

Apart from resistance in *M. tuberculosis* strains, the rest of the human microbiome can also be affected by TPT regimens. Nested studies that evaluate the effect of TPT on the human microbiome should particularly be considered for regimens that include broad-spectrum antibiotics (e.g. levofloxacin). Such antibiotics can reduce the diversity of gastrointestinal and respiratory tract flora and lead to the acquisition of resistance to medications that are used for treating other common infections (18, 44).

[Return to Section 3.1.8 in the main GEG document](#)

A6.1.9 Treatment duration

Patients may prefer shorter TPT regimens, particularly if such regimens are effective and well-tolerated. However, trials often collect few data on the value of shortening TPT duration for people who receive TPT. In noninferiority trials, the efficacy of the intervention may be similar to the SOC; however, the noninferiority margin expresses the trade-off between duration and efficacy from the perspective of the investigators, not the other stakeholders (patients, providers and NTP managers). In the absence of an outcome summarizing the benefit of a shorter regimen, GDG panels typically consider whether the shorter duration translates into higher rates of TPT completion and lower rates of AEs, which is an indirect judgement (Box A6.3 and Table A6.2). Direct measurement of the benefits of TPT shortening may be required to support strong recommendations for shorter treatments.

PRO surveys (covering acceptability, patient satisfaction or quality of life) may be able to capture the perceived benefits of shorter TPT. These will be considered in the evidence to decision (EtD) framework as part of an evaluation of the balance of effects, alongside other key considerations. To elucidate the advantages of TPT shortening for GDGs, future trials should consider incorporating the following in the trial design and implementation (4):

- metrics to express the cumulative effect of treatment shortening, including pill burden and indirect costs;
- embedded economic evaluations to capture both health system and patient-level costs across diverse settings;
- validated measurements to attribute improved adherence to shorter TPT duration rather than to confounding factors (e.g. health care support, incentives);
- validated measurements of PROs (acceptability, patient satisfaction or quality of life);
- composite outcomes that weigh both clinical and implementation benefits (i.e. not only noninferiority in efficacy but also the trade-off between slightly lower efficacy and much higher acceptability and adherence), which would be presented to the GDG along with the disaggregated outcomes; and
- multi-arm or adaptive designs that characterize the duration–efficacy curve (16, 45).

[Return to Section 3.1.9 in the main GEG document](#)



A6.1.10 Duration of follow-up

Longer-term follow-up in TPT trials is required to assess the durability of the preventive effect and late development of TB disease. Optimal selection of the follow-up duration must strike a balance: it should be long enough to capture a sufficient proportion of incident TB events but should also account for the risk that people may be reinfected with a new strain following randomization, particularly in high transmission settings. Shorter post-treatment monitoring periods can reduce the number of observed TB events, thereby decreasing the power to detect true differences, or a lack thereof, between study arms.

Given the low incidence of TB events, maximizing capture of such events and participant retention is critical to maintain statistical power and avoid biasing results towards the null. High participant retention during follow-up is essential to reduce the potential misclassification of TB outcomes (8).

The primary efficacy analysis should be anchored to a fixed time point from randomization, to avoid immortal time bias. To ensure comparability, participants in both intervention and comparator arms should be followed for an equivalent duration. This is typically 24 months after initiation or at least 12 months beyond completion of the longest TPT regimen. This approach aligns with evidence showing that most TB disease occurs within the first 1–2 years after TBI (46, 47). In addition to active follow-up, passive follow-up (with consent) may be undertaken by searching public health notification systems for individuals enrolled in TPT trials who have subsequently developed TB.

Variable follow-up times in TPT trials can be problematic, as the rate at which TB develops after treatment is nonlinear (i.e. more events occur early after end of treatment). Hence, combining different follow-up lengths can introduce bias. Innovative and adaptive trial designs offer a promising approach in this context (4). They allow for variable follow-up durations within the same trial, which is particularly useful in basket trials or stratified designs. Follow-up periods can be tailored to the needs of different subgroups; for example:

- children who are at higher risk of rapid progression may require shorter follow-up; and
- people with HIV may need stratified follow-up durations based on antiretroviral therapy (ART) status, immune recovery and local epidemiological conditions.

Such flexibility enhances the trial's responsiveness to biological and contextual variability, while maintaining methodological rigour.

[Return to Section 3.1.10 in the main GEG document](#)

A6.1.11 Sample size and statistical issues

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 a potentially effective treatment. When determining the sample size, it is important to consider the desired level of precision for each outcome measure that is usually considered critical or important (Box A6.3), not just for the primary outcome on which the trial hypothesis may have been based. Nonetheless, it is understood that a study cannot be powered for all possible secondary outcomes, owing to practical limitations on sample size.



A WHO consultation on TPT trial design highlighted key sample size considerations for TPT trials (4).

- Low event rates (i.e. progression from TBI to TB disease) may require large sample sizes to observe sufficient events, especially when using a noninferiority design.
- Selecting populations at highest risk (e.g. close contacts of patients with TB, people with HIV) can improve efficiency and reduce the required sample size.
- Misestimating the event rate in controls can lead to an underpowered trial.
- Novel strategies, including adaptive approaches to noninferiority margins, can address small numbers of outcome events and optimize statistical power without dramatically increasing sample size.
- Explicitly incorporating the imperfections of diagnostic tests for TBI into power and sample size calculations is essential.

The legacy TPT regimens, such as 6 or 9 months of daily isoniazid monotherapy, have been shown to have an efficacy of about 60–90% compared with no treatment (48). Demonstrating the superiority of a new TPT regimen is often impractical, given the low event rates and hence large sample size required. TPT trials therefore often employ a noninferiority trial design to evaluate whether a new TPT regimen is not worse than an existing one by a prespecified margin (6, 49, 50). One of the key considerations is the choice of noninferiority 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, while 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). Second, the choice of participants (e.g. those with a positive TBI test result) and random allocation with strict allocation concealment are essential to avoid selection bias and false declaration of noninferiority. Stratified randomization (e.g. by HIV status, age or geographical locations based on 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 noninferiority.

Treatment-shortening trials using noninferiority designs have contributed immensely to recent guidelines, leading to recommendations for effective TPT regimens of 1–4 months’ duration. These shorter regimens were found to offer substantial benefits over the previous 6–9-month isoniazid-based regimens. However, even if they may be preferred by people prescribed TPT, they are still too long and relatively toxic for an individual who has a very low baseline risk of disease. The efficacy of even shorter TPT regimens, including regimens under 2 weeks’ duration, therefore needs further exploration. The optimal duration of TPT regimens lasting less than 1 month may not be best explored using classic two-arm noninferiority trials. One approach is to evaluate the relationship between different durations and efficacy of TPT. This would require some participants to be randomly assigned to no (or delayed) TPT to characterize the number of TB cases averted by the shortest options. However, such designs would require careful ethical consideration and would need to ensure that participants not receiving treatment provided informed consent. Within treatment-shortening trials, additional research is required to identify patient views regarding the optimal trade-off between efficacy and acceptability (4, 51).

[Return to Section 3.1.11 in the main GEG document](#)



A6.1.12 Analysis and reporting

Trials should present outcomes in both the ITT and per-protocol populations, especially for noninferiority trials. All safety outcomes should be systematically reported. AEs (severity, frequency and patient-important outcomes), quality of life and patient-reported experience should be presented. Subgroup analyses should be prespecified and reported, with justification for inclusion (e.g. by age, HIV status, geography or risk profile). Reporting of other important outcomes, such as palatability, patient preferences, acceptability, feasibility of administration, cost-effectiveness, drug-drug interactions and acquisition of resistance, should be considered. Collection of data on the outcome of TPT for women who become pregnant after randomization is particularly valuable, as pregnant women are often excluded from trials because of an unfavourable risk-benefit analysis. These data allow GDGs to make better-informed decisions about treatment during pregnancy (4, 26).

Public confidence in scientific research evidence and associated guidelines relies on transparency and reproducibility. There are established international standards for registration of trials, publication of protocols and statistical analysis plans, and reporting of trial results according to the latest Consolidated Standards of Reporting Trials (CONSORT) guidelines (52). To establish the applicability of trial results, key subgroups of interest during guideline development include those based on age (≤ 5 , 5–10 and >10 years), HIV status (positive or negative and, ideally, receiving ART or not, and CD4 count) and level of adherence to TPT. The overall combined effect of these factors may be evaluated by carrying out stratified analyses.

Any per-protocol analyses should carefully consider the estimand of interest and use appropriate methods, avoiding exclusions after randomization where possible, as such exclusions can introduce bias (53). It is recommended to specify the estimand of interest, align analysis methods to this estimand, state all assumptions and plan sensitivity analyses to test robustness. Multiple estimands – including those that account for adherence – are often warranted to fully capture the clinical question, and sensitivity analyses are essential to explore how different ways of handling intercurrent events (e.g. nonadherence) might affect results.

[Return to Section 3.1.12 in the main GEG document](#)

A6.1.13 Data sharing

TPT trials should aim to harmonize key outcomes and data structures across studies to facilitate robust analyses of pooled data. Pooled studies typically involve individual participant data (IPD) meta-analysis, allowing for more detailed evaluation of treatment outcomes than meta-analysis of summary effect estimates (54). Anonymized data sharing should be included on informed consent forms during trials, with timely publication and sharing or placement of data in a public repository immediately after publication of primary trial results. This allows pooled analyses of the most up-to-date data for the common good (8).

Analyses that combine data from multiple trials can increase both the generalizability of the study findings and the directness of the evidence (Box A6.5). They may additionally enable verification of the conclusions, improve data validity and support testing of new hypotheses using existing data, maximizing the value of research investments. These analyses can also enable GDGs to better identify subgroups that may benefit most from TPT.



Box A6.5. Case study 5

The effectiveness of levofloxacin (Lfx) as tuberculosis (TB) preventive treatment (TPT) for contacts of individuals with multidrug-resistant and rifampicin-resistant TB was evaluated in two independent clinical trials, one in Viet Nam (Levofloxacin for the Prevention of Multidrug-Resistant Tuberculosis in Vietnam [V-QUIN]) and one in South Africa (Tuberculosis Child Multidrug-Resistant Preventive Therapy [TB-CHAMP]). The V-QUIN trial primarily enrolled adults, with some adolescents and children, while the TB-CHAMP trial predominantly investigated Lfx TPT in children and adolescents (55, 56). Collaboration between investigators during trial design ensured that participants of all ages were collectively recruited between the two trials. Researchers from both trials collaborated before unblinding the results to design a combined analysis strategy (57). Combining data from these two trials from two geographical locations, through a preplanned individual participant data (IPD) meta-analysis, increased the generalizability of the studies' findings. Bayesian analyses enabled direct estimates of the probability that Lfx reduced the incidence of TB. The separate trial findings, as well as an IPD meta-analysis combining the trial findings, were presented to the World Health Organization (WHO) guideline development group. WHO also commissioned a systematic review. These combined analyses contributed to the strong recommendation issued by WHO, with moderate certainty of the estimate of effects (12).

De-identified IPD should be made widely available through established data repositories. Sharing the statistical code used to calculate a trial's primary outcomes increases transparency of the analytical process. Having data publicly available can also facilitate alignment of outcome definitions and harmonized data collection in the future (58).

[Return to Section 3.1.13 in the main GEG document](#)

A6.1.14 Nonrandomized studies

Nonrandomized studies (NRS) are often undertaken to evaluate TPT in special populations or for rare outcomes, long-term safety, or policy-level interventions, where an RCT would be impractical or unethical (e.g. pregnancy outcomes in women receiving TPT, population-level TPT effects) (59). Evidence from NRS may be used to complement RCT evidence, especially in outcome synthesis for guidelines or policy decisions. However, the risk of biases, such as confounding and selection bias, is greater with NRS, which may result in misleading estimates of effectiveness (60). Therefore, establishing clear causal relationships between TPT and outcomes in NRS is difficult. Proper analytical methods, risk-of-bias tools and prespecified protocols are essential for maximizing the value of NRS (61, 62). A case study where NRS was used to complement RCT evidence for guideline development is shown in Box A6.6.



Box A6.6. Case study 6

The World Health Organization (WHO) first recommended the 3-month weekly rifapentine plus isoniazid (3HP) tuberculosis preventive treatment (TPT) regimen in 2015 and further updated its scope in the 2018 guidelines. The foundational evidence for 3HP came from a landmark 2011 randomized controlled trial (63), which demonstrated that it was noninferior to longer isoniazid regimens and had better completion rates and a more favourable safety profile. This formed the basis for the initial WHO guidelines.

However, subsequent real-world implementation and safety data, often derived from observational studies, were crucial for refining and expanding these guidelines. These studies confirmed high TPT completion rates in routine clinical settings, a key advantage of the 3HP regimen. A 2016 study provided valuable data on the 3HP regimen's safety and efficacy in people coinfecting with HIV, a critical subgroup not fully addressed in the initial trial (64). Furthermore, pragmatic implementation studies provided nuanced insights into real-world challenges by identifying clinical predictors of adverse drug reactions and different delivery strategies (e.g. directly observed therapy versus self-administered therapy) (65–67).

This overall body of evidence helped WHO move from a recommendation based on efficacy but limited to low-burden settings to a comprehensive recommendation applicable everywhere.

In the Grading of Recommendations Assessment, Development and Evaluation (GRADE) framework, randomized evidence is prioritized because of its strong protection against bias. However, historically, a lack of randomized evidence, gaps in such evidence or suboptimal precision for some outcomes have often been addressed using observational datasets from cohort studies for collaborative IPD meta-analyses (68). Detailed discussion about the possibility and direction of different biases is necessary when such data are used to support guideline recommendations or when historical nonconcurrent controls are used as a comparator. Recently, more advanced analytical methods have begun to be used in TB research to address confounding and other biases (69). Efforts have also been made to improve the quality and comparability of longitudinal data collection, particularly with a view to increasing the precision of safety analyses. These approaches may result in upgrading the certainty of the evidence, although not to the level of adequately powered RCTs (70).

Modern epidemiological approaches to analysis of observational data include propensity score matching, target trial emulation and causal inference. The aim of propensity score matching is to adjust the composition of prognostic covariates of participants receiving different treatment regimens, to make them as comparable as possible and address confounding by indication. In target trial emulation, the principles of RCT study design are applied to observational data, with an overall goal of estimating the causal effect of a treatment strategy or regimen (68, 71). This framework facilitates the specification of a clear causal question, amplification of key assumptions, and avoidance of biases (e.g. confounding and immortal time bias). A major strength of target trial emulation is that observational analyses grounded in this framework can incorporate other aspects of GEG for trials (e.g. related to estimands, duration of follow-up and specification of the intervention), which are described elsewhere in this document. Robust and comprehensive sensitivity analyses that evaluate the impact of assumptions made in observational analyses also play an important



role in understanding potential sources of bias. Even the most advanced analytical methods applied to high-quality datasets cannot fully account for residual or unmeasured confounding or selection bias. In some instances, it has been possible to compare the findings from observational cohort data with those of RCTs of similar or identical regimens. In other cases, observational data have remained the only source of evidence because of the use of outdated regimens or disruption in balance that prevents the mounting of a trial. In these cases, despite having relatively large datasets, the indirectness and risk of bias have generally precluded a strong WHO recommendation.

