Stop TB Partnership

New Diagnostics Working Group

TARGET PRODUCT PROFILE SURVEY: NEXT-GENERATION DRUG- SUSCEPTIBILITY TESTING AT PERIPHERAL CENTRES

Contacts:
Chiacchiaretta Matteo, m.chiacchiaretta@hotmail.it
Miotto Paolo, miotto.paolo@hsr.it

TARGET PRODUCT PROFILE SURVEY: NEXT-GENERATION DRUG- SUSCEPTIBILITY TESTING AT PERIPHERAL CENTRES

Contents

AIM AND PROCESS 2

DEFINITIONS AND CLARIFICATIONS 2

RESULTS 3

CHARACTERISTICS NOT REACHING AGREEMENT (Less than 80% "Mostly agree", "Fully agree") 7

Priority of anti-TB agents for testing – Minimal definition 7

Price individual test (cost of reagents and consumables only; after scale-up; ex-works; excluding shipping and subsidiary factors. Non-negotiated prices) 9

Capital costs for the instrument (non-negotiated prices) – Minimal definition 12

Limit of detection of minor variants – both Optimal and Minimal definitions 14

Indeterminate results during detection –Minimal definition 16

OTHER CHARACTERISTICS REACHING AGREEMENT BUT VERY CLOSE TO THE THRESHOLD (equal or above 80%, but below 85% "Mostly agree", "Fully agree") 17

Target population 17

Limit of detection (TB detection after a first reaction) –Minimal definition 19

Limit of detection (TB detection after a second reaction) – both Optimal and Minimal definitions 20

Sample type – both Optimal and Minimal definitions 22

Sample volume – Optimal definition 23

Time to results – both Optimal and Minimal definitions 24

Daily throughput – Minimal definition 26

ANNEX 1 – Survey detailed results 27

ANNEX 2 – Target Product Profile draft document submitted together with the survey 97

AIM AND PROCESS

The aim of the survey was to reach consensus on the proposed characteristics. For this survey a Delphi-like process was used; the Target Product Profile (TPP) "Next-Generation Drug-Susceptibility Testing at Peripheral Centres" draft document and the survey were sent to a list of prioritized invitees representative of broad spectrum of categories (technical agencies and researchers, funding organizations, reference tuberculosis laboratories, National tuberculosis programs, implementers and clinicians, representatives from industry, and advocate for patients), and to a broader audience through general dissemination channels such as the New Diagnostics Working Group web-site and the Global Tuberculosis Network. Additional working groups such as the European Laboratory Initiative (ELI-TB) and the Global Laboratory Initiative (GLI-TB) have been also invited to participate to the survey. Participants were requested to provide a reply on 47 specific questions (30 with "optimal" and "minimal" requirements, 47 with an unique "optimal/minimal" requirement, for overall 77 answers submitted for each participant), expressing their level of agreement on the proposed characteristics according to a predefined Likert scale. Participants were also asked to provide feedbacks and comments if they disagreed with the proposed characteristics.

DEFINITIONS AND CLARIFICATIONS

Characteristic – refers to a specific requirement or specification that is measurable.

Minimal – for a specific characteristic, refers to the lowest acceptable output for that characteristic. For clarification, solutions must meet the 'Minimal' characteristic in order to be acceptable (CAVEAT: a test may still be acceptable if shortcomings pertain to the soft targets and if specific hard targets (marked with an asterisk) are only missed marginally).

Optimal – for a specific characteristic, provides the best output for that characteristic that is believed to be realistically achievable. Meeting the 'Optimal' characteristics provides the greatest differentiation and the greatest impact for the end users, clinicians and patients. Developers would ideally design and develop their solutions to meet the 'Optimal' requirements for all characteristics.

Likert scale – 1 Disagree, 2 Somewhat disagree, 3 Neither agree nor disagree, 4 Mostly agree, 5 Fully agree

Participants were asked to provide written comments if they scored agreement equal or below 3.

Consensus was considered reached if at least 80% (≥80%) of the responders scored 4 (Mostly agree)

or higher. Since characteristics might be defined as "optimal" and "minimal", each parameter has been evaluated individually.

Abbreviations: AMK – amikacin; BDQ – bedaquiline; CAP – capreomycin; CLO – clofazimine; DLM – delamanid; DST – drug susceptibility testing; FQ – fluoroquinolones; INH – isoniazid; KM – kanamycin; LEV – levofloxacin; LZD – linezolid; MDR – multidrug resistant TB (RIF, INH); MOX – moxifloxacin; PZA – pyrazinamide; RIF – rifampin; TPP – Target Product Profile; WHO – World Health Organization; XDR – extensively drug resistant TB (INH, RIF, FQ, AMK, KAN, CAP).

RESULTS

In total, 46 individuals from 25 countries replied to the survey. Responders were stratified based on their **Organizational Affiliation** (Advocacy organization, Clinician Consultant, Funding agency, Implementer (e.g. NGO), Industry IVD, International Non-profit, Laboratory Technician, Microbiologist, National TB Program, PDP/Technical agency, Reference Laboratory, Researcher (e.g. University) and based on their **WHO Region**. Accordingly, researchers, clinicians and professionals linked to National TB Programs accounted for the most represented categories, involving 50% of the participants. Professionals from industry, technical agencies, and international non-profit (overall 25%) followed. The remaining categories accounted together for the 25% (individual range: 2-5%) of the participants (Figure 1). Concerning the stratification by geographical region, the majority of the participants were from EURO (61%), followed by PAHO (20%). SEARO, WPRO, AFRO, and EMRO were represented by 2 to 7% of participants (overall 19%) (Figure 2).

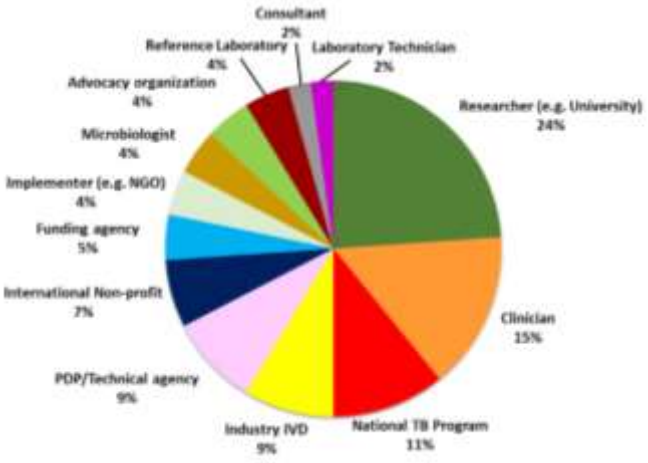


Figure 1. Stratification of survey participants by organizational profile.

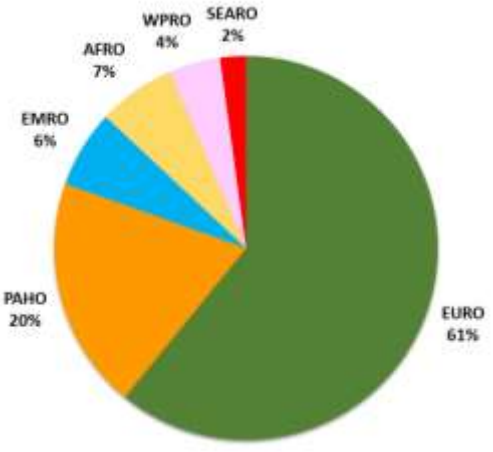


Figure 2. Stratification of survey participants by WHO region.

As described in the "Definitions and clarifications" section, consensus was defined as scoring 4 (Mostly agree) or higher in at least 80% (≥80%) of replies.

Out of 47 characteristics surveyed, agreement was reached for 42 (89%) components. Five characteristics did not reach consensus agreement for at least one of the two optimal or minimal requirement:

- Priority of anti-TB agents for testing (survey question nr. 2)
  - Minimal definition agreement: 67%
- Price of individual test (survey question nr. 7)
  - Optimal/Minimal definition agreement: 67%
- Capital costs for the instrument (survey question nr. 8)
  - Minimal definition agreement: 76%
- Limit of detection of minor variants (survey question nr. 17)
  - Optimal and Minimal definitions agreement: 78%
- Indeterminate results during detection (survey question nr. 19)
  - Minimal definition agreement: 78%

Overall results for minimal and optimal characteristics are reported in Figure 3 and Figure 4, respectively.

Figure 3. Overall consensus agreement on survey questions – minimal characteristics. Likert scale: 1 Disagree (red), 2 Somewhat disagree (orange), 3 Neither agree nor disagree (yellow), 4 Mostly agree (light green), 5 Fully agree (dark green). The consensus threshold (≥80%) is also highlighted (cyan line).

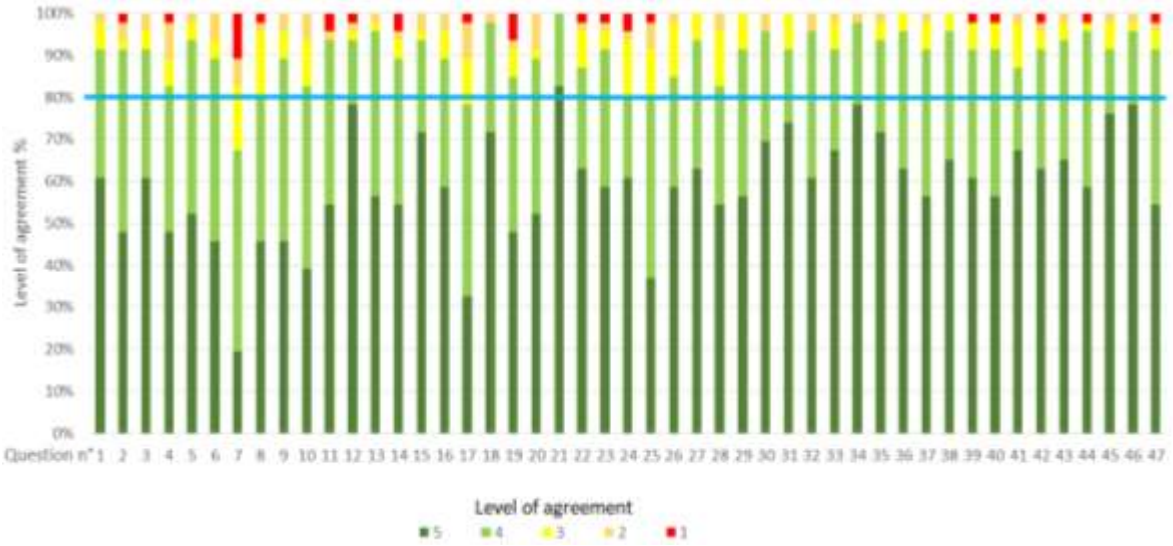


Figure 4. Overall consensus agreement on survey questions – optimal characteristics. Likert scale: 1 Disagree (red), 2 Somewhat disagree (orange), 3 Neither agree nor disagree (yellow), 4 Mostly agree (light green), 5 Fully agree (dark green). The consensus threshold (≥80%) is also highlighted (cyan line).

Despite reaching the consensus agreement threshold, the survey highlighted debate for an additional seven characteristics (80% ≤ consensus agreement < 85%):

- Target population (survey question nr. 4)
  - Optimal definition agreement: 83%
- Limit of detection, TB detection after first reaction (survey question nr. 9)
  - Minimal definition agreement: 83%
- Limit of detection, TB detection after second reaction (survey question nr. 10)
  - Minimal definition agreement: 80%
  - Optimal definition agreement: 83%
- Sample type (survey question nr. 24)
  - Optimal definition agreement: 80%
- Sample volume (survey question nr. 25)
  - Optimal definition agreement: 80%
- Time to result (survey question nr. 28)
  - Minimal definition agreement: 83%
  - Optimal definition agreement: 83%
- Daily throughput (survey question nr. 29)
  - Minimal definition agreement: 83%

The overview shows at a glance that the proposed draft was generally considered positively among the participants to the survey. To be noted that professionals from advocacy organization, industries, and international non-profit agencies showed some dissent with the proposed TPP document (rate of "mostly agree", "fully agree" between 70 and 80%). However, considering neutral answers ("neither agree nor disagree"), disagreement ("somewhat disagree", "fully disagree") was always below 20% (Figure 5).

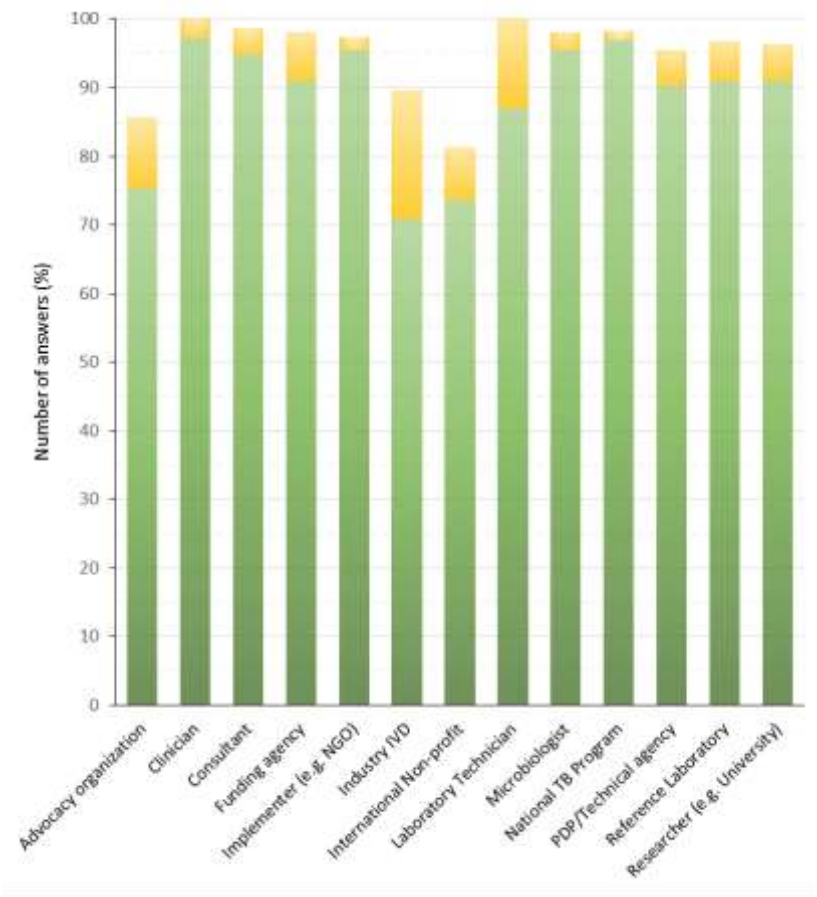


Figure 5. Overall consensus agreement on survey questions stratified by professional category. The percentages are calculated considering the total number of "4 – mostly agree" plus "5 – fully agree" (green) or "3 – neither agree nor disagree" (yellow) answers out of the 77 questions (30 with "optimal" and "minimal" requirements, 47 with an unique "optimal/minimal" requirement) submitted to each participant, whitin each professional category.

In the following sections, characteristics not reaching consensus agreement are discussed in details, and the revision proposed by the TF are reported.

