Which New Diagnostics for Tuberculosis, and When?

Frank Cobelens,[1] Susan van den Hof,[1,2] Madhukar Pai,[3] S. Bertel Squire,[4] Andrew Ramsay,[5] and Michael E. Kimerling[6] on behalf of the Evidence for Scale-up Group

[1]Department of Global Health, Academic Medical Center; and Amsterdam Institute of Global Health and Development, Amsterdam, and [2]KNCV Tuberculosis Foundation, The Hague, Netherlands; [3]Department of Epidemiology and Biostatistics, McGill University, Montreal, Canada; [4]Clinical Research Group, Liverpool School of Tropical Medicine, United Kingdom; [5]United Nations Children's Fund/United Nations Development Programme/World Bank/World Health Organization (WHO) Special Programme for Research and Training in Tropical Diseases, WHO, Geneva, Switzerland; and [6]Bill and Melinda Gates Foundation, Seattle, Washington

Recently, new diagnostic tools for tuberculosis detection and resistance testing have become available. The World Health Organization endorses new tuberculosis diagnostics by using the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) process. This endorsement process takes place when limited evidence beyond test accuracy is available. There is a need to provide guidance to tuberculosis programs about which new diagnostics to scale up and how best to position them in diagnostic algorithms. To speed adoption of new diagnostics for tuberculosis, the policy recommendation process should be revised to consist of 2 steps: technical recommendation and programmatic recommendation. Technical recommendation would follow the GRADE process and be based on accuracy with limited cost and feasibility data, while programmatic recommendation would include patient-important outcomes, cost-effectiveness when implemented under routine conditions, and factors critical to successful scale-up. The evidence for both steps should be systematically collected, but each requires different study designs.

The past decade has seen a sharp increase in the development of new tools for diagnosing tuberculosis and tuberculosis drug resistance. Until recently, the diagnostic tools available in low-income and middle-income countries were largely limited to microscopic examination of sputum smears and mycobacterial culture using solid media [1, 2]; drug susceptibility testing was performed by use of slow phenotypic methods and was often only done in periodic surveys. Since 2007, however, several tools with improved sensitivity and speed of diagnosis have been endorsed by the World Health Organization (WHO) (Figure 1) [4, 5].

Here we argue that the current process of endorsement is necessary but insufficient to help countries and donors decide whether and how a new diagnostic should be scaled up and that formal guidance should include the diagnostic tool's effect on patient-important outcomes when applied at a large scale; the most effective and cost-effective positioning within possible diagnostic algorithms; and factors critical to whether scale-up will be successful. We discuss such guidance in the context of a staged evaluation pathway for new diagnostics and the type of evidence that needs to be collected in each of these stages. Finally, we suggest how the required evidence can be collected in a systematic and reproducible way.

LIMITATIONS OF THE CURRENT ENDORSEMENT PROCESS

WHO issues recommendations and guidelines for tuberculosis control, including endorsement of new diagnostic tools and methods, for which it has adopted the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) process [6]. GRADE facilitates a transparent assessment and categorization
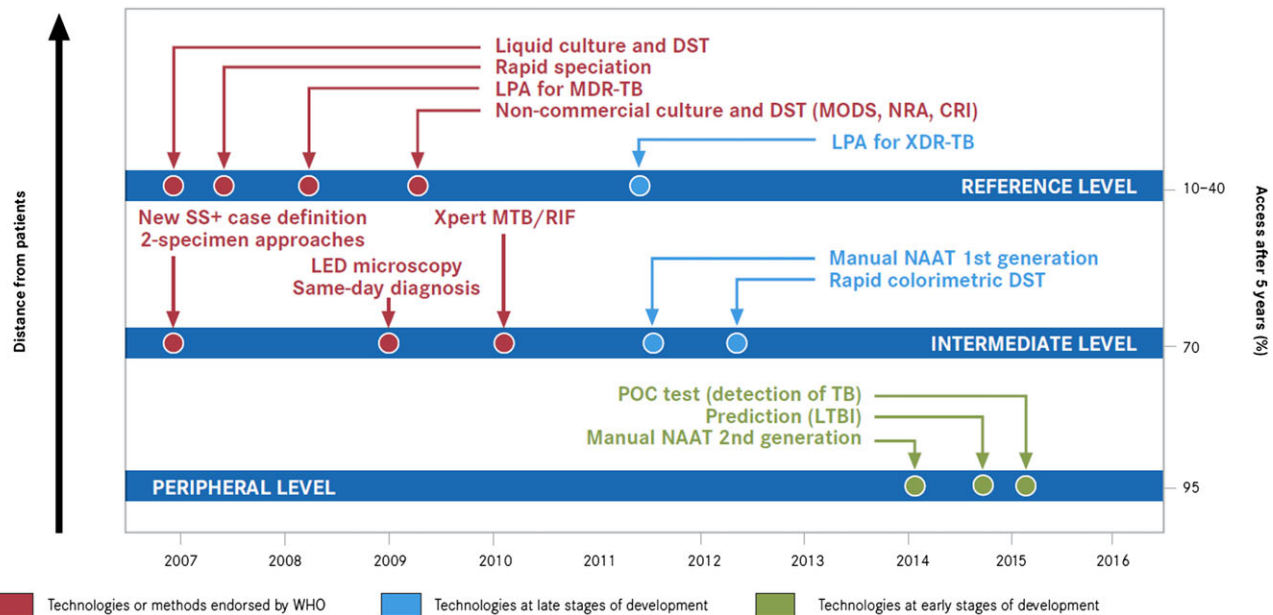
Figure 1. Tuberculosis diagnostics pipeline in 2011. "Distance from patients" refers to where the diagnostic technology is placed or used (reference laboratory or center, intermediate level, and peripheral level), with peripheral level being closest to point-of-care testing. "Access after 5 years (%)" refers to the proportion of patients who will be able to access the test at each level of the healthcare delivery system. Reproduced with permission from [3] Abbreviations: CRI, colorimetric redox assay; DST, drug susceptibility testing; LED, light-emitting diode; LPA, line-probe assay; LTBI, latent tuberculosis infection; MDR-TB, multidrug-resistant tuberculosis; MODS, microscopically observed drug-susceptibility assay; NAAT, nucleic acid amplification test; NRA, nitrate reductase assay; POC, point-of-care; SS+, sputum smear-positive; TB, tuberculosis; XDR-TB, extensively drug-resistant tuberculosis.

