APPLYING EVIDENCE MAPPING METHODS TO GUIDELINES

APPLYING EVIDENCE MAPPING METHODOLOGIES TO THE WORLD HEALTH ORGANIZATION'S TUBERCULOSIS GUIDELINES

By: ANISA HAJIZADEH, BHSc

A Thesis Submitted to the School of Graduate Studies in Partial Fulfilment of the Requirements for

the Degree Master of Public Health

McMaster University MASTER OF PUBLIC HEALTH (2020) Hamilton, Ontario (Health Research Methods Evidence and Impact- HEI)

TITLE: Applying Evidence Mapping Methodologies to the World Health Organization's Tuberculosis Guidelines

AUTHOR: Anisa Hajizadeh, BHSc (McMaster University)

SUPERVISOR: Professor H.J. Schünemann

PAGES: [xi], [78]

LAY ABSTRACT

The World Health Organization (WHO) issues guidelines to help clinicians, policy-makers, and researchers make informed decisions in their work. Guidelines contain recommendations that can be thought of as bottom-line answers to the questions we ask the scientific literature (based on the evidence available to us today). The WHO's Tuberculosis (TB) Department is partaking in a novel digital reorganization of their guideline recommendations using the evidence-mapping methods proposed in this thesis. This thesis uses the principles of evidence mapping to create recommendation maps that, like any map, chart the landscape in a given domain (in this case, TB recommendations). The recommendation map will help guide the WHO in setting priorities for future research and guideline development.

ABSTRACT

Background: Tuberculosis (TB) is the number one infectious disease killer in the world. TB is both preventable, and curable. Since 1997, the World Health Organization's (WHO) Global TB (GTB) Programme has released evidence-informed publications to guide member states. In their EndTB strategy, the WHO set a mandate to eradicate TB by 2035, in part by intensifying TB research and innovation. As an effort towards this goal, this project applies evidence mapping methodologies to published WHO TB recommendations, in an innovative process called "recommendation mapping" (RM).

Objectives: The prime objective of RM is to allow guideline developers and key stakeholders to identify gaps and clusters of recommendations across publications, serve as an instrumental tool in the sequence of guideline development (from intelligent priority setting, to the assembly of final recommendations) and increase the accessibility of key guideline components. The secondary objective of this work is to poise guideline components for live update and refinement in a rapidly learning health system.

Methods: In this mixed methods study, a methodological framework for mapping guideline components is proposed, with both a quantitative and narrative assessment of raw data and final map outputs. A qualitative analysis from the perspective of key stakeholders, policy-makers, researchers and WHO-GTB liaisons working in guideline development is also included. For the methodological piece, all publications containing WHO TB recommendations were eligible for the mapping exercise. Each recommendation was extracted according to all subdomains of their PICO backbone. Subsections of recommendations are coded using existing ontologies (SNOMED-CT, ATC, ICD-11). A centralized database containing extracted and coded recommendations was then presented in an online and interactive schematic. For the qualitative assessment of palatability of this approach within the organization, semi-structured interviews and a survey was delivered to eligible participants at two Guideline Development Group meetings for WHO tuberculosis treatment and screening guidelines.

Results: The notable result of this work is the development, refinement and application of recommendation mapping methodologies. 20 WHO-GTB guidelines underwent an application of the novel recommendation methodologies proposed in this thesis to create an interactive map, and a searchable database. In-depth interviews and survey results with 21 participants (WHO GTB staff, WHO TB- guideline development group members and technical experts) pointed to concerns in the current accessibility and organization of WHO-GTB guidelines.

Conclusions: Recommendation mapping may have utility in charting the terrain of recommendations, inform priority setting, and provide a scaffold for the future transition to living guidelines.

PREFACE

This thesis has been written as a "sandwich thesis" and consists of an individual manuscript that will be submitted to a journal for publication. The format is as follows:

CHAPTER1: Introduction

CHAPTER 2: Manuscript 1: "Applying Evidence Mapping to Guidelines: Piloting Recommendation Mapping Methods"

CHAPTER 3: Conclusion

At the time of submission, this manuscript has not yet been submitted to a journal for publication.

Acknowledgements

I would like to express my deep gratitude to Dr. Holger Schünemann, my supervisor, for his guidance into the field of clinical and public health guideline development, for the remarkable opportunity to work with the WHO, and for his generosity of time, academic/life expertise and invaluable mentorship. Witnessing the assiduous diligence of a high performer and serviceful clinician-researcher, making a big impact, is truly inspiring. I also extend a sincere appreciation to my supervisory committee, Dr. Robby Nieuwlaat, and Dr. Domink Mertz for their time, guidance and constructive feedback in the planning, development, and completion of this research work. Both brought their respective expertise in epidemiology and infectious disease, to provide critical feedback. Special thanks to our hard-working interdisciplinary team of collaborators; this project wouldn't have been possible without the continued assistance of Dr. Dennis Falzon, Dr. Ernesto Jaramilo at the WHO, the savvy technical support of Bart Dietl, Artur Nowak, Dr. Jan Brozek and the rest of the team at Evidence Prime Inc., our team of data abstractors from HEI, and Dr. Tamara Lotfi, project coordinator for the Quality Assurance project this thesis contributes to. The McMaster Master of Public Health program is unparalleled in facilitating the education of budding public health professionals, and for Dr. Emma Apatu and the rest of the faculty and staff (Angie, Stephanie), my gratitude cannot be left unnoted. I must also thank my family and friends who, while never quite understanding what I was doing exactly, nor ever really comprehending why I was "*still in school*" were tremendously supportive in their own right.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

LIST OF ABBREVIATIONS AND ACRONYMS USED

AMR: Antimicrobial Resistance

TB: Tuberculosis

DR-TB: Drug-Resistant Tuberculosis

MDR-TB: Multidrug Resistant Tuberculosis

XDR-TB: Extensively Drug Resistant Tuberculosis

EBM: Evidence Based Medicine

EIPH: Evidence Informed Public Health

KT: Knowledge Translation

CQI: Continuous Quality Improvement

WHO: World Health Organization

WHO-GTB: World Health Organization Global TB Programme

HEI: Department of Health Research Methods, Evidence and Impact

GRADE: Grading of Recommendations Assessment, Development and Evaluation

GDG: Guideline Development Group

GRC: Guideline Review Committee

NTP: National TB Programme

PICO: Population, Intervention, Comparator, Outcome

RM: Recommendation Mapping

DECLARATION OF ACADEMIC ACHIEVEMENT

I, Anisa Hajizadeh, declare this thesis to be my own work. Part of this work may be submitted for publication later.

To the best of my knowledge, the content of this document does not infringe on anyone's copyright.

My supervisor, Dr. Holger Schünemann, and the members of my supervisory committee, Dr. Robby Nieuwlaat and Dr. Dominik Mertz, have provided guidance and support at all stages of this project. Using their feedback, I drafted protocols, designed study materials, collected data with the assistance of a team of data abstractors, and conducted the final analysis presented. The outputs of the electronic maps are designed in collaboration with Evidence Prime Inc., with input from key stakeholders (mainly liaisons at the WHO).

CHAPTER 1: INTRODUCTION

1.1 Background

1.1.1 Tuberculosis

Tuberculosis (TB) represents a high disease burden as the number one infectious disease in the world today. It is a disease that is both preventable and curable, whose public health picture is eclipsed by the multifaceted dimensions of resistance, comorbidities, barriers to access to effective medication, and the efficacy of screening programs, among other contributions to disease burden. The need for innovative approaches to translating the strides in research to have pragmatic effect at the grassroots, can be bridged by a meticulous focus on guidelines. TB is one of the top 10 causes of death worldwide, with an estimated global incidence of 10.0 million (range; 9.0-11.1) in 2018, which equates to over 27,000 new cases per day (1). Antimicrobial resistance arising from natural Darwinian evolution, and accelerated by human behavior, poses a pressing concern, threatening to undo the hard-won public health gains of the last century, and erode treatment options for infectious diseases like TB. Since the development and introduction of antimicrobial agents like antibiotics in the last century, the microbes they target evolve according to biological evolution, and this process is quickened by the overuse or misuse of these agents within the human sphere. There are three levels of resistance in the context of TB; drug-resistant TB (DR-TB), multidrug resistant TB (MDR-TB) and extensively drug-resistant TB (XDR-TB). What differentiates the three is the degree of bacterial resistance to one of the first-line treatments, both first-line treatments, and additional second line treatments respectively. Those patients with drug-resistant TB cannot expect the relapse-free cure in a 6-month course offered by the first line medication rifampicin (2). People living with rifampicin-resistant TB (RR-TB), are living with a bacterial strain of TB that is resistant to a first line treatment; isoniazid , qualifying them as living with a multi-drug resistant strain of TB

(MDR-TB). In 2018, it was estimated that the incidence of RR-TB was 484,000 people (range;417,000-556,0000-639,000), and of these individuals, 78% had MDR-TB (1). Should an individual living with MDR-TB have additional strains resistant to fluoroquinolones and second-line injectable agents, they would be considered to be living with extensively drug-resistant TB (XDR-TB). Of the 2018 MDR-TB cases, 6.2% (95% CI: 4.4–8.2%) were estimated to have XDR-TB (1).

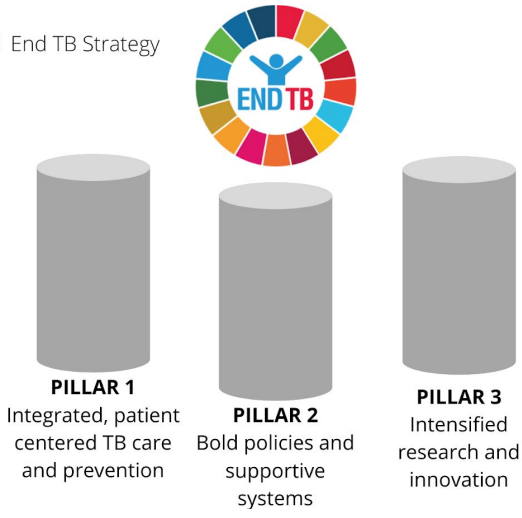
1.1.2 WHO-GTB Guidelines

In 2014, the WHO and its member states unanimously endorsed the WHO's End TB Strategy, resulting in a concerted global commitment to meet the 2035 target to halt the TB epidemic (3). In pragmatically assisting member states to eradicate TB at the grassroots level, the Global TB Programme (GTB) has issued high-level clinical guidelines since 1997 (4). Beginning in 2011, these guidelines have been developed according to the WHO handbook for guideline development, which includes, for relevant aspects of the development process, the GRADE (Grading of Recommendations, Assessment, Development and Evaluation) approach. They have since followed the logical sequence of steps to produce reliable and consistent guidelines based upon measured consideration of the certainty evidence and strength of recommendation (4,5). The resulting high-quality WHO TB treatment guidelines have been updated in 2016 and 2018 by Guideline Development Groups (GDG) following the GRADE method (4,6,29. The GRADE approach, with its transparency, offers unique advantages to quality assurance, user and consumer feedback and implementation at the country level.

Guidelines *should* offer end-users with a summary of the evidence based on systematic reviews or health technology assessments, to assist in meeting the evidence-based standard guiding health systems and care (7). *Evidence-based medicine* (EBM) is the ethos behind moving research into practice, and was defined by Gordon Guyatt, Dave Sackett and colleagues as the

"...the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients" (8). An additional term coined by Guyatt and colleagues, *evidence informed public health* (EIPH), builds upon EBM and articulates "[t]he process of integrating science-based interventions with community preferences to improve the health of populations" (9). The ways in which the evidence that informs health systems, policy, and clinical practice is generated, appraised, synthesized and disseminated, plays a vital role in its application. From a public health perspective, recommendations should be poised to guide clinical practice, policy and local EIPH interventions. As part of the priority setting principles that arose from a 2006 review on the use of evidence by the WHO, trustworthy, evidence-based guidelines must involve a process of priority setting that includes relevant stakeholder groups (10).

Fig. 1 End TB Strategy



Current WHO TB guidelines follow the aforementioned high methodological and evidence-based standards for their development. However, they contain scattered recommendations across dispersed publications. In order to serve their highly necessary purpose, recommendations and other key components contained across these publications, may benefit from evidence-informed organizational methods; which for the purposes of this thesis project, revolve around evidence mapping. This project aims to assess the utility of applying these evidence organization methods to all published WHO TB guidelines and accompanying publications. Further, this project complements all three pillars of the End TB strategy for National TB Programs (NTPs), with notable contributions to Pillar 3; intensified research and innovation through optimizing currently available tools, development and rapid uptake of new

tools, and encouraging discovery to deflect the current trajectory of the TB epidemic (see Fig.1) (11).

1.1.3 A Quality Assurance Project: McMaster x WHO

The McMaster GRADE and Michael G DeGroote Cochrane Canada centre has played a long term role in developing WHO recommendations, particularly within the realm of TB. In 2018, the team at McMaster University proposed a reorganization and altered presentation of WHO-GTB issued guidelines, using evidence-informed methods for the purposes of streamlining the guideline development process, and the increasing accessibility of their publications. I was involved in drafting the original protocols for this larger quality-assurance project, and assumed a core component of this project for my thesis work. From my perspective considering the methods piece, the larger project tackles the following conceptual questions:

1. Can guidelines be deconstructed to their essential building blocks, and reconstructed using novel technologies via evidence-informed organization methods, to shape a more useful, accessible, dynamic, and interactive living product? If so:
    a. What are the essential skeletal components of a guideline?
    b. How can they be identified and charted?
    c. By what measure can increased 'usefulness' and 'accessibility' be assessed?
    d. In what way can we capture engagement with recommendations through an interactive platform? What are the parameters that will qualify the product as 'living'?

This thesis makes a contribution to the following subquestions:

a. Which components of a guideline would benefit from the application of evidence-mapping methodologies?

b. How do you apply evidence-mapping to these components?

c. How can we assess their purported utility?

1.1.4 Review of Evidence Mapping Methodologies

Evidence-mapping is a method to broker a body of knowledge through curating a usable tool in conjugation with stakeholders. It is a means to schematically visualize the breadth and state of the evidence on a given topic. Mapping provides highly credible evidence synthesis for decision makers, improves communications around the most relevant and reliable reviews, and most importantly, supports the strategic synthesis/consolidation of evidence, by highlighting gaps and redundancies. These in turn, inform priority setting (where to go next) based on the charted territory of evidence. In epidemiology and in other fields, innovative mapping techniques have been used to chart the landscape of evidence. In the case of the former, Evidence and Gap Maps (EGMs) provide a visual display of evidence from individual studies and systematic reviews on a thematic area, and are structured around a framework (matrix) of relevant interventions and outcomes (12). In the environmental sphere, Evidence Review Maps are used by decision makers to gauge intervention efficacy and human impacts (13). In the context of guidelines, evidence maps, PICO maps, recommendation (formal and informal) maps, will have utility in gauging which linkages require greater evidentiary support, what key questions have yet to be asked, where is there opportunity for consolidation, and what further research is required. Taken together, this approach aims to lay the foundational organization of WHO TB publications to prime living guidelines. Moreover, the outputs of the mapping work may help inform the future guideline development process, aid in the presentation of pertinent guideline information, and contribute to improved utility of recommendations and other key guideline components.

Evidence maps are a fairly new systematic method used to identify the landscape of evidence in a certain topic area, highlighting both the clusters and gaps in the literature. Evidence

mapping was a term first introduced in 2003 by the *Yale Prevention Research Centre*, and over the last 15 years, had taken various forms in scientific research (18,19). In 2010, 3ie produced their first evidence map, calling it an 'Evidence Gap Map', and have since prepared maps for a variety of topics (12). Evidence mapping can be categorized as a type of scoping study, where relevant literature is mapped in a field of interest (20). According to Mays et al., scoping studies are defined by their "...aim to map rapidly the key concepts underpinning a research area and the main sources and types of evidence available, and can be undertaken as stand-alone projects in their own right, especially where an area is complex or has not been reviewed comprehensively before" (21).