[Return to Section 3.1.14 in the main GEG document](#)

A6.1.15 Extrapolation to excluded populations

Extrapolation can have utility in the generation of TPT recommendations in different ways, such as to judge whether a given regimen can be used in people with HIV, children or pregnant women. Additionally, current recommendations limit contact investigation and provision of TPT to people who are exposed to bacteriologically confirmed pulmonary TB because there is limited evidence of the value of this when the index patient has another form of TB. Mathematical modelling can sometimes provide insights into the population-level impacts of an intervention demonstrated to be effective in a clinical trial. 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 or transmission studies, it may provide useful information for a GDG to determine the effect of the intervention on health equity and the potential impact on public health and resources (72, 73). Evidence from modelling studies is usually graded low on the certainty scale when used for evidence-based guidelines.

TPT trials often exclude pregnant women because of concerns about fetal safety and the risk-benefit balance. However, extrapolation of recommendations may be feasible if there are satisfactory preclinical toxicology data and available postmarketing or opportunistic Phase 3 data (74, 75).

A more transparent alternative to simple extrapolation is the use of basket trial designs and Bayesian approaches. These methods allow for the borrowing of information from other groups within the same trial or from external trials to make estimations in subgroups that lack power. The degree of borrowing is determined by expert opinion and external evidence and is presented transparently alongside other analyses (4, 12, 76).

[Return to Section 3.1.15 in the main GEG document](#)

A6.1.16 Pharmacokinetic studies

Trials of novel drugs should be accompanied by pharmacokinetic (PK) studies in target populations where data are lacking, including among specific populations such as older people, children and pregnant women. These studies are crucial for informing dosing recommendations and understanding drug-drug interactions. PK studies typically use either intensive or sparse sampling methods. Intensive sampling enables detailed individual PK profiling but carries a higher burden on participants, especially among children and pregnant women. Sparse sampling, which uses fewer samples per participant combined with population PK modelling, offers an ethically favourable and practical alternative for generating robust dosing data in these groups.



PK studies are considered a requirement to inform dosing for trials involving children. Children achieve lower serum concentrations of TB medicines and eliminate them faster than adults when given the same dosage (mg/kg/day). Therefore, it is important to collect paediatric data early in the development of TPT regimens to inform dosing. Separate studies to determine safety (and acceptability) in children may also be needed (Box A6.7).

Box A6.7. Case study 7

Pharmacokinetic (PK) side-studies can add information to trial findings that can make drug dosing more precise. An open-label lead-in PK study was conducted in Cape Town, South Africa, as part of the Tuberculosis Child Multidrug-Resistant Preventive Therapy (TB-CHAMP) trial of the use of levofloxacin (Lfx) as tuberculosis preventive treatment (TPT) for children exposed to multidrug-resistant and rifampicin-resistant tuberculosis. This substudy specifically characterized the population PK profile, safety and dosing of novel paediatric Lfx dispersible tablets that facilitate administration in children. When given to children, the dispersible 100 mg Lfx tablet was found to have a higher oral bioavailability (and thus drug exposure) than the nondispersible 250 mg adult tablet that was used in TB-CHAMP (77).

The Assessment of the Safety, Tolerability, and Effectiveness of Rifapentine Given Daily for LTBI (ASTERoID) study, which is evaluating 6 weeks of daily rifapentine alone as TPT, compared with 12–16 weeks of rifampicin (78), and the Bedaquiline Roll-out Evidence in Contacts and People Living With HIV to Prevent TB (BREACH-TB) study, which will evaluate 1 month of bedaquiline as a “pan-TPT” regimen (79), will both enrol children. They will also be accompanied by PK studies in children to help optimize the paediatric dosage.

PK studies may be considered for evaluating how physiological changes in pregnancy affect dosing and, consequently, efficacy and safety, along with human lactation studies. For example, the Phase 1/2 IMPAACT 2001 trial specifically evaluated the PK parameters, safety and tolerability of 3 months of weekly rifapentine plus isoniazid in pregnant women (74). This study found that no dose adjustment was needed for rifapentine from the second to the third trimester.

WHO usually engages external experts to assess the findings from PK reviews (see Section 4.5) (80). On the strength of this evaluation, WHO updates its operational guidance with tabular drug dosage schedules (e.g. see Table 4 in the *WHO operational handbook on tuberculosis* (13)), as well as providing interactive tools to find the right dose by indication and body weight (81).

[Return to Section 3.1.16 in the main GEG document](#)



A6.2 Guidance on generating evidence on values, resources required, cost-effectiveness, equity, acceptability and feasibility

A 6.2.1 Values

The value criterion in the GRADE framework refers to systematically capturing and integrating the preferences of people receiving TPT (e.g. in relation to pill burden, dosing frequency, regimen duration and AEs), care providers and the community into the recommendations (25). For trial designers, this means prioritizing stakeholder engagement using structured interviews, focus group discussions or discrete choice experiments to determine which outcomes (e.g. efficacy, AEs, pill burden and quality of life) matter most to those receiving TPT (82). Clinical trials should incorporate PROs and satisfaction measures, and use composite endpoints reflecting these priorities, in addition to biomedical efficacy and safety information.

Various types of studies can be employed to inform the values criterion, including qualitative studies and patient surveys (83–85). Although qualitative evidence does not produce findings that can be statistically generalized to a particular geographical location or population, it does provide analytical generalizability in terms of understanding the range of possible aspects that can explain a phenomenon of interest – in this case, the range of possible value considerations with regard to the intervention. The utility of qualitative research is discussed further in Section 3.3 and Annex 8.

Transparent documentation of how value judgements influence trial endpoints, analysis and guideline recommendations ensures that research is both scientifically rigorous and aligned with what matters to patients and health systems, strengthening the real-world impact and uptake of TPT interventions (86).

The strength of a recommendation can be influenced by uncertainties about how those affected by a given intervention value its outcomes. GDGs are often faced with a situation where study evidence and qualitative research on values are limited or lacking for a given question. In such cases, this will be registered in the EtD framework and reflected as a research gap to help direct future priorities for studies.

[Return to Section 3.2.1 in the main GEG document](#)

A6.2.2 Resources required and certainty of evidence of required resources

Understanding resource requirements is essential for translating trial results into practice. It helps predict the feasibility of implementing TPT regimens in various health care settings and allows NTPs to better prepare for the adoption of new TPT regimens. The higher the costs of an intervention, and the more resources it consumes, the less likely it is that a strong recommendation is warranted.

Where feasible, data on costs to the health system and participants should be collected in both the intervention and comparator arms during trials (87). These costs may include:

- direct medical costs (e.g. drugs, diagnostics, clinical visits, staff time, service delivery platforms, infrastructure changes);



- nonmedical costs (e.g. transport, food, accommodation, logistics, training);
- indirect costs (e.g. loss of income due to illness, time off work or school); and
- intangible costs (e.g. impacts on quality of life).

Collaboration with health economists and system implementers from the beginning of the trial ensures resource data collection is targeted, complete and relevant to policy. Internationally recognized tools, such as the Consolidated Health Economic Evaluation Reporting Standards, may be used to improve the transparency, completeness and consistency of health economic evaluations (88).

Costs often vary substantially between countries; therefore, for multicountry studies, costing should be carried out across multiple settings for broad representativeness (89). 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 may be required, or visits may be integrated with other care (e.g. in HIV clinics). 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 cost of the investigational product, should be presented where possible. Costs to the health system should be presented for a range of settings. Health system costs that may differ between the treatment arms include not only the cost of drugs, but also other aspects of management (e.g. adherence support, outpatient visits, routine monitoring) (90). Nonmedical and indirect costs are borne by the person receiving TPT. Out-of-pocket costs, loss of gainful employment and other indirect costs represent a large proportion of total treatment costs but are often excluded from both data collection exercises on cost and TB cost-effectiveness analyses (CEAs). These differences may explain variation in cost estimates between studies. It is also helpful to obtain data on issues that have a longer-term impact, such as indebtedness, to give GDGs a more factual picture of the impact of care.

The certainty of evidence concerning resources typically depends on how well the costs and resource requirements are quantified and whether these estimates are based on solid, reliable evidence. This may also depend on how many settings were assessed when data were gathered (i.e. whether the evidence captures the full diversity of modes of delivery of a TPT intervention and the degree of integration with other health services that a person is frequenting). To grade the certainty of evidence regarding required resources using the GRADE approach, the same domains that are used to determine the certainty of evidence of health effects are used; namely, risk of bias, inconsistency, imprecision, indirectness and dissemination bias. The certainty of evidence for required resources is best presented in accordance with the GRADE EtD framework (see Annex 3), highlighting limitations and variability in resource estimates.

[Return to Section 3.2.2 in the main GEG document](#)

A6.2.3 Cost-effectiveness

CEA provides valuable insights for policy-makers and helps determine which TPT regimens offer the best value for money, guiding decisions about which treatments to implement or prioritize. This information is crucial for making evidence-based decisions in resource-limited settings where budget constraints are significant. CEA should use parameters from a range of high- and low-resource settings if the study is conducted in multiple settings, to make the findings more generalizable (Box A6.8). Where possible, researchers should



undertake modelling to determine whether a newer TPT regimen is cost-effective for health systems and patients.

CEA should include both health system and societal perspectives and use an appropriate time horizon that includes both short- and long-term costs and effects, such as TB cases prevented over the years. The CEA should include all costs and benefits, not just those incurred by the health system or programme, but also those that fall on patients, families and the broader community. This avoids a situation where an intervention appears cost-effective or efficient simply because it shifts significant costs (e.g. travel, lost income or nonmedical expenses) from the health care provider to the affected individual or household.

Risk groups, such as household contacts, should be appropriately represented in any modelling to ensure that estimates take into account the increased burden of disease and morbidity or mortality that these individuals face. In the calculation of incremental cost-effectiveness ratios (ICERs), comparisons should be made with the established SOC to ensure policy relevance. As best practice, all analytical inputs, epidemiology, adherence rates, diagnostic accuracy, regimen efficacy and management of AEs need to be specified in calculating and interpreting ICERs and cost-utility measures, such as cost per quality-adjusted life-year or disability-adjusted life-year averted. Sensitivity analyses (one-way, multiway) are valuable to identify variables that most influence results and test the robustness of conclusions under different assumptions.

Box A6.8. Case study 8

A systematic review of studies published between June 2016 and September 2023 identified one high-quality cost-effectiveness study of tuberculosis (TB) preventive treatment (TPT) with fluoroquinolone (FQ) for contacts of people with multidrug-resistant and rifampicin-resistant TB (MDR/RR-TB) (12). This study found that cost-effectiveness was highest when implementing levofloxacin (Lfx)–moxifloxacin (Mfx) for children aged <5 years and children aged <15 years with HIV (incremental cost-effectiveness ratio, US\$ 738 per disability-adjusted life-year), but it averted fewer total deaths and years of life lost than providing Lfx–Mfx to all children aged <15 years (870 versus 1240 deaths averted). The cost-effectiveness of Lfx–Mfx decreased in countries with higher FQ resistance, with a greater number of contacts aged below 15 years needing to be treated per TB episode averted.

This analysis was updated with the efficacy estimates from the Tuberculosis Child Multidrug-Resistant Preventive Therapy (TB-CHAMP) and Levofloxacin for the Prevention of Multidrug-Resistant Tuberculosis in Vietnam (V-QUIN) trials, and results were found to be similar. A substudy of the V-QUIN trial estimated that for every 1000 adult MDR/RR-TB contacts who were given Lfx as TPT, compared with monitoring alone, there would be a total health system cost saving of US\$ 2091, a total health gain of 40.1 quality-adjusted life-years, and 0.56 MDR/RR-TB cases prevented (91). The World Health Organization guideline development group judged that these data favoured the intervention.

Although methods for collecting health system and patient costs are relatively standard, clinical trial data for evaluating cost-effectiveness may be limited in their generalizability to programmatic conditions, particularly in terms of adherence, safety, risk of emergent resistance and capturing longer-term treatment effects. Best practices for trial-based



CEAs include designing trials to collect data on likely programmatic effectiveness (not just clinical efficacy) and measuring costs or use of resources and health-related quality of life as part of the trial. Measures of adherence collected during trials would also be particularly useful in modelling real-world outcomes, as would data from routinely collected programmatic sources (e.g. electronic health records, where available). Data from new technologies, such as digital adherence monitoring, which have been investigated extensively for curative TB treatment and less for TPT and tracking of patient resource use and time, would also be helpful.

The cost-effectiveness profile of TPT can be substantially improved by pairing CEA with transmission modelling. Transmission models extend the traditional trial-based perspective by capturing indirect and population-level effects, such as reduced secondary transmission, over extended time horizons that go beyond the immediate trial population. When resource requirements are assessed in this way, programmes and policy-makers can more accurately estimate the true impact and value of scaling up TPT interventions, justifying investments in settings where long-term benefits outweigh initial costs (92).

[Return to Section 3.2.4 in the main GEG document](#)

A6.2.4 Equity

The impact of interventions on health equity is rarely directly measured in TB trials, although any differences in effectiveness between subgroups of participants and the perspectives of community advisory boards involved in the research may be relevant. Operational research that measures the actual uptake of novel interventions may also be valuable. Qualitative research can provide suitable evidence and can add voices from the community, end users and health care professionals. In the case of a qualitative evidence synthesis (a systematic review of qualitative evidence), equity issues may be identified directly from the findings, which can then be summarized and used to develop a narrative text. When such a synthesis does not explicitly identify equity issues, the evidence reviewers or experts in the field may be able to deduce issues from the findings. However, they must state that the equity issues were hypothesized from the evidence rather than being described explicitly.

Quantitative data about the effect of an intervention on equity could help policy-makers make informed decisions about implementation and direct resource allocation to reach those most in need. Unlike the indices used to express desirable or undesirable effects, or cost-effectiveness, there is as yet no universal quantitative metric in common use to report equity of TB prevention and care in the EtD framework. Work is underway to express health equity through distributional cost-effectiveness analysis (DCEA) (93). DCEA assesses how costs and benefits are distributed across different population groups, such as by age, socioeconomic status or geographical area, highlighting implications for equity and targeting of TPT. The DCEA approach offers a way to explicitly present trade-offs between efficiency and equity, helping GDGs assess whether interventions promote fairness or reduce health disparities beyond any aggregate health gains.

[Return to Section 3.2.5 in the main GEG document](#)

A6.2.5 Acceptability

Acceptability is often neglected in trial design and implementation, even though it is well known that acceptability of a drug and its formulations can directly affect adherence,



intended use, efficacy and safety. Acceptability of TPT incorporates usability, receptivity and integration within the local context (94). It is important that the evaluation of acceptability is planned early and undertaken holistically, including in-depth qualitative work alongside quantitative analysis (95). When developing a TPT intervention, it is critical to think about what is likely to be acceptable to key stakeholders.

Acceptability should ideally be measured directly to provide better-certainty evidence about this important aspect. Methods for doing so include stakeholder surveys (including of patients, providers and national policy-makers) and qualitative research (e.g. interviews and focus groups). Although quantitative studies can provide information on acceptability (e.g. the proportion of a group finding the intervention acceptable), qualitative studies can also provide insights into why a particular intervention may be more or less acceptable and under what circumstances. During protocol development, community advisory boards can provide input into participant perception of the factors that affect acceptability (95, 96).