of quality of evidence and strength of recommendations on patient-important effects (such as reductions in duration and severity of disease) for a wide range of medical interventions (Table 1) [7]. WHO endorsement provides countries and donors with the stimulus to adopt and invest in a new diagnostic and is an important step in the process by which the diagnostic is introduced to tuberculosis control programs.

However, policy makers, tuberculosis program managers, and donors not only need to know that a new tool can improve the diagnosis of tuberculosis or tuberculosis drug resistance or that it can do so in specific patient populations (eg, children).

Table 1. Characteristics of the Grading of Recommendations Assessment, Development, and Evaluation System for Rating Quality of Evidence and Strength of Recommendations

Clear separation between quality of evidence and strength of recommendations
Explicit evaluation of the importance of outcomes of alternative management strategies
Explicit, comprehensive criteria for downgrading and upgrading quality of evidence ratings
Transparent process of moving from evidence to recommendations
Explicit acknowledgment of values and preferences
Clear, pragmatic interpretation of strong versus weak recommendations for clinicians, patients, and policy makers

Adapted from [7].

They also need evidence that it will do so when applied routinely in a tuberculosis program and that it provides the best option within limits of the available resources for the epidemiological setting in which they operate [8]. There is increasing recognition that WHO endorsement is not sufficient for a successful adoption and scale-up of new tuberculosis diagnostics and that a broader base of "evidence for scale-up" is needed [9, 10]. Four reasons stand out for a change in the approach to introducing new diagnostics.

First, a primary concern for introduction of a new test is whether it improves patient-important outcomes and avoids false-positive diagnoses. For example, in regions of endemicity, most cases of sputum-smear negative tuberculosis are diagnosed clinically, but these often include false-positive diagnoses [11, 12]; even when culture is available for bacteriological confirmation, it is often not used because the result comes too late to influence treatment decisions [13]. The evidence base for GRADE-based endorsements of new tuberculosis diagnostics has been largely limited to data on test accuracy (sensitivity and specificity) only [4] because other data are not available [14], even though test accuracy should only be regarded as a surrogate for patient-important outcomes and consequently is rated as "low-quality" evidence by the GRADE process.

Second, new diagnostics are used in the context of existing tests. They should not be considered in isolation but as part

of diagnostic scenarios or algorithms incorporating several tests. A simple example of such a scenario is when patients presenting with symptoms suggestive of tuberculosis first have a smear examination and, if results are negative, have a new molecular test performed ("add-on strategy"). Alternatively, the new test could replace smear examination ("replacement strategy"), or a new point-of-care test may be first done in the community, and if results are positive, referral can be made for a confirmatory test ("triage" or "screening strategy"). The optimum positioning of any test in a diagnostic scenario has several determinants, including accuracy in different patient groups (eg, individuals with HIV infection), costs, and where to embed it in the healthcare system (eg, at the primary care level or in specialized laboratories). The current endorsement process, however, is based on weighing evidence about a single test only. Although the WHO has made recommendations for positioning the Xpert MTB/RIF assay in diagnostic scenarios [5], these are based on expert opinion rather than on systematically collected evidence.