CHARACTERISTICS NOT REACHING AGREEMENT (Less than 80% "Mostly agree", "Fully agree")

Priority of anti-TB agents for testing – Minimal definition

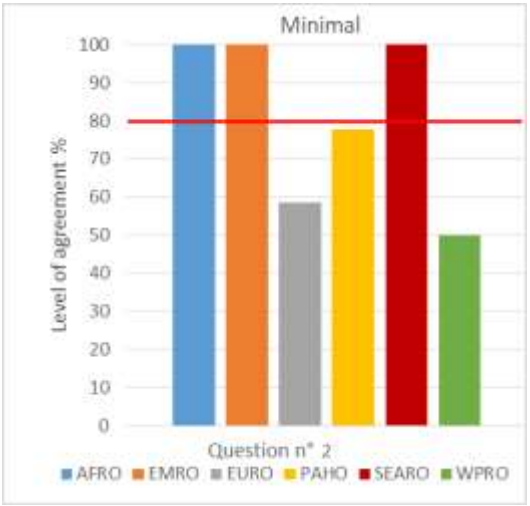
Originally proposed definition

Optimal	Minimal	Explanations
In order of decreasing importance: 1. RIF + INH + FQ 2. BDQ + LZD 3. CLO + DLM + pretomanid + AMK + PZA	In order of decreasing importance: 1. RIF + INH 2. FQ + KM* 3. AMK + CAP*	Drug prioritization considers universal DST access (END-TB Strategy) and that effective administration of anti-TB drugs can be achieved only by knowing susceptibility testing results. This is a general principle becoming crucial especially for MDR/XDR treatment. The proposed prioritization keep into account that FQs are relevant for both MDR and INH-R/RIF-S cases. In addition, new tests should be useful for triaging patients for short MDR regimen. The minimal requirements keep into account the transition time required for the complete drop-off from the regimens of the two injectable drugs no longer recommended (*), estimated in the range of 1-2 years depending upon the setting considered (especially where access to late generation drugs such as BDQ and LZD is strictly regulated and likely less available at peripheral level). The optimal requirements keep into account the full compliancy with the new WHO guidelines. To be also noted that treatment guidelines are moving forward to an "all oral" regimen; in this context, the relevance of AMK is expected to further decrease in the next 3-5 years. The differentiation of resistance among FQs is more a function of interpreting mutations (that is, evaluating the hierarchical structure of mutations) rather than detecting different mutations. **Sequence of detection and drug-resistance testing:** The proportion of patients with a diagnosis of TB who experienced pre-treatment loss to follow-up was found in the range of 4 to 38%. This scenario might vary substantially among countries. Initially testing for TB and DST might come at the expense of the sensitivity for TB detection, depending on the platform used and cost. However, a delay in DST might result in patients receiving inappropriate treatment until they return, assuming that the DST result will not be known in time to inform initial decisions about treatment. The acceptability of a longer wait time might vary among countries, and informing the patient of results on the same day if the result is not available during the first visit, might be associated with substantial costs.
(FQ always include LEV, MOX; any additional drug listed in the WHO treatment guidelines)	(FQ always includes LEV, MOX)	

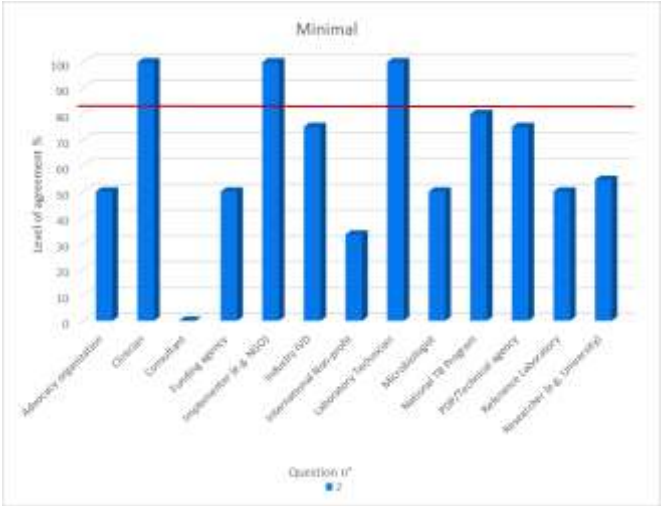
Survey results – Minimal definition

Overall level of agreement: 67%

Stratified by WHO region



Stratified by professional role



In general, since treatment guidelines are moving forward to an "all oral" treatment regimen, most of the participants asked for removing injectables from the priority list. In particular, comments were suggesting to remove KM and CAP from the prioritization, as these drugs are no longer recommended by WHO treatment guidelines. AMK can be useful under specific circumstances, and for this reason was included in the optimal situation. The TF members discussed these feedbacks, and considering that several countries are already taking up the latest WHO reccomendations, the TF agreed to remove KM and CAP from the priority list. Participants also suggested to include LZD, CLO, PZA and BDQ as minimal requirement; however, given the current knowledge of the molecular bases of resistance to these drugs, moving them among the minimal requirements would likely keep the development of new tests on hold as nobody would be able to reach minimal performances characteristics required. According to a similar rationale, PZA was not moved to the minimal requirements. The optimal requirement was also revised to follow some of the suggestions received from the survey.

Proposed revised definition

Optimal	**Minimal**	**Explanations**
In order of decreasing importance: 1. RIF + INH + FQ 2. BDQ + LZD + CLO 3. DLM + pretomanid + AMK + PZA 4. DCS	In order of decreasing importance: 1. RIF + INH 2. FQ 3. AMK	Drug prioritization considers universal DST access (END-TB Strategy) and that effective administration of anti-TB drugs can be achieved only by knowing susceptibility testing results. This is a general principle becoming crucial especially for MDR/XDR treatment. The proposed prioritization keep into account that FQs are relevant for both MDR and INH-R/RIF-S cases. In addition, new tests should be useful for triaging patients for short MDR regimen. The minimal requirements keep into account the lack of data on genetic basis of resistance for some high priority drugs for RR/MDR TB treatment (such as LZD CLO BDQ), and not because of prioritization of drugs over others. The optimal requirements keep into account the full compliancy with the new WHO guidelines. To be also noted that treatment guidelines are moving forward to an "all oral" regimen; in this context, the relevance of AMK is expected to further decrease in the next 3-5 years.
(FQ always include LEV, MOX; any additional drug listed in the WHO treatment guidelines)	(FQ always includes LEV, MOX)	The differentiation of resistance among FQs is more a function of interpreting mutations (that is, evaluating the hierarchical structure of mutations) rather than detecting different mutations. **Sequence of detection and drug-resistance testing:** The proportion of patients with a diagnosis of TB who experienced pre-treatment loss to follow-up was found in the range of 4 to 38%. This scenario might vary substantially among countries. Initially testing for TB and DST might come at the expense of the sensitivity for TB detection, depending on the platform used and cost. However, a delay in DST might result in patients receiving inappropriate treatment until they return, assuming that the DST result will not be known in time to inform initial decisions about treatment. The acceptability of a longer wait time might vary among countries, and informing the patient of results on the same day if the result is not available during the first visit, might be associated with substantial costs.

Price individual test (cost of reagents and consumables only; after scale-up; ex-works; excluding shipping and subsidiary factors. Non-negotiated prices)

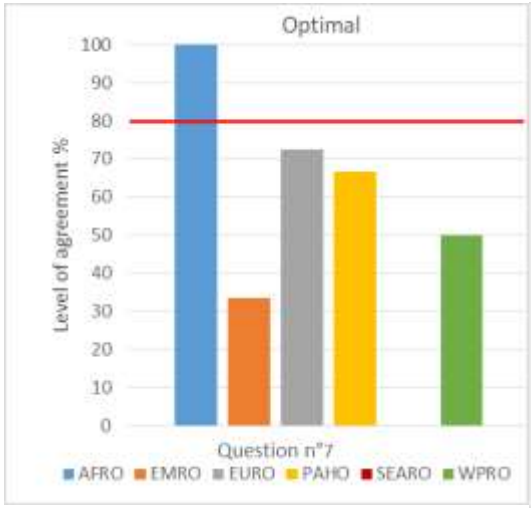
Originally proposed definition

Optimal/Minimal	**Explanations**
Detection of RIF + INH: 15-20 (±15%) USD; Detection of RIF +INH + FQ + AMK: 40-50 (±15%) USD; Detection of RIF + INH + FQ + AMK + KM* + CAP*: 50-60 (±15%) USD (FQ always includes LEV, MOX)	Meeting participants emphasized the critical need for the price to be kept within an affordable range. A price that is higher than available technologies would be justified only if the new test brings substantial added value in terms of vastly improved performance, greater suitability for decentralization, and the number of anti-TB agents for which resistance can be detected. Estimated ranges are based on current costs (expressed as min/max range) of the gold standard phenotypic DST that such new assays are expected to replace. We are currently unable to provide a price range for tests covering the "optimal" list of prioritized drugs because (i) the price might vary depending on the number of drugs considered, and (ii) there are no evidence data for foreseen what would be the cost of assays testing for new drugs such as DLM, LZD, and CLO. However, considering that phenotypic DST for first- and second-line drugs is estimated in the range of USD 50-100 (±30%), a new assay providing DST results for all the drugs listed among the "optimal" prioritization list should be within the same range, and any additional costs should be related to (i) the reduction of the turn-around-time, (ii) the increase of easiness of use and reduction of hands-on time, (iii) the reporting system, (iv) the reduction of need of extra reagents and/or equipment, and (v) the inclusion of additional drugs and/or additional testing (i.e. epidemiological details and/or comorbidities). Assay developers should consider cost-effectiveness in the context of willingness to-pay for disability-adjusted life year (DALY): screening tests should be cheaper enough for being cost-effective compared to current diagnostic algorithms, especially in high-burden, low-middle income settings. Finally, assay developers should consider that Global Drug Facility (GDF) negotiated prices for current assays for rapid DST endorsed by the WHO are <10 USD/test. (*) KM and CAP are no longer recommended, and are estimated to last in the clinical practice no longer than 1-2 years, depending upon the setting considered. Price estimates are considering that AMK, KM, and CAP share most relevant targets; thus, further implementation of an assay for detecting KM and CAP is not expected to be an expensive process.

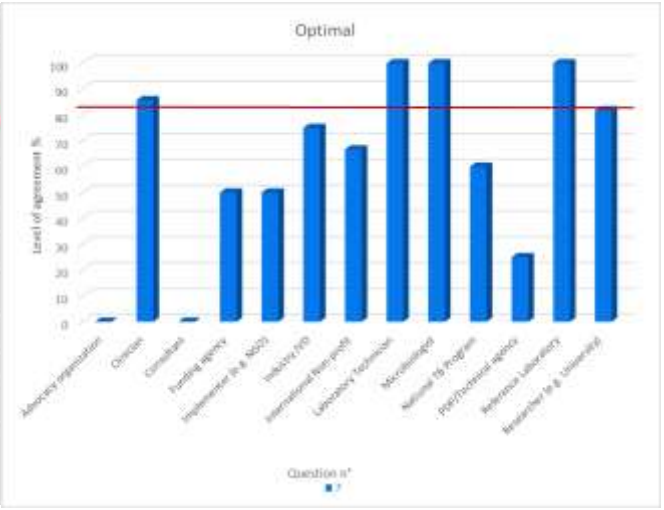
Survey results

Overall level of agreement: 67%

Stratified by WHO region



Stratified by professional role



Most of the participants asked to put target prices in line with the cost of current molecular diagnostic tests at peripheral centres, estimated in approx. 10.00 USD per tests in high burden countries. However, this is a GDF negotiated price, whereas as specified in the text and explanations, we are referring at *non-negotiated* prices. We also decided to consider as reference prices for both current existing rapid molecular assays (*non-negotiated* prices), and phenotypic drug susceptibility testing. Survey participants were also not agreeing on our choice about phenotypic testing as a reference; however, for some of the drugs listed in the priority for testing no molecular assay is available yet (i.e. BDQ, CLO, LZD). For this reason, the TF preferred to use as reference the assay that any new developed test should be able to replace. In general, survey feedbacks were asking to push for lowering prices as this should be achievable. The TF decided to split in "optimal" and "minimal" this characteristic requirement in order to keep the attractive for the industries alive, and to be inspirational toward reaching cheap solutions.

Proposed revised definition: Price of individual test (cost of reagents and consumables only; after scale-up; ex-works; excluding shipping and price subsidies. Non-negotiated prices)

Optimal	Minimal	Explanations
Detection of RIF+INH: ≤10 USD; Detection of RIF+INH+FQ+AMK: ≤25 USD (FQ always includes LEV, MOX)	Detection of RIF+INH: 15-20 (±15%) USD; Detection of RIF+INH+FQ+AMK: 40-50 (±15%) USD (FQ always includes LEV, MOX)	Meeting participants emphasized the critical need for the price to be kept within an affordable range. A price that is higher than available technologies would be justified only if the new test brings substantial added value in terms of vastly improved performance, greater suitability for decentralization, and the number of anti-TB agents for which resistance can be detected. Estimated ranges are based on current costs (expressed as min/max range) of the gold standard phenotypic DST that such new assays are expected to replace. We are currently unable to provide a price range for tests covering the "optimal" list of prioritized drugs because (i) the price might vary depending on the number of drugs considered, and (ii) there are no evidence data for foreseen what would be the cost of assays testing for new drugs such as DLM, LZD, and CLO. However, considering that phenotypic DST for first- and second-line drugs is estimated in the range of USD 50-100 (±30%), a new assay providing DST results for all the

	drugs listed among the "optimal" prioritization list should be within the same range, and any additional costs should be related to (i) the reduction of the turn-around-time, (ii) the increase of easiness of use and reduction of hands-on time, (iii) the reporting system, (iv) the reduction of need of extra reagents and/or equipment, and (v) the inclusion of additional drugs and/or additional testing (i.e. epidemiological details and/or comorbidities). Assay developers should consider cost-effectiveness in the context of willingness to-pay for disability-adjusted life year (DALY): screening tests should be cheaper enough for being cost-effective compared to current diagnostic algorithms, especially in high-burden, low-middle income settings. Finally, assay developers should consider that Global Drug Facility (GDF) negotiated prices for current assays for rapid DST endorsed by the WHO are <10 USD/test.

Capital costs for the instrument (non-negotiated prices) – Minimal definition

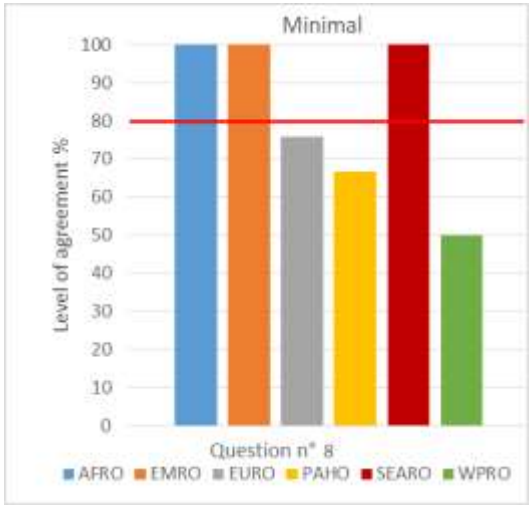
Originally proposed definition

Optimal	Minimal	Explanations
<5,000 USD (including warranties, service contracts and technical support)	<15,000 USD (including warranties, service contracts and technical support)	The lower the capital costs of the instrument are, the lower the initial cost would be, and thus the barrier to implementation would also be lower, particularly since the volume of instruments that would be distributed to microscopy centres is sizeable. The cost of the instrument should also include warranties, service contracts and technical support. Cost-effectiveness should be then evaluated during implementation according to the number of drugs/targets that a given technology can cover, the assay multiplexing, and the multipurpose options offered.

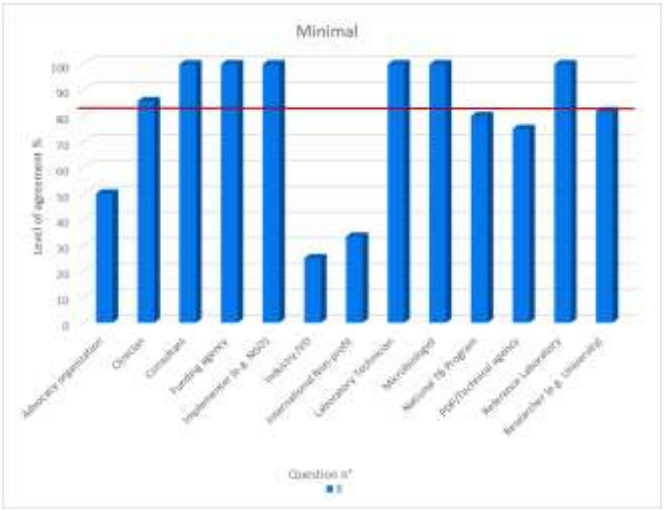
Survey results – Minimal definition

Overall level of agreement: 76%

Stratified by WHO region



Stratified by professional role



The question on the capital costs for the instrument lead to two main positions: participants asking for lowering the price (<10,000 USD, non-proofit and patient advocates), and participants asking for increasing the price (<20,000 USD, industries). The former clearly asks for lower prices to increase implementation programs feasibility, whereas the latter asks for higher prices because of realistic technological challenges. Also, participants suggested to mention here warranties and maintenance costs fro the initial investment. To be noted that the optimal requirement received some disagreement among survey participants as well (overall agreement: 80%).The TF discussed the feedback received and finally agreed to increase the minimal requirement in order to have industries interested to enter the field, with the inspirational goal to have instruments within 5,000.00 USD, as proposed in the optimal definition.