In 2016, Miake-Lye et al., set out to identify published evidence maps in the health sector, and compare/contrast the seemingly heterogenous definitions and approaches to this new methodological tool to review the evidence (22). They conducted a systematic review of 39 publications stated to have used evidence mapping methodology. Out of these, 31 offered elements of a definition; 67% stated that the maps reviewed the evidence to identify gaps and future research needs, 58% noted that the process engaged the target audience, and/or resulted in a user-friendly end product that made the literature accessible, digestible and useable. Other common threads that materialized from their work, were the use of domains to classify the data that comprised the evidence maps (such as PICO), visual representation of the map, the maps presented as online databases, or simply using EGM as a methodology without explicit presentation of a map. The distinguishing features of evidence mapping appeared to be "...involvement of stakeholders early in the research process, the rigor of the search strategy (i.e. all mapping publications describing systematic searches of online databases), and the production of a visual or searchable database, with the stated goal that such products are usable, and beneficial (36). In terms of visualizations of the maps, a "...cross tabular format, categorizing literature according to intervention and/or study design...[were common]...however, the

domains chosen to display and means of presentation will necessarily vary for any particular map according to the aims of the review" (36). In all, the authors proposed that the "...implied definition of what constitutes an evidence map is a systematic search of a broad field to identify gaps in knowledge and/or future research needs that present results in a user-friendly format, often a visual figure or graph, or a searchable database" (36).

In building upon the work of Miake-Lye et al., *The Campbell Collaboration* released a Methods Discussion Paper, in which they highlighted eight components contained in the various definitions available for evidence mapping (19). Evidence mapping definitions tend to include some information on the following eight points; systematic, type of evidence included, content of the map, structure of the map, transparency, visual/graphic display, accompanying description of the map, and intended uses (19). The Campbell Collaboration indicated that the missing element across all definitions is that evidence maps demonstrate the evidence that is there, but not what it says regarding the evidence. They suggest the following, comprehensive definition;

"an evidence and gap map is a systematic visual presentation of the availability of relevant evidence [of effects] for a particular policy domain. The evidence is identified by a search following a pre-specified, published search protocol. The map may be accompanied by a descriptive report to summarize the evidence for stakeholders such as researchers, research commissioners, policy makers, and practitioners. Evidence maps summarize what evidence there is, not what the evidence says" (19).

In terms of where EGMs fit into the evidence synthesis landscape, they can be likened unto the aforementioned scoping review, with a broad scope and narrow information range. In this way, they contrast systematic reviews which tend to provide deep content, with a

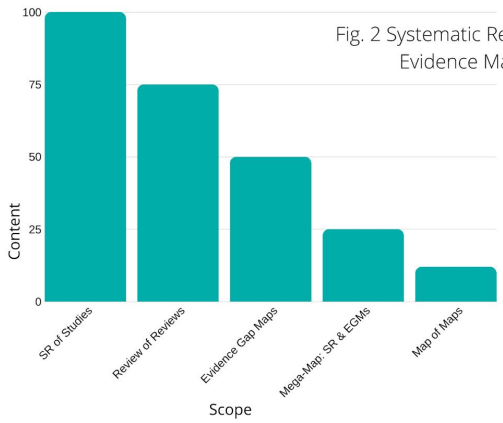


Fig. 2 Systematic Reviews and Evidence Maps

very narrower scope (Fig. 2). Evidence mapping can be thought of as a tool to broker a body of knowledge, and ease its accessibility. In terms of how EGMs compare and contrast other curated evidence platforms, the Campbell Collaboration proposed three differentiations, organized in successively higher levels of evidence brokering (Fig.3).

The first, Evidence Databases, "...contain evidence relating to a specific sector, e.g. ERIC for education, Epistimonikos for health, and the 3ie database for international development. Evidence databases differ from library catalogues and general databases, such as Google Scholar, since they are oriented to a particular audience, and possibly particular types of evidence. Both Epistimonikos and the 3ie database are restricted to systematic reviews and primary studies of effects" (19, 23-26). Next, Evidence Maps, "...classify the evidence which relates to a particular sector or issue, including some reporting on the features (but not the content) of the evidence" (19). Lastly, Evidence Portals, "...present evidence findings in a way in which is intended to be accessible to policy makers and practitioners. Examples are the Teacher and Learning Toolkit of the Education Endowment Foundation, and the IES's What Works Clearing House" (19, 27, 28).

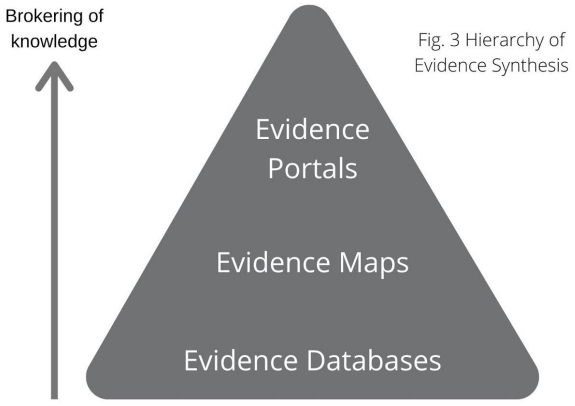


Fig. 3 Hierarchy of Evidence Synthesis

Perhaps the highest level of evidence curation, would be an evidence-based guideline that offers recommendations. The aforementioned three platforms take the database/map user back to the research studies, whereas guidelines offer the user with practical recommendations to inform practice, and decision-making. In this way, it is a preprocessed form of evidence that does not necessarily require the user to consult the literature directly. A potential fourth platform that we propose follows an even higher level of evidence curation/knowledge brokering:

recommendation maps. The following is a potential definition that will be refined throughout the project, and is offered here as a working definition:

Recommendation maps organize the recommendations from a series of guidelines on a certain condition, report features around the evidence that informs the recommendation (breadth and certainty of the evidence), highlights clusters and gaps of recommendations on certain domains (e.g. PICO), and status (e.g. publication date, updated/unchanged). The end product is a platform of mapped recommendations that are accessible to stakeholders (policy-makers, researchers, clinicians, patients and their families).

These evidence curation platforms are not incompatible, and do complement one-another. In the previous example of an evidence database - the Epistemonikos- stands as a prime example of this interaction. It contains over 115,0000 documents; much of which are systematic reviews, others are primary studies and structured summaries of the evidence. The database supports evidence maps which show which primary studies are reported in a given systematic review (24). In a similar fashion, other complementary platforms such as a searchable guideline database can support and house the recommendation maps to offer a comprehensive hub of guidance, with varying degrees of knowledge brokering.

1.1.5 Implementation Research for Public Health

Implementation strategies are defined as "methods or techniques used to enhance adoption, implementation, and sustainability of a clinical program or practice" (14). The literature on implementation strategies (both in terms of identifying, developing and testing) have been convoluted due to lack of conceptual clarity. The Expert Recommendations for Implementing Change (ERIC) study further refines former efforts by Powell et al. to devise an expert approved common nomenclature for implementation strategies (15,16). Implementation outcomes are conceptually and empirically different from service and clinical outcomes (17). Implementation

outcomes precede the latter two and influence their effectiveness (Fig. 4). Select conceptually discrete implementation outcomes presented by Proctor et al., served as endpoints for this study (17). We used the ontology of implementation research to gauge the perceptions, and key implementation and service outcomes of the outputs of this project. Participants included WHO GTB Programme staff, those involved stakeholders involved with WHO-GTB guideline development; WHO-GTB staff, policy-makers, researchers, and clinicians. The findings from this work informally informed the rationale for this thesis.

1.2 Objectives

The overarching objective of this thesis is to use evidence organization tools -mainly a novel process we term 'recommendation mapping'- to deconstruct WHO-GTB issued guidelines by their key building blocks (explicit recommendations), and chart the terrain of recommendations, to streamline the future guideline generation and ease of use. Primary and secondary and tertiary objectives are as follows:

Fig. 4 Implementation Measures

I. The prime objective of RM is to allow guideline developers and key stakeholders to identify gaps and clusters of recommendations across publications, and to provide an instrumental tool in the sequence of guideline development (from intelligent priority setting, to the assembly of final recommendations).

II. Provide the base work for a centralized, searchable database of recommendations (formal and informal), implementation considerations, good-practice statements, and accompanying evidence.

III. To poise guideline components (PICO questions and recommendations) for live update and refinement in a rapidly learning health system.

1.3 Rationale: Considering CQI (Continuous Quality Improvement) at two WHO-GTB GDGs

Apart from the purported justification for such quality-assurance measures given the global burden of TB and informal consultations with the WHO GTB Programme, the true rationale for this project has been largely informed by the results of a qualitative-assessment of the larger WHO TB project this thesis contributes to. In November and December of 2019, WHO GTB members, and WHO TB GDG members, observers and technical experts, were invited to participate in a mixed-methods study assessing the utility of evidence organization methods for WHO TB guidelines and accompanying publications. The study aimed to explore the perceptions of participants on the methods and initial outputs of the project.

1.3.1 Methods

1.3.1.1 Study Design

A mixed-methods study comprised of a survey and semi-structured interviews was designed to assess the quality-assurance project between HEI and the WHO. The study aimed to assess perceptions, implementation, and service outcomes of the quality assurance project. Participants included WHO GTB Programme staff,and those stakeholders involved with WHO-GTB guideline development; WHO-GTB staff, policy-makers, researchers, and clinicians. The strategies that were used in this part of the project included conducting interviews (local consensus discussions), and distributing a survey to tailor strategies.

Interviews served as a means to conduct local consensus discussions with participants. The WHO GTB team, GDG members, and relevant stakeholders were seen as identified champions/early adopters of the proposed outputs of the project, and have unique insight into the nature of the problem, and have a role within the institutions that will be impacted by this

work. The outcomes that served as endpoints were acceptability and appropriateness. of the quality assurance project Acceptability in this context is "...the perception among implementation stakeholders that a given....service, practice or innovation is agreeable, palatable or satisfactory" (17). Appropriateness on the other hand, is the "...perceived fit, relevance, or compatibility of the innovation of evidence-based practice for a given practice setting, provider, or consumer, and/or perceived fit of the innovation to address a particular problem (17). Appropriateness is useful in capturing 'pushback' to implementation, and can be measured using a rating scale. During local consensus discussions, semi-structured interviews were conducted to learn about barriers and facilitators to recommendation consolidation, contextualization, and the potential benefit of the proposed organizational intervention. These interviews followed a predetermined interview guide, were recorded, transcribed and coded, and thematically presented to inform the rationale of this thesis project. Please refer to Appendix A for the interview guide.

A survey was delivered to all attendees of the December2019? GDG meeting on *Molecular assays intended as initial tests for the diagnosis of pulmonary and extrapulmonary TB in adults and children -Policy Update, which was held near the beginning stages of the QA project, and served a convenient time to assess projected outcomes of the project*. On the second day of the four-day meeting, a brief 10 minute presentation on the quality-assurance project was delivered to attendees. A week later, an email with a link to a survey, was forwarded to all GDG members. Please refer to Appendix B for a copy of the survey script.

1.3.1.2 Study Population: Description, Size, Recruitment

Non-probability sampling was used to select members of the WHO TB Programme team who had been identified by a WHO liaison working with HEI on the quality-assurance project. Participation was limited to English speaking individuals. There were six individuals who were identified a-priori, and all were approached via e-mail via targeted- sampling.

Following the educational presentation displaying introducing the project and displaying the pilot work, receivers of said presentation were invited to participate in the survey using convenience-sampling. The survey was sent via email to all GDG members, and 15 responses are anticipated. Participation is limited to English speaking individuals.

1.3.1.3 Data Collection

Participants partook in a semi-structured interview. All interviews were audio-recorded using a password protected device. All interviews were conducted 1-1, with the exception of the first interview that was used for training purposes. Informed consent was obtained prior to all interviews (please see Appendix C for a copy of the consent form). Interviews were either conducted in-person, or by Skype/GoToMeeting.

The survey was delivered using Survey Monkey following the presentation at the December GDG meeting. The survey link was distributed 1 week following the presentation, and participants were given two weeks to complete it. The estimated time to complete the survey was 15 minutes. The recipients of the survey participated on a voluntary basis. The survey used a combination of closed and open ended questions, with branching logic where appropriate.

1.3.1.4 Data Analysis and Handling

Semi-structured interviews were audio-recorded and transcribed. Transcribed interviews were then entered into NVivo and coded. Concepts and themes were identified using a constant comparative method of analysis where new information is compared to previous information. Themes related to the implementation outcomes identified above (such as appropriateness and feasibility) were developed among others.

For the survey, anonymous data from Survey Monkey were entered and downloaded into SPSS by student researchers who were blinded to participant identifiers. Descriptive statistics

with total counts and percentages were accounted for.For example, were individuals from certain member-states more familiar with the guideline development process. Thematic analysis was used for open-ended questions.

All data was kept confidential, and a list of participant IDs and their associated names were kept in separate locations. All electronic information was kept on password protected computers, and only the research team had access to the data.

1.3.2 Results

In total, 21 people participated in this CQI component of the project. From the WHO GTB team, four members completed interviews during the December GDG meeting. An additional 6 individuals (researchers, GDG panelists, policy-makers) were interviewed at either the November and December GDG meetings. The survey was delivered following a brief 10 minute meeting to a GDG in Geneva in December 2019, updating a TB diagnostic guideline. From this meeting there were 11 responses.

The interviews allowed for a candid discussion on the internal workings of the WHO GTB guideline development process, and the survey allowed for participants to answer more pointed questions regarding the outputs of the project. Beginning with the interviews, there were three themes that can be drawn; 1) piece-meal guidelines compromise streamlined development and engagement, 2) guideline literacy may be improved through a centralized platform, 3) high-level guidance devoid of grassroots connection is a fallacy. Beginning with the first theme, multiple interviewees noted that the discrete publication of PDF guidelines is difficult to navigate. One participant involved in a recent curation of guidelines for a consolidated re-publication, admitted that they often had difficulty locating the appropriate sources, evidence profiles/EtDs in their attempt to congregate information. Consolidated guideline work tries to address the existence of multiple sources of similar guidance, but in their aggregation, they face challenges. For the

second theme, participants shared some concern around health literacy, and capacity for guideline-users to assimilate information provided. The GRADE process allows for a logical and transparent documentation of guideline development, but reading an Evidence to Decision (EtD) framework, understanding what the strength and certainty of a recommendation implies, is not always intuitive. When the searchable database was described, participants felt that such a platform with accessible explanation of terminology and process, would be beneficial. The last point was an emphasis on the context of guideline development. Participants stressed that the WHO does not publish guidance un-informed by the realities of the member states where they land. WHO-GTB team members frequently travel to, and converse with those 'on the ground' in NTPs, and in other areas where guidelines are being contextualized.

In terms of the survey responses that relate to the work captured by this thesis, these questions were answered by NTP members, researchers and GDG members. All participants have been working in TB- centered work for over 10 years. There was disparity in how accessible the current presentation of guidelines and recommendations are, and that it took over a minute to locate required information

Fig. 5 Accessibility of Guidelines

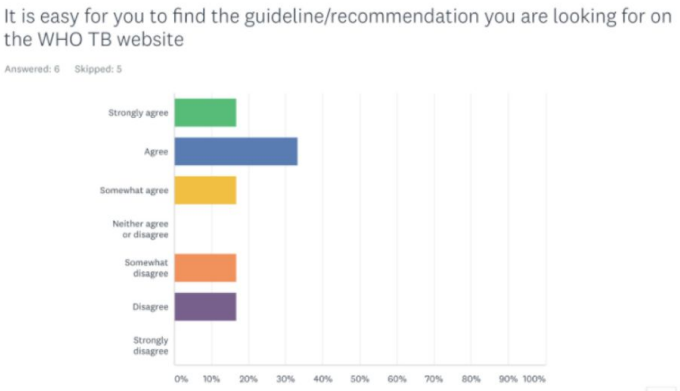
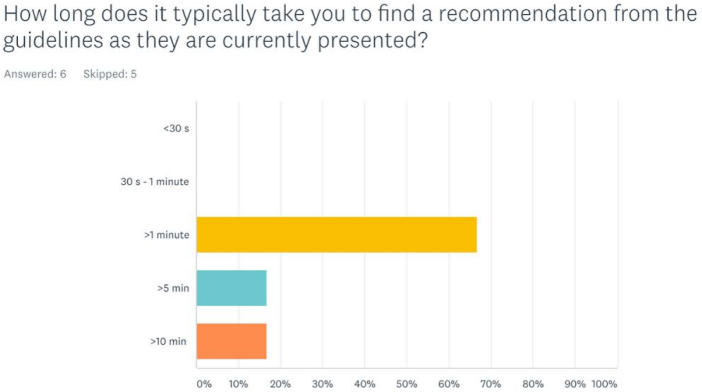


Fig. 6 Time to Access Recommendations



When asked to describe barriers and facilitators to the presentation of guidelines, one participant noted that in the past, "*[t]here were complaints in the recent past (2018-19) that MDR guidelines ended editing differently than decided under GDG, although this has not been my experience. The process has several levels, layers, and is not only dependent on GDG…*". In questions exploring the dissemination of WHO-GTB guidelines, one participant expressed a critique in the overlapping or varied quantity of recommendations, and suggested that recommendations should be able to be transferred to a one-pager that can be easily absorbed and incorporated by programs.