Third, while operational aspects such as infrastructure requirements and biosafety are considered in the GRADE process, the considerations involve limited or no evidence on what these operational aspects imply for large-scale implementation. A number of WHO-endorsed diagnostics have undergone evaluation in routine settings in so-called demonstration studies. These studies have contributed important evidence on test accuracy under program conditions, on operational aspects, and on some patient-important outcomes (eg, turnaround times of test results) [15], but most were done on a moderate scale at sites selected for their capacity to perform the study (ie, sites where staff training and quality assurance were intensified). Therefore, these demonstration studies tend to overestimate the effectiveness of the new diagnostic in routine use and have limited ability to identify critical factors for successful scale-up [16]. For some diagnostics, no such demonstration studies have been attempted [17].

Fourth, cost-effectiveness and affordability of a new diagnostic are major considerations for its introduction. Many of the newly developed tuberculosis diagnostics have higher per-patient costs than existing tests, so these need to be weighed against better diagnostic performance, improvement in patient outcomes, and potential savings in areas such as staff time, direct patient costs, and unnecessary treatment. Analyses that give definite answers about the cost-effectiveness of a new diagnostic should be based on observations in the settings in which the test is to be implemented and, ideally, should compare various scenarios in various epidemiological and economic settings. While data from demonstration studies have been used in prediction models [18], their outcomes can only be regarded as indicative of the cost-effectiveness that can be achieved when the diagnostic is implemented on a programmatic scale.

THE NEED FOR GUIDANCE ON SCALE-UP OF NEW DIAGNOSTICS

This lack of evidence is compounded by the arrival of increasing numbers of diagnostics, each with its own (often partially overlapping or even competing) potential place in the diagnosis of tuberculosis and/or tuberculosis drug resistance. As shown in Figure 1, there are now 6 WHO-endorsed tests for detection of drug resistance: liquid culture, line-probe assays, the microscopically observed drug-susceptibility (MODS) assay, nitrate reductase assays (NRAs), colorimetric redox indicator (CRI) methods, and Xpert MTB/RIF. Liquid culture, MODS assay, NRAs, and CRI methods can test for susceptibility to various drugs, but because of biosafety concerns these methods will typically be implemented at the referral level [17]. Line-probe assays could be placed at an intermediate level in the healthcare system but have low sensitivity on smear-negative sputum samples [15]. Xpert MTB/RIF could be placed in peripheral laboratories but only for rifampicin-resistance testing [13]. For policy makers, tuberculosis program managers, and donors, choosing between (combinations of) these tests is increasingly difficult, with the risk that no choices are made. Given the range of possible combinations, this would be true even if sufficient evidence for scale-up was available for each test. Furthermore, some tuberculosis programs (eg, the program in India) that began scale-up of liquid culture, line probe assays, and LED fluorescence microscopy are now confronted with the pressure to adopt and scale-up Xpert MTB/RIF, resulting in "new-tool fatigue."

Therefore, guidance is needed for countries to make evidence-based decisions on adoption and scale-up of new diagnostics. Initial guidance was issued in 2008 by the STOP tuberculosis Partnership (available at: http://www.stoptb.org/wg/dots_expansion/inatmeetings.asp), but it lacked detail on how new tools could be incorporated within algorithms in routine programmatic settings.

While initiatives such as the STOP tuberculosis Partnership have a role in promoting the uptake of new tools, it is our view that the WHO should take responsibility for issuing guidance as to when, where, and how countries should scale up new diagnostics for tuberculosis control. The WHO is the only international body with the authority to issue guidelines that will be followed by national disease control programs, and it does so for various aspects of tuberculosis control.

The process of issuing such guidance should be separated from the endorsement process. This is needed because collecting the "evidence for scale-up" needed to formulate guidance takes time and requires initial (pilot) scale-up for which, in many countries, WHO endorsement is a prerequisite. The sequence and key elements of these steps can be defined by analyzing the
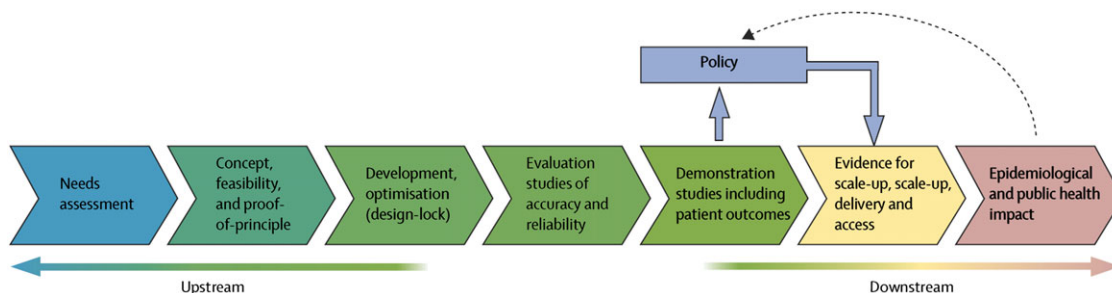
Figure 2. Schematic showing the pathway to tuberculosis diagnostics, formulated by the Stop tuberculosis Partnership's New Diagnostics Working group. Reproduced from [23].

current and ideal evaluation pathways for new tuberculosis diagnostics.