Proposed revised definition

Optimal	Minimal	Explanations
<5,000 USD (including warranties, service contracts and	<20,000 USD (including warranties, service contracts and	The lower the capital costs of the instrument are, the lower the initial cost would be, and thus the barrier to implementation would also be lower, particularly since the volume of instruments that would be distributed to microscopy centres is sizeable. The cost of the instrument

technical support – at least for 3 years)	technical support – at least for 3 years)	should also include warranties, service contracts and technical support. Cost-effectiveness should be then evaluated during implementation according to the number of drugs/targets that a given technology can cover, the assay multiplexing, and the multipurpose options offered.

Limit of detection of minor variants – both Optimal and Minimal definitions

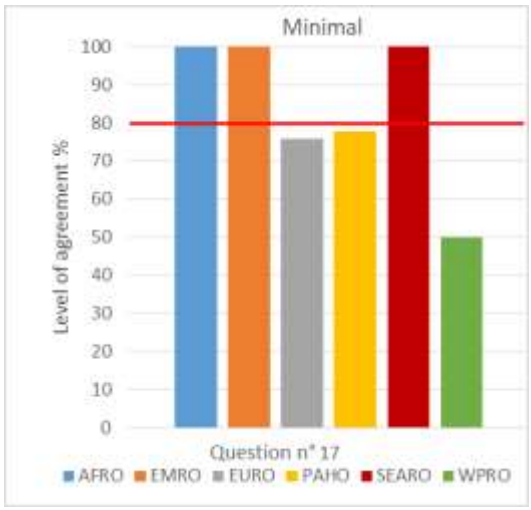
Originally proposed definition

Optimal	Minimal	Explanations
≤10% (that is 10 resistant bacteria out of 100)	≤20% (that is 20 resistant bacteria out of 100)	

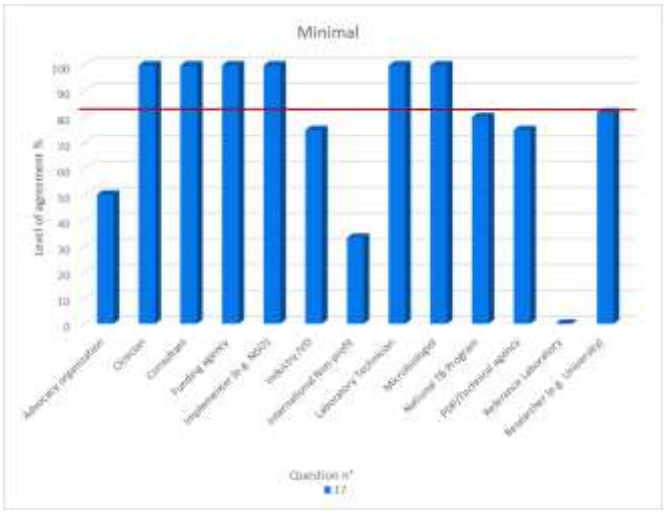
Survey results – Minimal definition

Overall level of agreement: 78%

Stratified by WHO region



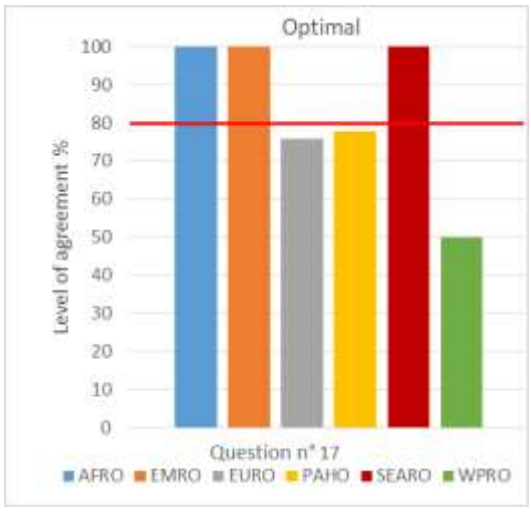
Stratified by professional role



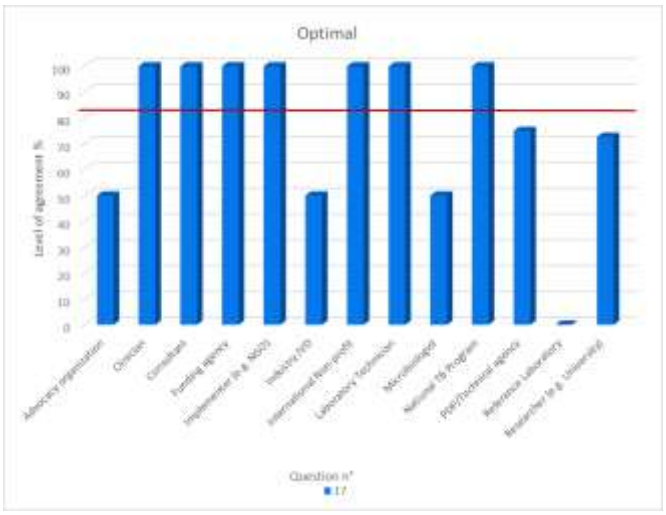
Survey results – Optimal definition

Overall level of agreement: 78%

Stratified by WHO region



Stratified by professional role



Most of the participants suggested to lowing the thresholds. However, current rapid diagnostics unlikely can reach <20 mimimum variants. The TF agreed to keep an attainable minimal threshold, and to push on better tests for the optimal threshold.

Proposed revised definition

Optimal	Minimal	Explanations
≤3% (that is 3 resistant bacteria out of 100)	≤20% (that is 20 resistant bacteria out of 100)	

Indeterminate results during detection –Minimal definition

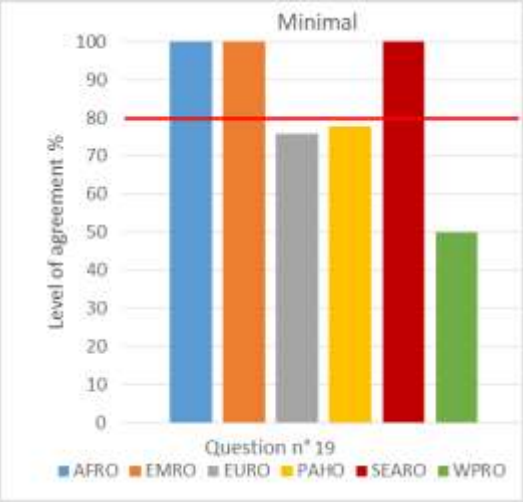
Originally proposed definition

Optimal	Minimal	Explanations
< 5%	≤ 10%	Indeterminate results may be caused by a lower sensitivity for detecting TB during the second reaction.

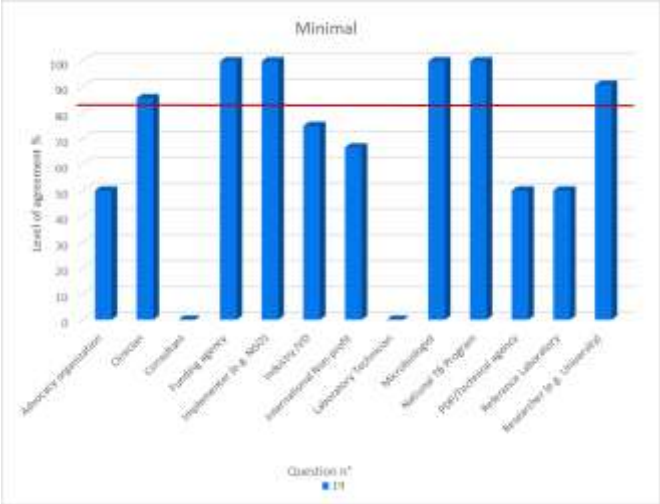
Survey results – Minimal definition

Overall level of agreement: 78%

Stratified by WHO region



Stratified by professional role



The participants suggested for lowing the threshold, however during the revision process the TF noticed that this is somehow already specified in the limit of detection and indeterminate rates clearly depends upon the starting bacillary load. Accordingly, the TF agreed to change this characteristic in a more appropriate form: “indeterminate results during DST”. Since also the optimal requirement received some disagreement among survey participants (overall agreement: 85%), the TF agreed to lower both the thresholds. To be noted that current proposed definition is in agreement with the “indeterminate” definition proposed for line probe assay (Line probe assays for drug-resistant tuberculosis detection. Interpretation and reporting guide for laboratory staff and clinicians. Global Laboratory Initiative, 2018. http://www.stoptb.org/wg/gli/assets/documents/LPA_test_web_ready.pdf).

Proposed revised definition: Indeterminate results during DST

Optimal	Minimal	Explanations
< 3%	≤ 5%	Indeterminate for a specific drug (or group of drugs) is defined as a valid test itself (i.e. generic procedural controls and TB detection controls are passed) but the internal controls for that specific drug (or group of drugs) is missing.

OTHER CHARACTERISTICS REACHING AGREEMENT BUT VERY CLOSE TO THE THRESHOLD (equal or above 80%, but below 85% "Mostly agree", "Fully agree")

Target population

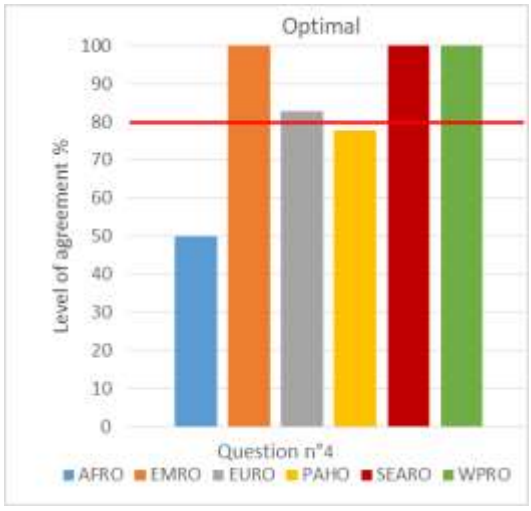
Originally proposed definition

Optimal/Minimal	Explanations
Target groups are all patients suspected of having TB, with a special focus on those at high risk of morbidity and mortality from drug-resistant TB, such as people living with HIV and those at high risk of having MDR-TB (for example, household contacts of patients diagnosed with MDR-TB, and persons with a history of TB, especially those for whom first-line therapy has failed) in countries with a medium incidence to a high incidence of TB as defined by WHO	The optimal target population should be all adults with signs and symptoms of, although the resource implications need to be considered. Children aged <11 years have limited ability to produce sputum for testing. Therefore, initial validation studies should focus on adults. WHO's categories: High-incidence countries are those with > 100 cases per 100 000 population; medium-incidence countries are those with 20–40 cases per 100 000 population; and low-incidence countries are those with < 10 cases per 100 000 population [5].

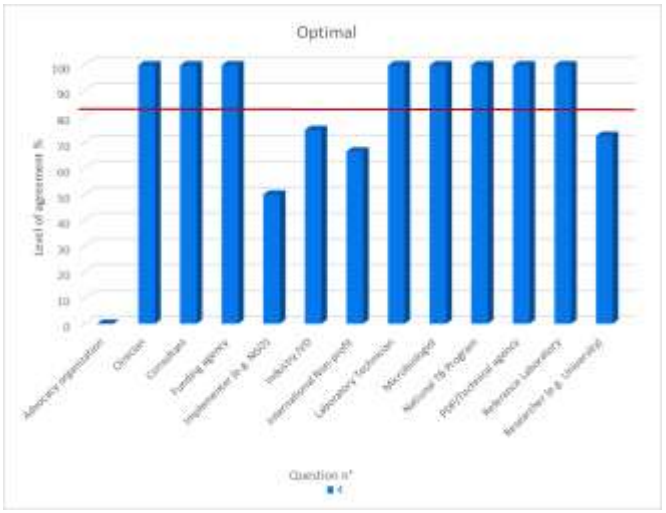
Survey results

Overall level of agreement: 83%

Stratified by WHO region



Stratified by professional role



The participants suggested to include evaluation for non-tuberculous mycobacteria (NTM), socially vulnerable groups as priority population, and to enphatize the importance of diagnosis in children as well. Also, participants suggested to consider countries with low TB incidence as well. Whereas targeting "any mycobacterial infection" would not be appropriate, it should be noted that the characteristic "multiuse platform" mentions the the usefulness of targeting NTM infections. THe TF agreed to wait for the stakeholders meeting feedback on this specific aspect of the TPP.

Proposed revised definition

Optimal/Minimal	Explanations
Target groups are all patients in need of evaluation for TB, with a special focus on those at high risk of morbidity and mortality from drug-resistant TB, such as people living with HIV and those at high risk of having MDR-TB (for example, household contacts of patients diagnosed with MDR-TB, and persons with a history of TB, especially those for whom first-line therapy has failed) in countries with a medium incidence to a high incidence of TB as defined by WHO	The optimal target population should be all adults with signs and symptoms of, although the resource implications need to be considered. Children aged <11 years have limited ability to produce sputum for testing. Therefore, initial validation studies should focus on adults. WHO's categories: High-incidence countries are those with > 100 cases per 100 000 population; medium-incidence countries are those with 20–40 cases per 100 000 population; and low-incidence countries are those with < 10 cases per 100 000 population [5].

Limit of detection (TB detection after a first reaction) –Minimal definition

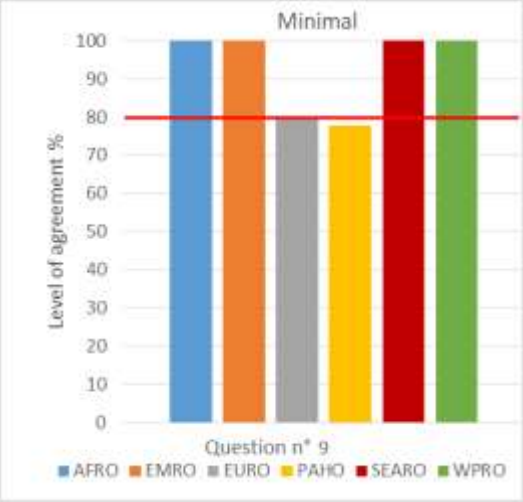
Originally proposed definition

Optimal	Minimal	Explanations
< 4.5 genome equivalents/reaction and < 10e2 CFU/assay using one sample	between 10e2 CFU/assay and 10e5 CFU/assay using one sample	Limit of detection testing should be performed as outlined in the United States Food and Drug Administration's guidance document.

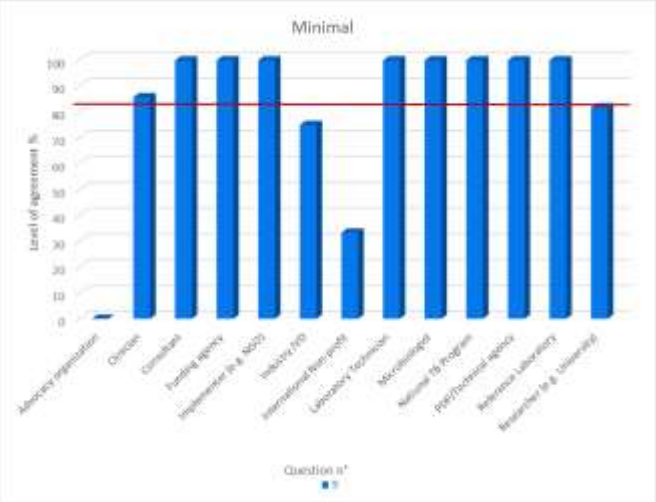
Survey results – Minimal definition

Overall level of agreement: 83%

Stratified by WHO region



Stratified by professional role



The participants suggested for lowing the threshold. The TF agreed to reduce the minimal threshold as proposed as this should be feasible with current available technologies.

Proposed revised definition

Optimal	Minimal	Explanations
< 4.5 genome equivalents/reaction and < 10e2 CFU/assay using one sample	between 10e2 CFU/assay and 10e4 CFU/assay using one sample	Limit of detection testing should be performed as outlined in the United States Food and Drug Administration's guidance document.