1.3.3 Discussion

In all, there seems to be a need to hone in on the translation between what happens during the GDG, how the final guideline is presented and how accessible guideline components are for developers, and end-users. Further, a centralized platform organizing WHO-GTB guidelines seems highly palatable to the internal organization, and those associated with WHO TB guideline development.

1.4 References

1. Global tuberculosis report [Internet]. World Health Organization. 2019 [cited 26 April 2020]. Available from: https://www.who.int/tb/publications/global_report/en/
2. WHO. Guidelines for treatment of tuberculosis. 4th edn (WHO/HTM/TB/2009.420). 2009. http://whqlibdoc.who. int/publications/2010/9789241547833_eng.pdf Date last accessed: November 7, 2016.
3. WHO End TB Strategy. World Health Organization; 2014. https://www.who.int/tb/post2015_strategy/en/. Accessed April 19, 2019.

4. Falzon D, Schünemann H, Harausz E et al. World Health Organization treatment guidelines for drug-resistant tuberculosis, 2016 update. European Respiratory Journal. 2017;49(3):1602308. doi:10.1183/13993003.02308-2016

5. Guyatt G, Oxman A, Vist G et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ. 2008;336(7650):924-926. doi:10.1136/bmj.39489.470347.a

6. WHO Treatment Guidelines For Multidrug- And Rifampicin-Resistant Tuberculosis 2018 Pre-Final Update. World Health Organization; 2018. https://www.who.int/tb/publications/2018/WHO.2018.MDR-TB.Rx.Guidelines.prefinal.text.pdf. Accessed April 19, 2019.

7. DiCenso A, Bayley L, Haynes R. Accessing pre-appraised evidence: fine-tuning the 5S model into a 6S model. Evidence-Based Nursing. 2009;12(4):99-101.

8. Sackett D, Rosenberg W, Gray J, Haynes R, Richardson W. Evidence based medicine: what it is and what it isn't. BMJ. 1996;312(7023):71-72.

9. Cullum, N., Ciliska, D. K., Marks, S., & Haynes, B. (2008). An Introduction to Evidence-Based Nursing. In: N. Cullum, D. Ciliska, R.B. Haynes, & S. Marks (Eds.), Evidence-Based Nursing: An Introduction (pp. 1-8). Oxford, U.K.: Blackwell

10. Oxman A, Schünemann H, Fretheim A. Improving the use of research evidence in guideline development: Priority setting. Health Research Policy and Systems. 2006;4(1).

11. Implementing The End TB Strategy: The Essentials. Geneva, Switzerland: World Health Organization; 2015. https://www.who.int/tb/publications/2015/The_Essentials_to_End_TB/en/. Accessed May 14, 2019

12. Snilstveit, Birte; Vojtkova, Martina; Bhavsar, Ami; Gaarder, Marie. 2013. Evidence Gap Maps : A Tool for Promoting Evidence-Informed Policy and Prioritizing Future Research. Policy Research Working Paper;No. 6725. World Bank, Washington, DC. © World Bank. https://openknowledge.worldbank.org/handle/10986/16941 License: CC BY 3.0 IGO

13. O'Leary B, Woodcock P, Kaiser M, Pullin A. Evidence maps and evidence gaps: evidence review mapping as a method for collating and appraising evidence reviews to inform research and policy. Environmental Evidence. 2017;6(1). doi:10.1186/s13750-017-0096-9

14. Proctor EK, Powell BJ, McMillen JC. Implementation strategies: recommendations for specifying and reporting. Implement Sci. 2013;8:1–11.

15. Waltz TJ, Powell BJ, Chinman MJ, Smith JL, Matthieu MM, Proctor EK, et al. Expert Recommendations for Implementing Change (ERIC): protocol for a mixed methods study. Implement Sci. 2014;9:1–12.

16. Powell B, Waltz T, Chinman M, Damschroder L, Smith J, Matthieu M et al. A refined compilation of implementation strategies: results from the Expert Recommendations for Implementing Change (ERIC) project. Implementation Science. 2015;10(1).

17. Proctor E, Silmere H, Raghavan R et al. Outcomes for Implementation Research: Conceptual Distinctions, Measurement Challenges, and Research Agenda. *Administration and Policy in Mental Health and Mental Health Services Research*. 2010;38(2):65-76. doi:10.1007/s10488-010-0319-7

18. Katz, D. L., Williams, A. L., Girard, C., & Goodman, J. (2003). The evidence base for complementary and alternative medicine: methods of evidence mapping with application to CAM. *Alternative therapies in health and medicine, 9(4),* 22.

19. Evidence and gap maps: a comparison of different approaches. Oslo, Norway: The Campbell Collaboration. Retrieved from: www.campbellcollaboration.org/
DOI: https://doi.org/10.4073/cmdp.2018.2

20. Arksey, H., & O'Malley, L. (2005). Scoping studies: towards a methodological framework. *International journal of social research methodology, 8(1),* 19-32.

21. Mays, N., Roberts, E., & Popay, J. (2001). Synthesising research evidence. In N. Fulop, P. Allen, A. Clarke, & N. Black (Eds.), *Studying the organisation and delivery of health services: Research methods.* London: Routiedge.

22. Miake-Lye, I. M., Hempel, S., Shanman, R., & Shekelle, P. G. (2016). *What is an evidence map? A systematic review of published evidence maps and their definitions, methods, and products.* Los Angeles, CA, USA: Syst Rev. Retrieved from https://www.ncbi.nlm.nih.gov/m/pubmed/26864942/

23. ERIC Institute of Education Services [Internet]. 2020 [cited 11 March 2020]. Available from: https://eric.ed.gov/

24. Epistemonikos: Database of the best Evidence-Based Health Care. Epistemonikos.org. https://www.epistemonikos.org/. Published 2019. Accessed June 18, 2019.

25. International Initiative for Impact Evaluation [Internet]. 3ieimpact.org. 2020 [cited 11 March 2020]. Available from: https://www.3ieimpact.org/

26. Google Scholar [Internet]. 2020 [cited 11 March 2020]. Available from: https://scholar.google.com/

27. Education Endowment Foundation [Internet]. 2020 [cited 11 March 2020]. Available from: https://educationendowmentfoundation.org.uk/evidence-summaries/teaching-learning-toolkit

28. What Works Clearinghouse: Standards Handbook [Internet]. 4th ed. 2020 [cited 11 March 2020]. Available from: https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_standards_handbook_v4.pdf

29. Schünemann H, Best D, Vist G, Oxman A. Letters, numbers, symbols and words: how to communicate grades of evidence and recommendations. CMAJ [Internet]. 2003 [cited 26 April 2020];:169(7): 677–680. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC202287/

Appendix A: Interview Guide

Thank you for choosing to participate in this interview. At this time, I would like to walk you through the consent form, to inform you of what the goals of this interview are, your rights as a participant, confidentiality, the impact your participation will have, and how you can access the results.

I have roughly 10 questions to ask you in this semi-structured interview. Most of the questions are open-ended, and relevant digressions are welcome. We will use the questions as a guide, but I may ask additional questions not listed here to further the conversation on points of interest.

1. In what capacity have you been involved with WHO TB recommendations in either guidelines or accompanying publications?
2. How long have you been involved with WHO TB recommendation formulation, publication and/or dissemination?
3. In your opinion, what are some barriers to priority setting for WHO TB recommendations?
4. In your opinion, what are some of the redundancies in either WHO TB recommendations across all publications, or in the development process?
5. What are the gaps in WHO TB recommendations across all publications, if there are any?
6. Based on your experience, what are some barriers to consolidating WHO TB recommendations?
7. As of now, what supports do end users (member-states, NTPs, other key stakeholders) have in contextualizing and using WHO TB recommendations in their context (i.e. adolopment- the adoption, adaption or creation of recommendations to fit context)?

8. If you could modify any one part of the WHO TB recommendation generation, publication, dissemination process, what would you change?
9. What are the barriers to using WHO TB recommendations as they are in your country and setting?
10. Would you or other colleagues you know be able to work with WHO to provide feedback on what was done to use recommendations?
11. Are there WHO TB recommendations that you have applied in your setting without changing them (ask for examples if come to mind)?

Appendix B: Survey Script/ Questionnaire

Thank you for following our presentation, and for taking the time to fill out this questionnaire. We would like to explore your perceptions of how to improve the presentation and organization of recommendations offered by the WHO global TB department. Our approach includes novel methods for recommendation mapping and analytical frameworks. We are also interested in learning about your experience with WHO TB recommendation development, publication, dissemination, organization, and engagement. This survey is estimated to take 20 minutes to complete. . We will be collecting some demographic information so that we can better understand from what perspective answers are coming from, but please bear in mind that this survey is anonymous. We will only collect personal identifying information (name and email at the end of the survey) if you agree to be contacted for follow-up questions.

Your participation in the survey is completely voluntary and all of your responses will be kept confidential.You are free to withdraw from the study at any point. If you choose to withdraw after you have completed the survey, we will only be able to remove your answers if you have provided us with you Name and Email. Otherwise, the information you provide us with as anonymous in the survey cannot be removed. You can choose to skip specific questions on the survey. However, it is helpful if you can answer as many questions as possible. No personally identifiable information will be associated with your responses to any reports of these data.

While this research might not be of direct benefit to you, your input will help us learn more about the WHO TB recommendation publication, consolidation, dissemination and engagement process. This information may influence the course of action taken in prioritizing areas, developing recommendations in guidelines and other documents, accessing and using the recommendations, and engaging with stakeholders.

This study has been reviewed by the Hamilton Integrated Research Ethics Board (HiREB). HiREB is responsible for ensuring that participants are informed of the risks associated with the research, and the participants are free to decide if participation is right for them. If you have any questions about your rights as a research participant, please call The Office of the Chair, HiREB at 1-905-521x42013.

Your filling out of the survey provides consent for participating in this study.

Thank you in advance for your time and cooperation.

Demographics

1. In what role are you participating in this project?
    a. WHO TB Programme staff

b. National TB Program Representative
    i. If so, what region? _____ Prefer not to respond.
c. Member-State Representative
    i. If so, what country? _____ Prefer not to respond.
d. Researcher
    i. In which country do you currently work? _____ Prefer not to respond.
e. Policy-Maker
    i. In which country do you currently work? _____ Prefer not to respond.
f. Clinician
    i. In which country do you currently work? _____ Prefer not to respond.
g. Other
    i. Please explain _____

2. What is the highest degree or level of school you have completed? If currently enrolled, highest degree received.?
    a. No schooling completed
    b. High school graduate, diploma or the equivalent (for example: GED)
    c. Some college credit, no degree
    d. Trade/technical/vocational training
    e. Associate degree
    f. Bachelor's degree
    g. Master's degree
    h. Professional degree
    i. Doctorate degree
    j. Other
    k. Prefer not to respond

3. How long have you been involved in TB-focused work?
    a. <1 year
    b. 1-2 years

c. 3-5 years

d. 6-9 years

e. >10 years

4. In what capacity have you been working in TB (i.e., what is your field of work?)?

    a. _____

Assessing Current WHO TB Guideline Recommendations Generation and Publication

5. Have you been involved in any part of the WHO TB guideline development process?
    a. Y/N; if Y:
        i. What barriers can you identify in WHO TB recommendation prioritization?
        ii. What barriers can you identify in WHO TB recommendation generation?
            1. If possible, please give an example with the PICO question asked. (PICO is an acronym for population, intervention, comparison and outcome components that frame a question) _____
        iii. What barriers can you identify to WHO TB recommendation dissemination?
            1. If possible, please give an example with the PICO question asked. _____
        iv. What barriers can you identify to WHO TB recommendation implementation?
            1. If possible, please give an example with the PICO question asked. _____
        v. What facilitators can you identify to WHO TB recommendation implementation?
            1. If possible, please give an example with the PICO question asked. _____

6. There are redundancies across WHO TB guidelines and accompanying publications (manuals, implementation guides, etc.).
    a. Strongly agree
    b. Agree
    c. Somewhat agree
    d. Neither agree or disagree

e. Somewhat disagree
f. Disagree
g. Strongly disagree

7. Current WHO TB recommendations are contained in discrete WHO publications available as PDF documents on the WHO TB Programme website. Please rate the following statements provided below.
    a. The current body of WHO TB publications are easy to navigate.
        i. Strongly agree
        ii. Agree
        iii. Somewhat agree
        iv. Neither agree or disagree
        v. Somewhat disagree
        vi. Disagree
        vii. Strongly disagree

    b. It is easy for you to find the guideline/recommendation you are looking for on the website
        i. Strongly agree
        ii. Agree
        iii. Somewhat agree
        iv. Neither agree or disagree
        v. Somewhat disagree
        vi. Disagree
        vii. Strongly disagree

    c. How long does it typically take you to find a recommendation from the guidelines as they are currently presented?
        i. <30 s
        ii. 30 s- 1 minute
        iii. >1 minute

8. Please tell us how the prioritization, generation, dissemination and implementation of WHO TB recommendations could be improved?
    a. _____

Adolopment is the epidemiological term used to encompass the contextualization process of adopting, adapting, or developing recommendations de novo. The cornerstones of adolopment are to:

- Identify and prioritize credible existing guidelines or evidence syntheses of interest and relevance. This step should involve the relevant stakeholders and proper priority setting.
- Evaluate and complete GRADE EtD (Evidence to Decision) Frameworks for each recommendation. This step involves identifying and reviewing information of existing EtD frameworks or identifying information that informs the EtD criteria and completing a new EtD for the adoloped recommendation.
    - EtD Frameworks helps organize panel members movement from evidence to decisions in the recommendation development process. EtD frameworks include background information, criteria for decision making (including judgements, research evidence, additional considerations), and conclusions.
- Final adoption, adaptation, or de novo creation of recommendations based on the extent of changes that contextualization or updating demand for the original recommendation or degree of work involved.

9. In your opinion, are end users currently supported in their implementation of recommendations?
    a. Y/N
    b. If Y: Please elaborate on how? _____
    c. If N: Please elaborate on why there is no support _____

10. From your experience, how are WHO recommendations being used? (please select all that apply, and kindly provide an example)
    a. Adopted as is
        i. Please provide an example_____
    b. Adapted/modified to fit the context
        i. Please provide an example_____
            1. Why was the recommendation adapted? _____
    c. New recommendation created to meet local needs
        i. Please provide an example_____
            1. Why was the recommendation changed? _____

11. What support do you currently use to ease/facilitate adolopment?
    a. Guideline
    b. Implementation manuals, operational guides, handbooks
    c. Consultation with experts
    d. Non-WHO resources
        i. Please elaborate _____

e. Other
    i. _____

12. What do you think is needed to ease/facilitate adolopment?
    a. _____

Recommendation Mapping and Analytical Frameworks

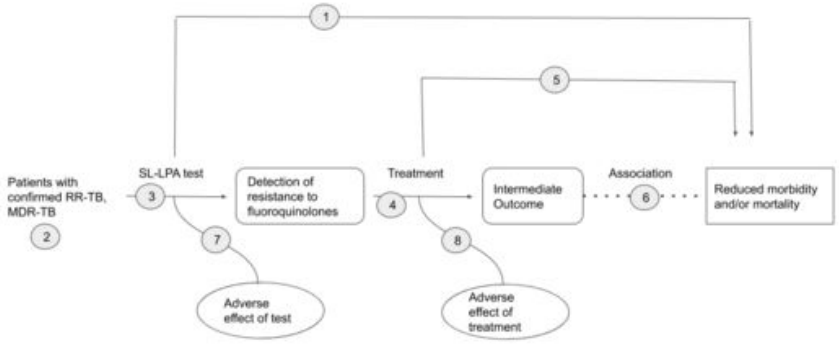
Recommendation Mapping is a novel form of evidence mapping, which organizes the breadth of evidence on a certain topic, aids in identifying clusters and gaps in the evidence, and sets the groundwork for consolidation and priority setting. It involves dividing recommendations by the population, interventions, controls, and outcome (PICO) they address. The groundwork of recommendation mapping will also set the stage for the creation of analytical frameworks, linking recommendations across guidelines, and highlighting questions (and relevant evidence) in each step of the causal pathway.