Comparative evidence regarding the acceptability of TPT regimens can be generated directly within a trial (97), but generating indirect evidence outside the trial context can also be of value (e.g. by studying the acceptability of different real or hypothetical regimen attributes). People affected by TB should be closely consulted, to ensure their views on acceptability are heard.

Acceptability can also be indirectly inferred through observed adherence and discontinuation rates. In the absence of other data, measures of adherence and rates of discontinuation during clinical trials or observational studies are often considered by GDGs as proxy measures of acceptability, as well as being important determinants of treatment outcome. A range of methods are available to quantify adherence, such as patient or supporter diaries, pill counting and digital adherence devices. Each of these measures provides a different level of assurance about successful delivery of doses and differing amounts of detail about levels and patterns of nonadherence. The value of adherence monitoring in trials, however, needs to be balanced against the risk of further inflating efficacy, given that some methods of gathering data on treatment adherence may also improve adherence (Hawthorne effect) (24).

[Return to Section 3.2.6 in the main GEG document](#)

A6.2.6 Feasibility

Evidence on feasibility can provide direct insights into the implementation considerations for an intervention, which are described further in WHO guidelines and operational handbooks.

Whether a GDG will consider an intervention feasible will depend on the complexity of the intervention and the settings in which it has been tested (the clinical trial conditions). The closer the trial conditions are to those of programmes in the settings where the trial was conducted, and the wider the range of those settings, the greater the confidence in the subsequent deployment of that intervention globally. In some cases, additional evidence may already be available from demonstration projects, operational research or concomitant cohort studies conducted alongside trials. Such evidence of feasibility is important to document and include in the EtD framework to guide the GDG's deliberations.

For some interventions, critical global logistical issues – such as limited supply of the new drug or shortage of TBI tests – may represent clear barriers. For other interventions, these



considerations may be more local, such as the regulatory requirements for a new drug or a new indication for a drug. Annex 3 provides an example of how findings from a survey of NTP managers on the feasibility of a new TPT regimen for contacts of people with MDR/RR-TB were summarized in the EtD framework presented to a recent GDG (see also Box 3.1 in Section 3.3). Qualitative evidence can highlight areas where feasibility challenges in different health systems will probably occur or hamper an immediate rollout of a novel TPT regimen (see Annex 8).

Return to Section 3.2.7 in the main GEG document

References for Annex 6

1. Catalogue of Bias Collaboration, Nunan D, Heneghan C. Lack of blinding. In: Catalogue of Bias [website]. University of Oxford; 2018 (<https://catalogofbias.org/biases/lack-of-blinding/>).
2. Mason AJ, Gonzalez-Maffe J, Quinn K, Doyle N, Legg K, Norsworthy P et al. Developing a Bayesian adaptive design for a phase I clinical trial: a case study for a novel HIV treatment. *Stat Med*. 2017;36:754-71 (<https://doi.org/10.1002/sim.7169>).
3. Chang V, Phillips PPJ, Imperial MZ, Nahid P, Savic RM. A comparison of clinical development pathways to advance tuberculosis regimen development. *BMC Infect Dis*. 2022;22:920 (<https://doi.org/10.1186/s12879-022-07846-w>).
4. Report of the technical consultation on innovative clinical trial designs for evaluating new TB preventive treatments, virtual meeting, 15-17 September 2021. Geneva: World Health Organization; 2022 (<https://iris.who.int/handle/10665/353393>). License: CC BY-NC-SA 3.0 IGO.
5. Thorlund K, Haggstrom J, Park JJ, Mills EJ. Key design considerations for adaptive clinical trials: a primer for clinicians. *BMJ*. 2018;360:k698 (<https://doi.org/10.1136/bmj.k698>).
6. Olliaro PL, Vaillant M. Designing noninferiority tuberculosis treatment trials: identifying practical advantages for drug regimens with acceptable effectiveness. *PLoS Med*. 2019;16:e1002850 (<https://doi.org/10.1371/journal.pmed.1002850>).
7. 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;380:1001-11 (<https://doi.org/10.1056/nejmoa1806808>).
8. du Cros P, Greig J, Alffenaar JC, Cross GB, Cousins C, Berry C et al. Standards for clinical trials for treating TB. *Int J Tuberc Lung Dis*. 2023;27:885-98 (<https://doi.org/10.5588/ijtld.23.0341>).
9. Protecting Households On Exposure to Newly Diagnosed Index Multidrug-Resistant Tuberculosis Patients (PHOENIX MDR-TB) [website]. National Library of Medicine; 2025 (<https://clinicaltrials.gov/study/NCT03568383>).
10. Stout JE, Turner NA, Belknap RW, Horsburgh CR, Sterling TR, Phillips PPJ. Optimizing the design of latent tuberculosis treatment trials: insights from mathematical modeling. *Am J Respir Crit Care Med*. 2020;201:598-605 (<https://doi.org/10.1164/rccm.201908-1606oc>).
11. Guidance for best practices for clinical trials. Geneva: World Health Organization; 2024 (<https://iris.who.int/handle/10665/378782>). License: CC BY-NC-SA 3.0 IGO.
12. WHO consolidated guidelines on tuberculosis: module 1: prevention: tuberculosis preventive treatment, 2nd ed. Geneva: World Health Organization; 2024 (<https://iris.who.int/handle/10665/378536>). License: CC BY-NC-SA 3.0 IGO.
13. WHO operational handbook on tuberculosis: module 1: prevention: tuberculosis preventive treatment, 2nd ed. Geneva: World Health Organization; 2024 (<https://iris.who.int/handle/10665/378535>). License: CC BY-NC-SA 3.0 IGO.



14. WHO consolidated guidelines on tuberculosis: module 2: screening: systematic screening for tuberculosis disease. Geneva: World Health Organization; 2021 (<https://iris.who.int/handle/10665/340255>). License: CC BY-NC-SA 3.0 IGO.
15. WHO consolidated guidelines on tuberculosis: module 3: diagnosis: tests for TB infection. Geneva: World Health Organization; 2022 (<https://iris.who.int/handle/10665/362936>). License: CC BY-NC-SA 3.0 IGO.
16. Shorter and Safer Treatment Regimens for Latent TB (SSTARLET) [website]. National Library of Medicine; 2025 (<https://clinicaltrials.gov/study/NCT06498414>).
17. Winters N, Belknap R, Benedetti A, Borisov A, Campbell JR, Chaisson RE et al. Completion, safety, and efficacy of tuberculosis preventive treatment regimens containing rifampicin or rifapentine: an individual patient data network meta-analysis. *Lancet Respir Med*. 2023;11:782-90 ([https://doi.org/10.1016/s2213-2600\(23\)00096-6](https://doi.org/10.1016/s2213-2600(23)00096-6)).
18. Fox GJ, Nguyen CB, Nguyen TA, Tran PT, Marais BJ, Graham SM et al. Levofloxacin versus placebo for the treatment of latent tuberculosis among contacts of patients with multidrug-resistant tuberculosis (the VQUIN MDR trial): a protocol for a randomised controlled trial. *BMJ Open*. 2020;10:e033945 (<https://doi.org/10.1136/bmjopen-2019-033945>).
19. Ntinginya NE, Te Brake L, Sabi I, Chamba N, Kilonzo K, Laizer S et al. Rifapentine and isoniazid for prevention of tuberculosis in people with diabetes (PROTID): protocol for a randomised controlled trial. *Trials*. 2022;23:480 (<https://doi.org/10.1186/s13063-022-06296-8>).
20. 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 [website]. BioMed Central; 2025 (<https://www.isrctn.com/ISRCTN37955921>).
21. Phillips PPJ, Mitnick CD, Neaton JD, Nahid P, Lienhardt C, Nunn AJ. Keeping phase III tuberculosis trials relevant: adapting to a rapidly changing landscape. *PLoS Med*. 2019;16:e1002767 (<https://doi.org/10.1371/journal.pmed.1002767>).
22. Puhan MA, Schunemann HJ, Murad MH, Li T, Brignardello-Petersen R, Singh JA et al. A GRADE Working Group approach for rating the quality of treatment effect estimates from network meta-analysis. *BMJ*. 2014;349:g5630 (<https://doi.org/10.1136/bmj.g5630>).
23. Gupta A, Montepiedra G, Aaron L, Theron G, McCarthy K, Bradford S et al. Isoniazid preventive therapy in HIV-infected pregnant and postpartum women. *N Engl J Med*. 2019;381:1333-46 (<https://doi.org/10.1056/nejmoa1813060>).
24. Vernon A, Fielding K, Savic R, Dodd L, Nahid P. The importance of adherence in tuberculosis treatment clinical trials and its relevance in explanatory and pragmatic trials. *PLoS Med*. 2019;16:e1002884 (<https://doi.org/10.1371/journal.pmed.1002884>).
25. Aschmann HE, Musinguzi A, Kadota JL, Namale C, Kakeeto J, Nakimuli J et al. Preferences of people living with HIV for features of tuberculosis preventive treatment regimens - a discrete choice experiment. *medRxiv* [preprint]. 2023 Sep 14:2023.09.13.23295043 (<https://doi.org/10.1101/2023.09.13.23295043>).
26. Gupta A, Hughes MD, Garcia-Prats AJ, McIntire K, Hesselning AC. Inclusion of key populations in clinical trials of new antituberculosis treatments: current barriers and recommendations for pregnant and lactating women, children, and HIV-infected persons. *PLoS Med*. 2019;16:e1002882 (<https://doi.org/10.1371/journal.pmed.1002882>).
27. ICH harmonised guideline: a selective approach to safety data collection in specific late-stage pre-approval or post-approval clinical trials (E19). International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use; 2022 (https://database.ich.org/sites/default/files/ICH_E19_Guideline_Step4_2022_0826_0.pdf).



28. CTCAE and adverse event reporting [website]. National Cancer Institute; 2025 (<https://dctd.cancer.gov/research/ctep-trials/for-sites/adverse-events>).
29. CIOMS Working Group VI on the Management of Safety Information from Clinical Trials, Council for International Organizations of Medical Sciences. Management of safety information from clinical trials: report of CIOMS Working Group VI. Geneva: Council for International Organizations of Medical Sciences; 2005 (<https://kohahq.searo.who.int/bib/29403>).
30. Division of AIDS (DAIDS) table for grading the severity of adult and pediatric adverse events, version 2.1. Bethesda, Md: Division of AIDS, National Institute of Allergy and Infectious Diseases; 2017 (<https://rsc.niaid.nih.gov/sites/default/files/daidsgradingcorrectedv21.pdf>).
31. Chihota V, Gombe M, Gupta A, Salazar-Austin N, Ryckman T, Hoffmann CJ et al. Tuberculosis preventive treatment in high TB-burden settings: a state-of-the-art review. *Drugs*. 2025;85:127-47 (<https://doi.org/10.1007/s40265-024-02131-3>).
32. Salazar-Austin N, Dowdy DW, Chaisson RE, Golub JE. Seventy years of tuberculosis prevention: efficacy, effectiveness, toxicity, durability, and duration. *Am J Epidemiol*. 2019;188:2078-85 (<https://doi.org/10.1093/aje/kwz172>).
33. Miladi QN, Pahria T, Pramukti I. Effectiveness of digital health interventions to enhance continuity of care in patients with pulmonary tuberculosis: a systematic review of randomized controlled trials. *Patient Prefer Adherence*. 2025;19:1807-23 (<https://doi.org/10.2147/ppa.s533210>).
34. Target product profiles for tuberculosis preventive treatment. Geneva: World Health Organization; 2020 (<https://iris.who.int/handle/10665/333947>). License: CC BY-NC-SA 3.0 IGO.
35. Asare-Baah M, Salmon-Trejo LAT, Venkatappa T, Garfein RS, Aiona K, Haas M et al. Factors associated with the discontinuation of two short-course tuberculosis preventive therapies in programmatic settings in the United States. *Open Forum Infect Dis*. 2024;11:ofae313 (<https://doi.org/10.1093/ofid/ofae313>).
36. Vesga JF, Lienhardt C, Nsengiyumva P, Campbell JR, Oxlade O, den Boon S et al. Prioritising attributes for tuberculosis preventive treatment regimens: a modelling analysis. *BMC Med*. 2022;20:182 (<https://doi.org/10.1186/s12916-022-02378-1>).
37. Martin D, Barbosa P, Ramos JP, Vieira M, Duarte R. People first: a participatory community approach on patient-reported outcomes in tuberculosis. *Breathe (Sheff)*. 2024;20:230138 (<https://doi.org/10.1183/20734735.0138-2023>).
38. Section 7.4. Indirect evidence and impact on patient-important outcomes. In: Schunemann H, Brozek J, Guyatt G, Oxman A, editors. *GRADE handbook for grading quality of evidence and strength of recommendations*. The GRADE Working Group; 2013 (<https://gdt.gradepro.org/app/handbook/handbook.html#h.t09m6hxd5iz>).
39. Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE) [website]. National Cancer Institute; 2025 (<https://healthcaredelivery.cancer.gov/pro-ctcae/>).
40. Kendall EA, Hussain H, Kunkel A, Kubiak RW, Trajman A, Menzies R et al. Isoniazid or rifampicin preventive therapy with and without screening for subclinical TB: a modeling analysis. *BMC Med*. 2021;19:315 (<https://doi.org/10.1186/s12916-021-02189-w>).
41. Zhou G, Luo S, He J, Chen N, Zhang Y, Cai S et al. Effectiveness and safety of tuberculosis preventive treatment for contacts of patients with multidrug-resistant tuberculosis: a systematic review and meta-analysis. *Clin Microbiol Infect*. 2024;30:189-96 (<https://doi.org/10.1016/j.cmi.2023.09.015>).
42. Chiang SS, Brooks MB, Jenkins HE, Rubenstein D, Seddon JA, van de Water BJ et al. Concordance of drug-resistance profiles between persons with drug-resistant tuberculosis and their household contacts: a systematic review and meta-analysis. *Clin Infect Dis*. 2021;73:250-63 (<https://doi.org/10.1093/cid/ciaa613>).