Limit of detection (TB detection after a second reaction) – both Optimal and Minimal definitions

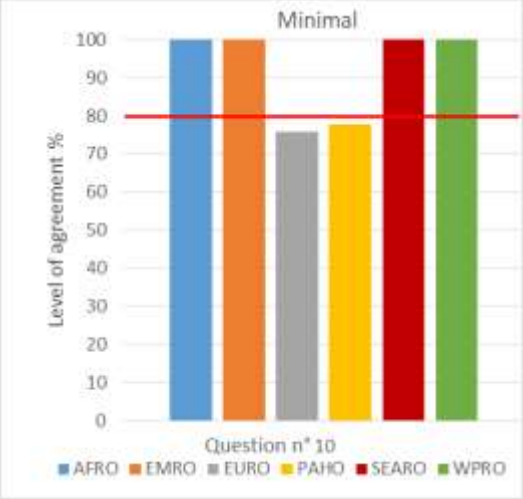
Originally proposed definition

Optimal	Minimal	Explanations
≥4.5 genome equivalents/reaction and 131 CFU/mL of sputum	between 10e2 CFU/assay and 10e5 CFU/assay using one sample	A slightly decreased analytical sensitivity for TB detection in the second reaction for resistance testing (in comparison with the first reaction) both for the optimal and minimal requirements will avoid resistance calls (for example, no M. tuberculosis but resistance present) but will come at the expense of a slightly lower sensitivity for DST.

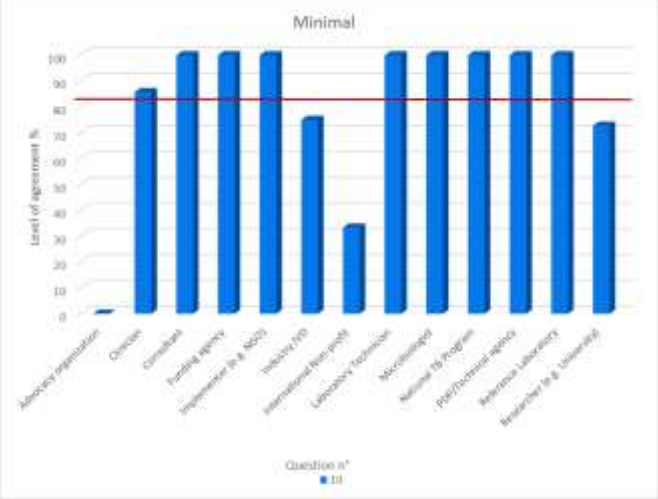
Survey results

Minimal overall level of agreement: 80%

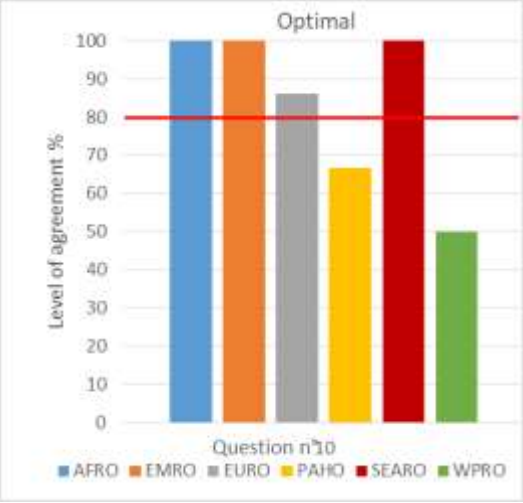
Stratified by WHO region

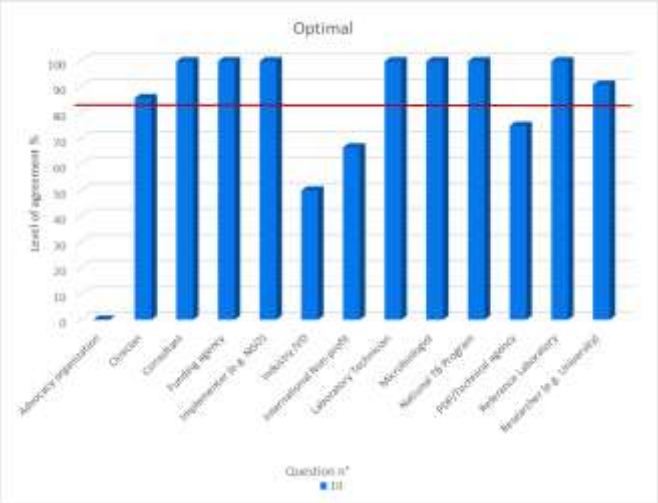


Stratified by professional role



Optimal overall level of agreement: 83%





The participants suggested for lowing the threshold, however the TF agreed to keep the minimal threshold as proposed to keep the development of new assays feasible (to be noted that this limit is referring at the second reaction).

Proposed revised definition

Optimal	Minimal	Explanations
≥4.5 genome equivalents/reaction and 131 CFU/mL of sputum	between 10e2 CFU/assay and 10e5 CFU/assay using one sample	A slightly decreased analytical sensitivity for TB detection in the second reaction for resistance testing (in comparison with the first reaction) both for the optimal and minimal requirements will avoid resistance calls (for example, no M. tuberculosis but resistance present) but will come at the expense of a slightly lower sensitivity for DST.

Sample type – both Optimal and Minimal definitions

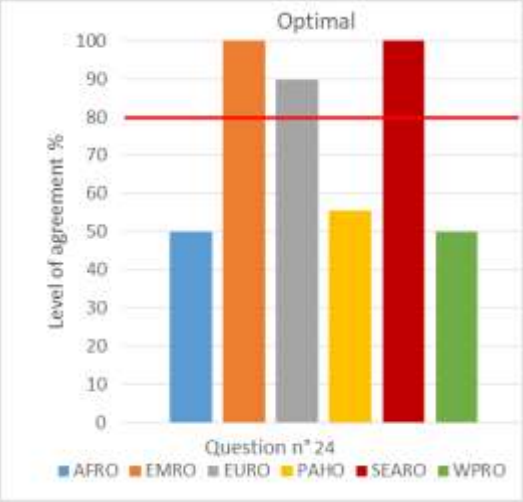
Originally proposed definition

Optimal/Minimal	Explanations
Unprocessed sputum	

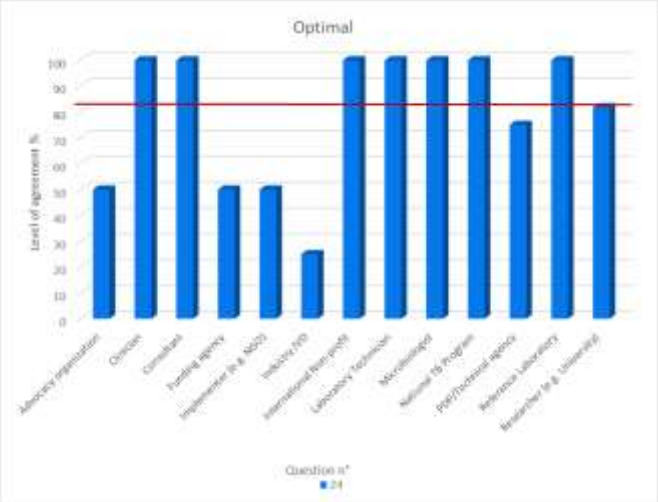
Survey results

Overall level of agreement: 80%

Stratified by WHO region



Stratified by professional role



The participants asked to consider additional sample types. The TF agreed to extend the range of sample type.

Proposed revised definition

Optimal	Minimal	Explanations
Unprocessed sputum, and additional clinically relevant specimens for TB or other targeted diseases (see "Multiuse platform")	Unprocessed sputum	Additional clinically relevant specimens for TB could be alternative easily collected sample types (especially for categories of patients where sputum is difficult to obtain), as well as specimens for extra-pulmonary TB.

Sample volume – Optimal definition

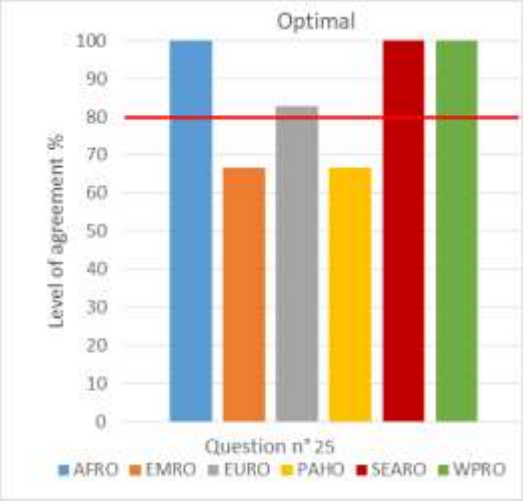
Originally proposed definition

Optimal	Minimal	Explanations
Up to 10 mL	< 0.5-2 mL	The lowest volume possible for all types of samples should be 0.1 ml, especially since HIV-positive patients may have difficulty providing a sample; however, this should not come at the expense of decreased sensitivity. If a higher volume is available, the test should be able to use it if doing so would increase sensitivity. Additionally, the ideal test would need only 1 sample even if requires 2 or more steps or reactions.

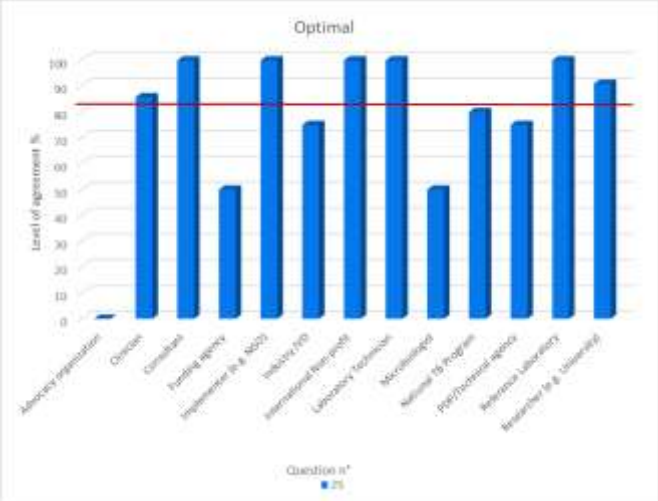
Survey results

Overall level of agreement: 80%

Stratified by WHO region



Stratified by professional role



The participants suggested for lowering the sample volume; however, this is likely to reduce the sensitivity on low bacterial burden samples (as defined in the explanations). The TF decided to keep current proposed values, and wait for final stakeholders meeting suggestions.

Proposed revised definition

Optimal	Minimal	Explanations
Up to 10 mL	< 0.5-2 mL	The lowest volume possible for all types of samples should be 0.1 ml, especially since HIV-positive patients may have difficulty providing a sample; however, this should not come at the expense of decreased sensitivity. If a higher volume is available, the test should be able to use it if doing so would increase sensitivity. Additionally, the ideal test would need only 1 sample even if requires 2 or more steps or reactions.

Time to results – both Optimal and Minimal definitions

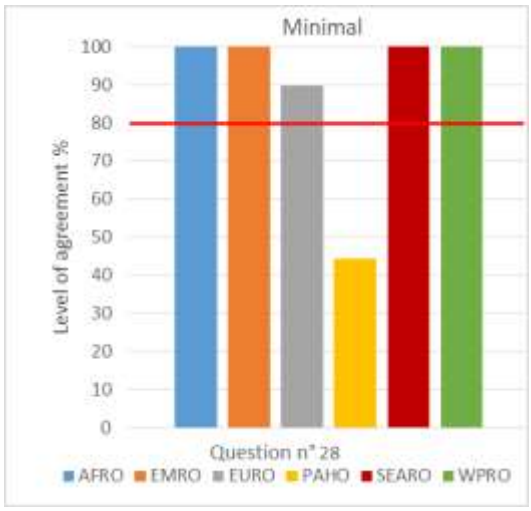
Originally proposed definition

Optimal	Minimal	Explanations
< 30 minutes for detection and DST	< 24 hours for detection and DST	The need for rapid turnaround, the possibility of batching or using random access for testing, and the ability to test multiple samples at the same time are interrelated. The time to result is probably the most important parameter since extending the wait time for patients may result in loss to follow-up.

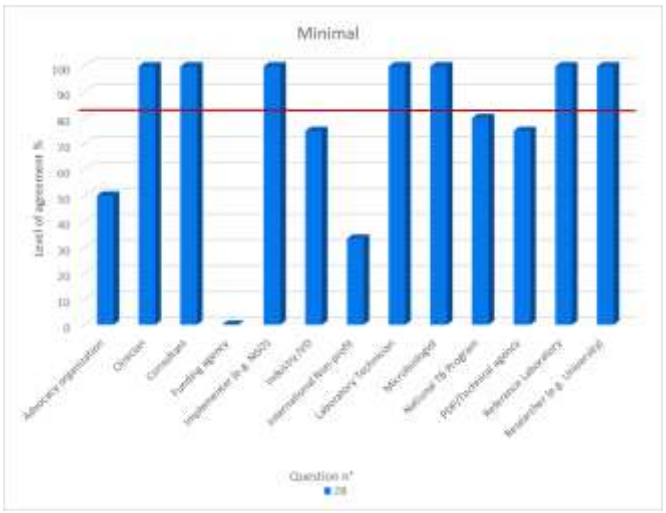
Survey results

Minimal overall level of agreement: 83%

Stratified by WHO region

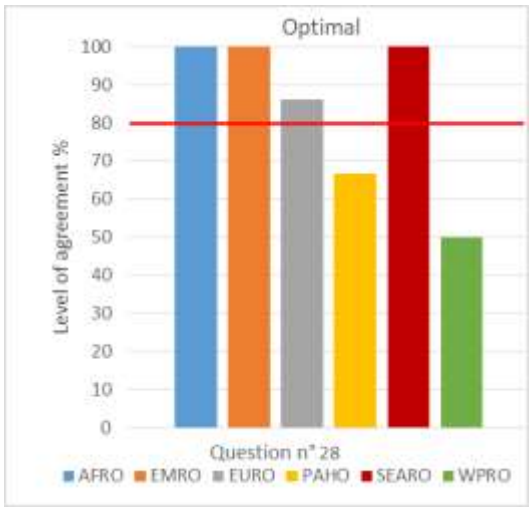


Stratified by professional role

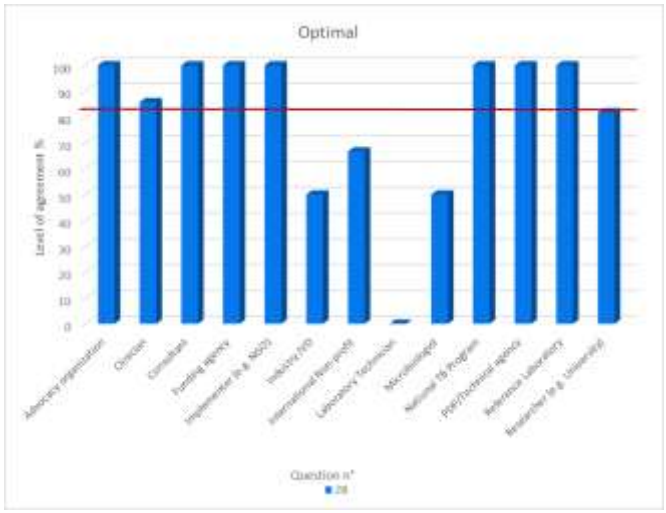


Optimal overall level of agreement: 83%

Stratified by WHO region



Stratified by professional role



The participants suggested for lowering the time to results; however, in order to include all current technologies, the TF recognized the need to keep reasonable time ranges. The TF decided to reduce the time to results, however the TF also included an acceptable threshold.

Proposed revised definition

Optimal	Minimal	Explanations
< 30 minutes for detection and DST (< 2 hours acceptable)	< 12 hours for detection and DST (< 24 hours acceptable)	The need for rapid turnaround, the possibility of batching or using random access for testing, and the ability to test multiple samples at the same time are interrelated. The time to result is probably the most important parameter since extending the wait time for patients may result in loss to follow-up.