13. Recommendation mapping allows for the identification of gaps and clusters of recommendations. Are you aware of any methods that already do this?
    a. Y/N
    b. If Y: Please elaborate _____
14. Recommendation mapping involves working with base-unit of guidelines; the recommendation. What other parts of the publication do you think should be included in the mapping? (Please check all that apply).
    a. Guideline question
    b. Recommendation remarks
    c. Support statements from implementation guides, operational manuals, handbooks
    d. Other _____
    e. None of the above

15. The evidence that informs WHO guideline recommendations as they are currently presented, is readily apparent.
    i. Strongly agree
    ii. Agree
    iii. Somewhat agree
    iv. Neither agree or disagree
    v. Somewhat disagree
    vi. Disagree
    vii. Strongly disagree

Analytical frameworks visually capture the logical sequence of questions and evidence that guide the overarching guideline development process; from question to recommendations. Analytical frameworks can be connected across screening, diagnostic, and treatment guidelines to create a successive flow of recommendations (see figure below):

Example: Analytical Framework for Recommendation 1 in "The use of molecular line probe assays for the detection of resistance to second-line anti-tuberculosis drugs"[15]



16. How would you, as an end-user of WHO TB recommendations, like to see analytical frameworks be used? [Check all that apply]
    a. One analytical framework per guideline
    b. Meta-analytical framework connected (where possible) across guidelines
    c. Clickable links to evidence/(EtDs?) informing causal linkages in analytical frameworks

Centralized Database/Hub of Recommendations

The centralized database/hub of recommendations provides a visualization of the mapping work, and provides end-users with the opportunity to explore and engage with recommendations (adolopment).

17. Roughly how long does it take you to search for a recommendation using the discrete publications organized by year that are currently on the WHO TB Programme website?
    a. 0- 10 s
    b. 10-30 s
    c. 30 s- 1 minute
    d. >1 minute
18. Roughly how long does it take you to search for a recommendation using this platform?
    a. 0- 10 s
    b. 10-30 s

c. 30 s- 1 minute
d. >1 minute

19. Do you think the presentation of recommendations is intuitive?
    a. Y/N
    b. Kindly suggest any improvements that could be made to the interface_____
20. Are the concentric circles organized by outcome, easy to navigate?
    a. Y/N/Neutral
    b. If N: kindly provide an alternative suggestion _____
21. Are the population categories organized by colour, easy to navigate?
    a. Y/N/Neutral
    b. If N: kindly provide an alternative suggestion _____
22. Are the tags for PICO components of the recommendation, easy to navigate?
    a. Y/N/Neutral
    b. If N: kindly provide an alternative suggestion _____
23. Are you familiar with ICD-11 codes for population?
    a. Y/N
    b. If Y: is this a fitting way for you to search for population? _____
24. Are you familiar with ATC classification for pharmaceuticals?
    a. Y/N
    b. If Y: is this a fitting way for you to search for interventions and/or controls? _____
25. Please provide any feedback you have on the searchable database (i.e. any other search functions)?_____

Perceptions

26. Did you learn about the quality improvement measures in the presentation on _____.
    a. Y/N/Unsure
27. Do you think the measures introduced in this project will have utility in organizing recommendations, and highlighting clusters and gaps?
    a. Y/N
    b. If N: Please elaborate _____
28. Do you think the measures introduced in this project will have utility in easing accessibility and engagement of end-users with WHO TB recommendations?
    a. Y/N
    b. If N: Please elaborate _____
29. Please provide any suggestions, lingering questions, comments or concerns regarding the project. _____

Priority Setting

30. What has been your involvement in priority setting for the WHO TB guideline development process?
    a. Very involved
    b. Somewhat involved
    c. Barely involved
    d. Not involved

31. Would you like to be involved in priority setting for WHO TB guideline development process?
    a. Y/N/Neutral

32. If you are not as involved as you would like to be, why?
    a. No clear opportunity to participate in priority setting
    b. Time constraints
    c. Not applicable
    d. Other _____

Appendix C: LETTER OF INFORMATION / CONSENT

Assessing the utility of evidence organization methods of WHO TB Guideline Recommendations: a mixed methods study

Investigators:

Local Principal Investigator:
Dr. Holger Schunemann
Department of HEI
McMaster University
Hamilton, ON, Canada
(905) 525-9140 ext. 24931
E-mail: holger.schunemann@mcmaster.ca

Student Investigator:
Anisa Hajizadeh
Department of HEI
McMaster University
Hamilton, ON, Canada
(905) 525-9140 ext. 26771
E-mail: hajizaa@mcmaster.ca

Purpose of the Study

The HEI Department at McMaster University is collaborating with the WHO to improve the accessibility, organization and consolidation of WHO TB guideline recommendations in manuals and supporting documents. The quality improvement measures mainly involve the creation of a centralized platform to host all WHO TB recommendations, for increased accessibility, engagement, consolidation and priority setting. To compliment this work, this mixed methods study aims to assess the utility of the outputs from the perspective of WHO Global TB Programme staff, guideline panel members and other key stakeholders.

You are invited to take part in this study on assessing the implementation and service outcomes of the quality improvement measures; namely recommendation mapping leading to centralized platform of recommendations. We want to gauge your thoughts on the utility of these measures, and potential areas of improvement in WHO TB guideline publication, dissemination, and consolidation. We are hoping to learn how the quality improvement project fairs from the perspective of users, barriers to organization/consolidation of recommendations, degree of engagement with key stakeholders, and areas of further need.

Procedures involved in the Research

In order to assess the quality improvement measures that will be taking place over the span of three years (2019-2021), we invite you to participate in a survey and interview in November-December 2019, and at the time of project completion in 2021.

Based on either the initial pilot work in November-December 2019, or the outputs of the entire project in 2021, we will be conducting a survey and interview with interested participants. The survey will be available through an online link, and will take approximately 20 minutes to complete. The interview will be conducted in person, and will take approximately 30 minutes. Your participation is completely voluntary, and you are free to withdraw at any point. Results from the survey and interview will be kept anonymous.

The Survey:
You will be shown the outputs from the pilot in a brief presentation. A link to an online survey will be made available to you immediately following the presentation, and can be completed within 48 hours. Should you choose to participate, you will provide consent to the survey on the first page of the online survey script, and will be asked to complete all questions that follow. Your answers and participation will remain anonymous.

The Interview:
You will be asked to complete an interview with research personnel from McMaster University. The interview will be roughly guided by a series of questions, with opportunities to elaborate and digress on relevant points of interest. The interview will be audio-recorded. Your participation in the interview is anonymous. We will ask questions about your experience with WHO TB recommendation development, publication, and use. We will then ask about your thoughts on the quality improvement measures, to understand your perceptions on its benefits, and opportunities for improvement.

Potential Harms, Risks or Discomforts:

It is not likely that there are any risks or harms associated with participating in this study.

You do not need to answer questions that you do not want to answer or make you feel uncomfortable. You can always stop to take a break. You can withdraw (stop taking part) at any time. I describe below the steps I am taking to protect your privacy.

Potential Benefits

The research may not benefit you directly. However, we hope to learn more about the WHO TB recommendation publication, consolidation, dissemination and engagement. This information may influence the course of action taken in prioritizing areas, producing recommendations in guidelines and other documents, accessing and using the recommendations, and engaging with stakeholders.

Confidentiality

You are participating in this study confidentially. We will not use your name or any information that would allow you to be identified. No one but the research team will know whether you participated unless you choose to tell them.

The information/data you provide will be kept on a password protected laptop that only research personnel have access to. Once the study is complete, an archive of the data, without identifying information, will be kept until results have been published.

Participation and Withdrawal

Your participation in this study is voluntary. It is your choice to be part of the study or not. If you decide to be part of the study, you can decide to stop (withdraw), at any time, even after signing the consent form or part-way through the study. If you decide to withdraw, there will be no consequences to you. Information provided up to the point where you withdraw will be kept unless you request that it be removed. If you do not want to answer some of the questions you do not have to, but you can still be in the study.

Information about the Study Results

We will be publishing the results of this study. If you would like to receive a summary of the results personally, please let us know how you would like me to send it to you.

Questions about the Study

If you have questions or need more information about the study itself, please contact us at: hajizaa@mcmaster.ca , (905) 525-9140 ext. 26771

This study has been reviewed by the Hamilton Integrated Research Ethics Board (HiREB). The HiREB is responsible for ensuring that participants are informed of the risks associated with the research, and that participants are free to decide if participation is right for them. If you have any questions about your rights as a research participant, please call the Office of the Chair, HiREB, at 905.521.2100 x 42013.

CONSENT

I have read the information presented in the information letter about a study being conducted by Dr. Schunemann and colleagues, of McMaster University.

I have had the opportunity to ask questions about my involvement in this study and to receive additional details I requested.

I understand that if I agree to participate in this study, I may withdraw from the study at any time.

I will be given a signed copy of this form. I agree to participate in the study.

_____	_____	
Name of Participant (Printed)	**Signature**	**Date**

Consent form explained in person by:

_____ _____

Name and Role (Printed)	Signature	Date

CHAPTER 2: Manuscript

Authors:
Anisa Hajizadeh[1], Dominik Mertz[1,4], Robby Nieuwlaat[1,2], Dennis Falzon[3], Ernesto Jaramillo[3], Tamara Lotfi[1,2], Jan Brozek[1,2], Tamara Lotfi[1,2], Andrea Darzi[1], Qi Wang[1], Al Subhi Mahmood[4], Praveen Saroey[4], Micayla Matthews[1], Finn Schünemann[1,5], Bart Dietl6, Artur Nowak6, Kuba Kulesza6, , Micayla Matthews1, Giovanna Muti Schunemann[2], Antonio Bognanni[2], Rana Charide, Holger Schünemann[1,2]

Affiliations:
1 Department of Health Research Methods, Evidence and Impact, McMaster University, Hamilton (ON), Canada, 2 Michael G. DeGroote Cochrane Canada and GRADE Centre, McMaster University, Hamilton (ON), Canada, 3 Global TB Programme World Health Organization, Geneva, Switzerland, 4 Department of Medicine, McMaster University, Hamilton (ON), Canada, 5 University of Freiburg, Breisgau, Germany, 6 Evidence Prime Inc, Poland, xx (Giovanna & Antonio-TBD)

Contacts:
Anisa Hajizadeh: hajizaa@mcmaster.ca, Dominik Mertz mertzd@mcmaster.ca, Robby Nieuwlaat: nieuwlr@mcmaster.ca, Dennis Falzon: falzond@who.int, Ernesto Jaramillo: jaramilloe@who.int, Jan Brozek: jan.brozek@mcmaster.ca, Tamara Lotfi: lotfit@mcmaster.ca, Andrea Darzi darzia@mcmaster.ca, Qi Wang wangq87@mcmaster.ca, Al Subhi Mahmood alsubhi@hhsc.ca, Praveen Saroey saroeyp@mcmaster.ca, Micayla Matthews matthm9@mcmaster.ca, Finn Schünemann finnschuenemann@googlemail.com, Holger Schünemann: holger.schunemann@mcmaster.ca

2.1 Abstract

Background: Evidence-mapping is a method to broker a body of knowledge through curating a usable tool in collaboration with stakeholders. Here, the methods of evidence mapping are novelly applied to guidelines through a process of recommendation mapping.

Objectives: The objectives of this work was to describe the principles of evidence mapping, which is typically reserved to single studies and systematic reviews, to the pinnacle of preprocessed evidence syntheses; guideline recommendations. The primary objective is to establish methodologies for RM. Secondary objectives include the application of evidence mapping

methods to other guideline components (PICO questions), the modelling of an interactive, dynamic, electronic WHO-TB recommendation map, and a template for guideline developers on the final articulation of recommendations following PICO ontology.

Methods: Methods were iteratively developed and piloted through the creation of a recommendation map for the World Health Organization (WHO). All guidelines containing explicit recommendations developed following international evidence-based and transparent standards were eligible for inclusion. Recommendations were extracted, deconstructed by their Population, Intervention, Comparison, Outcome (PIC(O)) subdomains, and coded using standard ontologies (ICD-11, ATC, SNOMED-CT). Coded recommendations were visually displayed in a dynamic electronic map, organized according to evidence-mapping principles, with stakeholder input. Raw data of extraction was analyzed quantitatively, and final map outputs underwent a narrative analysis to identify gaps, clusters and trends.

Results: All WHO-GTB issued guidelines published since the introduction of GRADE methodology into the organization (2007), were included, 83 records were screened resulting in 20 guidelines eligible for the mapping exercise. From these guidelines, 211 recommendations were extracted according to Cochrane PIC(O) ontology, for a total of 548 possible entries per recommendation. From these 20 guidelines, 42 PICO questions guiding the development of guidelines published ≥2014 were also extracted. Analysis of the raw data of extraction revealed missing and multiple PICO subdomains across recommendations and PICO questions. A recommendation map produced as a trial for World TB Day on March 24th was qualitatively analyzed to reveal gaps and clusters of recommendation for TB management across several populations.

Conclusions: Recommendation maps may allow guideline developers and users to identify gaps and clusters of recommendations across publications, serve as an instrumental tool in the sequence of guideline development (from intelligent priority setting, to the assembly of final recommendations) and increase the accessibility of key guideline components, including the rationale for a recommendation. The data yield of RM may also serve as a scaffold, along with other guideline components extracted and coded, for live update and refinement in a rapidly learning health system.

2.2 Introduction

The exponential growth of health research may give rise to rapid learning health systems marked by a culture of deliberate knowledge acquisition, optimized evidence generation, and skillful application. This paper introduces novel methods to poise guidelines in a given domain, towards a rapid learning health system characterized by digital capture, real time access and continuous learning via the foundational work of recommendation mapping (1). This work introduces methods for recommendation mapping, developed and applied in tandem for the World Health Organization's (WHO) tuberculosis (TB) guidelines.

Scoping studies have broadly been used to comprehensively document the breadth of evidence in a given topic. Evidence maps are more detailed than scoping reviews, and the scope of the latter may benefit from the insight gleaned from the former (2,3). Evidence maps are used to identify the landscape of evidence in a certain topic area, highlighting both clusters and gaps in the literature, for the purpose of priority setting in research (3-5). Within the health sector, there are heterogenous definitions of what constitutes evidence mapping, and the role they play in the review of evidence. Evidence maps usually employ systematic methods to search for evidence, identify gaps in knowledge/future research needs, and are presented in a user-friendly visual or searchable database (6). We believe that the concepts of evidence mapping applies to guideline

recommendations are .....[define] recommendations. Recommendation maps chart the terrain of recommendations for a given condition to inform priority setting, streamline guideline development, and enhance the accessibility of recommendations through a user-friendly schematic in a user friendly way. We define recommendation mapping as a tool to organize recommendations from a series of guidelines on a certain condition/topic, report features around the evidence that informs the recommendation, highlights clusters and gaps of recommendation by PIC(O) question domain (according to the PIC(O) components selected for the cross-tabulation of the map), and other features (i.e. publication date, updated, unchanged). The end product is a platform of mapped recommendations that are accessible to stakeholders (policy-makers, researchers, clinicians, patients and their families).

As an international public health authority, the WHO contributes to public health through providing robust high-level guidance at the level of policy, as well as support in the implementation and contextualization at the level of health systems, care, and individual health through handbooks and operational guides. Since 2003, the WHO has used the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) approach to develop its guidelines and the GRADE summary of findings (SoF) to evaluate the quality of evidence, and grade the strength of recommendations. The WHO Global TB Programme (WHO-GTB) has issued TB publications since 1996. In 2007, the WHO-GTB adopted the international standard for guideline development, GRADE (Grading of Recommendations Assessment, Development and Evaluation) to transparently produce reliable, evidence-based guidelines which weigh the certainty of the evidence and propose a strength for the recommendation (7-10). The WHO-GTB Programme adheres to the guideline development methods of the GRADE approach, as indicated in their handbook for guideline review committees, with a second edition updated in 2014 (11). The standardized development of WHO-GTB guidelines, results in publications transparent in development. This spans from the a-priori PICO questions guiding the search for evidence, to the

conception of recommendations, all by an interdisciplinary guideline development group, with disclosure and management of conflicts of interest.

The eradication of TB is of high global priority, given its station as the number one infectious disease worldwide- a sobering reality despite it being both preventable and curable (12). The WHO has set a mandate to eradicate TB by 2035, in their End TB Strategy proposed in 2014. The strategy is composed of three pillars; integrated patient-centered TB care and prevention, bold policies and supportive systems, and intensified research and innovation (13). Taken together, WHO issued TB guidelines serve as an opportune starting point for the application of this work, given the high quality of the publications, the impending eradication goal, and urgent call for innovative research strategies.