43. Seid G, Alemu A, Diriba G, Zerihun B, Tadesse G, Mariam SH et al. Drug resistance profile of Mycobacterium tuberculosis complex isolated from pulmonary tuberculosis patients and their household contacts in central Ethiopia. BMC Infect Dis. 2025;25:806 (<https://doi.org/10.1186/s12879-025-11220-x>).
44. Patangia DV, Anthony Ryan C, Dempsey E, Paul Ross R, Stanton C. Impact of antibiotics on the human microbiome and consequences for host health. Microbiologyopen. 2022;11:e1260 (<https://doi.org/10.1002/mbo3.1260>).
45. Bratton DJ, Phillips PP, Parmar MK. A multi-arm multi-stage clinical trial design for binary outcomes with application to tuberculosis. BMC Med Res Methodol. 2013;13:139 (<https://doi.org/10.1186/1471-2288-13-139>).
46. Menzies NA, Swartwood N, Testa C, Malyuta Y, Hill AN, Marks SM et al. Time since infection and risks of future disease for individuals with Mycobacterium tuberculosis infection in the United States. Epidemiology. 2021;32:70-8 (<https://doi.org/10.1097/ede.0000000000001271>).
47. Behr MA, Edelstein PH, Ramakrishnan L. Revisiting the timetable of tuberculosis. BMJ. 2018;362:k2738 (<https://doi.org/10.1136/bmj.k2738>).
48. Comstock GW. How much isoniazid is needed for prevention of tuberculosis among immunocompetent adults? Int J Tuberc Lung Dis. 1999;3:847-50 (<https://www.ncbi.nlm.nih.gov/pubmed/10524579>).
49. Tweed CD, Quartagno M, Clements MN, Turner RM, Nunn AJ, Dunn DT et al. Exploring different objectives in non-inferiority trials. BMJ. 2024;385:e078000 (<https://doi.org/10.1136/bmj-2023-078000>).
50. Walker J. Non-inferiority statistics and equivalence studies. BJA Educ. 2019;19:267-71 (<https://doi.org/10.1016/j.bjae.2019.03.004>).
51. Walker TM, Watson JA, Moore DAJ, Frick M, Jamrozik E. Tuberculosis preventive therapy: scientific and ethical considerations for trials of ultra-short regimens. Lancet Infect Dis. 2025;25:e432-e8 ([https://doi.org/10.1016/s1473-3099\(25\)00083-0](https://doi.org/10.1016/s1473-3099(25)00083-0)).
52. Hopewell S, Chan AW, Collins GS, Hrobjartsson A, Moher D, Schulz KF et al. CONSORT 2025 statement: updated guideline for reporting randomised trials. BMJ. 2025;389:e081123 (<https://doi.org/10.1136/bmj-2024-081123>).
53. Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research. E9(R1) statistical principles for clinical trials: addendum: estimands and sensitivity analysis in clinical trials. Rockville: Food and Drug Administration; 2021 (<https://www.fda.gov/regulatory-information/search-fda-guidance-documents/e9r1-statistical-principles-clinical-trials-addendum-estimands-and-sensitivity-analysis-clinical>).
54. Ross JM, Badje A, Rangaka MX, Walker AS, Shapiro AE, Thomas KK et al. Isoniazid preventive therapy plus antiretroviral therapy for the prevention of tuberculosis: a systematic review and meta-analysis of individual participant data. Lancet HIV. 2021;8:e8-e15 ([https://doi.org/10.1016/s2352-3018\(20\)30299-x](https://doi.org/10.1016/s2352-3018(20)30299-x)).
55. 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;391:2304-14 (<https://doi.org/10.1056/nejmoa2314325>).
56. Hesselink 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;391:2315-26 (<https://doi.org/10.1056/nejmoa2314318>).
57. Duong T, Brigden J, Simon Schaaf H, Garden F, Marais BJ, Anh Nguyen T et al. A meta-analysis of levofloxacin for contacts of multidrug-resistant tuberculosis. NEJM Evid. 2025;4:EVIDoa2400190 (<https://doi.org/10.1056/evidoa2400190>).



58. Modi ND, Kichenadasse G, Hoffmann TC, Haseloff M, Logan JM, Veroniki AA et al. A 10-year update to the principles for clinical trial data sharing by pharmaceutical companies: perspectives based on a decade of literature and policies. *BMC Med.* 2023;21:400 (<https://doi.org/10.1186/s12916-023-03113-0>).
59. Kherabi Y, Tunesi S, Kay A, Guglielmetti L. Preventive therapy for contacts of drug-resistant tuberculosis. *Pathogens.* 2022;11:1189 (<https://doi.org/10.3390/pathogens11101189>).
60. Salcher-Konrad M, Nguyen M, Savovic J, Higgins JPT, Naci H. Treatment effects in randomized and nonrandomized studies of pharmacological interventions: a meta-analysis. *JAMA Netw Open.* 2024;7:e2436230 (<https://doi.org/10.1001/jamanetworkopen.2024.36230>).
61. Sterne JAC, Hernán MA, McAleenan A, Reeves BC, Higgins JPT. Chapter 25: Assessing risk of bias in a non-randomized study. In: Higgins JPT, Thomas J, Chandler J, Cumpston M et al., editors. *Cochrane handbook for systematic reviews of interventions version 6.5.* Cochrane; 2024 (<https://www.cochrane.org/authors/handbooks-and-manuals/handbook/current/chapter-25>).
62. Sterne JA, Hernan MA, Reeves BC, Savovic 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 (<https://doi.org/10.1136/bmj.i4919>).
63. 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;365:2155-66 (<https://doi.org/10.1056/nejmoa1104875>).
64. Sterling TR, Scott NA, Miro JM, Calvet G, La Rosa A, Infante R et al. Three months of weekly rifapentine and isoniazid for treatment of *Mycobacterium tuberculosis* infection in HIV-coinfected persons. *AIDS.* 2016;30:1607-15 (<https://doi.org/10.1097/QAD.0000000000001098>).
65. Musinguzi A, Kasidi JR, Kadota JL, Welishe F, Nakitende A, Akello L et al. Evaluating the implementation of weekly rifapentine-isoniazid (3HP) for tuberculosis prevention among people living with HIV in Uganda: a qualitative evaluation of the 3HP Options Trial. *medRxiv [preprint].* 2024 Aug 22:2024.08.19.24308041 (<https://doi.org/10.1101/2024.08.19.24308041>).
66. Kadota JL, Musinguzi A, Aschmann HE, Akello L, Welishe F, Nakimuli J et al. Adverse events reported during weekly isoniazid-rifapentine (3HP) tuberculosis preventive treatment among people with human immunodeficiency virus in Uganda. *Open Forum Infect Dis.* 2024;11:ofae667 (<https://doi.org/10.1093/ofid/ofae667>).
67. Belknap R, Holland D, Feng PJ, Millet JP, Cayla JA, Martinson NA et al. Self-administered versus directly observed once-weekly isoniazid and rifapentine treatment of latent tuberculosis infection: a randomized trial. *Ann Intern Med.* 2017;167:689-97 (<https://doi.org/10.7326/m17-1150>).
68. Fox GJ, Benedetti A, Mitnick CD, Pai M, Menzies D, Collaborative Group for Meta-Analysis of Individual Patient Data in MDR-TB. Propensity score-based approaches to confounding by indication in individual patient data meta-analysis: non-standardized treatment for multidrug resistant tuberculosis. *PLoS One.* 2016;11:e0151724 (<https://doi.org/10.1371/journal.pone.0151724>).
69. Mathur MB, VanderWeele TJ. Methods to address confounding and other biases in meta-analyses: review and recommendations. *Annu Rev Public Health.* 2022;43:19-35 (<https://doi.org/10.1146/annurev-publhealth-051920-114020>).
70. Campbell JR, Falzon D, Mirzayev F, Jaramillo E, Migliori GB, Mitnick CD et al. Improving quality of patient data for treatment of multidrug- or rifampin-resistant tuberculosis. *Emerg Infect Dis.* 2020;26:e190997 (<https://doi.org/10.3201/eid2603.190997>).
71. Fu EL. Target trial emulation to improve causal inference from observational data: what, why, and how? *J Am Soc Nephrol.* 2023;34:1305-14 (<https://doi.org/10.1681/asn.0000000000000152>).
72. Basu S, Andrews J. Complexity in mathematical models of public health policies: a guide for consumers of models. *PLoS Med.* 2013;10:e1001540 (<https://doi.org/10.1371/journal.pmed.1001540>).



73. Sumner T, White RG. The predicted impact of tuberculosis preventive therapy: the importance of disease progression assumptions. *BMC Infect Dis.* 2020;20:880 (<https://doi.org/10.1186/s12879-020-05592-5>).
74. Mathad JS, Savic R, Britto P, Jayachandran P, Wiesner L, Montepiedra G et al. Pharmacokinetics and safety of 3 months of weekly rifapentine and isoniazid for tuberculosis prevention in pregnant women. *Clin Infect Dis.* 2022;74:1604-13 (<https://doi.org/10.1093/cid/ciab665>).
75. Gupta A, Mathad JS, Abdel-Rahman SM, Albano JD, Botgros R, Brown V et al. Toward earlier inclusion of pregnant and postpartum women in tuberculosis drug trials: consensus statements from an international expert panel. *Clin Infect Dis.* 2016;62:761-9 (<https://doi.org/10.1093/cid/civ991>).
76. Turner RM, Turkova A, Moore CL, Bamford A, Archary M, Barlow-Mosha LN et al. Borrowing information across patient subgroups in clinical trials, with application to a paediatric trial. *BMC Med Res Methodol.* 2022;22:49 (<https://doi.org/10.1186/s12874-022-01539-3>).
77. White YN, Solans BP, Denti P, van der Laan LE, Schaaf HS, Vonasek B et al. Pharmacokinetics and optimal dosing of levofloxacin in children for drug-resistant tuberculosis: an individual patient data meta-analysis. *Clin Infect Dis.* 2024;78:756-64 (<https://doi.org/10.1093/cid/ciae024>).
78. Assessment of the Safety, Tolerability, and Effectiveness of Rifapentine Given Daily for LTBI (ASTERoid) [website]. National Library of Medicine; 2024 (<https://clinicaltrials.gov/study/NCT03474029>).
79. Bedaquiline Roll-out Evidence in Contacts and People Living With HIV to Prevent TB (BREACH-TB) [website]. National Library of Medicine; 2025 (<https://clinicaltrials.gov/study/NCT06568484>).
80. WHO operational handbook on tuberculosis: module 1: prevention: tuberculosis preventive treatment: web annex A: pharmacokinetics modelling and simulation studies for dosages of 6Lfx and 3HP, 2nd ed. Geneva: World Health Organization; 2024 (<https://doi.org/10.2471/B09127>). License: CC BY-NC-SA 3.0 IGO.
81. TB Knowledge Sharing: TB Drug Dose Finder [website]. World Health Organization; 2025 (<https://tbksp.who.int/en/tbdrugdosefinder>).
82. Kerkhoff AD, West NS, Del Mar Castro M, Branigan D, Christopher DJ, Denkinger CM et al. Placing the values and preferences of people most affected by TB at the center of screening and testing: an approach for reaching the unreachable. *BMC Glob Public Health.* 2023;1:27 (<https://doi.org/10.1186/s44263-023-00027-0>).
83. Selva A, Sola I, Zhang Y, Pardo-Hernandez H, Haynes RB, Martinez Garcia L et al. Development and use of a content search strategy for retrieving studies on patients' views and preferences. *Health Qual Life Outcomes.* 2017;15:126 (<https://doi.org/10.1186/s12955-017-0698-5>).
84. Foppiano Palacios C, Moll AP, Shrestha R, Kompala T, Davis JL, Charalambous S et al. Evaluating community knowledge of tuberculosis preventive therapy in rural South Africa. *BMC Public Health.* 2025;25:918 (<https://doi.org/10.1186/s12889-025-21719-2>).
85. Gray AT, Surey J, Esmail H, Story A, Harris M. "It's too hard" - the management of latent TB in under-served populations in the UK: a qualitative study. *BMC Health Serv Res.* 2022;22:1464 (<https://doi.org/10.1186/s12913-022-08855-w>).
86. Farrell A, PLOS Medicine Editors. Improving transparency in clinical trial reporting. *PLoS Med.* 2025;22:e1004588 (<https://doi.org/10.1371/journal.pmed.1004588>).
87. Malhotra A, Bedru A, Mulatu F, Nonyane BAS, Cohn S, Mulder C et al. Cost and cost-effectiveness of pediatric home-based versus facility-based TB preventive treatment in Ethiopia (CHIP-TB). *PLOS Glob Public Health.* 2025;5:e0004466 (<https://doi.org/10.1371/journal.pgph.0004466>).



88. Husereau D, Drummond M, Augustovski F, de Bekker-Grob E, Briggs AH, Carswell C et al. Consolidated Health Economic Evaluation Reporting Standards 2022 (CHEERS 2022) statement: updated reporting guidance for health economic evaluations. *Health Policy Open*. 2022;3:100063 (<https://doi.org/10.1016/j.hpopen.2021.100063>).
89. Shin H, Jo Y, Chaisson RE, Turner K, Churchyard G, Dowdy DW. Cost-effectiveness of a 12 country-intervention to scale up short course TB preventive therapy among people living with HIV. *J Int AIDS Soc*. 2020;23:e25629 (<https://doi.org/10.1002/jia2.25629>).
90. Mafirakureva N, Mukherjee S, de Souza M, Kelly-Cirino C, Songane MJP, Cohn J et al. Cost-effectiveness analysis of interventions to improve diagnosis and preventive therapy for paediatric tuberculosis in 9 sub-Saharan African countries: a modelling study. *PLoS Med*. 2023;20:e1004285 (<https://doi.org/10.1371/journal.pmed.1004285>).
91. Hasan T, Thu Anh N, Binh Hoa N, Viet Nhung N, Yapa HM, Graham SM et al. Evaluating the cost-effectiveness of levofloxacin therapy for household contacts of multidrug-resistant tuberculosis in Vietnam. *Lancet Reg Health West Pac*. 2025;61:101666 (<https://doi.org/10.1016/j.lanwpc.2025.101666>).
92. Ryckman T, Weiser J, Gombe M, Turner K, Soni P, Tarlton D et al. Impact and cost-effectiveness of short-course tuberculosis preventive treatment for household contacts and people with HIV in 29 high-incidence countries: a modelling analysis. *Lancet Glob Health*. 2023;11:e1205-e16 ([https://doi.org/10.1016/s2214-109x\(23\)00251-6](https://doi.org/10.1016/s2214-109x(23)00251-6)).
93. Cookson R, Griffin S, Norheim OF, Culyer AJ, Chalkidou K. Distributional cost-effectiveness analysis comes of age. *Value Health*. 2021;24:118-20 (<https://doi.org/10.1016/j.jval.2020.10.001>).
94. 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:e0001267 (<https://doi.org/10.1371/journal.pgph.0001267>).
95. Purchase SE, Wademan DT, Tshetu NL, Rafique M, Hoddinott G, Seddon JA et al. Holistic acceptability of an adult levofloxacin formulation in children and adolescents on a tuberculosis preventive treatment trial. *PLOS Glob Public Health*. 2024;4:e0003381 (<https://doi.org/10.1371/journal.pgph.0003381>).
96. Mwinga A, Moodley K. Engaging with community advisory boards (CABs) in Lusaka Zambia: perspectives from the research team and CAB members. *BMC Med Ethics*. 2015;16:39 (<https://doi.org/10.1186/s12910-015-0031-y>).
97. Prevent TB: Choice Architecture for TPT Delivery [website]. National Library of Medicine; 2025 (<https://clinicaltrials.gov/study/NCT04466488>).



Annex 7. Randomized controlled trials investigating tuberculosis preventive treatment: a systematic review

A7.1 Rationale

Published clinical trials evaluating the effectiveness of tuberculosis (TB) preventive treatment (TPT) report a range of primary and secondary trial outcomes. This systematic review aimed to summarize the outcomes reported by recent Phase 3 and 4 clinical trials of TPT, to identify and quantify the key gaps and inform the discussion on developing this guidance on evidence generation document.

A7.2 Methods

This systematic review was performed in accordance with the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) (1). Electronic searches were conducted in PubMed, Cochrane Library and EMBASE databases for English-language articles published between 1 January 2000 and 30 March 2025.

Inclusion criteria were:

- the study population comprised people taking TPT, regardless of confirmed TB infection or presumed drug resistance of the source case;
- the study design was a randomized clinical trial (Phase 3 or 4), including publications reporting secondary outcomes of clinical trials and published protocols; and
- the study was published in the English language.

Exclusion criteria were:

- trials describing treatment of TB disease only;
- trials where protocols or registries lacked clear definitions of study outcomes;
- Phase 1 or 2 trials; and
- trials of TB vaccines.