Daily throughput – Minimal definition

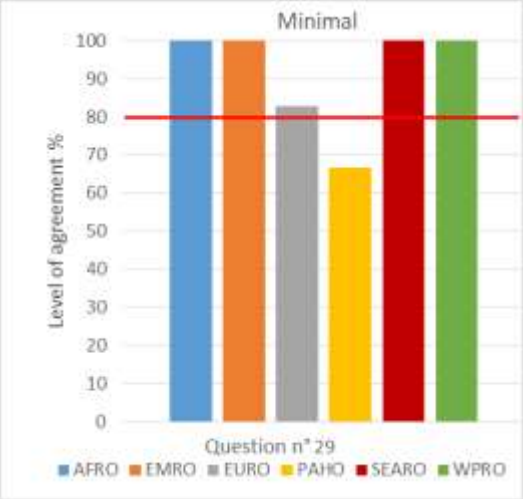
Originally proposed definition

Optimal	Minimal	Explanations
> 25 tests	> 10 tests	The daily throughput needed in most microscopy centres is <10 tests per day. Daily throughput requirements are considering currently available technologies as baseline.

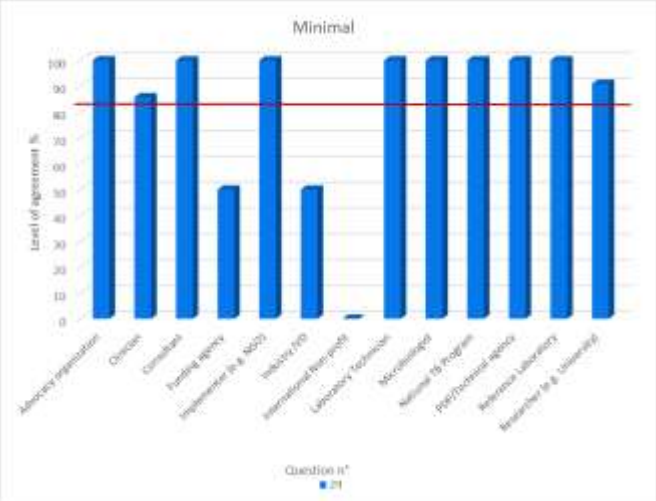
Survey results

Minimal overall level of agreement: 83%

Stratified by WHO region



Stratified by professional role



Some of the participants suggested to increase minimal daily throughput. The TF considered current daily throughput at microscopy centres. The TF decided to keep current proposed values, and wait for final stakeholders meeting suggestions.

Proposed revised definition

Optimal	Minimal	Explanations
> 25 tests	> 10 tests	The daily throughput needed in most microscopy centres is <10 tests per day. Daily throughput requirements are considering currently available technologies as baseline.

ANNEX 1 – Survey detailed results

NOTE: Survey answers and comments are reported as provided by the participants.

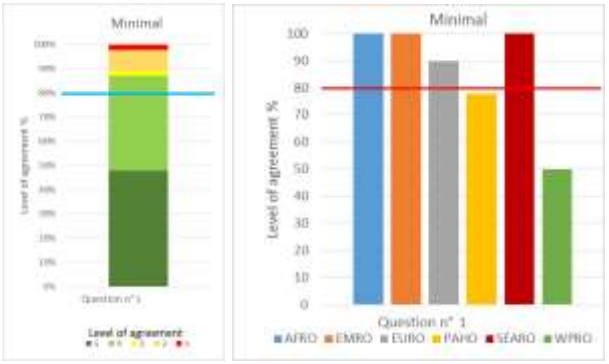
1) Goal of test

To provide support for effective anti-TB therapy in the context of the roll-out of new regimens; to provide the characteristics and qualities of a test that would have a sufficiently rapid turnaround time for TB detection and would provide data about DST that can be used to inform treatment decisions.

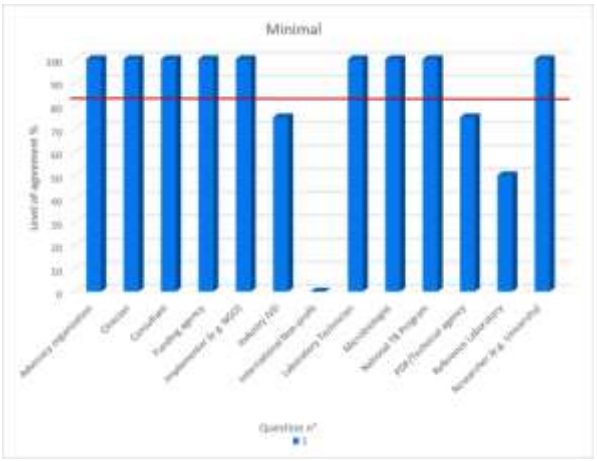
Minimal: Diagnosis of TB disease and detection of drug resistance to provide rapid triage of patients and identification of adequate treatment regimen (1st line treatment vs 2nd line treatment)

Level of agreement 87%

Stratified by WHO region



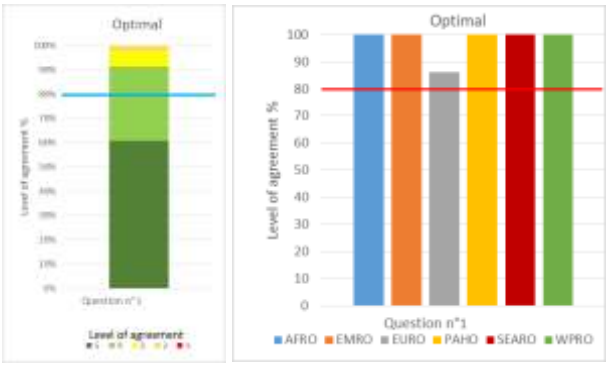
Stratified by professional role



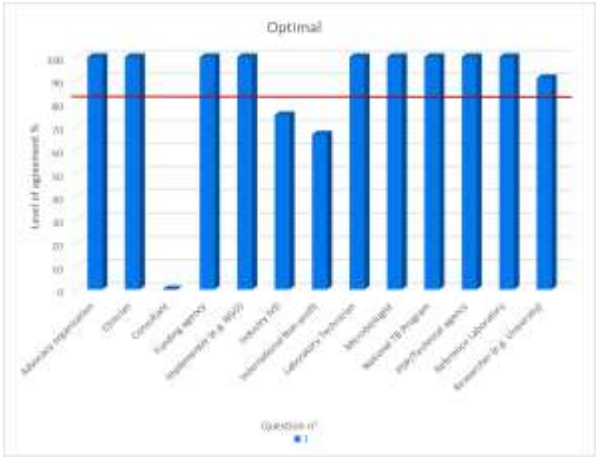
Optimal: Diagnosis of TB disease and detection of drug resistance to inform decision making about the optimal (individualized) regimen

Level of agreement 91%

Stratified by WHO region



Stratified by professional role



Comments:

- We should be moving towards individualized regimens. Minimal and optimal criteria should be identical.
- Diagnosis of TB disease and detection of drug resistance to provide rapid triage of patients and avoiding inappropriate treatment regimen
- Minimal: diagnosiss of ACTIVE TB disease (identifying viable bacilli) and detection of drug resistance....
- The minimal would be met by existing Xpert assay and NGS needs to do more
- Would be good to specify what the target turnaround time is
- You need to consider for the optimal diagnosis of TB or other Mycobacteria or Respiratory infection (with NGS this is possible). Also, having baseline info on bacterial load (genomic or cell) within the sample to be definitive is critical for consideration under OPTIMAL, in order to provide possibly a measure for patient follow-up.
- Remove "Diagnosis of TB disease"
- Minimal needs to address the issues with current programs which it doesn't. A triage assay will not solve the problems providing actionable information will but the actionable information must be able to be dealt with at the point of testing. This also glosses over the bigger ethical and moral issues which it should not as the individual being tested is paramount to the design and development of the assay
- triage test may be very expensive and technical if it gives the optimal information mentioned, you will be paying for a test that is not needed in the majority of cases being triaged?
- optimal needs to assess susceptibility to the drug that will be given in addition to resistance detection

2) Priority of anti-TB agents for testing

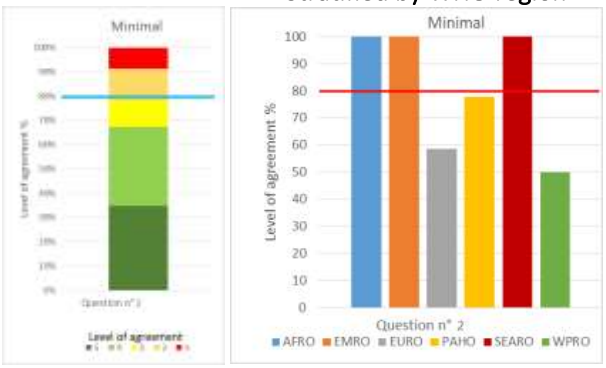
(in order of decreasing importance)

The proposed prioritization keep into account that FQs are relevant for both MDR and INH-R/RIF-S cases.

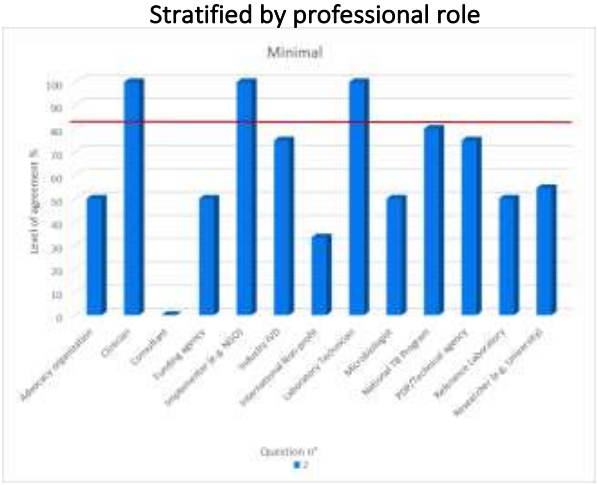
Minimal: RIF, INH, FQ, KM, AMK, CAP

Level of agreement 67%

Stratified by WHO region



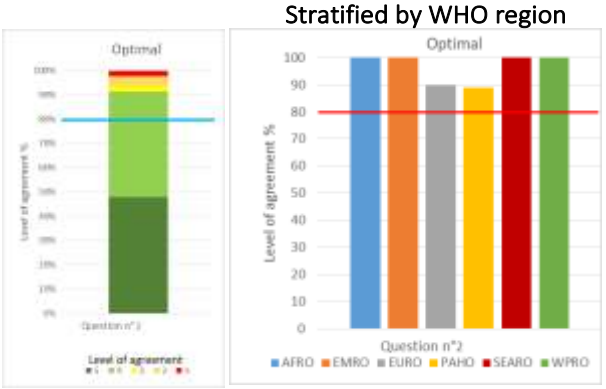
Stratified by professional role



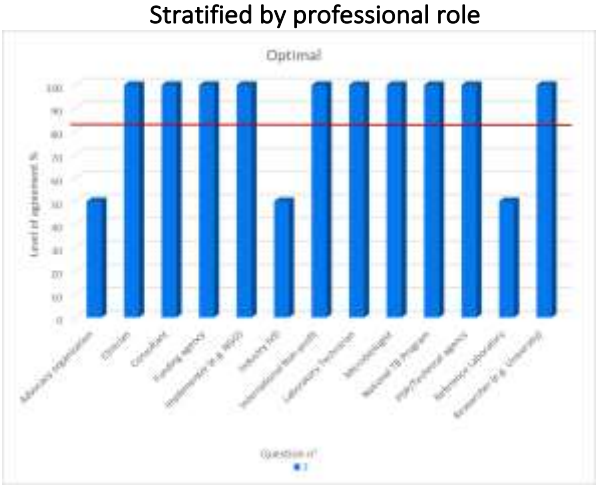
Optimal: RIF, INH, FQ, BDQ, LZD, CLO, DLM, pretomanid, AMK, PZA

Level of agreement 91%

Stratified by WHO region



Stratified by professional role



Comments:

- PZA must be in minimal.
- As kanamycin is no longer recommended by WHO, I see no value to include this drug on DST.
- Would qualify that the reason for prioritization of minimal is lack of data on genetic basis of resistance and not because of prioritization of drugs over others
- Given the drig development horizon BDQ, Pretominid & LZD should be included
- High cost may limit the use in country context like Pakistan we agree that this is needed in long run

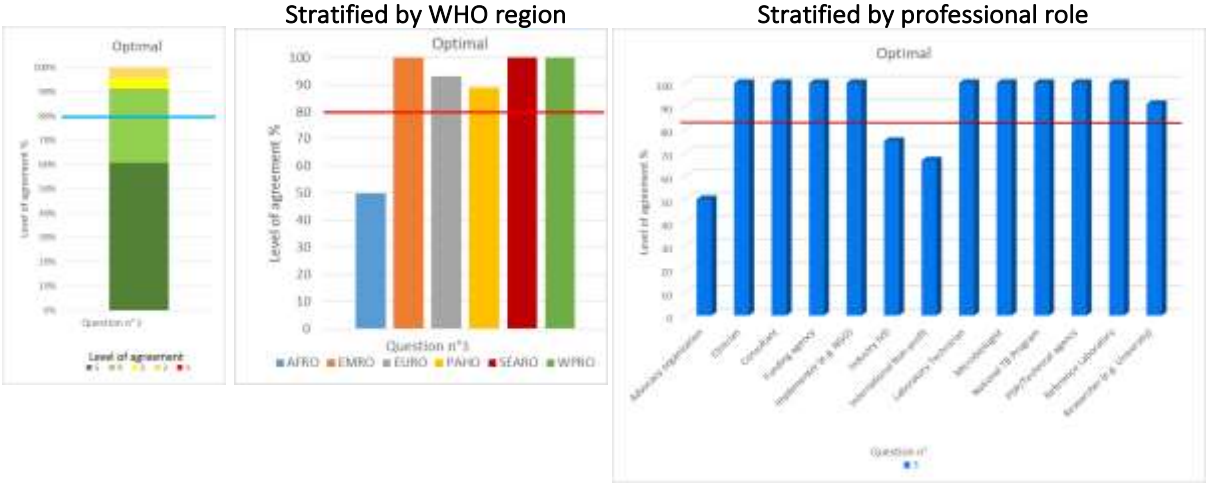
- KM and AMK are not relevant and should not be prioritized. Linezolid, bdq and clofazimine should be prioritized.
- With the recent removal of KM and CAP by WHO, it doesn't make sense to include these. Plus the new and re-purposed drugs are being adopted
- clofazimine and cycloserine should be higher priority in the optimal, given guidelines as well as their risk of side effects: 1) RIF/INH/FQ, 2) BDQ+LZD+CLO+Cycloserine 3) DLM+pretomanid+AMK+PZA. Minimum: 1) RIF+INH, 2)FQ+Clofazimine+Cycloserine 3) AMK/KM/CAP . ALso in the explanations, it is inappropriate to specify a 1-2 year dropoff period of injectable-containing regimens for a variety of reasons, including the fact that this final TPP will be available already a year after WHO communicated it was deprioritizing injectables, but also bc amikacin may still need to be used in some circumstances. Also disagree with including that sensitivity might decrease-- this is more germane to the later point on sensitivity anyway
- You need to include for MIN: MTBC detection and OPT: Mycobacterial (including MTBC detection), and possibly Bacteria vs Viral component for infection
- AMK/CAP are likely to be lower priority going forward so may no longer need to be minimal
- Too many resitance markers unkown
- we can ony know if we fully evaluate the available drugs
- To provide all anti-TB agents
- Minimal: question whether injectables should be included given new txt regimens. In optimal: can FQ be differentiated, MOX and others?
- Problem is that our understanding of resistance is not at a level where we can bundle drugs. KM and CAP don't even appear in the Optimal so are they being ignored? We have minimal knowledge of BDQ, DLM and some of the others so how do I develop a test for them
- the minimal should not necessarily include aminoglycosides as these are third line agents whereas you need linezolid, clofazimine and bedaquiline in minimal as they will be firstline for RR-MDR/TB
- n the minimal I would delete KM and CAP
- SLIDs no longer as minimal?
- Minimal: RIF, INH, FQ, AMK (I would leave out CAP and KAN) - Optimal: I would choose between DLM and pretomanid

3) Assay design

Optimal/Minimal: The assay should be designed in such a manner that the addition of or removal of analytes does not require extensive analytical and clinical re-verification and revalidation of the assay

Level of agreement 91%

Stratified by WHO region | **Stratified by professional role**



Comments:

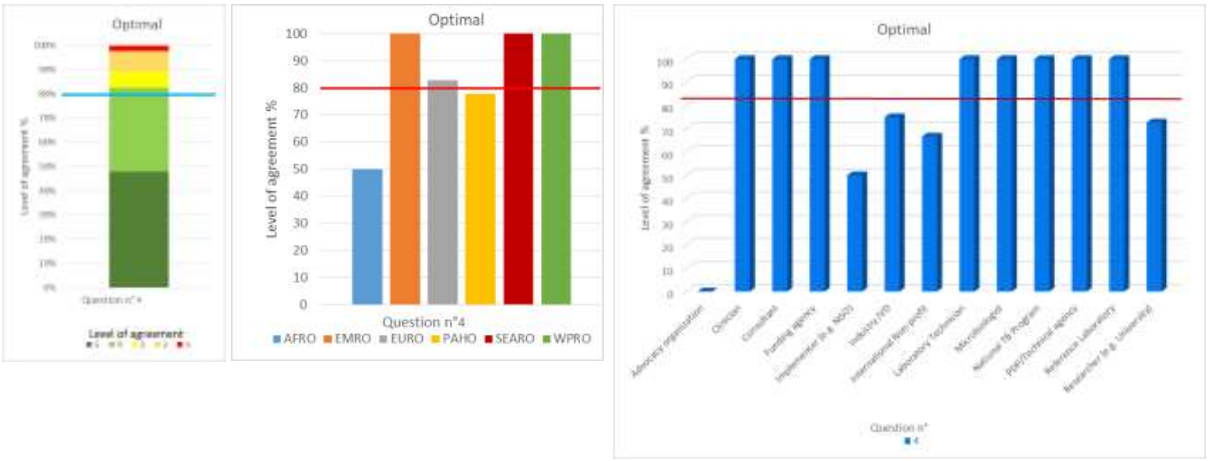
- QA/QC needs to be maintained to ensure accuracy of results so a cartrige based system would be optimal to avoid this issue
- I think that for OPT: NO analytes should be additionally needed, but yes to the fact that additions do not require re-verification or revalidation
- Impossible in the current regulatory environment
- Your biggest problem is the technology and the adaptability of the technology followed by the manufacturers not understanding this aspect NOR the process by which the amended product would be cleared for use

4) Target population

Optimal/Minimal: Target groups are all patients suspected of having TB, with a special focus on those at high risk of morbidity and mortality from drug-resistant TB, such as people living with HIV and those at high risk of having MDR-TB (for example, household contacts of patients diagnosed with MDR-TB, and persons with a history of TB, especially those for whom first-line therapy has failed) in countries with a medium incidence to a high incidence of TB as defined by WHO

Level of agreement 83%

Stratified by WHO region | **Stratified by professional role**



Comments:

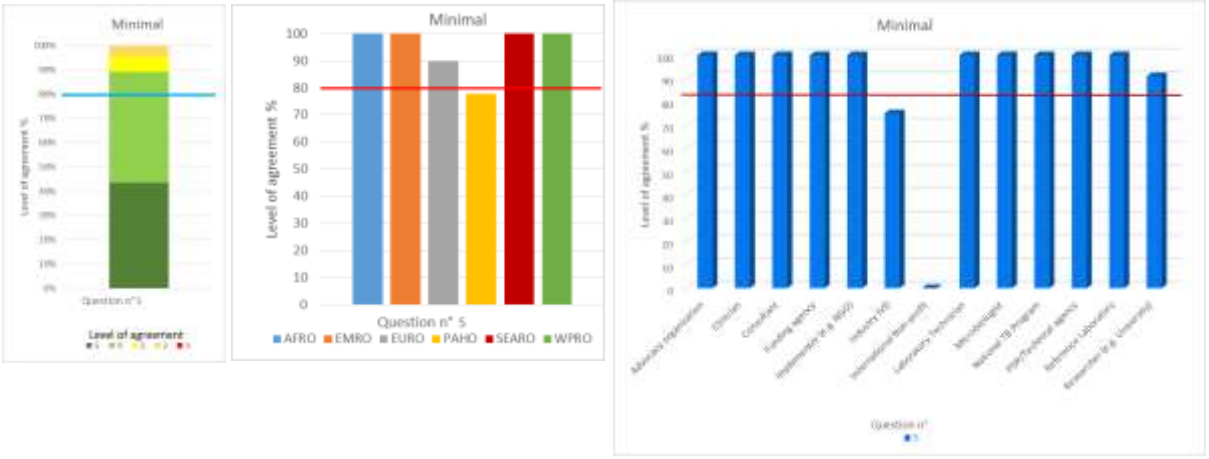
- Also non-tuberculous mycobacteria should be evaluated
- this is relevant also in countires with low TB incidence....
- Add socially vulnerable groups as a priority pop. e.g homeless, slum dwellers, excluded migrant communities
- It would be useful to acknowledge the importance of a diagnostic for children, perhaps in optimal and not in minimal. Would also be good to note countries with a medium to high incidence of TB and/or MDR-TB. Also please use appropriate language, ie "patients in need of evaluation for TB" as opposed to stigmatizing use of "suspected"
- For OPT- ALL cases presumptive of TB infection - END STOP. No additional delineations for special focus. JUST everyone.
- This should be thought of from a triage standpoint. 1 in 10 symptomatic's could have TB, why in all circumstances would you spend a lot more money with a comprehensive resistance test? We need to spend the money wisely and this will not allow it to happen or will force all manufacturers to do what is easiest
- Availability of NGS DST also in low-incidence settings

5) Target user of test

Minimal: Health-care workers with minimal/moderate training

Level of agreement 89%

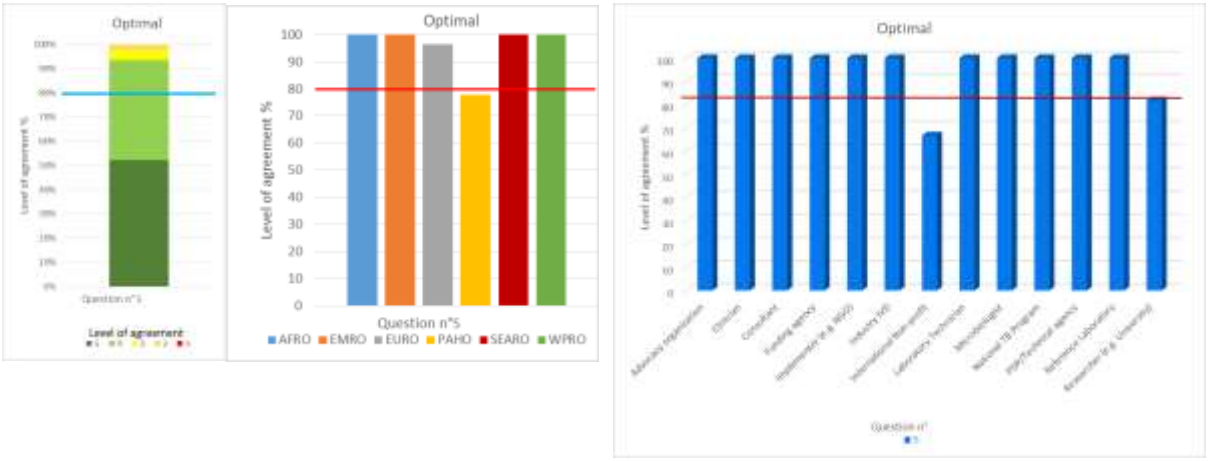
Stratified by WHO region **Stratified by professional role**



Optimal: Health-care workers with minimal training necessary

Level of agreement 93%

Stratified by WHO region **Stratified by professional role**



Comments:

- Minimal training needs to be defined. NGS vs. LPAs would have very different 'minimal' training needs.
- Training will be essential as is the case for any diagnostic
- MIN: HCW or Laboratorians min/mod training; OPT: ANY HCW in all settings for placement min training
- Make it Optimal for both as that is reality, unless ethically and morally comprehesive resistnce testing should not occur near the individual

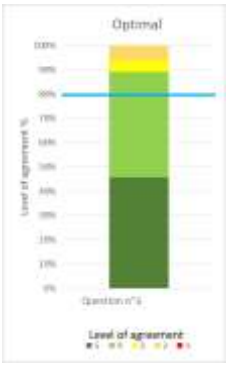
6) Setting (level of the health-care system)

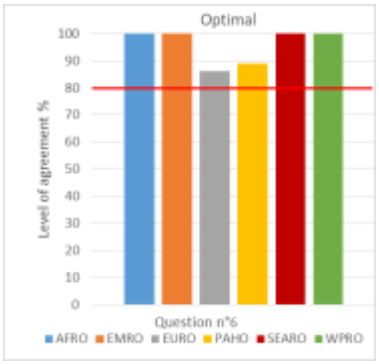
Implementation at the microscopy-centre level should be feasible using the specifications as outlined. This would embed the test in an infrastructure that is based around smear microscopy. However, the test could be implemented at higher levels of care as well. Testing for resistance to the anti-TB agents included in second-line therapy could be incorporated into separate reactions, but ideally it would be feasible to test the same specimen.

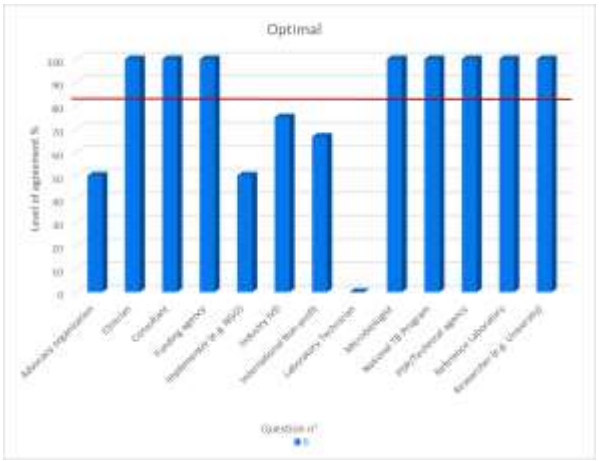
Optimal/Minimal: Peripheral and/or microscopy-centre level of the health-care system

Level of agreement 89%

Stratified by WHO region **Stratified by professional role**







Comments:

- Experience with GeneXpert has shown that it only works in microscopy-only labs if electricity is also taken care off (large UPSs or inverter/chargers on large battery packs). Very often lectricity in these settings is irregular & unreliable.
- Optimal would be a POCT that could be used by low skilled outreach workers
- I would implement at higher levels first then take the lessons learned to see if it could be rolled out at lower levels. Use of transport media could be used to bridge as well.
- I think we need to stop using the term microscopy centers, but rather refer more specifically to infrastructure and HR requirements for settings. Microscopy centers in vary significantly from country to country. Maybe suggest to use Lower peripheral HC facility levels with min infrastructure vs. lower level laboratories which may include Microscopy centers. Needs revision descriptive for placement
- Settings of use re where drugs are available
- You have to face the ethical and moral issues first before this can be addressed

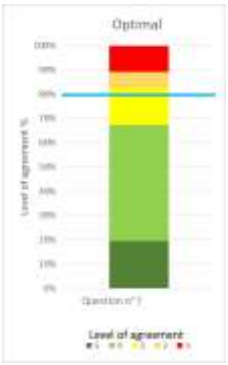
7) Price of individual test

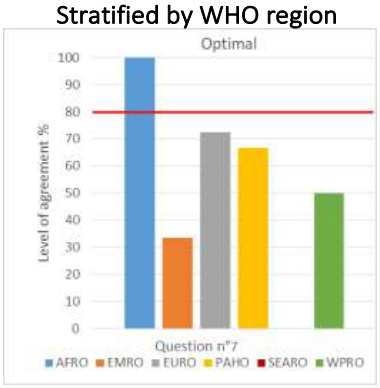
Estimated ranges are based on current costs (expressed as min/max range) of the gold standard phenotypic DST that such new assays are expected to replace.

Optimal/Minimal: Detection of RIF + INH: 15-20 (±15%) USD; Detection of RIF +INH + FQ + AMK: 40-50 (±15%) USD; Detection of RIF + INH + FQ + AMK + KM + CAP: 50-60 (±15%) USD (FQ always includes LEV, MOX)

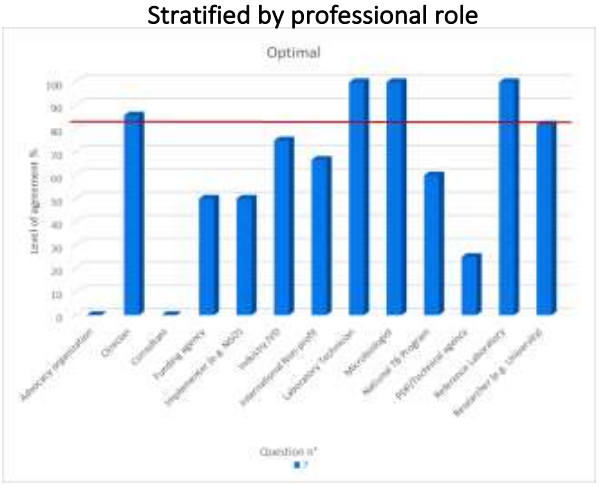
Level of agreement 67%

Stratified by WHO region





Stratified by professional role



Comments:

- Marked as mostly agree, as long as taking all costs into consideration (technician time, associated consummables, etc.)
- RIF and INH must be less than 10USD.
- TB is mainly in developing countries. So cost should be cheap (affordable) for routinely use.
- Agree with option-1 ie Optimal but do not agree with combination and cost of option 2 (RIF +INH + FQ + AMK: 40-50) and option 3 (RIF + INH + FQ + AMK + KM + CAP: 50-60)
- Optimally or ideally it could be even cheaper. Again, the kanamycin is no longer needed.
- I do not agree with the costs as they should be based on currently available molecular tests such as LPAs and not phenotypic DST. Also it should consider what is feasible and 20-25USD for RIF+INH+FQ+AMK for an integrated molecular test is feasible.
- The cost implication will limit the use in resource constraint but high burden countries
- No reasons to pay for unuseful tests (inj)
- pinning the price to existing tests is inappropriate, given the far too low uptake of DST currently. Even Xpert MTB/RIF is due for a price reduction from $10, so detection of RIF+INH should be max $10 and I would advocate for less ($6-7). We are talking about large volumes of scale here and most manufacturers receive public investment for R&D so are not in need of fully recouping cost of R&D through price. Also really don'tfind it meaningful to list costs for the minimal assay when we are hoping for optimal.

- I do not understand where these prices the drugs (SIRESMfxLfxAmk, CLO,LAD, BDQ, DLM,etc) then 50-60USD for this is my OPT. IF we have only RHFQAm, then 15-20 is optimal. LPA is currently for FL/SL around 20, but this should be lower as we move forward.
- Further efficiencies in test pricing over current phenotypic options would help generate demand and support countries to provide more individual DSTs within current budget lines
- The costs seem reasonable when comapred to the standard phenotypic DST, however, for the intended settings of use (microscopy-centre level), the estimated price ranges appear to be too high.
- clarify whether target pricing range is ex-works (or not)
- Over priced, I would ask for transparency and then factor in a cost plus model but industry will not what to do that. In general this is priced way too high; RIF/INH - $10, RIF/INH/FQ/AMK - $25, RIF/INH/AMK/KM/CAP - $35. This is also not in agreement with the previous ranking and doesn't address reflex testing so it is very confusing
- Dont agree with RIF + INH + FQ + AMK + KM + CAP: 50-60 (±15%) USD
- Should all be below $15
- You will need to go back and have a look at your tests and look at the new treatment guidelines.. this would have been great a year ago but now things have changed

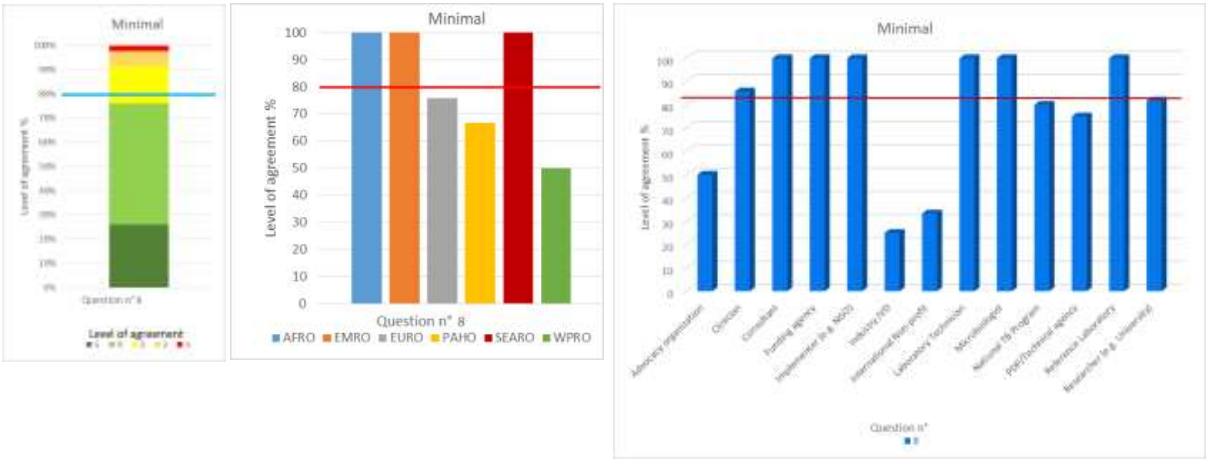
8) Capital costs for the instrument

The cost of the instrument should also include warranties, service contracts and technical support. Cost-effectiveness should be then evaluated during implementation according to the number of drugs/targets that a given technology can cover, the assay multiplexing, and the multipurpose options offered.