The primary objective of this work is to describe the concept of evidence mapping methodologies in the context of guidelines through "recommendation mapping" (RM), developed and applied in tandem with WHO-TB recommendations. The secondary objectives of this work include the mapping of guideline components (PICO questions), the curation of an electronic recommendation map and database for the WHO-GTB, and the development of a template for guideline developers for the articulation of recommendations using PICO ontology.

2.3 Methods

2.3.1 General Methods

A mixed methods approach was employed to develop and apply recommendation mapping methods through: 1) iterative consensus on a RM definition, an ontology framework, and fitting terminology selection for coding, 2) extraction of recommendations by chosen ontology, 3) coding of recommendations using a combination of bottom-up and top-down approaches, 4) quantitative analysis of recommendations and PICO questions by raw data of extraction, 5) qualitative analysis of final mapping outputs to identify gaps and clusters.

2.3.2 Eligibility Criteria

All recommendations contained in published WHO TB guidelines published ≥2007 were eligible for inclusion in the recommendation mapping (RM) exercise. All types of WHO guidelines were included; standard guidelines, consolidated guidelines, interim guidelines and guidelines produced in response to an emergency or urgent need. All files were available in English. Guidelines which did not use the GRADE approach in their development (i.e. no grades of final recommendations, and/or no evidence profiles, and/or evidence to decision tables), were not included. Further, recommendations in collaborative guidelines assessing other populations (e.g. physical conditions in those with severe mental disorders, HIV/TB guidelines),were only included if they contained recommendations that considered TB in their PICO.

2.3.3 Information Sources and Selection

All publications were accessed through the WHO-GTB Programme website, as well as through documents provided by WHO-GTB staff. Links to all publications were collected in a spreadsheet organized by year. The WHO GTB team verified the final collection of guidelines eligible for RM.

2.3.4 Data Collection

Data extraction was completed in two phases. In lieu of duplicate extraction, abstractors trialed the recommendation extraction with 3 recommendations against an example of extraction collaboratively agreed upon and came to agreement. The first phase of extraction involved an exclusive focus on the verbatim recommendation, by extracting elements of recommendations using PICO ontology (14,15). Integral to the guideline development process, is framing questions by PICO components that guide the search for evidence and focus the formation of

recommendations (16). Each recommendation then, has an inherent PICO backbone; with varying degrees of complexity and completeness. In this first phase, recommendations were extracted by the verbatim PIC(O)elements as published. The next step involved the assignment of a standardized nomenclature using both a top-down, and bottom-up approach to assigning codes to the extracted PICO subdomains. To decide upon the top-down vocabulary, an initial pilot of RM from ≥2014 was scanned for common terms that could be comprehensively captured by a standardized vocabulary (Appendix A). A similar exercise was conducted for the bottom-up approach for a vocabulary, to identify the WHO TB-specific terms used consistently across the TB guidelines in the pilot. For the top-down assignment of codes, ICD-11 and SNOMED-CT were decided upon for Population, and for Intervention, ATC or SNOMED-CT codes could be referenced. If a PIC(O) element was eligible for more than one coding system, then only one vocabulary was manually coded, and the other was assigned through data-linking.

The second phase of extraction involved consulting the original guideline to a) supplement the verbatim recommendation, and b) abstract the referenced evidence that informed each recommendation. To supplement recommendations as they were finally presented in a standardized fashion, additional information that was eligible to be extracted and coded was found in three places in the guideline; the accompanying remarks, superscripts in the recommendation referencing additional information in a footnote, and key definitions available in the glossary of a guideline. Evidence that informed the guideline development process (and subsequent recommendations) was pulled by referencing available GRADE evidence profiles, Evidence to Decision (EtD) tables, and/or references to systematic reviews and/or primary studies if available.

PICO questions from the last 5 years (≥2014) were extracted to match the pilot extraction. The methods (eligibility criteria, data extraction) were congruent with....

2.3.5 Data Items

The recommendation extraction form was designed in collaboration with Evidence Prime Inc., collaborators at HEI, and the WHO GTB program. Please see the recommendation extraction guide in the Appendix B. All PICO elements were first extracted verbatim by how it appeared in the publication. Each component was then further expanded upon, to capture the minutiae of each element, for a total of 548 possible entries per recommendation. Should there have been more than one P, I, C contained in the recommendation, additional entries were added. Please refer to Table 1 in Appendix C for the PICO subdomains that were of interest for extraction. Other elements that were extracted included the date of publication, whether it was a new/updated/unchanged recommendation (if it was from a consolidated guideline), and the evidence that informed it. Guideline definitions were also abstracted in a separate sheet, and coded according to the guideline ID they corresponded to (as elements like age, and regimen duration, could be defined differently from one guideline to the next). The Outcome subdomain of PICO in the form is the least-comprehensive in terms of data items of interest. Outcomes really should not be part of the final articulation of a recommendation because there always are many and it may mislead decision makers that only some are considered if not all are listed (furthermore, recommendations are about all desirable and undesirable consequences beyond health benefits and harms), and so in earlier iterations, and could be completely omitted from the extraction form. For recommendations, P, I, C are the main elements of interest. There are variances in outcome definitions, and further, any one recommended intervention will have *multiple* outcomes that cannot be accounted for in the text of the recommendation. Outcomes are infrequently present in recommendations, are not emphasized in our work, and as a result, will not be weighted heavily in the outputs (i.e. it will be of more interest to have a recommendation map by the axes of P x I, than by O).

2.3.6 Synthesis of Results: Mapping Visualization

A cross- tabular representation is used to plot recommendations against two variables of interest from the extraction; population and intervention/comparison. *Who* qualifies for the recommendation, and *what* they are to do/receive/be eligible for, are of prime importance in organizing recommendations. Focusing on the P and I/C for mapping does not preclude other parts of the recommendation from being included in the interactive map; O are of lesser importance in regards to organization. The x-axis hosts all extracted elements of I/C according to PICO ontology. Similarly, the y-axis contains all extracted elements of P. The final maps are created by web designers and IT specialists; namely Jan Brozek, Bart Dietl, Artur Nowak and Kuba Kulesza. A dynamic- as opposed to a static- online presentation is opted for, for end-users to manipulate the specific PxI/C elements of interest to them. A dynamic and interactive map will not only allow the end-user to pick which broad axes to display (i.e. by P,I,C), but will also allow viewers to select any combination of subdomains P,I,C,O subdomains to plot. For instance, if one is interested in viewing PxI, for the y-axis of P, one could select P combinations of P_{age}, $P_{condition}$ to be plotted against I_{name} and I_{dosage}. This option allows for a much wider range of many RMs to become available for the purposes of priority setting, as intricacies in gaps and clusters of recommendations based on all subdomains, can be considered. As with some evidence gap maps, the recommendation maps are accompanied by a searchable database/portal which houses the complete collection of recommendations included in the mapping exercise, and provides end-users with a more detailed access to the recommendations and resources (i.e. a breakdown of the PICO in the recommendation, the strength of the recommendation and certainty of the evidence, a link to the guideline, supplementary evidence tables, and related recommendations).

2.3.7 Analysis

Analysis was conducted in two parts. First, a quantitative analysis of missing and multiple PICO subdomains in raw data of recommendation and PICO question extraction respectively, and then in association with one-another. The former allows for a commentary on the final articulation of recommendations and PICO questions, and the latter provides insight into the recommendations that finally arise from a-priori PICO questions guiding their development, and the relationship between missing/multiple PICO subdomains in the question and resulting recommendations. Secondly, a narrative analysis on the gaps and clusters of the recommendation map, and a commentary on the implications of such trends is provided.

2.4 Results

2.4.1 General Results

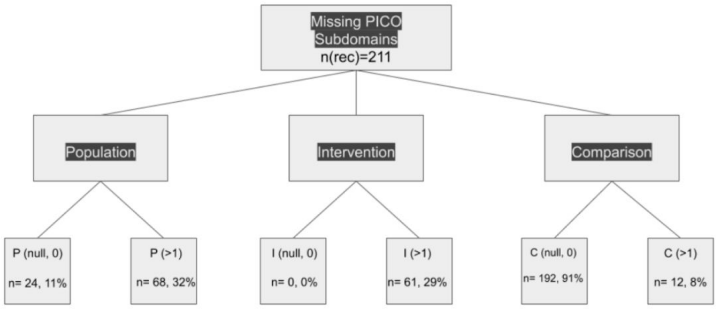
In total, there were 56 guidelines containing explicit recommendations published ≥2007. Only 20 guidelines met the eligibility criteria for inclusion into the mapping exercise. From these 20 guidelines, there were 211 recommendations, and PICO 104 questions, with 42 PICO questions that were extracted from guidelines published ≥2014 to create the recommendation map (please see Fig. 5).

2.4.2 Missing and Multiple PICO Elements

From the verbatim extraction of recommendations and PICO questions by PICO element, it is evident that there are stylistic differences in how final recommendations are articulated and finally presented. Results from raw data extraction of both mapping exercises trialed on WHO-GTB guidelines presented us with recommendations and questions with entirely missing and/or multiple PICO elements (see Fig. 1 and Fig.2). As to the populations,P 11% of recommendations had missing populations in their verbatim actionable statement, whereas 32% had more than one population, with one recommendation containing a high of 5. There were no

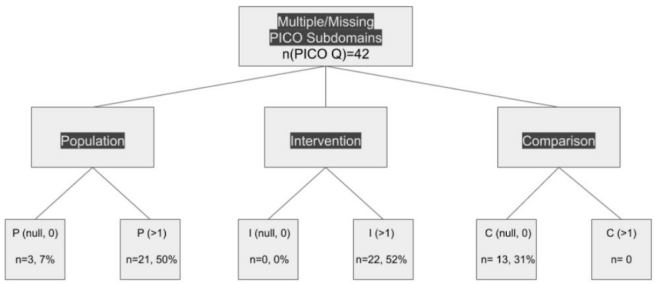
recommendations with an absent intervention or exposure, and 29% of the recommendations included more than one intervention in their statement. A vast majority of recommendations omitted mention of a comparison (91%), with the majority of those remaining presenting more than one comparison (8%). Of these, 4 recommendations presented multiple grades within the same recommendation (see Appendix E Table 2). Outcomes were not assessed for recommendations, as they were almost never presented, and for PICO questions they were almost always included, indicative of the appropriate form and fashion of recommendation and PICO question reporting.

Fig. 1 Missing & Multiple PICO in Recommendation



While the sample size was smaller for PICO questions, as only PICO questions in the last 5 years were extracted, a similar trend followed for population, intervention and comparison. For the population, 7% did not include a population in their verbatim question statement, and 50% considered multiple populations in the stated questions. Like recommendations, 0% contained missing interventions, and over half (52%) contained more than one intervention. For comparisons, 31% failed to mention a comparison, with no question making explicit mention of more than one comparison (comparisons were either to another test, standard of care, or no intervention- all of which were considered as 1 comparison).

Fig. 2 Missing & Multiple PICO in Guidelines Questions



Missing PICO subdomains in the verbatim text of a recommendation or a PICO question is a misnomer, raising stylistic concerns around articulation. If a population is missing from the recommendation, it is usually found elsewhere in the guideline, and writers may have felt disinclined to include it in the final statement. In Phase 2 of extraction when recommendations and questions are supplemented with further information pertaining directly to the recommendations/questions (from the guideline definitions, remarks, footnotes, and evidence to decision tables respectively), the final input of PICO data becomes complete; for all population components for recommendations and questions, and for all comparisons used in questions. This supplementation does not usually lead to the addition of a comparison for the recommendation, as this is often not found in the aforementioned places. The value of including comparisons in a recommendation, albeit a simple comparison to standard of care, or no intervention, is unclear for end-users. Likewise, the supplementation of PICO

Fig. 3 Multiplicity in Population

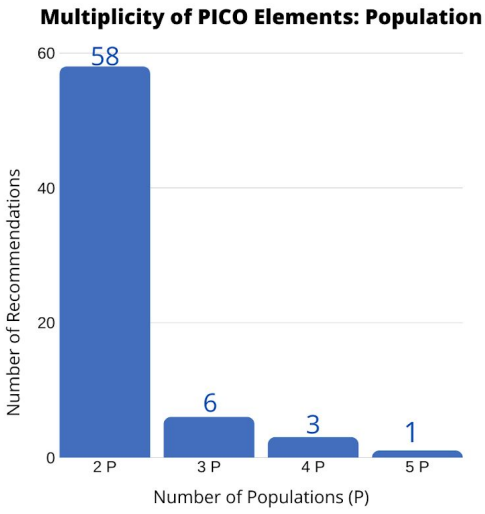
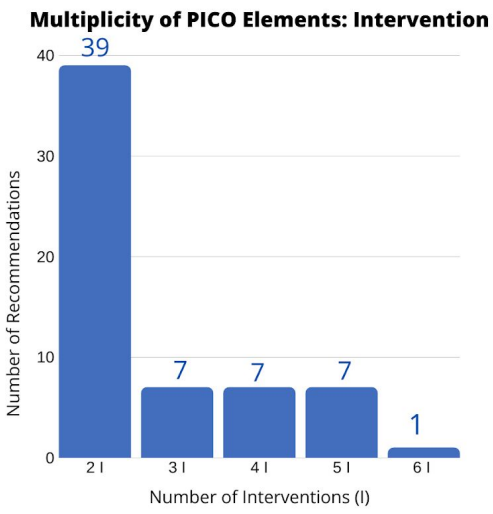


Fig. 4 Multiplicity in Interventions



questions with information from the PICO tables found in the appendices of the guideline, alleviates the concern of missing PICO elements, and usually introduces the concept of multiplicity in PICO subdomains.

2.4.2 Multiplicity of PICO Subdomains: Recommendations and PICO Questions

For recommendations, there was mainly duplicity in the population and intervention domains of the PIC(O) extraction. . From the 68 recommendations with multiple populations, the majority had two populations included, with a high of one recommendation with five discrete populations (see Fig. 3). For interventions, the majority of the 62 recommendations with multiple interventions likewise had two stated interventions, with a high of 6 (see Fig. 4).

Multiplicity in PICO subdomains in recommendations arise for thre reasons; 1) the PICO question guiding the development of the recommendations is loaded with multiple P,I ,C , 2) the way a recommendation is finally written to combine P,I,C elements discussed by the GDG, and 3) the way the statement is extracted. Only the lattermost cause was within the control of mapping. Multiple PICO subdomains in a verbatim recommendation or PICO question, or that arise in the supplementation of the recommendation or PICO question during the extraction process, introduces interesting implications for guideline mapping work. The very obvious implication being the redundancy in the final map (i.e. in a 2D cross tabular plot, a recommendation addressing two populations with the same intervention will appear in two places along the population axis, but within the same intervention-axis). For example, many recommendations targeted MDR/RR-TB patients, which represent two populations, multi-drug and monoresistance (17). This recommendation will appear twice along the population-axis, and once along the intervention axis (given the recommendation only included one intervention). Other implications of multiplicity include non-standardization in the presentation of recommendations. This is pronounced if most recommendations contain one clear progression of PICO elements in the

general form of "*in **P**, **I** is recommended in place of **C** (strength, certainty of evidence)*", closely mirroring the PICO question underpinning its development. Like missing PICO elements, it is not so clear how a multi-PICO recommendations or multi-GRADE recommendation for more than one comparison (see Table 2 in Appendix D) is received by the end-user.

2.4.3 Linking Recommendations to PICO Question Extraction

Upon a closer examination of **PICO questions** guiding the development of a guideline and the subsequent *recommendations* that arise, there is no clear trend between multiplicity of PICO subdomains in a PICO question and multiplicity in the final recommendation, pointing perhaps, to the stylistic differences in the development and writing of a guideline. There are however, at least three apparent patterns linking the recommendation mapping extraction to the extraction for PICO question mapping.

The first is the direct translation from question to recommendation. In a 2019 guideline on infection control, one of the 4 PICO questions posed is, **"In health workers or other persons attending health care or congregate settings, can respiratory hygiene and/or cough etiquette in people with presumed of confirmed TB reduce TB transmission when compared with settings where these interventions are not implemented?"** (18). The resulting recommendation mirrors the PICO question, as a direct response, "*Respiratory hygiene (including cough etiquette) in people with presumed or confirmed TB is recommended to reduce M. tuberculosis transmission to health workers, persons attending health care facilities or other persons in settings with a high risk of transmission. (Strong recommendation based on low certainty in the estimates of effects)*" (18). The PICO subdomains extracted for each statement could be superimposed upon one another with nearly perfect precision.