EVALUATION PATHWAY OF NEW DIAGNOSTICS IN TUBERCULOSIS

Several pathways and frameworks have been proposed for the development and evaluation of new diagnostics [19–22]. One of these was developed especially for tuberculosis (Figure 2) [23] and conceptualizes research needs as stages in a linear pathway of development, evaluation, policy recommendation, delivery, and impact. Although the latter pathway has elucidated the process for new tuberculosis diagnostics, some deficiencies and inconsistencies remain.

"Policy," or the recommendations and decisions associated with the introduction and scale-up of diagnostics, comes before evidence for scale-up is collected, and feedback of evidence into policy only occurs in the final stage of epidemiological and public health impact evaluation. In addition, "policy" has both international and country-specific dimensions. At the global level, the main policy consideration is whether a new test should be scaled up at all, whereas regional and national considerations are based on feasibility, added value to existing diagnostic practice, cost-effectiveness of the test in a diagnostic algorithm, and affordability. Consequently, the value chain (ie, all activities and evidence leading to policy decisions) at the global level is largely independent of local variations in epidemiology and health system, the value chain at the country level is setting specific and has larger implications for health systems. There is also wide variation between countries in their endorsement or policy processes; few countries have clearly worked out a process for adoption, policy, and scale-up of new tools for tuberculosis. Finally, the private sector value chain is not considered within the current global value chain. The widespread use of tuberculosis antibody–detection assays by the private sector in many countries where the tuberculosis burden is high, despite lack of evidence of benefit [24, 25], and the recent negative policy recommendation by WHO [26] show

that user perceptions of the private and public sectors can be very different.

Further complicating the understanding of the pathway is the lack of consensus on the exact meaning of some of the terms that denote the various types of evidence needed. The term "demonstration study" has no clear definition with regard to its scope and the required level of representativeness for routine healthcare settings. Neither is it clear what precisely comprises "evidence for scale-up." Finally, the term "impact" is widely used but interpreted differently. While the classical epidemiological definition of "impact" refers to the effect that risk factors or interventions have at the (affected) population level [27], GRADE requires evidence on the clinical impact at the individual level [6]. A recently developed impact assessment framework uses the term in even broader sense, with various layers of information relevant for policy decisions about new diagnostics, including cost-effectiveness from a patient and a health system perspective [10].

We propose a revised pathway (Figure 3) for new tuberculosis diagnostics that describes these steps in more detail. It is characterized by a distinction between technical and programmatic policy recommendation, with a distinction between global and country value chains.