Minimal: <15.000 USD (including warranties, service contracts and technical support)

Level of agreement 76%

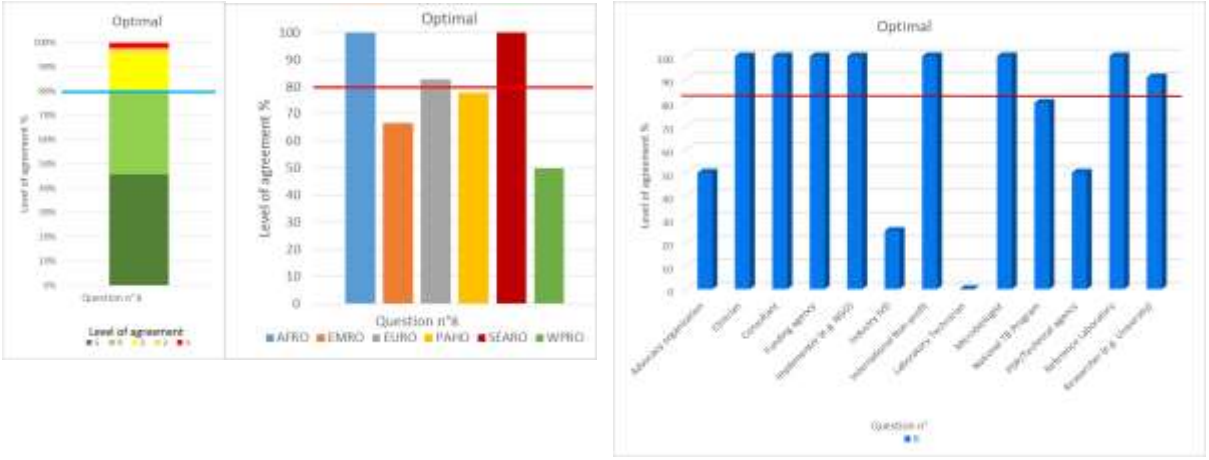
Stratified by WHO region **Stratified by professional role**



Optimal: <5.000 USD (including warranties, service contracts and technical support)

Level of agreement 80%

Stratified by WHO region **Stratified by professional role**



Comments:

- minimal: <10,000 Optimal: <3,000
- Given the present prices, I do not think it is possible to produce an apparatus capable of doing such essays for less than 5'000 USD
- To be feasible to implement at peripheral level, the minimal should be even less expensive.
- low entry costs will make the transition more palatable

- Minimal<10.000
- Depends on number of colors and number of instruments sold
- The capital costs are too high again considering the distribution that this instruments will need to have. Additionally, for optimal, we should be targeting much lower anyway-- $1,000 for optimal and $10,000 for minimal.
- Currently, GX is 17,500 for HBDC so I would hope 10K would be feasible in 5 years.
- Depends on size and output
- As stated, cost-effectiveness needs to be evaluated as for some technologies it may not be possible to bring the instrument costs down to the criteria stated. As the test intends to replace phenotypic DST instrument costs need to be compared to the the cost of phenotypic DST set-up, as it was done with the individual test costs.
- Recommend to clarify period of warranty, service and support needed for optimal and minimal (e.g. 3 years for minimal)
- Clarification required. How is this tied to throughput? How are service contracts etc factored in as they are a yearly process?
- <20,000 USD
- <15k is maybe too strict as minimal requirement. consider cost of the existing Illumina iSeq 100

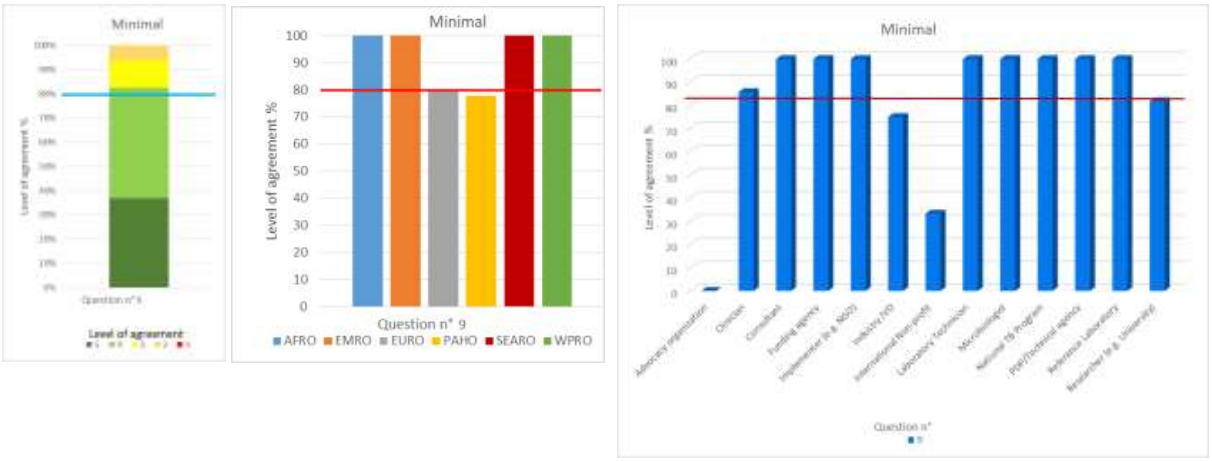
9) Limit of detection, TB detection after first reaction

Estimates are based on current commercially available assays endorsed for direct testing on sputum specimens.

Minimal: Between 10^2 CFU/assay and 10^5 CFU/assay using one sample

Level of agreement 83%

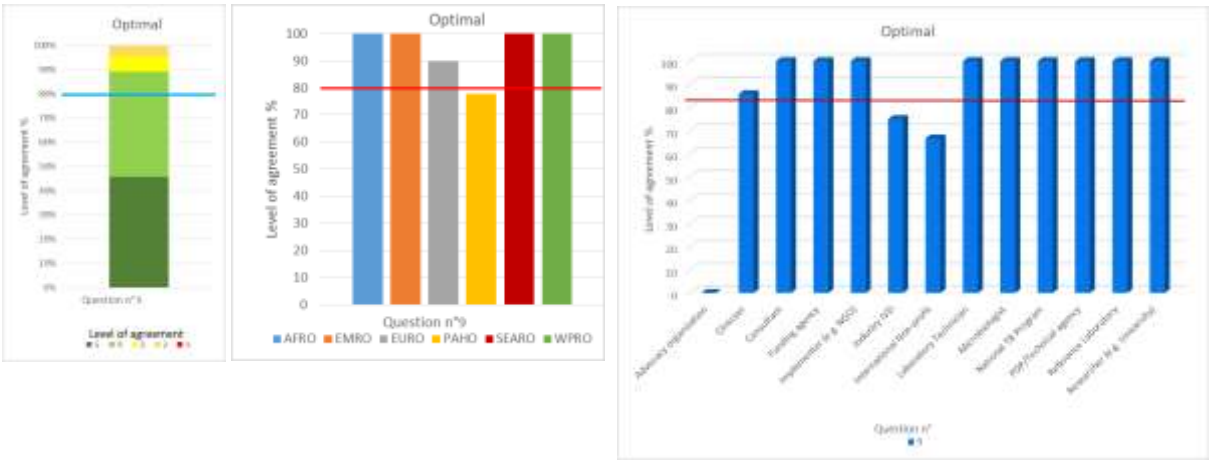
Stratified by WHO region **Stratified by professional role**



Optimal: < 4.5 genome equivalents/reaction and < 10^2 CFU/assay using one sample

Level of agreement 89%

Stratified by WHO region **Stratified by professional role**



Comments:

- Optimal is too low to be easily feasible- even Xpert MTB/RIF is ~130cfu/mL and out of this range.
- Higher sensitivity should be achievable and will educe the need for repeat testing.
- No change recommended-- I am just not informed about this to be able to give a meaningful opinion, hence neither agreeing nor disagreeing
- Minimal should be what GX Ultra is today. <100 CFU: OPT - CFU should be <20/assay
- i am not expert in this question

- Needs clarification as this is highly dependent upon the process used. This also doesn't address reflex testing and how it is all tied together. What happens when I give you a highly sensitive assay for detection?
- 10^5 is not enough

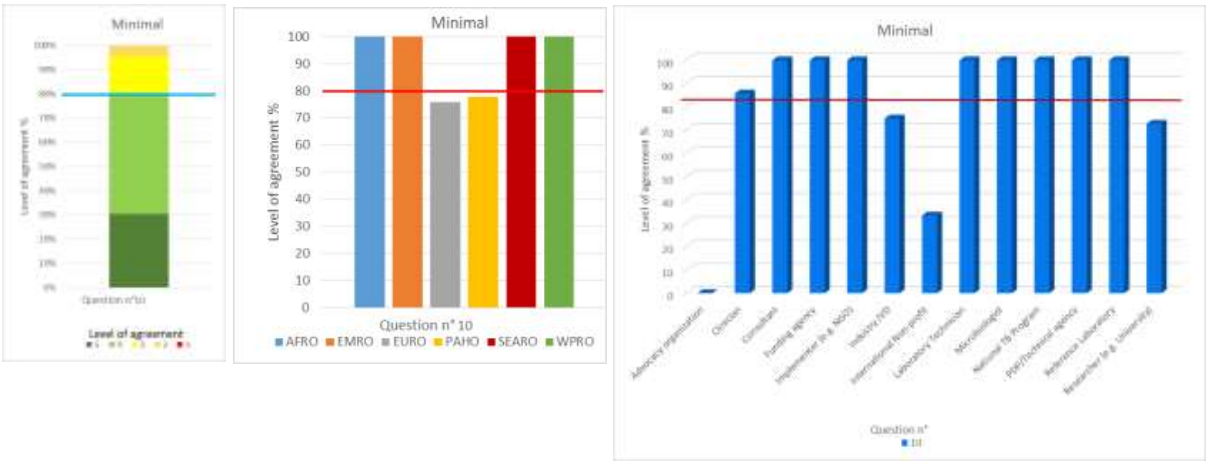
10) Limit of detection, TB detection after second reaction

Estimates are based on current commercially available assays endorsed for direct testing on sputum specimens.

Minimal: Between 10^2 CFU/assay and 10^5 CFU/assay using one sample

Level of agreement 80%

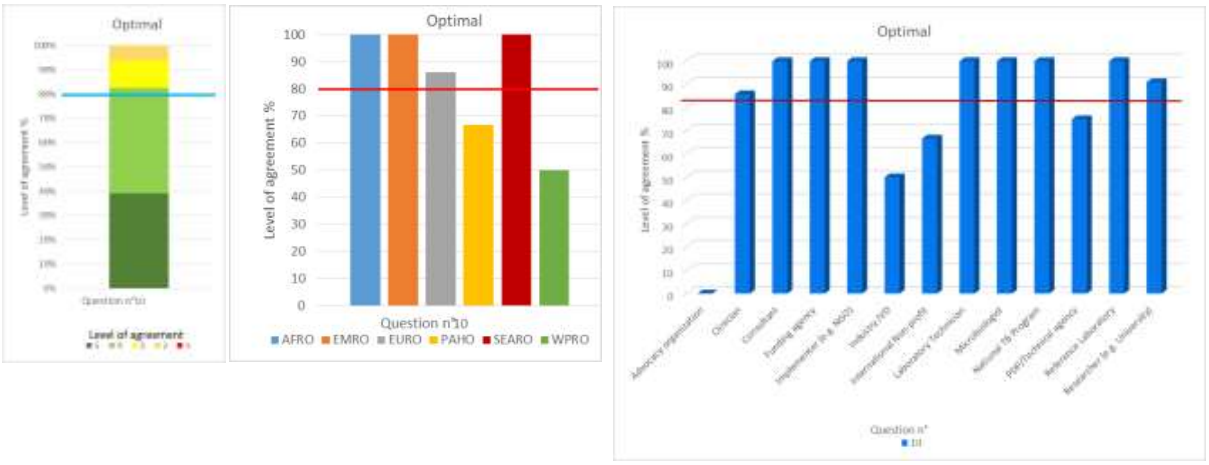
Stratified by WHO region **Stratified by professional role**



Optimal: ≥4.5 genome equivalents/reaction and 131 CFU/mL of sputum

Level of agreement 83%

Stratified by WHO region **Stratified by professional role**



Comments:

- Optimal: 50 CFU/ml
- A single assay is all that should be required if QA is in place and assay is optimized
- Ditto
- I do not understand why there needs to be a second reaction?
- I am not expert in this question
- Needs to be thought through. This all needs to be tied together as there are consequences
- still not enough

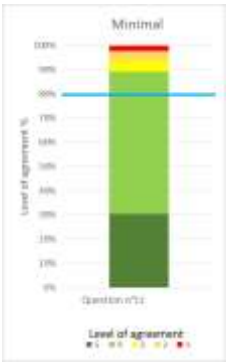
11) Diagnostic sensitivity for TB detection

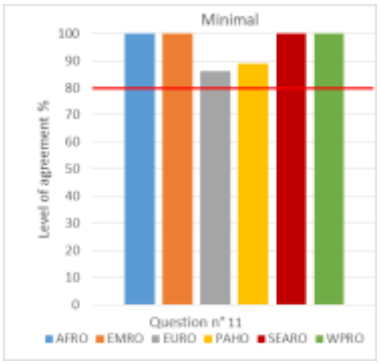
Estimates are based on current commercially available assays endorsed for direct testing on sputum specimens.