A7.3 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 these 46 randomized trials are presented in Table A7.1.



A7.3.1 Populations included

Geographically, 14 trials (30.4%) were conducted in Africa and an equal proportion in Asia, five (10.9%) in Europe, and seven (15.2%) across multiple regions. Regarding study population, 25 studies (54.3%) enrolled people with HIV, with 16 (34.8%) exclusively enrolling this group. Children were included in 12 studies (26.1%), and seven (15.2%) of these focused solely on children. Only one study (2.2%) included pregnant women, and another one (2.2%) reported outcomes in an older population (aged 50–69 years) in rural China. Comorbidities studied included diabetes (eight studies, 17.4%), hepatitis C virus infection (seven studies, 15.2%) and renal transplantation (two studies, 4.3%).

A7.3.2 Regimens used

Of the 46 included studies, 14 (30.4%) evaluated use of isoniazid treatment of different durations. Other interventions included rifampicin and pyrazinamide (three studies, 6.5%), isoniazid and rifampicin (one study, 2.2%), rifampicin alone (seven studies, 15.2%) and various combinations of weekly or twice-weekly isoniazid with rifapentine for 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 people with multidrug-resistant and rifampicin-resistant TB.

A7.3.3 Outcomes assessed

The most frequently assessed primary outcome was the incidence of microbiologically or clinically confirmed TB, reported in 26 studies (56.5%). Overall, 31 studies (69.6%) assessed TB incidence as a primary or secondary outcome. Adverse events (AEs) were the primary outcome in nine 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 Common Terminology Criteria for Adverse Events or Division of AIDS tables, was used in 24 studies (52.2%). AEs of special interest were assessed in 33 studies (71.7%), with hepatotoxicity (29 studies, 63.0%) being the most common. Serious AEs 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 seven studies (15.2%) and was reported as a primary or secondary outcome in 29 studies (63.0%). 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 included patient-reported outcomes. Only four studies (8.7%) investigated the pharmacokinetics of the drugs used.

A7.3.4 Limitations of included studies

The 46 studies included in this systematic review provide valuable insight into the evidence from clinical trials of TPT that is relevant to World Health Organization (WHO) guidelines. However, several methodological limitations have a bearing on the utility of the study findings to future guidelines. Further studies should consider addressing these gaps. First, the studies included had limited reporting on AEs of special interest and severity. Several studies either did not classify AEs by grade (1–5) or did not report on Grade >3 AEs. This makes comparisons between studies difficult. Furthermore, included AEs were often generalized (e.g. “hepatotoxicity”), and there was limited information on non-hepatotoxic AEs. Second, the studies had inconsistent definitions, measurement and reporting of TPT adherence.



Some trials reported the use of pill count or directly observed treatment, and others reported measurement of biomarkers, such as urine metabolites. Third, there was also variability in the definition of treatment completion in terms of the proportion of pills taken, as well as the time taken to complete therapy. Such variations make comparisons difficult, limiting applicability in guideline development and programmatic use.

Among the included studies, there was underrepresentation of specific subpopulations, particularly pregnant women, young children, older people and those with comorbidities. There was significant heterogeneity in the assessment of prevalent and incident TB. The specific diagnostic tests used (phenotypic versus genotypic), follow-up duration and diagnostic algorithms varied between trials. The true incidence of TB across trials during the follow-up period is therefore difficult to ascertain. Data on long-term outcomes beyond 2 years of follow-up were also sparse. Finally, several trials compared newer short-course regimens (e.g. 1 month of daily rifapentine plus isoniazid, 3 months of weekly rifapentine plus isoniazid, and 4 months of daily rifampicin monotherapy) with traditional isoniazid monotherapy (6 or 9 months of daily isoniazid monotherapy), but few head-to-head comparisons between regimens exist.

Overall, the heterogeneity between studies makes direct comparisons and generalizability difficult. Harmonization of study methods can improve the evidence available for the WHO guideline development process and increase the strength of recommendations.



Table A7.1. Demographic features of included TPT trials published between 1 January 2000 and 30 March 2025

Author, year of publication	Setting(s)	Regimens (intervention vs control) ^a	Inclusion criteria	Age range (years)	% children (aged <15 years)	% pregnant women	% people with HIV	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 (2)	Haiti	12H vs placebo	Adults aged ≥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 (3)	Brazil, Haiti, Mexico, USA	2RZ vs 12H	People with HIV aged ≥13 years with positive TST	16–70	0	0	100	NA	1583	71.5	TB incidence	AEs, mortality, probable TB	Mean: 37 months (range: NA)
Johnson et al., 2001 (4)	Uganda	3HR, 3HRZ vs 6H	People with HIV with positive TST	18–50	0	0	100	NA	2736	NA	TB incidence	Mortality, acquired drug resistance	2 years
Quigley et al., 2001 (5)	Zambia	6H, 3RZ vs placebo	People with HIV aged ≥15 years	≥15	0	0	100	NA	1053	NA	TB incidence	Mortality	Mean: 3 years (up to 7 years)
Leung et al., 2003 (6)	Hong Kong	2RZ vs 6H	People with silicosis and positive TST	Mean in 2RZ arm: 61.6	0	0	0	Silicosis	76	98.7	Hepatotoxicity of any grade	NA	Up to 10 years
Agarwal et al., 2004 (7)	India	12H vs no treatment	People with renal transplant for ESRF	15–58	0	0	0	Renal transplant	85	92.6	TB incidence	Hepatotoxicity grade ≥3	30 months
Menzies et al., 2004 (8)	Canada	4R vs 9H	Adults with positive TST	≥18	0	0	0	NA	116	56	Percentage of TPT doses taken	AEs causing discontinuation, treatment costs	Treatment duration
Tortajada et al., 2005 (9)	Spain	2RZ vs 6H	Contacts of confirmed TB cases with positive TST	≥1	17.3 (aged ≤19 years)	0	0	NA	352	50.9	Safety and tolerability	AEs	Treatment duration





Author, year of publication	Setting(s)	Regimens (intervention vs control) ^a	Inclusion criteria	Age range (years)	% children (aged <15 years)	% pregnant women	% people with HIV	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 (10)	India	12H vs no treatment	People aged ≥14 years on renal replacement therapy	16–53 in 12H arm; 14–62 in control arm	0	0	0	ESRF, HBV, HCV	109	84.4	TB incidence	Hepatotoxicity grade ≥3	Up to 3 years
Naqvi et al., 2006 (11)	Pakistan	9H vs no treatment	Renal allograft recipients	NA	0	0	0	Renal transplant, HCV	480	NA	TB incidence	NA	2 years
Mohammed et al., 2007 (12)	South Africa	12H vs placebo	People with HIV with negative TST	≥18	0	0	100	NA	118	49.2	TB incidence	AEs, adherence, death, change in CD4+ count	2 years
Rivero et al., 2007 (13)	Spain	6H, 3HR, 2RZ vs no treatment	People with HIV with negative TST	18–65	0	0	100	NA	319	72	TB incidence	AEs, treatment completion, mortality	2 years
Spyridis et al., 2007 (14)	Greece	3HR, 4HR vs 9H	Children aged ≤15 years with TBI	0–15	100	0	0	NA	926	51.4	Adherence, treatment completion	AEs, TB incidence	3 years
Zar et al., 2007 (15)	South Africa	H vs placebo	Neonates aged ≥8 weeks with HIV who were contacts of TB cases	≥8 weeks	100	0	100	NA	263	56	TB incidence	Grade 3–4 AEs, mortality	Median: 5.7 months (IQR: 2.0–9.7 months)
Menzies et al., 2008 (16)	Brazil, Canada, Saudi Arabia	4R vs 9H	Adults with TBI	≥18	0	0	1.5	NA	847	52.7	AEs causing discontinuation	Treatment completion	Treatment duration
Madhi et al., 2011 (17)	Botswana, South Africa	H vs placebo	HIV-exposed infants aged 91–120 days	91–120 days	100	0	40.4	NA	1352	48	TB incidence	Mortality	96–108 weeks

Author, year of publication	Setting(s)	Regimens (intervention vs control) ^a	Inclusion criteria	Age range (years)	% children (aged <15 years)	% pregnant women	% people with HIV	Included subpopulations	Total sample size (n)	Male sex (%)	Primary outcome measured	Secondary outcomes measured	Duration of follow-up (from treatment initiation)
Martinson et al., 2011 (18)	South Africa	3HP, 3HR, 6 years of H vs 6H	People with HIV with positive TST	IQR: 26.4–34.7	0	0	100	NA	1148	16.7	TB incidence	AEs causing discontinuation, adherence	Mean: 4 years (up to 6 years)
Samandari et al., 2011 (19)	Botswana	36H vs 6H + placebo	People with HIV aged ≥18 years	IQR: 28–39	0	0	100	NA	1995	28	TB incidence	Mortality	36 months
Sterling et al., 2011 (20)	Brazil, Canada, Spain, USA	3HP vs 9H	Close contacts of TB cases aged ≥2 years with positive TST	IQR: 25–46	0	0	2.6	HBV, HCV	7731	54.5	TB incidence	AEs causing discontinuation, Grade 3–5 AEs, mortality, acquired drug resistance, treatment completion	33 months
Chan et al., 2012 (21)	Taiwan	4R vs 6H	HIV-negative prison inmates	≥18	0	0	0	HCV, diabetes	373	100	AEs causing discontinuation	Hepatotoxicity of any grade, other causes of treatment cessation	Treatment duration
Swaminathan et al., 2012 (22)	India	6H vs 36H	People with HIV aged ≥18 years	≥18	0	0	100	NA	712	37	TB incidence	AEs, mortality	36 months
White et al., 2012 (23)	USA	4R vs 9H	Prison inmates with TBI	NA	0	0	0	HCV	364	93	Toxicity, treatment completion	AEs, adherence	Treatment duration
Jimenez et al., 2013 (24)	Spain	3HR vs 6H	Immigrants with TBI	12–40	0	0	0	NA	590	67.8	Adherence, AEs	TB incidence at 5 years	5 years
Gray et al., 2014 (25)	South Africa	H vs placebo	Infants aged ≥8 weeks with HIV	IQR: 17–63 months	100	0	100	NA	167	49.7	TB disease, mortality	AEs, hospital admissions, adherence	34 months





Author, year of publication	Setting(s)	Regimens (intervention vs control) ^a	Inclusion criteria	Age range (years)	% children (aged <15 years)	% pregnant women	% people with HIV	Included subpopulations	Total sample size (n)	Male sex (%)	Primary outcome measured	Secondary outcomes measured	Duration of follow-up (from treatment initiation)
Rangaka et al., 2014 (26)	South Africa	12H vs placebo	People with HIV aged ≥18 years on ART	IQR: 30–40	0	0	100	NA	1329	24.9	TB incidence	AEs, mortality, loss to follow-up	Median: 2.5 years (up to 3.7 years)
Kim et al., 2015 (27)	South Korea	9H vs no treatment	IGRA-positive kidney/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)
TEMPRANO ANRS 12136 Study Group, 2015 (28)	Ivory Coast	6H + ART vs ART	People with HIV aged ≥18 years with CD4+ count <800 cells/mm ³	IQR: 29–42	0	0	100	NA	2056	21.5	Mortality, AIDS-defining illness, non-AIDS-defining cancer or invasive bacterial disease	Grade 3–4 AEs	30 months
Villarino et al., 2015 (29)	Brazil, Canada, Hong Kong, Spain, USA	3HP vs 9H	Children aged 2–17 years with TBI	2–17	25 (aged ≤12 years)	0	0.5	NA	905	50.9	AEs causing discontinuation	AEs of any grade, mortality, TB incidence at 33 months	33 months
Denholm et al., 2017 (30)	Australia	3HP vs 9H	Adults aged ≥18 years with TBI	18–77	0	0	0	NA	80	48.8	Health system costs	Treatment completion	Treatment duration
Diallo et al., 2018 (31)	Australia, Benin, Brazil, Canada, Ghana, Guinea, Indonesia, Saudi Arabia, South Korea	4R vs 9H	Children with TBI	0–17	100 (aged ≤17 years)	0	0	NA	844	49.7	TB incidence	Grade 3–5 AEs, treatment completion	16 months

Author, year of publication	Setting(s)	Regimens (intervention vs control) ^a	Inclusion criteria	Age range (years)	% children (aged <15 years)	% pregnant women	% people with HIV	Included subpopulations	Total sample size (n)	Male sex (%)	Primary outcome measured	Secondary outcomes measured	Duration of follow-up (from treatment initiation)
Gao et al., 2018 (32)	China	3HP, 2H ₂ P ₂ vs no treatment	Rural residents aged 50–69 years	50–69	0	0	0	Diabetes	3738	54.9	TB incidence	Grade 3–5 AEs, treatment completion, mortality, AEs causing discontinuation	2 years
Menzies et al., 2018 (33)	Australia, Benin, Brazil, Canada, Ghana, Guinea, Indonesia, Saudi Arabia, South Korea	4R vs 9H	Adults with TBI	18–90	0	0	4	Immuno-suppression	6012	40.9	TB incidence	Grade 3–5 AEs, treatment completion	28 months
Sun et al., 2018 (34)	Taiwan	3HP vs 9H	Close contacts of TB cases aged ≥12 years with positive TST	≥12	0	0	0	NA	263	57.8	Treatment completion	AEs	2 years
Gupta et al., 2019 (35)	Botswana, Haiti, India, South Africa, Tanzania, Thailand, Uganda, Zimbabwe	6H during pregnancy vs 6H postpartum	Pregnant women with HIV aged ≥18 years	IQR: 24–33	0	100	100	NA	956	0	Pregnancy-related AEs	Grade 3–5 AEs, mortality, TB incidence	48 weeks
Swindells et al., 2019 (36)	Ten countries	1HP vs 9H	People with HIV aged ≥1 year	IQR: 28–43	0	0	100	NA	3000	46	TB incidence, mortality	AEs	Median: 3.3 years (range: NA)





Author, year of publication	Setting(s)	Regimens (intervention vs control) ^a	Inclusion criteria	Age range (years)	% children (aged <15 years)	% pregnant women	% people with HIV	Included subpopulations	Total sample size (n)	Male sex (%)	Primary outcome measured	Secondary outcomes measured	Duration of follow-up (from treatment initiation)
Churchyard et al., 2021 (37)	Ethiopia, Mozambique, South Africa	3HP vs 6H	People with HIV on ART	IQR: 35–49	0.3 (aged ≤18 years)	0	100	NA	4014	30.4 (in 3HP arm)	Treatment completion	Grade 3–5 AEs, TB incidence, mortality, discontinuation	24 months
LaCourse et al., 2021 (38)	Kenya	12H vs no treatment	HIV-exposed infants	6–10 weeks	100	0	0	NA	298	52.7	TB infection incidence	Grade 3–5 AEs, mortality	12 months
Ruan et al., 2021 (39)	China	3HP vs no treatment	HIV-negative adults aged ≥18 years with silicosis	18–65	0	0	0	NA	513	100	TB incidence	AEs, treatment completion, mortality, acquired drug resistance	37 months
Surey et al., 2021 (40)	United Kingdom	3HP vs 3HR	Adults aged 16–65 years with TBI	16–65	0	0	0	Diabetes, immuno-suppression, alcohol use, smoking	52	50	Treatment completion	AEs	Treatment duration
Chaisson et al., 2023 (41)	Uganda	3HP + ART vs ART	ART-naive people with HIV with CD4+ count ≤350 cells/mm ³	IQR: 25–37	0	0	100	NA	453	36.6	CD4+ count before and after treatment	NA	2 years
Tamez-Torres et al., 2023 (42)	Mexico	3R vs 6H	Adults aged ≥18 years with diabetes and positive TST	IQR: 50–62	0	0	0	Diabetes, alcohol use, smoking, renal disease	131	41.2	AEs causing discontinuation	AEs of special interest, tolerability, adherence, treatment completion	Treatment duration
Zhang et al., 2023 (43)	China	6 weeks of H ₂ P ₂	Adults aged 18–75 years with TBI	18–75	0	0	0	HBV, diabetes, alcohol use, smoking	677	58.1	TB incidence	AEs, treatment completion, AEs causing discontinuation, mortality	24 months