Technical policy recommendation would be similar to what is currently called endorsement of a new diagnostic and would be based on available information about accuracy, potential for improving patient-important outcomes, costs, and feasibility. Programmatic policy recommendation would be issued about 2 years after technical recommendation, with the objective of advising countries and donors about whether and how the new diagnostic tool should be scaled up and how it should be positioned in the diagnostic process, taking into account specific epidemiological patterns, such as HIV coinfection and resistance to ≥1 drug. Criteria should be based on information about the performance and cost-effectiveness of the test when incorporated into ≥1 diagnostic algorithms, as well as its feasibility and resource needs when implemented at scale under routine programmatic conditions [10].
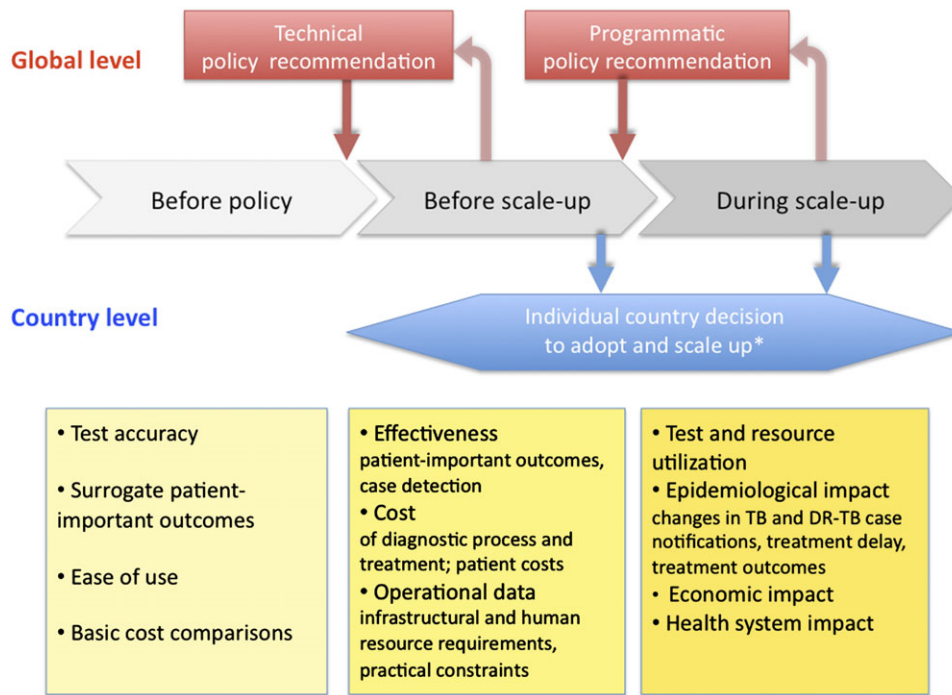
Figure 3. Proposal for a revised pathway focused on the postaccuracy phase of tuberculosis diagnostics, showing the proposed value chain for new diagnostic tests for tuberculosis. The grey arrows in the middle represent the stages in the evaluation pathway, and the colored boxes represent policy decisions at the global level (red) and the country level (blue). Countries would adopt implementation at different points and should provide feedback about their experiences (* in the blue box). In the stages before scale-up and during and after scale-up, evaluation data would be collected on diagnostic algorithms incorporating the new test.

This 2-step process breaks up the evaluation pathway of a new diagnostic into 3 distinct stages with regard to the evidence needed and the types of studies required: before policy, before scale-up, and during and after scale-up (Table 2). It also acknowledges that global and country value chains, while closely linked, are distinct, particularly with regard to their phasing (Figure 3). While most countries will probably scale up a new diagnostic tool only after global programmatic guidance has been issued, some countries (ie, "early adopters," particularly those that do not depend on international donors) may do so before global guidance is issued but ideally after having collected their own data supporting feasibility and scale-up.

EVIDENCE NEEDED IN THE VARIOUS STAGES OF EVALUATION

In the stage before policy, data need to be collected to inform technical policy recommendation only. The main question is whether the test has the technical requirements and operational potential to improve the diagnosis of tuberculosis, which is primarily relevant for making policy decisions at the global level. The data should come from field studies in tuberculosis-endemic countries, focusing on test accuracy, short-term patient-important outcomes, and ease of use.

Basic comparisons of total diagnostic costs (eg, costs that include sample transport) show whether the tool can be affordable and cost-effective. These studies would be done in routine care settings selected primarily for their ability to perform the study and would be based on comparison against a reference standard (eg, mycobacterial culture). The field studies would be limited with regard to their number, sample size, and duration.

In the stage before scale-up, data need to be collected in various settings to inform programmatic guidance. The main questions to be answered in this stage are about "evidence for scale-up," which we define as the effectiveness, cost-effectiveness, and operational requirements of the test when used at large scale in routine practice. This stage is relevant to both national and global policy decisions. Data should come from studies performed under routine programmatic conditions at the level in the healthcare system at which the tool is to be used (referred to as "implementation by research") [28], focusing on diagnostic algorithms or scenarios incorporating this tool. How the diagnostic is added to existing algorithms would be determined on the basis of the test's properties, such as the health system levels where it can be used (Figure 1), the presence of improved sensitivity in specific groups (such as HIV-infected patients), and whether it tests for drug resistance.

Table 2. Description of the Main Research Questions, Designs, and Outcomes, per Stage, in the Proposed Revised Pathway Focused on the Postaccuracy Phase of Tuberculosis Diagnostics