The second pattern involves one recommendation addressing more than one question, or, more than one recommendation addressing one question. In the case of the former,

multiplicity of PICO subdomains arise in the recommendation, but not in the question. The questions are; 1) **"In people of all ages at risk of active TB, does a 4-month daily rifampicin regimen safely prevent TB disease compared to other recommended TB preventive treatment regimens? 2) In people of all ages at risk of active TB, does a 1-month daily rifapentine plus isoniazid regimen safely prevent TB disease compared to other recommended TB preventive treatment regimens?"** (19). The resulting recommendation is published as, *"The following options are recommended for the treatment of LTBI regardless of HIV status: 6 or 9 months of daily isoniazid, or a 3-month regimen of weekly rifapentine plus isoniazid, or a 3 month regimen of daily isoniazid plus rifampicin. (Strong recommendation, moderate to high certainty in the estimates of effect). A 1-month regimen of daily rifapentine plus isoniazid or 4 months of daily rifampicin alone may also be offered as alternatives. (Conditional recommendation, low to moderate certainty in the estimates of effect)"* (19). The final recommendation also contains more than one certainty of the evidence ratingassessment to account for the multiple comparisons. On the flipside, one broad question posed in a 2019 MDR-TB treatment guideline reads, **"In patients with TB, are any interventions to promote adherence to TB treatment more or less likely to lead to the outcomes listed below?"** with three recommendations set to address it (19):

1. *Health education and counselling on the disease and treatment adherence should be provided to patients on TB treatment (strong recommendation, moderate certainty in the evidence)*
2. *A package of treatment adherence interventions may be offered to patients on TB treatment in conjunction with the selection of a suitable treatment administration option (conditional recommendation, low certainty in the evidence)*

3. *One or more of the following treatment adherence interventions (complementary and not mutually exclusive) may be offered to patients on TB treatment or to health-care providers:*
    a. *tracers and/or digital medication monitor (conditional recommendation, very low certainty in the evidence);*
    b. *material support to the patient (conditional recommendation, moderate certainty in the evidence);*
    c. *psychological support to the patient (conditional recommendation, low certainty in the evidence);*
    d. *staff education (conditional recommendation, low certainty in the evidence)*

Likewise, in the second example of this disjointed flow of number of PICO questions to number of recommendations, there are multiple GRADE assessments that arise, to accommodate the multiple comparisons being made in one recommendation. Multi GRADE assessments arise when one question is answered by multiple recommendations and vice versa (one recommendation sufficiently answers more than one PICO question).

In the third pattern, we see multiple PICO subdomains in a question that are not addressed in the final recommendation, or when no recommendation can be made for the entire PICO question. The question in a 2019 diagnostic guideline reads **"What is the diagnostic accuracy of LF-LAM for the diagnosis of TB in adults with advanced HIV disease irrespective of signs and symptoms of TB?**

a) **in inpatient settings CD4 cell count ≤ 200**
b) **in outpatient settings CD4 cell count ≤ 200**
c) **in all settings CD4 cell count ≤ 200**
d) **in inpatient settings CD4 cell count ≤ 100**

e) **in outpatient settings CD4 cell count ≤ 100**

f) **in all settings CD4 cell count ≤ 100 (20)".**

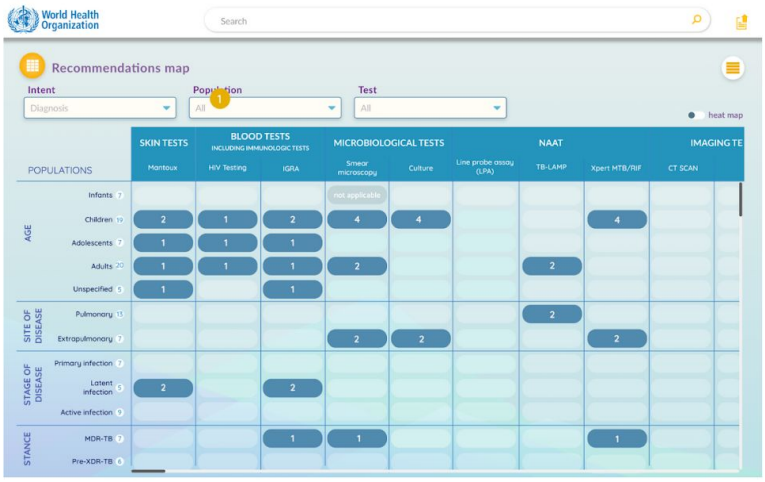
The recommendation in turn, only addresses one population considered, "*In inpatient settings, WHO strongly recommends using LF-LAM to assist in the diagnosis of active TB in HIV-positive adults, adolescents and children: irrespective of signs and symptoms of TB and with a CD4 cell count of less than 200 cells/mm (strong recommendation; moderate certainty in the evidence about the intervention effects)" (20).* Further, some questions will not result in a recommendation tailored to fit any of the PICO subdomains of the question, due to the inability of the guideline development group to form a recommendation based on the evidence presented. In the most recent 2020 consolidated guideline, the new question posed iswa: **"In pregnant and postpartum women, is isoniazid preventive treatment for TB as safe as other preventive treatment regimens?"**, with the recommendation from the previous guideline remaining unchanged and only considering the pregnancy component of the question peripherally; *"In settings with high TB transmission, adults and adolescents living with HIV who have an unknown or a positive LTBI test and are unlikely to have active TB disease should receive at least 36 months of daily isoniazid preventive treatment (IPT). Daily IPT for 36 months should be given whether or not the person is on ART, and irrespective of the degree of immunosuppression, history of previous TB treatment and pregnancy in settings considered to have a high TB transmission as defined by national authorities. (Conditional recommendation, low certainty in the estimates of effect)*" (19). The guideline notes that "[b]ased upon these findings the GDG concluded that there were insufficient grounds to change previous guidance or to develop a separate recommendation for the use of IPT in pregnant women with HIV. The GDG considered that systematic deferral of IPT to the postpartum would deprive women from its protective effect at a point when they are more vulnerable to TB" (20). Within this pattern, it is also important to consider the possibility of

unaddressed PICO domains in the question, or entirely unaddressed questions,finding answers in the guideline outside of the recommendations (*i.e. informal recommendations*).

2.4.4 Narrative Analysis of Map

On the date when this thesis was finalized, only a sample of the map was published by the WHO-GTB programme, which contains 18 recommendations from a consolidated and embargoed 2020 TB preventive guideline released for World TB Day on March 24th, 2020 (please see Fig. 10) (19). The map is an interactive display of recommendations by an intervention (y-axis) and population (x-axis), with ability to manipulate outputs along different axes to display desired cross tabulation of recommendations. The map portrays where there are clusters of recommendations (e.g. preventative treatment, for all age groups), and where there are gaps. Gaps can either be representative or true gaps in recommendations where recommendations should be made (e.g. screening for TB in 'other' populations), or false gaps that need not be filled (e.g. LTBI testing in those with confirmed TB such as HIV-associated TB, drug-susceptible and drug-resistant TB). True gaps requiring recommendations, as visible on the map, may provide needed direction for guideline priority setters. Further, newly available evidence with an outcome of TB prevention, and covering any cross-section of the population/intervention points on the map, can be considered for future updates of existing recommendations.

Fig. 5 Recommendation Map



2.5 Discussion

2.5.1 Summary

We developed a recommendation map for WHO-GTB guidelines. Until now, WHO developed 20 WHO-GTB guidelines with recommendations informed by systematic reviews which we included in the mapping exercise. From these 20 guidelines, 211 recommendations were extracted in full, with data contributing to the recommendation map and searchable database. From the guidelines published ≥2014, 42 PICO questions were identified which were also extracted, to inform the analysis of extracted recommendations.

2.5.2 Strengths and Limitations

The strengths of this work lie in the evidence-informed development, high collaboration in a multidisciplinary team including GRADE methodologists with over 20 years of experience in guideline development and methodological approaches to guideline innovation. In terms of limitations, aswith the application of any new methodological approach, there is lesser reliability of results, as this is the first time attempt to developing a recommendation map using PIC(O) elements which is planned as a next step.

2.5.3 Implications for practice and research

There are general implications of recommendation mapping as a method, and specific implications of this TB exercise that may inform future guideline creation. In terms of the former, a recommendation map, like an evidence map, presents clusters of recommendations and

apparent gaps. As a tool for priority setting, it is tempting to view clusters as areas of sufficiency, and gaps as areas of need. However, like any concentration of evidence in the literature, it may be representative of a super-saturation of knowledge, or, a highly complex area that is insufficiently addressed by the cluster of evidence, requiring multiple questions, fueling multiple studies, and many forms of pre-processed evidence syntheses (from reviews, to guidelines). Rather than place unduly emphasis on clusters and gaps for priority setting, the distribution of recommendations can alternatively provide a schematic on the breadth of recommendations made to date, to inform intelligent priority setting based on this cohesive view. Further, guideline developers should not only view spots within the confines of the axes for guideline development, but should consider the addition of populations, and interventions not yet considered/explored by guidelines of the past. In this way, all parts of the map should be used for guideline development (i.e. *what is there, and what could be*).

Specific takeaways that can be insinuated from the analysis of the raw extraction data revolve around the articulation of PICO questions and subsequent recommendations. It is not so clear that missing PICO subdomain are a pressing issue, if they are obviously covered in and around the recommendation in supporting definitions, remarks, and footnotes. The implications of multiplicity in PICO domains of a given question/recommendation and multiple grades for several comparisons, also remains unclear. The question that remains for future work in the mapping of guidelines, is, is it preferable to formulate single PICO questions that can be addressed by a single recommendation with a clear progression of PICO subdomains? This has the added advantage of being simple in contrast to recommendations with multiplicity in any given subdomain, as well as appearing discreetly in the recommendation map. Simplicity is not necessarily desirable. A recommendation with loaded PIC(O) elements may provide fruitful information for comprehensive contextualization by the end-user. Future research assessing the utility of the recommendation map and its use by the end-user through contextualization may

elucidate grassroot preferences. Tools to facilitate the construction of an initial PICO guideline question with an end-vision of the PIC(O) elements to be mirrored in the recommendation may also be helpful. . Such a tool may use elements of the PICO extraction and recommendation extraction form, to 1) construct the PICO question noting the specific PICO subdomain entries that should be addressed, 2) using a similar template to articulate the final recommendation inclusive of desired PICO subdomains.

2.5.4 Conclusions

The application of evidence mapping methodologies to guidelines, in a process of recommendation and PICO question mapping, is possible to achieve the mapping-specific objectives of this scoping review method. The WHO-GTB guidelines included a diverse group of recommendations which, despite largely using the same methodology, differed greatly in how they were finally articulated and presented across the guidelines considered, pointing to some stylistic liberty. Based on the extraction for mapping, which very closely examines each PICO element and all possible subsections in a recommendation, the final formulation of a recommendation (the PICO progression in the statement, and degree of completeness) should be closely considered by guideline developers in the future. In all, recommendation mapping is a productive exercise to chart the landscape of recommendations in a given condition domain, for the purposes of guideline development (from priority setting to final presentation of recommendations), facilitating centrally accessible and interactive guideline data, and providing institutions with a scoping review of its work, organized by evidence-informed methods.

2.6 References

1. Rapid Synthesis Creating a Rapid-learning Health System in Ontario [Internet]. McMaster Health Forum; 2018

[cited 11 March 2020]. Available from: https://www.mcmasterforum.org/docs/default-source/product-documents/rapid-responses/creating-a-rapid-learning-health-system-in-ontario.pdf

2. Mays, N., Roberts, E., & Popay, J. (2001). Synthesising research evidence. In N. Fulop, P. Allen, A. Clarke, & N. Black (Eds.), *Studying the organisation and delivery of health services: Research methods.* London: Routiedge.
3. Arksey, H., & O'Malley, L. (2005). Scoping studies: towards a methodological framework. *International journal of social research methodology, 8(1),* 19-32.
4. Katz, D. L., Williams, A. L., Girard, C., & Goodman, J. (2003). The evidence base for complementary and alternative medicine: methods of evidence mapping with application to CAM. *Alternative therapies in health and medicine, 9(4),* 22.
5. Evidence and gap maps: a comparison of different approaches. Oslo, Norway: The Campbell Collaboration. Retrieved from: www.campbellcollaboration.org/ DOI: https://doi.org/10.4073/cmdp.2018.2
6. Miake-Lye, I. M., Hempel, S., Shanman, R., & Shekelle, P. G. (2016). *What is an evidence map? A systematic review of published evidence maps and their definitions, methods, and products.* Los Angeles, CA, USA: Syst Rev. Retrieved from https://www.ncbi.nlm.nih.gov/m/pubmed/26864942/
7. Falzon, D. *et al.* World Health Organization treatment guidelines for drug-resistant tuberculosis, 2016 update. *Eur Respir J* **49**, doi:10.1183/13993003.02308-2016 (2017).
8. Alonso-Coello, P. *et al.* GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 2: Clinical practice guidelines. *BMJ* **353**, i2089, doi:10.1136/bmj.i2089 (2016).
9. Alonso-Coello, P. *et al.* GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 1: Introduction. *BMJ* **353**, i2016, doi:10.1136/bmj.i2016 (2016).
10. Guyatt, G. H. *et al.* Going from evidence to recommendations. *BMJ* **336**, 1049-1051, doi:336/7652/1049 [pii] 10.1136/bmj.39493.646875.AE (2008).
11. WHO handbook for guideline development. (2014). 2nd ed. [ebook] WHO. Available at: https://www.who.int/publications/guidelines/handbook_2nd_ed.pdf

12. Global Tuberculosis Report 2018. World Health Organization; 2018.

13. WHO End TB Strategy. World Health Organization; 2014.

14. Cochrane. (2019). *PICO ontology*. [online] Available at: https://linkeddata.cochrane.org/pico-ontology#d4e38 [Accessed 1 Oct. 2019].

15. Schünemann, H., Tugwell, P., Reeves, B., Akl, E., Santesso, N., Spencer, F., Shea, B., Wells, G. and Helfand, M. (2013). Non-randomized studies as a source of complementary, sequential or replacement evidence for randomized controlled trials in systematic reviews on the effects of interventions. *Research Synthesis Methods*, 4(1), pp.49-62.

16. Guyatt, G., Oxman, A., Kunz, R., Atkins, D., Brozek, J., Vist, G., Alderson, P., Glasziou, P., Falck-Ytter, Y. and Schünemann, H. (2011). GRADE guidelines: 2. Framing the question and deciding on important outcomes. *Journal of Clinical Epidemiology*, 64(4), pp.395-400.

17. Definitions and reporting framework for tuberculosis [Internet]. 4th ed. World Health Organization; 2013 [cited 11 March 2020]. Available from: https://apps.who.int/iris/bitstream/handle/10665/79199/9789241505345_eng.pdf;jsessionid=72F7032620320C6F0B55EAC7684B77D9?sequence=1

18. WHO guidelines on tuberculosis infection prevention and control, 2019 update, Geneva: World Health Organization; 2019.