Author, year of publication	Setting(s)	Regimens (intervention vs control) ^a	Inclusion criteria	Age range (years)	% children (aged <15 years)	% pregnant women	% people with HIV	Included subpopulations	Total sample size (n)	Male sex (%)	Primary outcome measured	Secondary outcomes measured	Duration of follow-up (from treatment initiation)
Chen et al., 2024 (44)	China	3H ₂ P 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 (45)	Viet Nam	6Lfx vs placebo	Household contacts (any age) of MDR/RR-TB cases	IQR: 28–52	2.90	0	0.4	Diabetes, renal disease, HBV, HCV, HIV, lung disease	2041	36	TB incidence	Grade 3–5 AEs, mortality, acquired drug resistance	30 months
Hesseling et al., 2024 (46)	South Africa	6Lfx vs placebo	Children who were contacts of MDR/RR-TB cases	IQR: 1.3–4.2	100	0	2.1	NA	922	49.2	TB incidence	Death, Grade 3–5 AEs	72 weeks
Huang et al., 2024 (47)	Taiwan	1HP vs 3HP	Close contacts of TB cases aged ≥12 years	≥12	0	0	0	Diabetes, HBV, HCV	490	49.2	Drug reactions	Treatment completion	Treatment duration

AE: adverse event, AIDS: acquired immunodeficiency syndrome; ART: antiretroviral therapy; ESRF: end-stage renal failure; H: isoniazid; HP: isoniazid and rifapentine; H₂P₂: 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; Lfx: levofloxacin; MDR/RR-TB: multidrug-resistant and rifampicin-resistant tuberculosis; NA: not applicable; P: rifapentine; R: rifampicin; TB: tuberculosis; TBI: tuberculosis infection; TPT: tuberculosis preventive treatment; TST: tuberculin skin test; United Kingdom: United Kingdom of Great Britain and Northern Ireland; USA: United States of America; Z: pyrazinamide.

^a Unless otherwise specified, numbers indicate months of treatment (e.g. 12H: 12 months of treatment with isoniazid). Except for H₂P₂ (twice weekly) and 3HP (weekly), all regimens are daily doses.



References for Annex 7

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 (<https://doi.org/10.1136/bmj.n71>).
2. 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;356:1470-4 ([https://doi.org/10.1016/s0140-6736\(00\)02870-1](https://doi.org/10.1016/s0140-6736(00)02870-1)).
3. 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. *JAMA*. 2000;283:1445-50 (<https://doi.org/10.1001/jama.283.11.1445>).
4. Johnson JL, Okwera A, Hom DL, Mayanja H, Mutuluza Kityo C, Nsubuga P et al. Duration of efficacy of treatment of latent tuberculosis infection in HIV-infected adults. *AIDS*. 2001;15:2137-47 (<https://doi.org/10.1097/00002030-200111090-00009>).
5. 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;15:215-22 (<https://doi.org/10.1097/00002030-200101260-00011>).
6. 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;124:2112-8 (<https://doi.org/10.1378/chest.124.6.2112>).
7. Agarwal SK, Gupta S, Dash SC, Bhowmik D, Tiwari SC. Prospective randomised trial of isoniazid prophylaxis in renal transplant recipient. *Int Urol Nephrol*. 2004;36:425-31 (<https://doi.org/10.1007/s11255-004-6251-6>).
8. 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. *Am J Respir Crit Care Med*. 2004;170:445-9 (<https://doi.org/10.1164/rccm.200404-478oc>).
9. 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? *Int J Tuberc Lung Dis*. 2005;9:276-81 (<https://www.ncbi.nlm.nih.gov/pubmed/15786890>).
10. 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. *Transpl Infect Dis*. 2005;7:99-108 (<https://doi.org/10.1111/j.1399-3062.2005.00103.x>).
11. Naqvi R, Akhtar S, Noor H, Saeed T, Bhatti S, Sheikh R et al. Efficacy of isoniazid prophylaxis in renal allograft recipients. *Transplant Proc*. 2006;38:2057-8 (<https://doi.org/10.1016/j.transproceed.2006.06.010>).
12. 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. *Int J Tuberc Lung Dis*. 2007;11:1114-20 (<https://www.ncbi.nlm.nih.gov/pubmed/17945069>).
13. Rivero A, Lopez-Cortes L, Castillo R, Verdejo J, Garcia MA, Martinez-Marcos FJ et al. [Randomized clinical trial investigating three chemoprophylaxis regimens for latent tuberculosis infection in HIV-infected patients]. *Enferm Infecc Microbiol Clin*. 2007;25:305-10 (<https://doi.org/10.1157/13102265>).
14. 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. *Clin Infect Dis*. 2007;45:715-22 (<https://doi.org/10.1086/520983>).



15. 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;334:136 (<https://doi.org/10.1136/bmj.39000.486400.55>).
16. 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. *Ann Intern Med*. 2008;149:689-97 (<https://doi.org/10.7326/0003-4819-149-10-200811180-00003>).
17. Madhi SA, Nachman S, Violari A, Kim S, Cotton MF, Bobat R et al. Primary isoniazid prophylaxis against tuberculosis in HIV-exposed children. *N Engl J Med*. 2011;365:21-31 (<https://doi.org/10.1056/nejmoa1011214>).
18. Martinson NA, Barnes GL, Moulton LH, Msandiwa R, Hausler H, Ram M et al. New regimens to prevent tuberculosis in adults with HIV infection. *N Engl J Med*. 2011;365:11-20 (<https://doi.org/10.1056/nejmoa1005136>).
19. 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;377:1588-98 ([https://doi.org/10.1016/s0140-6736\(11\)60204-3](https://doi.org/10.1016/s0140-6736(11)60204-3)).
20. 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;365:2155-66 (<https://doi.org/10.1056/nejmoa1104875>).
21. 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. *Int J Tuberc Lung Dis*. 2012;16:633-8 (<https://doi.org/10.5588/ijtld.11.0504>).
22. 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:e47400 (<https://doi.org/10.1371/journal.pone.0047400>).
23. White MC, Tulskey 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;18:131-42 (<https://doi.org/10.1177/1078345811435973>).
24. 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. *Int J Tuberc Lung Dis*. 2013;17:326-32 (<https://doi.org/10.5588/ijtld.12.0510>).
25. 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. *Int J Tuberc Lung Dis*. 2014;18:322-7 (<https://doi.org/10.5588/ijtld.13.0354>).
26. Rangaka MX, Wilkinson RJ, Boule 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;384:682-90 (<https://www.ncbi.nlm.nih.gov/pubmed/24835842>).
27. 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-gamma-releasing assay: an exploratory randomized controlled trial. *J Antimicrob Chemother*. 2015;70:1567-72 (<https://doi.org/10.1093/jac/dku562>).
28. TEMPRANO ANRS 12136 Study Group, Danel C, Moh R, Gabillard D, Badje A, Le Carrou J et al. A trial of early antiretrovirals and isoniazid preventive therapy in Africa. *N Engl J Med*. 2015;373:808-22 (<https://doi.org/10.1056/nejmoa1507198>).



29. 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 Pediatr.* 2015;169:247-55 (<https://doi.org/10.1001/jamapediatrics.2014.3158>).
30. 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;47:1433-6 (<https://doi.org/10.1111/imj.13601>).
31. 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;379:454-63 (<https://doi.org/10.1056/nejmoa1714284>).
32. 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. *Eur Respir J.* 2018;52:1801470 (<https://doi.org/10.1183/13993003.01470-2018>).
33. 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;379:440-53 (<https://doi.org/10.1056/nejmoa1714283>).
34. 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 (Edinb).* 2018;111:121-6 (<https://doi.org/10.1016/j.tube.2018.05.013>).
35. Gupta A, Montepiedra G, Aaron L, Theron G, McCarthy K, Bradford S et al. Isoniazid preventive therapy in HIV-infected pregnant and postpartum women. *N Engl J Med.* 2019;381:1333-46 (<https://doi.org/10.1056/nejmoa1813060>).
36. 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;380:1001-11 (<https://doi.org/10.1056/nejmoa1806808>).
37. Churchyard G, Cardenas V, Chihota V, Mngadi K, Sebe M, Brumskine W et al. Annual tuberculosis preventive therapy for persons with HIV infection: a randomized trial. *Ann Intern Med.* 2021;174:1367-76 (<https://doi.org/10.7326/m20-7577>).
38. 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. *Clin Infect Dis.* 2021;73:e337-e44 (<https://doi.org/10.1093/cid/ciaa827>).
39. 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. *Clin Microbiol Infect.* 2021;27:576-82 (<https://doi.org/10.1016/j.cmi.2020.06.008>).
40. 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 Infect Dis.* 2021;21:90 (<https://doi.org/10.1186/s12879-021-05766-9>).
41. 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;37:1097-101 (<https://doi.org/10.1097/qad.0000000000003508>).
42. Tamez-Torres KM, Mongua-Rodriguez N, Ferreyra-Reyes L, Torres-Gonzalez P, Delgado-Sanchez G, Martinez-Hernandez 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;11:1917 (<https://doi.org/10.3390/microorganisms11081917>).



43. 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:e2169195 (<https://doi.org/10.1080/22221751.2023.2169195>).
44. 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. *Int J Infect Dis.* 2024;147:107196 (<https://doi.org/10.1016/j.ijid.2024.107196>).
45. 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;391:2304-14 (<https://doi.org/10.1056/nejmoa2314325>).
46. Hesselning 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;391:2315-26 (<https://doi.org/10.1056/nejmoa2314318>).
47. 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. *Clin Microbiol Infect.* 2024;30:1410-7 (<https://doi.org/10.1016/j.cmi.2024.06.024>).



Annex 8. The role of qualitative research in supporting WHO tuberculosis preventive treatment policy development: approaches and methods

A8.1 The role of qualitative research in supporting the development of WHO guidelines on tuberculosis preventive treatment

Qualitative research can help to diversify the evidence that is used to make recommendations on health care. The World Health Organization (WHO) uses qualitative research, with its focus on understanding meaning and lived experience, to inform guideline development groups (GDGs) about the views and preferences of diverse stakeholders, such as people receiving the intervention, affected communities, programme staff and health care workers, including nurses, community health workers, laboratory staff and clinicians (1, 2). This evidence particularly contributes to the Grading of Recommendations Assessment, Development and Evaluation (GRADE) evidence to decision (EtD) criteria of values, equity, acceptability and feasibility. In the areas of tuberculosis (TB) and TB preventive treatment (TPT) policy, qualitative research results have been used to inform GDGs on:

- decision-making processes – how choices are made to order and start TB medication;
- communication practices – how providers and patients discuss TB medications, possible side-effects and treatment support;
- managing uncertainty – how different actors deal with unclear diagnostic results or treatment outcomes;
- interpretation – how clinicians, nurses, laboratory staff and patients translate test results into decisions about treatment and care;
- use of interventions – by both affected people and providers, such as tests for TB infection and how they are applied;
- contextual influences – the relationship between people's broader environment and the interventions delivered;
- values, attitudes and lived experiences – how these shape perceptions of health conditions and interventions for both providers and recipients; and
- intervention complexity and impact – how interventions affect different subgroups, and how individual and contextual factors influence results in different settings (3).

For instance, the following findings could inform the EtD criteria of values and acceptability.



- Most people at risk would prioritize a TPT regimen with fewer and less severe side-effects above all else, because it would allow them to continue routine activities.
- Patients value receiving an accurate diagnosis of infection and avoiding delays in testing.

A8.1.1 Tenets of qualitative research

Qualitative research is rooted in naturalistic inquiry; it is based on the framing of phenomena as having social meaning and relevance. Such research can uncover the social dimensions of phenomena that might otherwise be seen as medical in nature, such as disease, death, infection or disability, and is an in-depth study of social phenomena (e.g. health problems, treatments and interventions). Key qualitative data collection methods – for example, interviews, focus group discussions (FGDs), observations and document reviews – help to uncover the meanings that people apply to the relevant social phenomena and how they act on those meanings. This is important because meanings are a gateway to people’s feelings, values, motives and actions. Some typical examples include what people think about an illness (perceptions and attitudes), where those thoughts come from (experiences, assumptions, interactions with others and social norms) and where those thoughts take them (actions, practices and processes).

Qualitative data analysis involves a systematic approach that comprises careful reading of data, coding and reflection (with the research question and theoretical framework in mind); identification of patterns and recurrent themes; and, finally, identification of relationships between themes, concepts and theory generation or adaptation. Data analysis can be:

- theory-based, using sensitizing concepts (deductive);
- empirically driven, with the aim of building theory using, for example, grounded theory (inductive); or
- a combination of inductive and deductive.

In contrast to quantitative methods, the reporting of qualitative results relies less on numerical summarizing of the findings across different study participants and more on narrative description. In quantitative research, methodological quality is ensured through internal or external validity, reliability and objectivity. In qualitative research, the corresponding concepts used to assess rigour are credibility, transferability, dependability and confirmability (4). Key to ensuring rigour in qualitative research is the use of continuous critical reflexivity about field engagement, thick descriptions,² questioning of the researcher’s own assumptions and positionality, interrogation of trends, analysis of outliers or negative cases, and triangulation of different data sources. In addition, qualitative research requires an assessment of the certainty of evidence (using GRADE-CERQual [Confidence in the Evidence from Reviews of Qualitative Research]) (5).

A8.2 Approaches and methods of qualitative and mixed-methods research for generating evidence

This section describes how quantitative and qualitative evidence synthesis (QES), primary qualitative research studies and mixed-methods studies can generate evidence to support WHO policy-making. It also discusses how qualitative research can address the value, equity, acceptability and feasibility criteria of the EtD framework, using TPT-specific examples related

² A thick description of a phenomenon is one that does not simply list facts or events but instead provides a deeper interpretation of the meanings, contexts and implications behind that phenomenon.



to treatment, diagnosis and screening. The extent that qualitative evidence affects or drives a decision about a specific recommendation will vary across the questions considered by the guideline (1).

Primary qualitative studies involve collection of raw data directly from the source. Such studies include case studies, ethnographies, participatory action research, grounded theory studies, phenomenological studies and discourse analysis. Secondary studies include data that have already been collected and are readily available. Among such studies, QES involves a systematic review of qualitative studies. Mixed-methods research comprises study designs that systematically integrate qualitative and quantitative methods (e.g. techniques, concepts and approaches) in a single study to obtain different perspectives about a research issue. The three core designs for such research are convergent parallel design, explanatory sequential design and exploratory sequential design.