Stage	Global or Local	Main Questions (Type of Evidence)	Type of Study Designs	Outcomes Evaluated
Before policy	Global	Does the test have the technical requirements and operational potential to improve the diagnosis of tuberculosis?	Controlled field validation and demonstration studies in a limited number of well-controlled settings.	Test accuracy (sensitivity and specificity), surrogate patient-important outcomes (turnaround times, time to diagnosis and treatment, and improvement of case detection in the study population), ease of use, and basic cost comparisons.
Before scale-up	Global and local	What are the test's effectiveness, cost-effectiveness, and operational requirements when used in routine practice?	Designs involve pilot implementations under routine programmatic conditions at the level in the healthcare system at which the tool is intended to be used, focusing on diagnostic algorithms or scenarios in which the tool is incorporated, in a number of countries/settings selected for their representativeness for major epidemiological settings and resource levels. Designs include pragmatic, randomized, or cluster-randomized trials, such as those involving a phased-implementation design.	Effectiveness data to be collected would include improvement in case detection, time until treatment initiation, pretreatment morbidity and mortality, and treatment outcomes. Cost and effectiveness data would cover the entire process and include diagnostic and treatment costs, both for the health system and for the patient. Operational data would include infrastructural and human resource requirements and practical constraints to implementation.
During and after scale-up	Local	Is the new diagnostic tool implemented to its optimum effect? What are constraints in scale-up, cost, and resource projections to reach and sustain full scale-up? What is epidemiological impact?	Designs primarily involve monitoring and evaluation of data sets from routine recording and reporting systems. Targeted operational studies may be needed, as well as larger studies to estimate the epidemiological impact of the new diagnostic scenario, such as population surveys of the prevalence of tuberculosis or tuberculous mycobacterial infection or representative surveys of drug resistance.	Indicators involve use of the test and associated resources; changes in case notifications of tuberculosis and drug-resistant tuberculosis, in tuberculosis incidence and prevalence, in proportions of patients who test positive, in treatment delay, and in treatment outcomes; changes in patient expenditures for tuberculosis diagnosis; and broader effects on the health system, such as integration of diagnostic services for various disorders. These indicators would be monitored over time and between relevant segments of the population.

The postaccuracy phase of tuberculosis diagnostics is depicted in Figure 3.

Initial decision modeling can show the likely cost-effectiveness of various algorithms, which may be different depending on factors such as the HIV infection prevalence among patients with tuberculosis and the prevalence of multidrug resistance [18].

These studies should be done in several countries, primarily those selected for their epidemiological representativeness (eg, prevalence of HIV infection and multidrug-resistant tuberculosis), resources, and ability to collect quality data within the specified time frame. Effectiveness, cost-effectiveness, and operational data to be collected would take account of the entire process of implementation.

Comparison against a reference standard should not be required, and data collection should be based mainly on routine recording and reporting. Study designs should be comparative but focus on the intervention rather than on the individual test and therefore should be pragmatic, community randomized or clinic randomized, and could involve features such as phased implementation [29, 30]. Care needs to be taken to separate setting-specific effects (eg, quality of health services) from those that are generalizable across settings [16]. Coordination between the studies would be important to allow cost-effectiveness comparisons between the various diagnostic algorithms across the studies, through systematic reviews [31] or decision, operational, and transmission modeling to predict their cost-effectiveness in other settings [32, 33]. Also the cost-effectiveness of introducing the new diagnostic should be compared with that of alternative tuberculosis control interventions, and its impact on government budgets should be analyzed.

The final stage, during and after scale-up, acknowledges that scale-up of new diagnostics within a country takes time and often happens in a phased manner. This stage is meant to inform tuberculosis programs, policy makers, and donors about whether the new diagnostic tool is implemented to its

optimum effect, about the constraints in scale-up that need to be overcome, about cost and resource projections to reach and sustain full scale-up, and about the true epidemiological impact of the tools when brought to scale. This stage is thus primarily relevant to national policy decisions, although the synthesis of evidence from multiple countries may prompt adjustment of programmatic guidelines issued earlier. Data would be primarily collected through monitoring and evaluation of routine recording and reporting systems focusing on the entire diagnostic scenario. Larger studies may be done to estimate the epidemiological impact of that scenario, such as population surveys of the prevalence of tuberculosis or tuberculous mycobacterial infection or representative surveys of drug resistance. Results must be fed back to the WHO for review and modification of existing policies.

IMPLEMENTING THE PROPOSED EVALUATION PATHWAY

The proposed pathway for evaluating new tuberculosis diagnostics has several advantages. The separation of technical and programmatic policy recommendation allows all stakeholders to distinguish between the statement about whether a particular test has sufficient potential to be used in tuberculosis control, and the statement about how, where, and under which conditions this test should be implemented. It acknowledges that these statements require different types of evidence that can only be collected in a phased manner. Clarity for policy makers, tuberculosis program managers, and donors about this 2-step process will allow them to choose between scaling up after technical policy recommendation only (early adoption) and waiting until the WHO has issued a detailed programmatic policy recommendation. Countries that adopt the new technology early can play an important role in collecting the evidence needed in the stage before scale-up.