19. EMBARGOED: WHO/UCN/TB/2020.1

20. Lateral flow urine lipoarabinomannan assay (LF-LAM) for the diagnosis of active tuberculosis in people living with HIV. Geneva: World Health Organization; 2019

Appendix B: Extraction Guide

Appendix A:

1 The most frequent words

```
In [6]: all_freqs = nltk.FreqDist([w for doc in tokenized for w in doc if not word_filter(w)])

In [7]:
all_freqs.most_common(0)

Out[7]:
[('tb',
103),
    ('treatment',
    50), ('hiv',
    39),
    ('children',
    37),
    ('months',
    29), ('may',
    25),
    ('patients',
    24), ('living',
    23), ('used',
    22),
    ('pulmonary',
    20), ('test',
    17),
    ('suspected',
    16), ('isoniazid',
```

16), ('regimen',
15), ('contacts',
15), ('mg/kg',
14), ('low', 14),
('high', 14),
('people', 14),
('preventive',
14), ('offered',
13),
('recommended'
, 13),
('household',
13), ('disease',
12), ('settings',
12), ('active',
12),
('confirmed',
12), ('given',
12),
('symptoms',
12), ('years',
11), ('adults',
11), ('ltbi', 11),
('testing', 10),
('prevalence',
10), ('treated',
10), ('ipt', 10),
('countries', 9),
('tuberculous',
9), ('incidence',
9),
('adolescents',
9), ('initial', 8),
('aged', 8),
('clinical', 8),
('art', 8),
('range', 7),
('dose', 7),
('age', 7),
('conditional',

```
7), ('certainty',
7), ('evidence',
7)]
```

2 Two-word phrases that are unlikely to occur by chance

```
In [8]: finder = BigramCollocationFinder.from_documents(tokenized)
        finder.apply_freq_filter(3) finder.apply_word_filter(word_filter)
        finder.nbest(n=50, score_fn=nltk.collocations.BigramAssocMeasures.pmi)

Out[8]: [('bcg',
'vaccine'),
         ('culture-based', 'dst'),
         ('current', 'cough'),
         ('per', 'day'),
         ('phenotypic',
         'culture-based'),
         ('tuberculin', 'skin'),
         ('xpert', 'mtb/rif'),
         ('detect', 'resistance'),
         ('index', 'cases'),
         ('peripheral',
         'lymphadenitis'),
         ('comprehensive',
         'package'), ('cd4',
         'cell'), ('conventional',
         'microscopy'),
         ('maximum', 'dose'),
         ('range', '715'), ('aged',
         ''), ('tuberculous',
         'peripheral'), ('clinical',
         'evaluation'),
         ('appropriate', 'clinical'),
         ('tuberculous',
         'meningitis'),
         ('bacteriologically',
         'confirmed'), ('mg/kg',
         'per'), ('715', 'mg/kg'),
         ('close', 'contacts'),
         ('isoniazid',
```

```
'monotherapy'), ('skin',
'test'), ('low',
'certainty'),
('drug-susceptible',
'pulmonary'),
('household', 'contacts'),
('initial', 'test'),
('confirmed',
'pulmonary'), ('hiv',
'prevention'),
('adolescents', 'living'),
('treatment', 'adherence'),
('preventive', 'treatment'),
('hiv', 'status'), ('hiv',
'prevalence'), ('children',
'aged'),
('rifampicin-resistant',
'tb'), ('tb', 'incidence'),
('people', 'living'),
('active', 'tb'),
('children', 'suspected'),
('pulmonary', 'tb'), ('tb',
'disease'), ('high', 'tb'),
('children', 'living'),
('low', 'tb'), ('tb',
'treatment')]
```

3

3 Two-word phrases that are most frequent

In [9]: finder.nbest(n=50, score_fn=nltk.collocations.BigramAssocMeasures.raw_freq)

Out[9]:

```
[('pulmonary',
'tb'),
('preventive', 'treatment'),
('active', 'tb'), ('tb',
'disease'), ('tb',
```

'incidence'), ('tb',
'treatment'), ('confirmed',
'pulmonary'), ('high',
'tb'), ('household',
'contacts'), ('close',
'contacts'), ('adolescents',
'living'), ('children',
'suspected'), ('clinical',
'evaluation'), ('low',
'certainty'), ('low', 'tb'),
('maximum', 'dose'),
('715', 'mg/kg'),
('comprehensive',
'package'),
('conventional',
'microscopy'), ('hiv',
'prevalence'), ('initial',
'test'), ('isoniazid',
'monotherapy'), ('range',
'715'), ('skin', 'test'),
('tuberculin', 'skin'),
('tuberculous',
'meningitis'), ('xpert',
'mtb/rif'), ('aged', ''),
('appropriate', 'clinical'),
('bacteriologically',
'confirmed'), ('bcg',
'vaccine'), ('cd4',
'cell'), ('children',
'aged'), ('children',
'living'), ('culture-based',
'dst'), ('current',
'cough'), ('detect',
'resistance'),
('drug-susceptible',
'pulmonary'), ('hiv',
'prevention'), ('hiv',
'status'), ('index',
'cases'), ('mg/kg',
'per'), ('people',

```
'living'), ('per', 'day'),
('peripheral',
'lymphadenitis'),
('phenotypic',
'culture-based'),
('rifampicin-resistant',
'tb'), ('treatment',
'adherence'),
('tuberculous',
'peripheral')]
```

4 Three-word phrases that are most unlikely to occur by chance

```
In [10]: tri_finder = TrigramCollocationFinder.from_documents(tokenized)
         tri_finder.apply_freq_filter(3) tri_finder.apply_word_filter(word_filter)
         tri_finder.nbest(n=50, score_fn=nltk.collocations.TrigramAssocMeasures.pmi)

Out[10]: [('phenotypic', 'culture-based',
'dst'),
          ('mg/kg', 'per', 'day'), ('tuberculous',
          'peripheral', 'lymphadenitis'),
          ('appropriate', 'clinical', 'evaluation'),
          ('tuberculin', 'skin', 'test'), ('range',
          '715', 'mg/kg'), ('bacteriologically',
          'confirmed', 'pulmonary'), ('high', 'tb',
          'incidence'), ('drug-susceptible',
          'pulmonary', 'tb'), ('low', 'tb',
          'incidence'), ('confirmed', 'pulmonary',
          'tb'), ('active', 'tb', 'disease')]
```

5 Three-word phrases that are most frequent

```
In [11]: tri_finder.nbest(n=50, score_fn=nltk.collocations.TrigramAssocMeasures.raw_freq)

Out[11]: [('confirmed', 'pulmonary',
'tb'),
          ('high', 'tb', 'incidence'), ('low', 'tb',
          'incidence'), ('range', '715', 'mg/kg'),
          ('tuberculin', 'skin', 'test'), ('active',
```

'tb', 'disease'), ('appropriate', 'clinical',
'evaluation'), ('bacteriologically',
'confirmed', 'pulmonary'),
('drug-susceptible', 'pulmonary', 'tb'),
('mg/kg', 'per', 'day'), ('phenotypic',
'culture-based', 'dst'), ('tuberculous',
'peripheral', 'lymphadenitis')]

Appendix B: Extraction Guide

Methods: A Step by Step Guide

1. Determining Eligibility Criteria
    a. Included: All guidelines published ≥2007
    b. Excluded: Non-GRADE guidelines, Non-TB recommendations
        - Non-GRADE guidelines: guidelines that do not include an evidence profile/evidence to decision framework. Note, some guidelines follow GRADE but did not provide a grade for each recommendation. These guidelines are still included.
        - Non-TB recommendations: guidelines considering more than one condition are eligible, if they include TB-specific recommendations (i.e. one or more PICO element is related to TB). Only the TB-recommendations should be extracted from these guidelines.

2. Definitions:
    a. Copy and past definitions from the glossary/definitions section of guideline in the "Definitions in Guideline" tab of extraction form. Please be sure to link definitions with the Guideline ID they correspond to.
        - Please note, reading through and pulling definitions will help you in making informed choices in extraction. (e.g. the age-range for a "child" will be different from one guideline to the next).

3. Pull out and label Recommendations:
    a. Copy and paste verbatim recommendations as they are published.
        - These may be located near the beginning in an Executive Summary, in a separate "Recommendations" section, or written throughout the main text of the guideline. They will always be clearly indicated as a recommendation!
    b. Label each recommendation with a unique identifier in the "Recommendation Identifier" row 5, using the following format: Guideline ID_Rec# . Ex) the third recommendation for the 2019 consolidated MDR guideline is: WHO/CDS/TB/2019.7_3

4. Pull out Remarks:
    a. A remarks section is found at the end of the column for each recommendation. Plot remarks beneath the recommendation it corresponds to. Please note, not all Recs have Remarks and some Recs share the same remarks, so there may be duplications (i.e. insert remarks (albeit the same) for each recommendation it corresponds to).

5. Complete Extraction Form Entries for Recommendations:
    a. Refer to Data Extraction Sheet CLOSELY for explanation as to what should be extracted (verbatim entries, codes, etc.). The Extraction Sheet will tell you when to select for an item, and when to enter a code, name, number, etc.
    b. For Population, codes refer to either:
        - SNOMED-CT
        - ICD-11
    c. For Intervention/Comparison codes, refer to either:
        - ATC
        - SNOMED-CT
    d. Special instructions for certain rows in the form:
        - "Stage of Recommendation" (Note that for some consolidated guidelines will provide information on whether the rec is new, updated-previous, or unchanged previous. If this info is not available, simply select 'consolidated' option).
        - "Evidence Profile" (Look in Annex, sometimes a separate document)
        - "EtD" (Look in Annex, sometimes a separate document)
        - "Evidence synthesis for Evidence Profile" (Look in Annex, sometimes a separate document)
        - "Reference to evidence synthesis document" (Look in Annex, sometimes a separate document, or a link is provided in a reference)
        - "Intent of the Recommendation" Can be thought of as being synonymous with the intent of the entire guideline (if entire guideline is screening, diagnostics, etc.), or synonymous with the section of the guideline it corresponds to in a consolidated guideline (i.e., treatment, prevention, etc.). Please note, there is more than one option, as there may be more than one correct categorization (e.g. screening & TB Preventative Treatment).
        - "Intended Population" Categorize the recommendation by the main population(s) considered by the recommendation. This can be easily identified by the population extracted. Like Intent of Recommendation, there may be more than one intended population.

6. Supplement Verbatim Recommendation: Use BLUE INK.
    a. Insert additional information that is referred to/implied in the recommendation. **Remember to assign CODES to the additional information added using BLUE INK

b. Supplementation of a Rec should be done by consulting THREE sources in a guideline::

- 1) Information contained in superscripts & subscripts (1)
  - Rec:

> Enhanced adherence counselling[4] is suggested for all individuals initiating HIV post-exposure prophylaxis (conditional recommendation, moderate-quality evidence).

  - Superscript for 4 found in footnotes below:

[1] Provisional definition of substantial risk is defined as HIV incidence higher than 3 per 100 person-years in the absence of PrEP.
[2] Backbone regimen refers to the two-NRTI component of an ART regimen (normally comprising 3 ARV drugs).
[3] NVP should not be used in children above the age of two years.
[4] Enhanced adherence counselling includes baseline individual needs assessment, adherence counselling and education sessions and follow-up telephone calls.

  - See example in Rec. 6.
- 2) Information contained in Recommendation Remarks (2)
  - Rec:

Recommendations

2.1. In MDR/RR-TB patients on longer regimens, all three Group A agents and at least one Group B agent should be included to ensure that treatment starts with at least four TB agents likely to be effective, and that at least three agents are included for the rest of treatment after bedaquiline is stopped.[19] If only one or two Group A agents are used, both Group B agents are to be included. If the regimen cannot be composed with agents from Groups A and B alone, Group C agents are added to complete it (conditional recommendation, very low certainty in the estimates of effect).

  - Remarks found later in guideline:

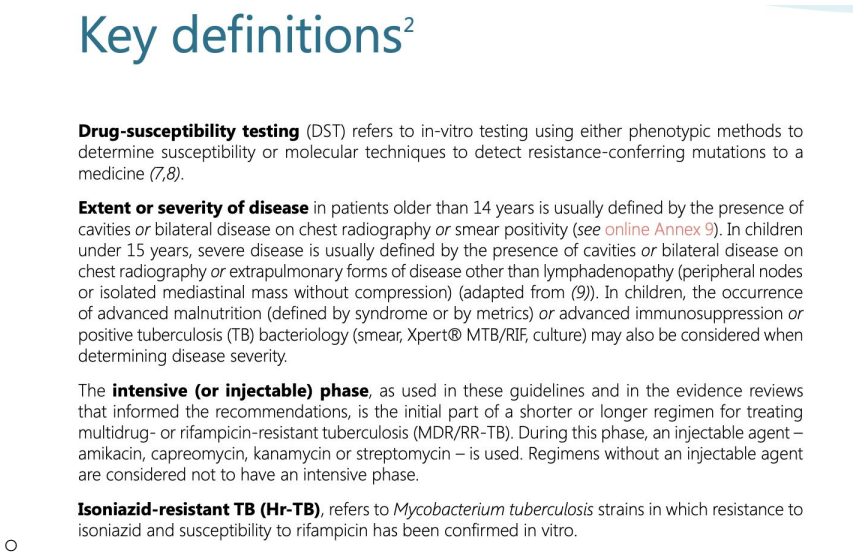
Remarks

The GDG assessed the individual contribution to patient outcomes of medicines used in longer MDR-TB regimens using primarily the estimates of effect from the 2018 IPD-MA and Trial 213 (delamanid) for PICO question 2 (MDR/RR-TB, 2018; *see* online Annexes 7 and 8 for the respective GRADE [Grading of Recommendations Assessment, Development and Evaluation] summaries of evidence for each medicine as well as the evidence-to-decision framework). Following a thorough assessment of the relative benefits and harms, recommendations were made for each medicine and they were classified into three groups (*see* Tables 2.1–2.3).

- **Group A**: fluoroquinolones (levofloxacin and moxifloxacin), bedaquiline and linezolid were considered highly effective and strongly recommended for inclusion in all regimens unless contraindicated.
- **Group B**: clofazimine and cycloserine or terizidone were conditionally recommended as agents of second choice.
- **Group C**: included all other medicines that can be used when a regimen cannot be composed with Group A and B agents. The medicines in Group C are ranked by the relative balance of benefit to harm usually expected of each.

  - See coded example in Rec. 10 in Extraction Form

- 3) Information contained in Guideline Definitions (1)



Key definitions[2]

Drug-susceptibility testing (DST) refers to in-vitro testing using either phenotypic methods to determine susceptibility or molecular techniques to detect resistance-conferring mutations to a medicine *(7,8)*.

Extent or severity of disease in patients older than 14 years is usually defined by the presence of cavities *or* bilateral disease on chest radiography *or* smear positivity (*see* online Annex 9). In children under 15 years, severe disease is usually defined by the presence of cavities *or* bilateral disease on chest radiography *or* extrapulmonary forms of disease other than lymphadenopathy (peripheral nodes or isolated mediastinal mass without compression) (adapted from *(9)*). In children, the occurrence of advanced malnutrition (defined by syndrome or by metrics) *or* advanced immunosuppression *or* positive tuberculosis (TB) bacteriology (smear, Xpert® MTB/RIF, culture) may also be considered when determining disease severity.

The **intensive (or injectable) phase**, as used in these guidelines and in the evidence reviews that informed the recommendations, is the initial part of a shorter or longer regimen for treating multidrug- or rifampicin-resistant tuberculosis (MDR/RR-TB). During this phase, an injectable agent – amikacin, capreomycin, kanamycin or streptomycin – is used. Regimens without an injectable agent are considered not to have an intensive phase.

Isoniazid-resistant TB (Hr-TB), refers to *Mycobacterium tuberculosis* strains in which resistance to isoniazid and susceptibility to rifampicin has been confirmed in vitro.

- What Extractors SHOULD NOT Do:

The program that pulls extracted information and codes requires streamlined input in the spreadsheet. Any slight deviation will result in an erroneous reading, and result in problems for our IT collaborators. Please avoid the following:

1. Do not merge cells, even if the information from one cell is the same for the next. Each cell requires its own input.
2. The 'Notes' section should not include any data that can be extracted in the form
3. Definitions should guide data entry (for instance, age categories, regiment length, etc.)
4. Do not try to fill sections based on what you think it might be. Leave it blank if it is not clear from the recommendation.

References

1. WHO consolidated guidelines on drug-resistant tuberculosis treatment. Geneva: World Health Organization; 2019. Licence: CC BY-NC-SA 3.0 IGO.
2. Consolidated guidelines on the use of antiretroviral drugs for treating and preventing HIV infection: recommendations for a public health approach – 2nd ed.