A8.2.1 QES

Within the GRADE framework, a QES typically provides the highest certainty of evidence on qualitative questions, in much the same way that systematic reviews and meta-analyses of quantitative evidence typically provide the best evidence to evaluate the desirable and undesirable health effects of an intervention. A QES can provide evidence that is richer and more robust and practical than that obtained from individual primary qualitative studies, by gathering, analysing and comparing evidence from multiple studies (1).

Findings from a QES may enter a guideline process through syntheses that are:

- already published and that address the guideline questions directly or indirectly; or
- commissioned for a guideline; these may include broad QES covering multiple guideline interventions and “mini-QES” focusing on a specific intervention.

Benefits and drawbacks

A QES is considered the highest quality of evidence. What allows a QES to strengthen evidence is not just greater sample size than a primary study, but identification of overlaps between themes or concepts across different contexts, situations and settings. Combined with the GRADE-CERQual assessment, identification of overlaps increases the confidence that decision-makers can have in a review finding. As with all secondary analyses, a QES is only as good as the published primary studies. If there are no published primary studies on the exact intervention that a GDG assesses, a QES can synthesize evidence that is indirect and focuses on a comparable context or intervention.

In the case of qualitative research, some of the findings might go beyond a particular technology or intervention (i.e. they might be transferable to different technologies and interventions). A decision-maker can gauge whether it would be better to synthesize qualitative evidence on a comparable intervention or on interventions for different diseases, rather than conducting a primary qualitative study on the specific intervention and disease that the GDG is assessing.

A8.2.2 Primary qualitative research studies

If a QES does not provide sufficient information, or if specific aspects of a novel intervention are not sufficiently covered by prior evidence, primary qualitative research could be undertaken with key stakeholders (e.g. professional users, including health care providers,



laboratory staff, community health workers and programme officers; and end users, including affected communities). This research could use data collection methods such as interviews, FGDs, observations or document review to generate important insights for the GDG. Such studies could either be embedded into a clinical study evaluating the intervention (see the examples in Section A8.4.5) or undertaken as standalone studies; the Social Science & Health Innovation for Tuberculosis (SSHIFTB) website (6) provides an overview of social scientists working on TB.

Qualitative research can be used:

- before, during or after trials (i.e. for formative, process or outcome evaluations) (7);
- to support the design of interventions (e.g. Example 1 in Section A8.4.5 illustrates how interviews with health care providers helped to inform the design of a community-based intervention);
- to support better understanding of the mechanisms and effects of complex health care interventions (e.g. Example 2 in Section A8.4.5 shows how qualitative research helped in understanding practical issues in the administration of 3 months of weekly rifapentine plus isoniazid [3HP] in children);
- to understand how complex health care interventions are experienced by recipients (e.g. Example 3 in Section A8.4.5 describes a family-supported strategy to enhance adherence to the 3HP TPT regimen among people with HIV, showing how affected people and staff experienced the complexities of taking treatment);
- to improve the external validity of trials and facilitate interpretation of findings from trials or other quantitative studies (e.g. Example 4 in Section A8.4.5 describes a qualitative process evaluation that showed why a strong recommendation for an intervention that would provide TPT to household contacts, which was supported by trial evidence of efficacy and safety, did not get implemented as intended); and
- to help trialists to be sensitive to trial participants and save resources by guiding researchers towards interventions that are more likely to be effective in future trials (8).

Benefits and drawbacks

Particularly for novel interventions, qualitative research performed alongside trials has the advantage of generating insights on the intervention in use (albeit use in research and not in routine use); also, standalone qualitative research can gather important insights from stakeholders with direct experience of giving or receiving routine care. Qualitative research can also be used to rank participant preferences. The confidence in the findings from primary qualitative studies is typically not assessed using GRADE-CERQual; instead, it is judged through discussions among the GDG members and authors of the study or the team conducting the independent reviews during the GDG meeting.

A8.2.3 Mixed methods: discrete choice experiments

Survey studies might be helpful in cases where information is already available about the factors influencing feasibility, acceptability, user preferences and experiences in a specific context; in such instances, the aim of survey studies is to identify or rank these factors in a statistically representative sample. A discrete choice experiment (DCE) is a mixed-methods research method, originally developed in economic psychology, that can quantify the strength of user preferences to determine what matters most to users and identify preference heterogeneity among the group (9, 10).



Conducting a DCE involves undertaking formative qualitative research to discover what people think about a product or intervention, then using those insights to design choice scenarios with different intervention attributes. These discrete choices are presented in quantitative surveys that ask participants to choose between two or more alternative products or services with simultaneously varying attribute levels. In choosing, participants need to make trade-offs between preferred and less preferred attribute levels presented in each scenario. Typically, a DCE requires at least 200 participants, with 1000 being better for identifying user heterogeneity; a practical guide for DCE sample size calculations is available (11). Econometric analysis of participant choices allows researchers to retrieve the value (utility) derived from each attribute level relative to the others. Researchers can also test whether preferences vary among participants (e.g. owing to age or sex) (10).

For example, a DCE was carried out among people with HIV receiving care at an urban HIV clinic in Kampala, Uganda. Eligible participants completed a series of nine random choices, each presenting two hypothetical TPT regimens that varied across key attributes (pills per dose, frequency, duration, need for antiretroviral therapy dosage adjustment, and side-effects). After being presented with each pair of regimens (labelled “A” and “B”), participants chose the option they preferred, then compared their choice with the idea of getting no treatment at all. Data were analysed with hierarchical Bayesian estimation, latent class analysis and willingness-to-trade simulations to quantify preferences and assess which treatment features mattered most and whether different groups of people had different preferences. The results showed that participants placed the greatest importance on regimens with fewer pills per dose, less frequent dosing schedules, shorter durations of treatment and no requirement to adjust their antiretroviral therapy. Reducing pills per dose and switching to weekly dosing instead of daily dosing were both highly valued features (12).

Benefits and drawbacks

A DCE could be particularly useful if there might be reporting bias during interviews. This might arise, for instance, when clinicians do not want to reveal what is going on, or when laboratory technicians are worried about losing their job (but would not say so openly) if a particular diagnostic test is introduced.

A DCE assesses the intentions of participants in relation to an intervention versus the actual experience of using that intervention. This can be helpful when decision-makers need to prioritize specific service characteristics (e.g. the job characteristics that influence the choice of health workers for rural postings) and want to understand the acceptability of each characteristic (13). Therefore, qualitative research is likely to be sufficient when the intention is to understand the preferences of different users regarding an intervention, to understand factors that could affect (positively or negatively) participants accessing services or adhering to treatment, or to rank different options. This is because each DCE requires formative qualitative research to inform scenario design (10). DCEs are resource-intensive (requiring one full-time researcher for at least 9–10 months); also, each DCE is highly specific, requiring a focus on one single type of decision-maker in a specific setting (e.g. patients, clinicians or laboratory technicians) for a specific decision.



A8.3 How qualitative research can contribute to GRADE EtD frameworks

Qualitative evidence can be either directly or indirectly relevant to a guideline question. Although EtD criteria – such as values, cost-effectiveness, equity, acceptability and feasibility – are ideally informed by quantitative research, qualitative findings have a crucial role when:

- no relevant quantitative data exist to inform one or more EtD criteria; or
- there is a need to understand underlying phenomena, such as why individuals place a high value on TPT, as opposed to merely determining that TPT is valued and quantifying its importance. For example, qualitative studies can explore patient or caregiver reasoning for prioritizing completion of TPT despite its challenges – a dimension not easily captured by quantitative approaches (14).

Qualitative evidence about interventions or contexts that are related but not identical to the question of interest may be included as indirect evidence. For a QES, the relevance and confidence of such evidence can be formally assessed using GRADE-CERQual (Table A8.1) (15).

Table A8.1. GRADE-CERQual components, definitions and mapping to related terms in GRADE

CERQual component	Definition	GRADE equivalent
Confidence	Graded as high, moderate, low or very low	Certainty
QES	Systematic review of qualitative evidence	Systematic review
Qualitative evidence profile	Provides details of each finding and the CERQual assessment for each of the components of CERQual	GRADE evidence profile
Summary of qualitative findings table	Summary listing of the findings and the overall CERQual assessment	Summary of findings table
Findings		Outcomes
Methodological limitations	This refers to the extent to which there are concerns about the design or conduct of the primary studies that contributed evidence to an individual review finding. Have any methodological limitations been identified that are serious enough to lower confidence in the review finding? A critical appraisal tool for qualitative studies should be applied to all primary studies and used to make this assessment; for example, the CASP checklists (16) or EPPI Centre tools (17).	Risk of bias (assessed using, for example, the Cochrane RoB tool (18), QUADAS (19) or ROBINS-I (20))



Coherence	This is an assessment of how clear and cogent the fit is between the data from the primary studies and a review finding that synthesizes those data. To what extent were the data transformed when summarizing the findings? Are there concerns about how well the findings are grounded in the primary data (e.g. when data contradict the findings or some of the data are ambiguous)?	Inconsistency
Adequacy	This is an assessment of the degree of richness, as well as the quantity, of data supporting a review finding. Review authors need to judge in the context of the specific review finding what constitutes data that are not sufficiently rich (i.e. studies that do not provide sufficient detail to gain an understanding of the phenomenon described in the review finding) or that are too few (i.e. one or very few studies, participants or observations). The idea is not to count studies contributing to a finding but to focus on where data may be thin or limited. A small number of conceptually rich studies contributing to a finding may be more powerful than a larger number of thin, descriptive studies.	Imprecision
Relevance	This refers to the extent to which the body of data from the primary studies supporting a review finding applies to the context specified in the review question. Authors should consider which contextual factors are important to the review question, and identify similarities and differences between the context of the studies supporting a review finding and the context specified in the review question; for example, differences in the focus of the study question, perspective of particular actor groups or populations (or subpopulations), phenomenon of interest or type of intervention, setting, location or time frame.	Indirectness
Dissemination bias	This refers to a systematic distortion of the phenomenon of interest due to selective dissemination of studies or individual study findings.	Publication bias

CASP: Critical Appraisal Skills Programme; CERQual: Confidence in the Evidence from Reviews of Qualitative Research; EPPI: Evidence for Policy & Practice Information; GRADE: Grading of Recommendations Assessment, Development and Evaluation; QES: qualitative evidence synthesis; QUADAS: Quality Assessment of Diagnostic Accuracy Studies; RoB: risk of bias; ROBINS-I: Risk of Bias in Non-randomised Studies – of Interventions.

Although direct qualitative evidence is preferable, indirect evidence can be valuable for guiding decisions on new TPT interventions that have not yet been widely implemented in programmatic settings. For example, a qualitative study among caregivers and health workers in Ethiopia found that family-centred, home-based delivery of TPT for children addressed major barriers, such as time and financial constraints, and improved community acceptability (21). This provided grounds for integrating TPT with other home-based services



to increase uptake and favoured the engagement of the community to help resolve concerns about TPT for children.

A8.4 How to generate high-certainty qualitative evidence to inform global TPT policy decision-making

A8.4.1 Use of theory in qualitative research

Social science theories 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 (22). 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 behaviour or experiences);
- reconnect the data to the research question or spur new questions; and
- improve the transferability or applicability of the data.

However, theories can also be reductive, limiting and normative. Various theoretical frameworks can be used to help guide qualitative evidence generation for the EtD criteria. Greenhalgh and Papoutsi (2019) (23) differentiate between mechanistic (implementation science), ecological (complexity science) and social (social science) approaches to investigating interventions, although there is substantial overlap between these approaches.

Implementation science

Implementation science (mechanistic) takes a phased, structured and somewhat top-down approach to developing, replicating and evaluating an intervention at multiple sites (23). It aims to increase the uptake of evidence-based practices and thereby enhance their public health impact. Implementation science is similar to clinical research because it explores the effectiveness of clinical innovations; however, it is distinct because it includes identifying and addressing barriers and facilitators to the uptake of evidence-based clinical innovations (24). This type of science uses theoretical approaches to provide a better understanding and explanation of how and why implementation succeeds or fails (22).

Complexity science

Complexity science (ecological) encourages a flexible and adaptive approach, acknowledging that change occurs in a complex system involving things, people and processes that are interconnected (23, 25). Few studies use complexity science in TB research, but one example relevant to TPT is an empirical study conducted by Perera et al. (2022) that used system dynamics modelling, a key tool from complexity science, to examine infection prevention in South Africa's TB programme (26). The authors organized participatory workshops with policy-makers, practitioners and patient advocates to develop models representing the complex and interconnected factors influencing the implementation of TPT. Their analysis showed how complexity informed planning – by mapping relationships between policy, organizational contexts, resource allocation and individual behaviours – can enhance decision-making for TPT in dynamic, real-world settings.



Social science

Social science approaches can identify and explain social mechanisms, such as why people behave the way they do. Sociological and anthropological conceptual frameworks used in TB research have been grouped into frameworks that address structure and settings, practices and processes, and experience and enculturation. For example, a study in Papua New Guinea analysed how community beliefs, stigma and support networks shape the uptake of TPT, directly demonstrating the relevance of sociocultural conceptual frameworks in understanding both practices and experiences in TB prevention efforts (27).

A8.4.2 Selection of study design

As with all research, selection of the appropriate design depends on the research question, the state of relevant knowledge (i.e. how much is known about similar interventions and technologies) and logistical considerations (e.g. resources, timelines and capacities).

A8.4.3 Sampling strategy

Clinical trials often recruit large numbers of participants. Qualitative research methods typically work with small sample sizes, using the principle of maximizing the potential for “saturation” (i.e. the point at which new interviews do not meaningfully add to the data already collected). Participants are therefore purposefully sampled based on the unique insights they can provide in answering a research question; this could be based on factors such as sex or gender, age, experienced vulnerabilities, experiences with TPT and professional or social roles. WHO GDGs appreciate insights from multiple stakeholders, such as end users (e.g. recipients of the intervention, affected communities, survivors) and professional users (e.g. providers, implementers, developers, policy-makers), and people from diverse geographical locations. Sampling for different groups of end users and professionals will increase sample size.

A8.4.4 Data collection

In-depth interviews and FGDs allow researchers to explore the factors that shape user preferences, lived experiences and decisions regarding TPT in greater detail. These methods can be used both in formal research studies and in routine programme monitoring and are particularly valuable for uncovering barriers and facilitators that may not emerge from quantitative surveys. For example, interviews with household contacts of people with TB can shed light on why some people complete TPT while others discontinue, providing insights into the impact of pill burden, side-effects, stigma or competing priorities. FGDs with health workers in high TB burden areas can explore implementation challenges, such as drug supply interruptions or clinic workflows, and strategies that have been effective in sustaining TPT delivery (28). They can also explore the perceptions of affected populations and health workers regarding new approaches in TPT delivery, such as the use of long-acting injectables and other technologies.

Observation of TPT consultations, screening encounters or medication dispensing can complement data from interviews and FGDs by highlighting real-world interactions between patients and providers and the practical use of tools such as symptom checklists, digital adherence monitors or blister-pack medications. Although such observations are useful, they are rarely sufficient on their own to understand deeper motivations, attitudes or values.