Furthermore, defining the evidence needed in each of the evaluation stages will bring clarity to countries, researchers, and funders about which studies need to be done and where and when the studies should be performed. For example, in the stage before policy, it will help avoid confusion about the required scope and type of settings of currently called "demonstration studies," since pragmatic trials to address cost-effectiveness, effectiveness, and determinants of large-scale implementation now only need to be done in the next stage, before scale-up. Similarly, studies of epidemiological impact, such as tuberculosis prevalence surveys, would not be a prerequisite for programmatic policy recommendation. However, all countries should be encouraged to perform the pragmatic and ongoing country-specific evaluations of new tools during and after their scale-up, and this should be regarded as a key activity within national tuberculosis programs and ministries of health.

Finally, the process allows planning of the subsequent stages. When a new diagnostic tool is endorsed, the 2-year process of developing the programmatic policy recommendation, as well as the time needed to obtain the evidence for that recommendation, can be projected. The current WHO endorsement process includes planned revisions that are based on information obtained after implementation, but such revisions have not taken place for any of the new diagnostics endorsed since 2007. This is partly because, before endorsement of Xpert MTB/RIF in December 2010, no postendorsement data collection had been planned.

We therefore expect that the proposed pathway will expedite the adoption and scale-up of the new diagnostics that are badly needed to improve tuberculosis control. There are however important conditions that need to be met to make it work. Programmatic policy recommendation on the most cost-effective and feasible use of a new diagnostic can only be formulated and issued if the evidence needed in the stage before scale-up is indeed collected. This means that countries, researchers, and funders must be willing to invest in the large-scale implementation studies proposed for this stage and to do so in a concerted way. As the studies would ideally be done in different epidemiological and economic settings, coordination is needed to ensure that efforts and investments are directed to studies in selected countries, and to align study protocols so that effectiveness and costing data from the various studies are comparable and feed into multicountry economic, operational, and transmission models. Finally, coordination between assay developers and funders is needed to plan these studies in an early stage in order to limit delays due to preparations and clearance procedures. Funders should be asking how and by whom evidence for scale-up of products that they support will be collected.

CONCLUSION

The current GRADE-based endorsement process for new diagnostics for tuberculosis and tuberculosis drug resistance is necessary for many countries but insufficient for making informed decisions about their adoption and scale-up. The WHO needs to issue policy recommendations on how new diagnostics should be implemented in the most cost-effective and feasible way, focusing on the diagnostic algorithm or scenario rather than on the individual test. These recommendations generally cannot be formulated at the time of endorsement because further evidence from implementation studies is needed. Such definitive guidance requires a 2-step process involving technical policy recommendation, followed in about 2 years by programmatic policy recommendation. In the stage between technical and programmatic policy recommendation, data need to be collected from large-scale, programmatic implementation studies that ideally have comparative, pragmatic trial designs in

selected settings. Once a test is scaled up, it is important to assess the epidemiological impact, which can inform revisions of existing policies.

Global policy makers, funding agencies, program managers, researchers, and implementers must work together to ensure coordination of data collection to answer the crucial questions required for accelerating widespread implementation of new diagnostics. The WHO and other partners are taking up this challenge for Xpert MTB/RIF already, which, together with this proposed revised pathway, should serve as an example for future tools. If done collaboratively, it will lead to improved case detection and better care and outcomes, ultimately resulting in an accelerated decline in the worldwide burden of tuberculosis.