Appendix C: Table 1. Data Items

Population	Intervention/Comparisons	Outcome	Other
Age	Route of administration (Oral p.o., intravenous i.v., intramuscular i.m., subcutaneous s.c., other)	Intended outcomes	Evidence Profile
Condition	Single dose amount	Intended Outcome Impact (reduced detriment, increased benefit, increased benefit + reduced detriment, undefined, not applicable)	EtD
Confirmation (Present undefined, present confirmed, present assumed, probably present, suspected, probably absent, absent undefined, absent confirmed, not suspected, not applicable)	Dose unit (milliliter ml, liter l, kilogram kg, gram g, milligram mg, microgram ug, gram per square meter g/m2, milligram per square meter mg/m2, gram per kg body mass g/kg, millimole mmol, micromole umol, unit U, international unit IU, thousand unit kU, other)		Stated Direction of the Recommendation (positive for intervention, positive for comparison, negative against intervention, negative against comparison, either intervention or comparison, unclear, not stated)
Activity (undefined episode, first episode, relapse, recurrence, remission, resolved, other, not-applicable)	Schedule Number (per second, per minute, per hour, per 12 hours, per 24 hours, per day, per week, per month, per year, other)		GRADE Strength (strong, conditional, unclear, not applicable, other, not stated)
Risk Group (Not applicable, low-susceptibility/risk to a condition, high-susceptibility (risk) to a condition), undefined susceptibility to a condition, high risk for exposure to a condition, other)	Treatment Duration		GRADE Certainty (very low, low, moderate, high, unclear, other, not stated)
Setting- Income Category (high income, LMIC, UMI, LMI, low income, unclear)	Treatment Duration Unit (single dose, minute, house, day, week, month, year, other)		Intent of the Recommendation (screening, diagnostic, preventative, therapeutic, palliative)
Setting- Statistics (TB low prevalence, TB low/medium prevalence, TB medium prevalence, TB medium/high prevalence, TB high prevalence, TB endemic, HIV low prevalence, HIV high prevalence, TB low incidence, TB high incidence, TB high incidence & transmission, TB high transmission, INH resistance low prevalence, INH resistance high prevalence, other, not applicable)	Treatment Phase (intensive phase, continuation phase, total treatment intensive + continuation, unspecified phase, other)		Evidence Synthesis for Evidence Profile
	Intended Intervention Provider (public TB control program, non-program public sector, private providers, local TB control program, NTP, patient her/himself self administered therapy, family member, lay person, community health worker, lay person OR community health worker, healthcare worker, note stated, not applicable, other)		Reference to Evidence Synthesis Document
	Intervention Setting (home, community, health care facility, undefined setting, not stated, not applicable)		

Appendix D: Table 2. Multi-GRADE Recommendations

Bedaquiline should be included in longer MDR-TB regimens for patients aged 18 years or more (strong recommendation, moderate certainty in the estimates of effect). Bedaquiline may also be included in longer MDR-TB regimens for patients aged 6–17 years (conditional recommendation, very low certainty in the estimates of effect)	WHO/CDS/TB/2019.7
One or more of the following treatment adherence interventions (complementary and not mutually exclusive) may be offered to patients on TB treatment or to health-care providers: a) tracers and/or digital medication monitor (conditional recommendation, very low certainty in the evidence); b) material support to the patient (conditional recommendation, moderate certainty in the evidence); c) psychological support to the patient (conditional recommendation, low certainty in the evidence); d) staff education (conditional recommendation, low certainty in the evidence)	WHO/CDS/TB/2019.7
The following treatment administration options may be offered to patients on TB treatment: a) Community- or home-based DOT is recommended	WHO/CDS/TB/2019.7

over health facility-based DOT or unsupervised treatment (conditional recommendation, moderate certainty in the evidence). b) DOT administered by trained lay providers or health-care workers is recommended over DOT administered by family members or unsupervised treatment (conditional recommendation, very low certainty in the evidence). c) Video-observed treatment (VOT) may replace DOT when the video communication technology is available, and it can be appropriately organized and operated by health-care providers and patients (conditional recommendation, very low certainty in the evidence).	
The following options are recommended for the treatment of LTBI regardless of HIV status: 6 or 9 months of daily isoniazid, or a 3-month regimen of weekly rifapentine plus isoniazid, or a 3 month regimen of daily isoniazid plus rifampicin. (Strong recommendation, moderate to high certainty in the estimates of effect). A 1-month regimen of daily rifapentine plus isoniazid or 4 months of daily rifampicin alone may also be offered as alternatives. (Conditional recommendation, low to moderate certainty in the estimates of effect).	WHO/UCN/TB/2020.1

CHAPTER 3: Conclusion

3.1 Summary of Findings

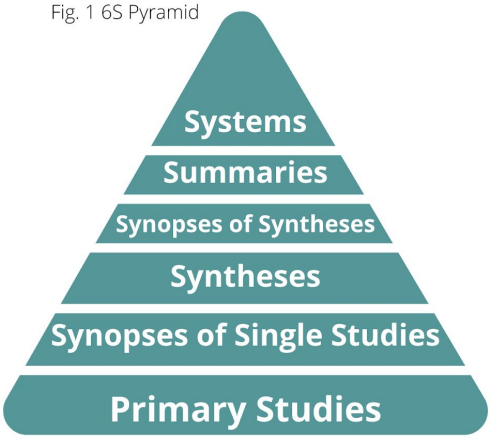
This work focuses on the introduction of evidence mapping methods to guidelines in the form of RM, and also briefly accounts for the related concept of mapping the PICO questions leading the development of a guideline. We define RM as; the organization of recommendations from a series of guidelines targeted towards a certain condition, based on the features of the recommendation, the evidence informing the recommendation, and by the PICO subdomains included. The end product is an interactive platform of recommendations, organized by a cross-tabulation of the PICO subdomains, a complementary searchable database, serving as a centrally accessible hub of all guideline recommendations.

The methods developed in this thesis were applied to WHO-GTB issued guidelines. Upon examining all WHO-GTB publications, 83 guidelines were screened, and 20 were finally included. From these, 211 recommendations were extracted in full, along with 42 PICO questions from the last 5 years. Based on the raw data of extraction, both a recommendation map and a searchable

database were built in collaboration with a health technology company affiliated with McMaster University; Evidence Prime Inc.

Recommendations culminate a large body of evidence into a concise, actionable, and informed statement. Many attempts to standardize the semantics and formatting of recommendations in the writing process have been undertaken but they are neither formalized nor quality assured with regards to the utility for this mapping exercise. There may be many characteristics that need to be considered in writing a recommendation such as brevity, clarity, intention, and content. Problems with these features can deter recommendations from serving their ultimate purpose in being implemented. Issues around clarity and specificity for instance, may impede the incorporation of recommendations into practice although we did not evalute this for the GTB recommendations specifically in this project (3,4). With regards to the emphasis and intention of a recommendation, the useof deontic logic (logic that concerns notions of obligation and permission) must be commonly understood in order to implement recommendations consistent with the developers' intention (5). In a 2006 review from the WHO Advisory Committee on Health Research to the WHO, initial guidance was issued as to how clinical, public health and health systems recommendations should be reported (6). Using a checklist from the Conference on Guideline Standardization (COGS), recommendations and rationale were initially advised to "[s]tate the recommended action previously and the specific circumstance under which to perform it. Justify each recommendation by describing the linkage between the recommendation and its supporting evidence. Indicate the quality of evidence and the recommendation strength..."(7).

Fig. 1 6S Pyramid



Since then, most of the literature around articulating the final recommendation, revolves around one of the aforementioned features; intent. Concerns around the implied emphasis of deontic logic in recommendations was considered by Lomotan et al., who found that 'must' conveys the highest level of obligation, 'may' and 'may consider' conveys the lowest levels of obligation, and 'should' and all other deontic terms conveyed intermediate levels of obligation (8). In a similar vein, Akl et al.'s 2012 RCT on conveying the strength of a recommendation found that wording for strength and direction (for or against an intervention), found that no one wording approach performed better in conveying the strength of clinical recommendation (8). In 2013 the GRADE Working Group outlined an approach to categorizing, labeling and wording health care recommendations. Again here, there was an emphasis on the articulation of recommendations as it pertains to the intention behind the assigned strength; 'we recommend...' and 'we suggest...' for strong and weak recommendations respectively (8).

Through this recommendation mapping work, a feature of recommendation articulation that was visibly heterogenous and problematic during extraction, was the content of recommendations. Our findings are in line with previous work in this field. A representative sample of guideline recommendations used to elucidate how recommendations are written, called the *Yale Guideline Recommendation Corpus*, found that many recommendations were not executable as they were written. Inconsistency in strength of recommendation reporting was also apparent in their 2009 publication (9,10). However, if the GRADE approach is assumed to cover the bases for reporting the strength and certainty of evidence, then what remains is a not a need to understand the semantics and formatting of a recommendation's *emphatic intent*, but the semantics and formatting of recommendation *content*.

From the rough data (single extraction, not duplicate) from the pilot of all WHO TB recommendations published ≥2014, some insight into the heterogeneity of formatting of recommendations can be gleaned. One of the assumptions of recommendation mapping is that

recommendations have a PICO backbone that reflects the PICO question from which they arose. Many recommendations were missing P, C, and O. On the flipside, other recommendations were densely filled with multiple PICO components. The semantics of PICO components had varying degrees of clarity. For example, a recommendation from the 2018 *Latent tuberculosis infection (11). Updated and consolidated guidelines for programmatic management* guideline mentions children and adult in the population parameters, but presumably the age range in the recommendation could also apply to adolescents.

> *Isoniazid monotherapy for 6 months is recommended for treatment of LTBI in both adults and children in countries with high and low TB incidence. (Strong recommendation, high-quality evidence. Existing recommendation) (11)*

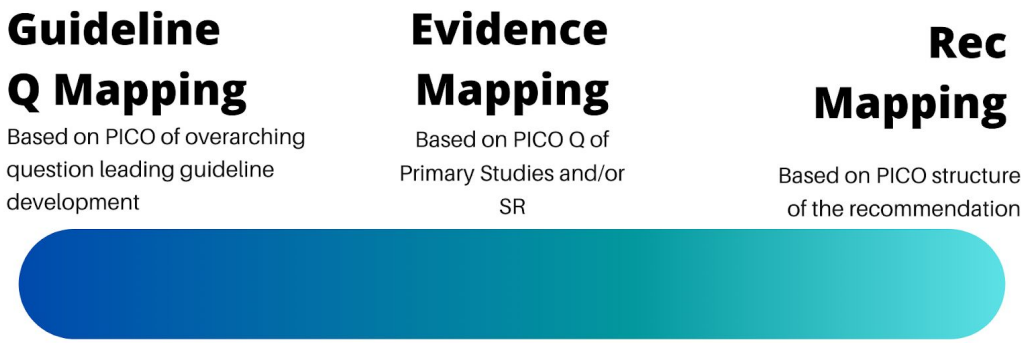
Some redundancies were apparent in the extraction process. For instance, in the 2018 *Latent tuberculosis infection. Updated and consolidated guidelines for programmatic management* guideline, the first recommendation below outlines what a test for LTBI entails, which could perhaps be added to other recommendations that reference LTBI testing (11). In the second recommendation below, what the intervention (LTBI test) actually entails is a pseudo gap in the comprehensiveness of the verbatim intervention extraction, as this information is available in another recommendation.

> *Either a tuberculin skin test (TST) or interferon-gamma release assay (IGRA) can be used to **test for LTBI**. (Strong recommendation, very low-quality evidence. New recommendation) (11)*

*Systematic **testing for LTBI** is not recommended for people with diabetes, people with harmful alcohol use, tobacco smokers and underweight people unless they are already included in the above recommendations. (Conditional recommendation, very low-quality evidence. Existing recommendation) (11).*

Complementing the recommendation extraction with guideline PICO question extraction was fitting. The PICO questions that guide the development of a guideline, can be thought of as the most upstream, outstanding questions we have to ask the literature and guideline developers. If evidence mapping methods could be placed along a continuum it would appear as the following:

Fig. 2 Evidence Mapping Continuum



Recommendations can be thought of as the finalmost, bottom-line answers we have at a given date, based on the available evidence considered. If RM is purported to be of benefit to guideline developers, then a complementary PICO question map may also prove useful for initial stages of guideline development. Both RM and PICO Q mapping may provide a comprehensive overview of the direction guideline developers have taken, provide a schematic of the charted terrain of questions and recommendations, and hopefully inform their future path. Similarly, plotting the

evidence that informs guidelines through standard evidence mapping methods may also complete the picture.

3.2 Implications

The main implication of this work, are suggestions around guideline structuring and recommendation articulation, in light of the learning from in-depth data extraction for the mapping. In general, there is a narrow range of stylistic freedom between GRADE guideline development and the final writing and publication of the resulting guideline. GRADE provides not only the methodological underpinning for development, but also direction for structure (GRADE Handbook to be updated soon)). There is a very small gap that exists between this direction and its implementation, which facilitates room for just enough liberty to compromise the rigour of development with poorly articulated end-products. It is within this space where multiple and missing PICO elements arise, where evidence documents (evidence syntheses, profiles, EtDs) are difficult to trace/link, where unstandardized organization of guideline components (such as remarks) are completely absent or inconsistently detailed, and additional actionable statements are haphazardly included, convoluting the clarity and station of recommendations. For this reason, we are developing a tool using the quintessential components of the extraction forms developed for mapping, to assist in the formulation of a final recommendation. This will assist guideline writers in paying special attention to the omission of key PICO subdomains in the question and/or recommendation, and similarly, multiplicity in these elements.

3.3 Limitations and Strengths of Work

Many methods employed, and outputs developed in this thesis are novel. This allowed for great opportunity to be creative in the development of the methods, and also presented challenges. This work was strengthened by a strong collaboration with all involved in the larger

project this thesis contributes to. It was through this continued process of consulting, and learning, that the work was strengthened. As with appropriate evidence-informed public practice, ensuring integrated and continuous quality improvement measures were of paramount importance. This thesis contributed to a larger international collaboration between the McMaster University, the WHO, Evidence Prime Inc., and opportunities for feedback from all key players were continuously sought to create a continued pattern of implementation and evaluation. Ultimately, discussions with WHO-GTB staff, GDG, and among collaborators, were of paramount importance to streamlining the methods presented. In terms of limitations, extraction of the materials was not completed in duplicate, which could threaten the validity of the results. Additionally, focusing on guidelines conducted following the GRADE approach led to preferential inclusion of clinical over public health recommendations, due to a misclassification of health-system guidance as either best-practice statements or informal recommendations. The methods presented are not solely for clinical guidelines, however, monitoring/evaluation, programmatic management, and models of care, were not trialed in this mapping exercise of WHO-GTB recommendations.

3.4 Further Research

Future research testing evidence mapping methods in the realm of guidelines, is needed. Further, studies assessing the utility of such outputs for guideline developers, stakeholders, and end-users, will also be warranted. Measures of utility may include a comparative assessment between accessibility of WHO-GTB recommendations as formerly presented versus the digital representation, qualitative assessment of perceptions around usefulness of a visual of gaps and clusters of recommendations for guideline priority setting.

3.5 Final Remarks

The application of evidence mapping methodologies to guidelines is both possible and necessary in light of the demands of a rapidly learning health system. In thinking about the need for living guidelines based on live, continually updated evidence syntheses, and the potential for atomization of question development, converting the key components of guidelines (questions, evidence, recommendations) to data through the RM and related methods, will be a prerequisite step towards achieving these living systems. Further, the insights gleaned through an application of RM to WHO-GTB guidelines, provides productive implications for guideline developers, in terms of the construction of guidelines with the long-term vision of integrating artificial intelligence in mind.

4.6 References

1. 6S Search Pyramid | National Collaborating Centre for Methods and Tools [Internet]. Nccmt.ca. 2020 [cited 11 March 2020]. Available from: http://nccmt.ca/capacity-development/6s-search-pyramid
2. Shekelle P, Kravitz R, Beart J, Marger M, Wang M, Lee M. Are nonspecific practice guidelines potentially harmful? A randomized comparison of the effect of nonspecific versus specific guidelines on physician decision making. Health Serv Res [Internet]. 2000 [cited 11 March 2020];. Available from: https://www.ncbi.nlm.nih.gov/pubmed/10737446
3. Michie S, Johnston M. Changing clinical behaviour by making guidelines specific. BMJ. 2004;328(7435):343-345.
4. VON WRIGHT G. I. DEONTIC LOGIC. Mind. 1951;LX(237):1-15.
5. Oxman, A.D., Schünemann, H.J. & Fretheim, A. Improving the use of research evidence in guideline development: 14. Reporting guidelines. *Health Res Policy Sys* 4, 26 (2006) doi:10.1186/1478-4505-4-26
6. Shiffman RN, Shekelle P, Overhage JM, Slutsky J, Grimshaw J, Desh- pande AM: **Standardized Reporting of Clinical Practice Guide- lines: A Proposal from the Conference on Guideline Standardization.** *Ann Intern Med* 2003, **139:**493-8.

7. Lomotan EA, Michel G, Lin Z, Shiffman RN. How "should" we write guideline recommendations? Interpretation of deontic terminology in clinical practice guidelines: survey of the health services community. *Qual Saf Health Care*. 2010;19(6):509–513. doi:10.1136/qshc.2009.032565

8. Akl E, Guyatt G, Irani J, Feldstein D, Wasi P, Shaw E et al. "Might" or "suggest"? No wording approach was clearly superior in conveying the strength of recommendation. Journal of Clinical Epidemiology. 2012;65(3):268-275.

9. Hussain T, Michel G, Shiffman RN. The Yale Guideline Recommen- dation Corpus: a representative sample of the knowledge content of guidelines. Int J Med Inform 2009;78:354e63.

10. Hussain T, Michel G, Shiffman R. The Yale Guideline Recommendation Corpus: A representative sample of the knowledge content of guidelines. International Journal of Medical Informatics. 2009;78(5):354-363.

11. Latent tuberculosis infection: updated and consolidated guidelines for programmatic management. Geneva: World Health Organization; 2018.