Topic guides for interviews or FGDs usually start with a set of open-ended questions, followed by targeted prompts to probe more deeply for explanations, examples, contextual factors and outliers. For example, if the opening question is “How feasible is it for you to complete this TPT regimen?”, and the participant responds positively, the interviewer might probe with “What has made it feasible for you?”, encouraging discussion of concrete factors, such as family support, short regimen duration or flexible clinic hours. Similarly, for participants reporting difficulties, follow-up probes can elicit details of specific barriers, such as work conflicts, transport costs or fear of medication side-effects. Topic guides for TPT-related evidence generation, especially for the GRADE criteria (values, acceptability, feasibility, equity and implementation considerations), should be tailored to the population and context. They may include questions on:

- the participant’s current role (e.g. contact of a person with TB, caregiver, community health worker, programme manager);
- step-by-step descriptions of how the participant was screened, counselled and initiated on TPT;
- the participant’s experiences with adherence support (e.g. SMS reminders, pill boxes, home visits);
- interactions between patients and providers, and teamwork among health care workers;
- challenges encountered (e.g. drug shortages, misinformation, stigma) and strategies used to overcome them;
- specific acceptability and feasibility considerations (e.g. preference for shorter TPT regimens);
- access constraints, including travel distance, clinic hours and costs; and
- broader contextual or political factors influencing implementation (e.g. policies on eligibility criteria, integration with HIV services).

In keeping with good qualitative research practice, topic guides should be developed in collaboration with affected communities or individuals who are knowledgeable about the local health and social context, ensuring that language and examples are clearly relevant. Typical durations are 25–60 minutes for one-on-one interviews and 60–90 minutes for FGDs, allowing sufficient time for in-depth discussion and probing.

A8.4.5 Data analysis

Qualitative data analysis is time consuming and requires strong analytical skills. Steps typically include data familiarization, coding (inductive or open, or deductive from a framework), code clustering and theme development, and review and refinement of themes in relation to the framework and entire dataset. Qualitative analysis software (e.g. NVivo or ATLAS.ti) can aid the organization of the analysis process but cannot do the actual analysis. The software can support analysts’ efficiency in storing and retrieving data, coding, editing and revising coding, organizing data and sharing files among researchers.

The remainder of this section provides four examples of the use of qualitative research embedded in clinical trials involving TPT.



Example 1: Qualitative research to help inform the design and delivery of a community-based intervention

The Preventing Childhood Tuberculosis in Lesotho (PREVENT) Study was a mixed-methods, cluster-randomized implementation study that evaluated the effectiveness and acceptability of a community-based intervention (CBI) versus standard of care (SOC) to identify child contacts of adult patients with TB, screen them for TB and provide eligible children with 3HP as TPT (29). At SOC-assigned health facilities, usual care for child contact management (CCM) was delivered; this included contact tracing, screening and provision of TPT. At CBI-assigned health facilities, CBI was delivered; this involved:

- mentoring of health care workers in CCM;
- visits to all households of adult patients with TB;
- referral of all HIV-negative children aged below 5 years and all HIV-positive children to health facilities;
- intensive education on the importance of TB prevention, TPT provision and adherence support; and
- monitoring and review of CCM data.

To inform the CBI design and delivery in the PREVENT Study, Hirsch-Moverman et al. (2021) (29) conducted formative research using interviews with health care providers to explore their attitudes about childhood TB prevention in Lesotho. Using a sequential explanatory design phase, quantitative data collection and analysis were followed by collection and analysis of qualitative data, including semistructured in-depth interviews with 20 purposively selected health care providers (interviewed in six groups) and 28 caregivers. The aim was to explain the quantitative results and to explore the acceptability of the CBI and the use of intervention components.

This study is a good example of how qualitative research can be used to help inform the design and delivery of interventions and to explore the acceptability of the intervention.

Example 2: Exploring the acceptability of a novel fixed-dose combination TPT regimen

A qualitative substudy within an abbreviated child TPT trial tested the acceptability of a dispersible formulation for children receiving the 3HP regimen. Semistructured interviews were conducted with 15 caregivers and 10 children (aged ≥ 6 years) enrolled in the trial. Caregivers appreciated the shorter regimen and the simplified dosing of the fixed-dose combination but reported taste and administration challenges, especially for younger children. These qualitative findings supported the need for palatable formulations and caregiver education about administration techniques and were used to refine programme messaging before broader rollout (30).

Example 3: Using mixed methods to contextualize TPT adherence challenges

The qualitative component of an explanatory mixed-methods study, which also involved a quantitative cohort study, evaluated the acceptability and implementation of a family-supported strategy to enhance adherence to the 3HP TPT regimen among people with HIV in Tanzania (31). Researchers conducted in-depth interviews and FGDs with treatment supporters and community health workers in HIV clinics across six regions. The family-supported strategy combined three elements: family member or close friend support, SMS



reminders for weekly dosing, and targeted health education sessions. The following key findings describe the acceptability of this approach.

- Most participants appreciated the shorter 3HP regimen, felt SMS reminders supported regular dosing, and reported that involving a trusted supporter was a strong motivator for completing the treatment.
- Flexibility in treatment supporter selection improved programme acceptability.
- Financial and logistical barriers affected the attendance of treatment supporters at clinic-based education, suggesting that remote or alternative support may be helpful.
- Mild side-effects were mostly self-limiting, and participants valued counselling and practical support.
- Alignment with work schedules, lifestyle and social support contributed to high rates of TPT completion.

This qualitative study demonstrates how integrating person-centred support mechanisms – combining SMS reminders with family or community-based encouragement – can optimize TPT adherence and completion among people with HIV in routine programmatic settings.

Example 4: Explaining qualitative study results through a qualitative process evaluation

A mixed-methods implementation study in rural Karnataka, India, evaluated the delivery of TPT for household contacts of patients with pulmonary TB (32). Public primary health care clinics identified, screened and initiated eligible contacts on TPT according to national guidelines.

Findings from the quantitative evaluation showed moderate success with TPT initiation and completion rates, but substantial gaps remained, particularly in the early stages of the contact management cascade (identification, testing and initiation). To better understand these outcomes, the researchers conducted a qualitative process evaluation towards the end of the study. The qualitative arm of the study involved interviews with index patients, household contacts and health care workers. Manual thematic analysis was conducted, identifying key enablers, challenges and solutions from multiple stakeholder perspectives.

The evaluation identified several key barriers contributing to implementation gaps:

- documentation issues – contact tracing data were often recorded in paper logbooks and not entered into the national digital system (Nikshay), limiting follow-up and monitoring;
- staff engagement and training gaps – some health care workers were unfamiliar with national TPT guidelines, leading to missed or delayed initiation of TPT for eligible contacts;
- logistical constraints – delays in TB infection testing (with interferon-gamma release assay or chest X-ray) in rural areas slowed TPT initiation and discouraged follow-up; and
- weak adherence monitoring: monthly follow-ups for TPT adherence were inconsistent, with no systematic reminders in place.



References for Annex 8

1. Downe S, Finlayson KW, Lawrie TA, Lewin SA, Glenton C, Rosenbaum S et al. Qualitative evidence synthesis (QES) for guidelines: Paper 1 – using qualitative evidence synthesis to inform guideline scope and develop qualitative findings statements. *Health Res Policy Syst.* 2019;17:76 (<https://doi.org/10.1186/s12961-019-0467-5>).
2. Lewin S, Glenton C, Lawrie TA, Downe S, Finlayson KW, Rosenbaum S et al. Qualitative evidence synthesis (QES) for guidelines: Paper 2 – using qualitative evidence synthesis findings to inform evidence-to-decision frameworks and recommendations. *Health Res Policy Syst.* 2019;17:75 (<https://doi.org/10.1186/s12961-019-0468-4>).
3. Noyes J, Booth A, Cargo M, Flemming K, Harden A, Harris J et al. Chapter 21: Qualitative evidence. In: Higgins JPT, Thomas J, Chandler J, Cumpston M et al., editors. *Cochrane handbook for systematic reviews of interventions* version 6.5. Cochrane; 2024 (<https://www.cochrane.org/authors/handbooks-and-manuals/handbook/current/chapter-21>).
4. Denzin NK, Lincoln YS. *Handbook of qualitative research*. Thousand Oaks: Sage Publications; 1994.
5. GRADE-CERQual [website]. The GRADE Working Group; 2018 (<https://www.cerqual.org/>).
6. SSHIFT [website]. Social Science & Health Innovation for Tuberculosis; 2025 (<https://sshiftb.org/>).
7. Lewin S, Glenton C, Oxman AD. Use of qualitative methods alongside randomised controlled trials of complex healthcare interventions: methodological study. *BMJ.* 2009;339:b3496 (<https://doi.org/10.1136/bmj.b3496>).
8. O’Cathain A, Thomas KJ, Drabble SJ, Rudolph A, Hewison J. What can qualitative research do for randomised controlled trials? A systematic mapping review. *BMJ Open.* 2013;3:e002889 (<https://doi.org/10.1136/bmjopen-2013-002889>).
9. Lancsar E, Louviere J. Conducting discrete choice experiments to inform healthcare decision making: a user’s guide. *Pharmacoeconomics.* 2008;26:661-77 (<https://doi.org/10.2165/00019053-200826080-00004>).
10. Terris-Prestholt F, Neke N, Grund JM, Plotkin M, Kuringe E, Osaki H et al. Using discrete choice experiments to inform the design of complex interventions. *Trials.* 2019;20:157 (<https://doi.org/10.1186/s13063-019-3186-x>).
11. de Bekker-Grob EW, Donkers B, Jonker MF, Stolk EA. Sample size requirements for discrete-choice experiments in healthcare: a practical guide. *Patient.* 2015;8:373-84 (<https://doi.org/10.1007/s40271-015-0118-z>).
12. Aschmann HE, Musinguzi A, Kadota JL, Namale C, Kakeeto J, Nakimuli J et al. Preferences of people living with HIV for features of tuberculosis preventive treatment regimens - a discrete choice experiment. *medRxiv* [preprint]. 2023 Sep 14:2023.09.13.23295043 (<https://doi.org/10.1101/2023.09.13.23295043>).
13. Ryan M, Kolstad JR, Rockers PC, Dolea C. How to conduct a discrete choice experiment for health workforce recruitment and retention in remote and rural areas: a user guide with case studies. Geneva: World Health Organization; 2012 (<https://documents.worldbank.org/en/publication/documents-reports/documentdetail/586321468156869931>).
14. Sebastian MS, Bothamley GH. Tuberculosis preventive therapy: perspective from a multi-ethnic community. *Respir Med.* 2000;94:648-53 (<https://doi.org/10.1053/rmed.1999.0877>).
15. Lewin S, Booth A, Glenton C, Munthe-Kaas H, Rashidian A, Wainwright M et al. Applying GRADE-CERQual to qualitative evidence synthesis findings: introduction to the series. *Implement Sci.* 2018;13:2 (<https://doi.org/10.1186/s13012-017-0688-3>).
16. CASP Checklists [website]. Critical Appraisal Skills Programme (CASP); 2025 (<https://casp-uk.net/casp-tools-checklists>).



17. Evidence for Policy & Practice Information: Tools [website]. EPPI Centre, Social Science Research Unit, UCL Institute of Education: 2019 (<https://eppi.ioe.ac.uk/cms/Resources/Tools/tabid/184/Default.aspx>).
18. Sterne JAC, Savovic J, Page MJ, Elbers RG, Blencowe NS, Boutron I et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ*. 2019;366:l4898 (<https://doi.org/10.1136/bmj.l4898>).
19. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155:529-36 (<https://doi.org/10.7326/0003-4819-155-8-201110180-00009>).
20. Sterne JA, Hernan MA, Reeves BC, Savovic 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 (<https://doi.org/10.1136/bmj.i4919>).
21. Salazar-Austin N, Bergman AJ, Mulder C, Tudor C, Mulatu F, Conradie G et al. Improving access to tuberculosis preventive treatment for children in Ethiopia: designing a home-based contact management intervention for the CHIP-TB trial through formative research. *BMC Health Serv Res*. 2024;24:1043 (<https://doi.org/10.1186/s12913-024-11451-9>).
22. Nilsen P. Making sense of implementation theories, models and frameworks. *Implement Sci*. 2015;10:53 (<https://doi.org/10.1186/s13012-015-0242-0>).
23. Greenhalgh T, Papoutsi C. Spreading and scaling up innovation and improvement. *BMJ*. 2019;365:l2068 (<https://doi.org/10.1136/bmj.l2068>).
24. Bauer MS, Kirchner J. Implementation science: What is it and why should I care? *Psychiatry Res*. 2020;283:112376 (<https://doi.org/10.1016/j.psychres.2019.04.025>).
25. Braithwaite J, Churrua K, Long JC, Ellis LA, Herkes J. When complexity science meets implementation science: a theoretical and empirical analysis of systems change. *BMC Med*. 2018;16:63 (<https://doi.org/10.1186/s12916-018-1057-z>).
26. Perera S, Parkhurst J, Diaconu K, Bozzani F, Vassall A, Grant A et al. Complexity and evidence in health sector decision-making: lessons from tuberculosis infection prevention in South Africa. *Health Policy Plan*. 2022;37:1177-87 (<https://doi.org/10.1093/heapol/czac059>).
27. Jops P, Cowan J, Trumb RN, Kupul M, Kuma A, Bell S et al. The role and value of counsellors in the treatment journeys of people with tuberculosis and their families: qualitative insights from the South Fly District of Papua New Guinea. *PLOS Glob Public Health*. 2024;4:e0002572 (<https://doi.org/10.1371/journal.pgph.0002572>).
28. Lestari T, Graham S, van den Boogard C, Triasih R, Poespoprodjo JR, Ubra RR et al. Bridging the knowledge-practice gap in tuberculosis contact management in a high-burden setting: a mixed-methods protocol for a multicenter health system strengthening study. *Implement Sci*. 2019;14:31 (<https://doi.org/10.1186/s13012-019-0870-x>).
29. Hirsch-Moverman Y, Howard AA, Mantell JE, Lebelo L, Frederix K, Wills A et al. Improving child tuberculosis contact identification and screening in Lesotho: results from a mixed-methods cluster-randomized implementation science study. *PLoS One*. 2021;16:e0248516 (<https://doi.org/10.1371/journal.pone.0248516>).
30. Marthinus AJ, Wademan DT, Saule Z, Hirsch-Moverman Y, Viljoen L, Winckler J et al. Children and providers' perspectives on once-weekly rifapentine and isoniazid TB preventive therapy. *IJTLD Open*. 2025;2:13-8 (<https://doi.org/10.5588/ijtldopen.24.0250>).
31. Pamba D, Sanga E, Mlalama K, Maganga L, Mangu C, Lwilla A et al. Optimizing delivery strategies for 3HP TB preventive treatment in Tanzania: a qualitative study on acceptability of family approach in HIV care and treatment centers. *PLOS Glob Public Health*. 2024;4:e0003142 (<https://doi.org/10.1371/journal.pgph.0003142>).



32. Samudyatha UC, Soundappan K, Ramaswamy G, Mehta K, Kumar C, Jagadeesh M et al. Outcomes and challenges in the programmatic implementation of tuberculosis preventive therapy among household contacts of pulmonary TB patients: a mixed-methods study from a rural district of Karnataka, India. Trop Med Infect Dis. 2023;8:512 (<https://doi.org/10.3390/tropicalmed8120512>).



For further information, please contact:
Department for HIV, Tuberculosis, Hepatitis
and Sexually Transmitted Infections
World Health Organization
20, Avenue Appia CH-1211 Geneva 27 Switzerland
Web site: www.who.int/tb