Notes

References

1. McNerney R. J Infect Dis **2012**; XX:XX–XX. [to be completed].
2. Nahid. J Infect Dis **2012**; XX:XX–XX. [to be completed].
3. World Health Organization. Global tuberculosis control. Geneva: World Health Organization, **2011**. http://www.who.int/tb/publications/global_report/en/index.html Accessed 21 June 2011.
4. Pai M, Minion J, Steingart K, Ramsay A. New and improved tuberculosis diagnostics: evidence, policy, practice, and impact. Curr Opin Pulm Med **2010**; 16:271–84.
5. World Health Organization. WHO endorses new rapid tuberculosis test, **2010**. http://www.who.int/tb/laboratory/new_rapid_test/en/index.html. Accessed 10 June 2011.
6. Schünemann HJ, Schünemann AH, Oxman AD, et al. Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. BMJ **2008**; 336:1106–10.
7. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ **2008**; 336:924–6.
8. Schito M. J Infect Dis **2012**; XX:XX–XX. [to be completed].
9. van Kampen SC, Ramsay AR, Anthony RM, Klatser PR. Retooling national TB control programmes (NTPs) with new diagnostics: the NTP perspective. PLoS One **2010**; 5:e11649.
10. Mann G, Squire SB, Bissell K, et al. Beyond accuracy: creating a comprehensive evidence base for TB diagnostic tools. Int J Tuberc Lung Dis **2010**; 14:1518–24.
11. Soto A, Solari L, Gotuzzo E, Acinelli R, Vargas D, Van der Stuyft P. Performance of an algorithm based on WHO recommendations for the diagnosis of smear-negative pulmonary tuberculosis in patients without HIV infection. Trop Med Int Health **2011**; 16:424–30.
12. Koole O, Thai S, Khun KE, et al. Evaluation of the 2007 WHO guideline to improve the diagnosis of tuberculosis in ambulatory HIV-positive adults. PLoS One **2011**; 6:e18502.
13. Boehme CC, Nicol MP, Nabeta P, et al. Feasibility, diagnostic accuracy, and effectiveness of decentralised use of the Xpert MTB/RIF test for diagnosis of tuberculosis and multidrug resistance: a multicentre implementation study. Lancet **2011**; 377:1495–505.
14. Pai M, Ramsay A, O'Brien R. Evidence-based tuberculosis diagnosis. PLoS Med **2008**; 5:e156.
15. Barnard M, Albert H, Coetzee G, O'Brien R, Bosman ME. Rapid molecular screening for multidrug-resistant tuberculosis in a high-volume public health laboratory in South Africa. Am J Respir Crit Care Med **2008**; 177:787–92.
16. Squire SB, Ramsay AR, van den Hof S, et al. Making innovations accessible to the poor through implementation research. Int J Tuberc Lung Dis **2011**; 15:862–70.
17. World Health Organization. Noncommercial culture and drug-susceptibility testing methods for screening patients at risk for multi-drug-resistant tuberculosis: policy statement. WHO/HTM/TB/2011.9. Geneva, Switzerland: World Health Organization, 2011.
18. Vassall A, van Kampen S, Sohn H, et al. Rapid diagnosis of tuberculosis with the Xpert MTB/RIF assay in high burden countries: a cost-effectiveness analysis. PLoS Med **2011**; 8:e1001120.
19. World Health Organization. TB diagnostics and laboratory strengthening. http://www.who.int/tb/laboratory/en/. Accessed 10 June 2011.
20. Stop TB Partnership's New Diagnostics Working Group, World Health Organization. Pathways to better diagnostics for tuberculosis: a blueprint for the development of TB diagnostics. Geneva: World Health Organization, **2009.**
21. Lijmer JG, Leeflang M, Bossuyt PM. Proposals for a phased evaluation of medical tests. Med Decis Making **2009**; 29:E13–21.
22. Ramsay A, Steingart KR, Pai M. Assessing the impact of new diagnostics on tuberculosis control. Int J Tuberc Lung Dis **2010**; 14:1506–7.
23. Wallis RS, Pai M, Menzies D, et al. Biomarkers and diagnostics for tuberculosis: progress, needs, and translation into practice. Lancet **2010**; 375:1920–37.
24. Steingart KR, Flores LL, Dendukuri N, et al. Commercial serological tests for the diagnosis of active pulmonary and extrapulmonary tuberculosis: an updated systematic review and meta-analysis. PLoS Med **2011**; 8:e1001062.
25. Dowdy DW, Steingart KR, Pai M. Serological testing versus other strategies for diagnosis of active tuberculosis in India: a cost-effectiveness analysis. PLoS Med **2011**; 8:e1001074.
26. World Health Organization. Policy statement: commercial serodiagnostic tests for diagnosis of tuberculosis. WHO/HTM/TB/2011.5. Geneva: World Health Organization, 2011. http://whqlibdoc.who.int/publications/2011/9789241502054_eng.pdf. Accessed 21 June 2011.
27. Porta MA. Dictionary of epidemiology. 5th ed. New York, NY: Oxford University Press, **2008**.
28. Rusen ID, Squire SB, Billo NE. Poverty and lung health. Expert Rev Respir Med **2010**; 4:163–5.
29. Treweek S, Zwarenstein M. Making trials matter: pragmatic and explanatory trials and the problem of applicability. Trials **2009**; 10:37.
30. Moulton LH, Golub JE, Durovni B, et al. Statistical design of THRio: a phased implementation clinic-randomized study of a tuberculosis preventive therapy intervention. Clin Trials **2007**; 4:190–9.
31. Mowatt G, Grimshaw JM, Davis DA, Mazmanian PE. Getting evidence into practice: the work of the Cochrane Effective Practice and Organization of care Group (EPOC). J Contin Educ Health Prof **2001**; 21:55–60.
32. Uys PW, Warren R, van Helden PD, Murray M, Victor TC. Potential of rapid diagnosis for controlling drug-susceptible and drug-resistant tuberculosis in communities where *Mycobacterium tuberculosis* infections are highly prevalent. J Clin Microbiol **2009**; 47:1484–90.
33. Roos BR, van Cleeff MR, Githui WA, et al. Cost-effectiveness of the polymerase chain reaction versus smear examination for the diagnosis of tuberculosis in Kenya: a theoretical model. Int J Tuberc Lung Dis **1998**; 2:235–41.