Minimal: Sensitivity should be >80% for a single test when compared with 2 liquid cultures; for smear-negative TB it should be >60%; for smear-positive TB it should be 99%

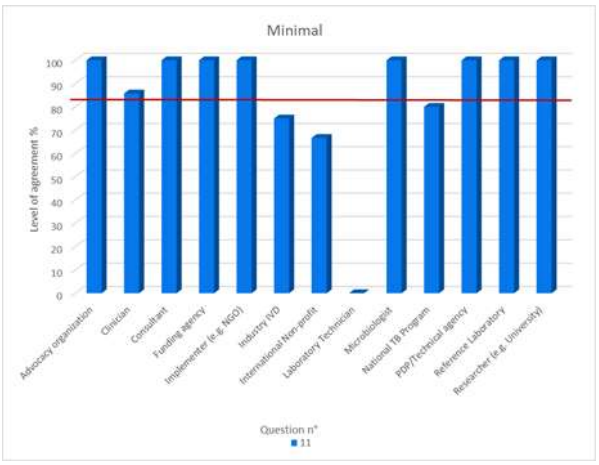
Level of agreement 89%

Stratified by WHO region





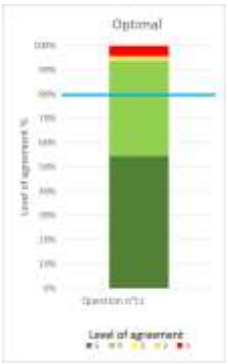
Stratified by professional role

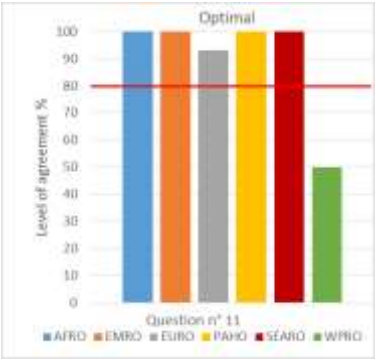


Optimal: Sensitivity for detecting TB should be >95% for a single test when compared with 2 liquid cultures; for smear-negative TB it should be >68%. For smear-positive TB it should be 99%

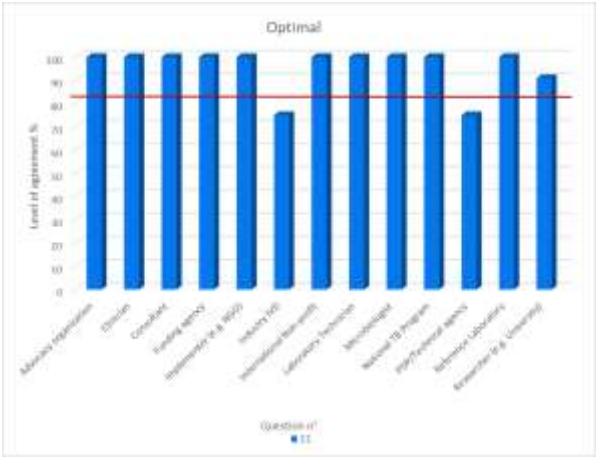
Level of agreement 93%

Stratified by WHO region





Stratified by professional role



Comments:

- Not 68% but 80%
- 60% for smear negative TB seems too low, to introduce such a new and expensive testing method
- Diagnostic sensitivity needs to be higher range to justify the added costs

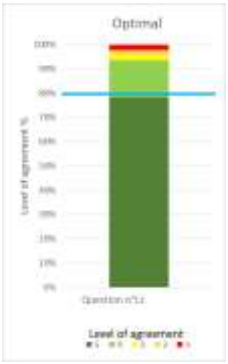
- Smear negative OPT should be 80% (not 68%), we need to push harder for increased sensitivity - also What about OPT: having capacity to test OTHER specimens for Cell Free DNA (blood/Urine) or even consider EPTB samples in the TPP.
- Why 2 cultures when standard of care is one? Performance should be assessed versus standard of care. TRUTH is a different matter
- if only 70% for smear negative, still keep culture then

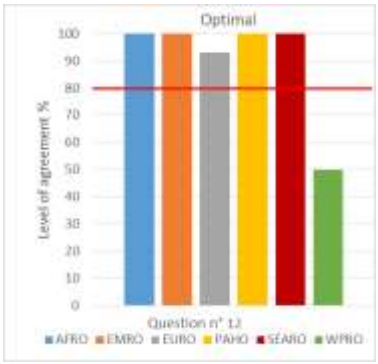
12) Diagnostic specificity for TB detection

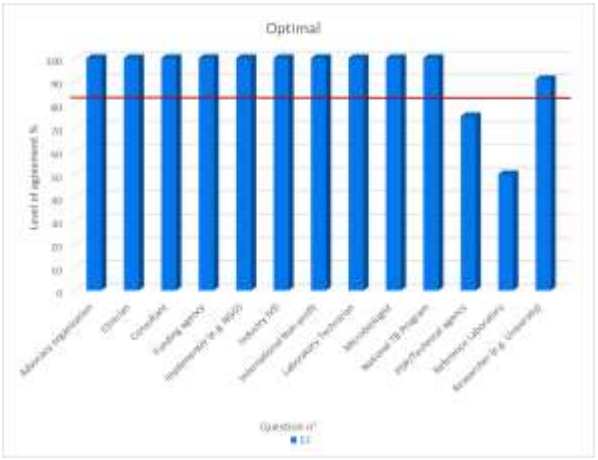
Optimal/Minimal: Specificity should be >98% for a single test when compared with culture

Level of agreement 93%

Stratified by WHO region **Stratified by professional role**







Comments:

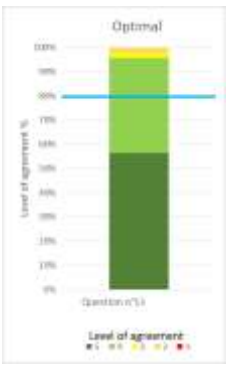
- >99%
- Difficult to achieve with a more sensitive molecular test. I think >=97% should be acceptable
- Specificity should be >99%

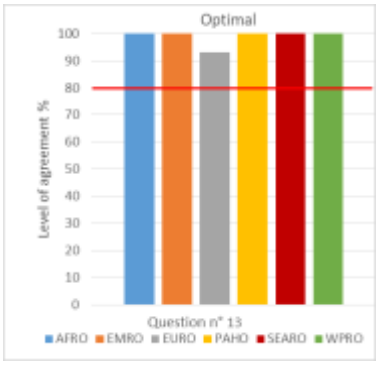
13) Diagnostic sensitivity for DST compared against genetic sequencing as the reference standard

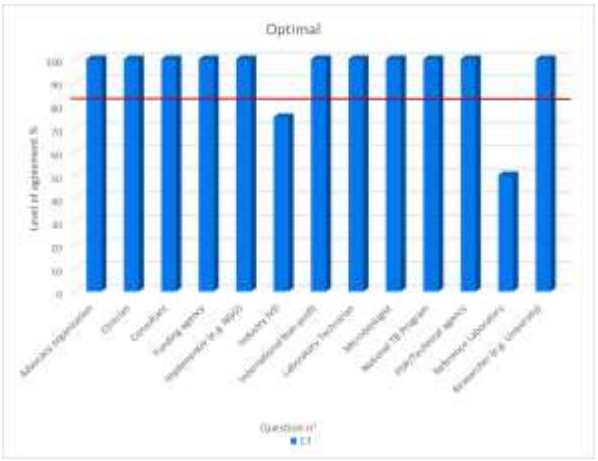
Optimal/Minimal: Sensitivity should be >98% for detecting targeted SNPs for resistance when compared with genetic sequencing

Level of agreement 96%

Stratified by WHO region | **Stratified by professional role**







Comments:

- Sensitivity should be the same than genetic sequencing
- Is this for every claimed mutation? This will take years of testing to prove as written here. Samples or plasmids? The concept is correct but the wording needs clarification
- may be even higher

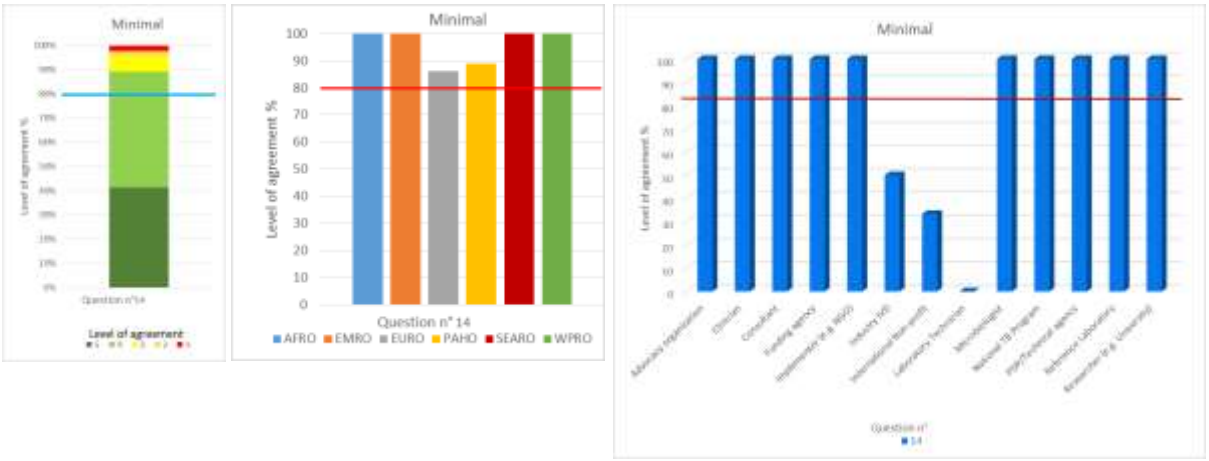
14) Diagnostic sensitivity for DST compared against phenotypic DST as a reference standard

Modelling data suggest that for rapid DST to be more cost effective than culture, on a currently available platform it must attain an aggregated sensitivity of 88% for all clinically relevant mutations. A lower sensitivity could be tolerated for a test with high specificity.

Minimal: RIF: >95% sensitivity for detection of phenotypic resistance. INH, FQ: >90% sensitivity for detection of phenotypic resistance. BDQ, LZD, CLO, DLM, pretomanid, AMK, PZA, KM*, CAP*: ≥80% sensitivity for detection of phenotypic resistance

Level of agreement 89%

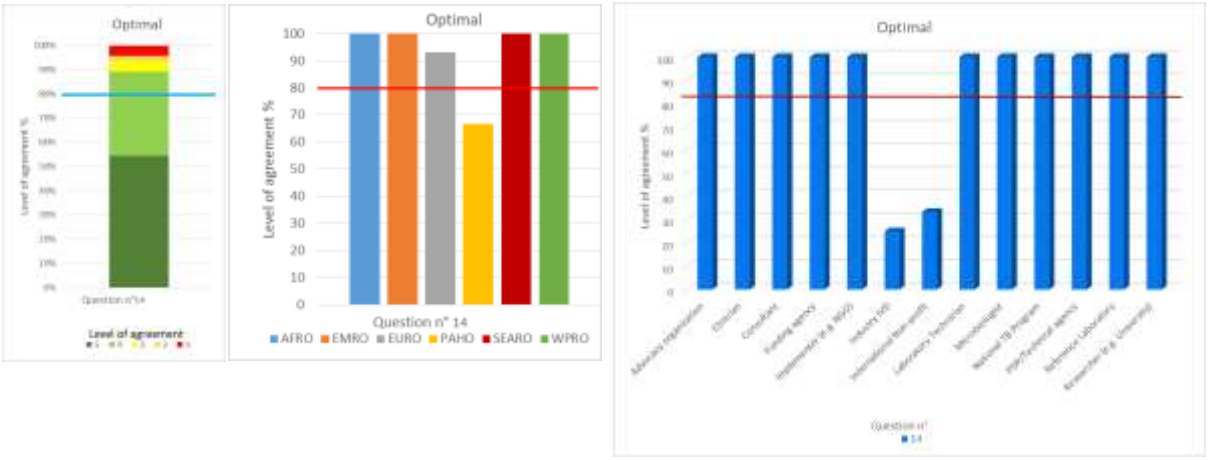
Stratified by WHO region **Stratified by professional role**



Optimal: RIF, INH, FQ, BDQ, LZD, CLO, DLM, pretomanid, AMK, PZA, KM, CAP: >95% sensitivity for detection of phenotypic resistance

Level of agreement 89%

Stratified by WHO region **Stratified by professional role**



Comments:

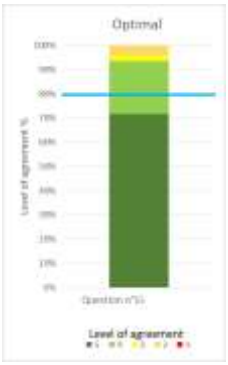
- Given that the molecular basis of resistance is not well defined for newer drugs and even CAP, this would be unattainable
- 80% seems too low for most of drugs mentioned
- Additional work is still needed to ensure sensitivity of >95% for all drugs. Minimal is achievable now and will help obtain the data needed to increase sensitivity of some drug-specific assays
- Our tests for TB work in a background of 20,000 human genomes
- Seems very ambitious
- Needs clarification and we already know based upon retorspective analysis that this is not achievable as stated here. How would this be studied as I can cherry pick samples and you could never be achieve this

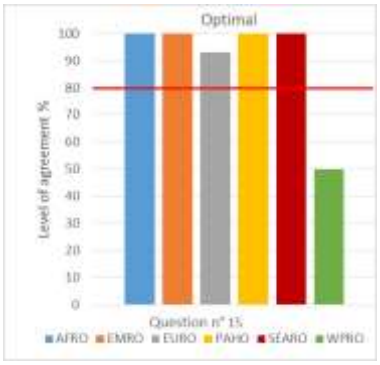
15) Diagnostic specificity for DST compared against genetic sequencing as the reference standard

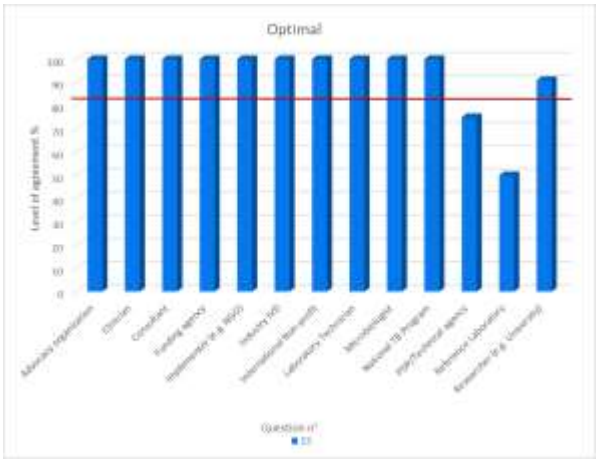
Optimal/Minimal: Specificity should be ≥98% for any anti-TB agent for which the test is able to identify resistance when compared against genetic sequencing as the reference standard

Level of agreement 93%

Stratified by WHO region **Stratified by professional role**







Comments:

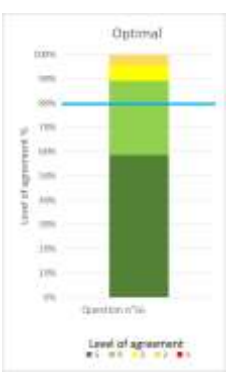
- ≥99%
- Specificity should be the same
- should be as close to 100% as possible

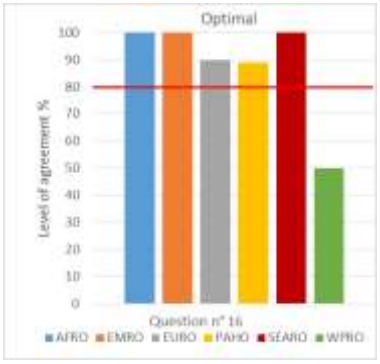
16) Diagnostic specificity for DST compared against phenotypic DST as a reference standard

Optimal/Minimal: The specificity of targeted sequencing for the mutations included for any anti-TB agent for which the test is able to identify resistance should be ≥ 98% when compared against the phenotypic reference standard recommended for each anti-TB agent

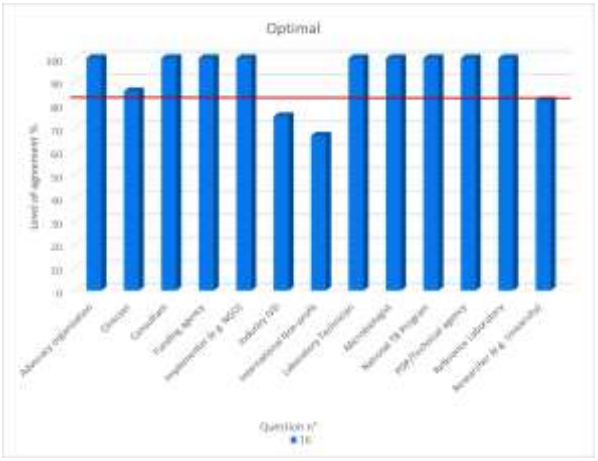
Level of agreement 89%

Stratified by WHO region





Stratified by professional role



Comments:

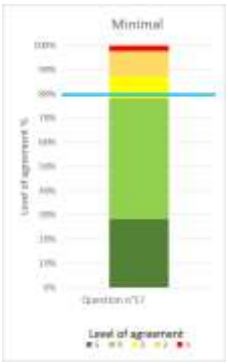
- Should consider possibility of disputed mutations and allow for lower specificity if the presence of such mutations can be proven
- This depends on the culture based assay and the drug tested. There are some discrepancies such as the disputed rpoB mutations.
- phenotype DST to some drugs is not reliable such as PZA
- We know current standards for DST are inadequate, this should be better defined
- should be as close to 100% as possible

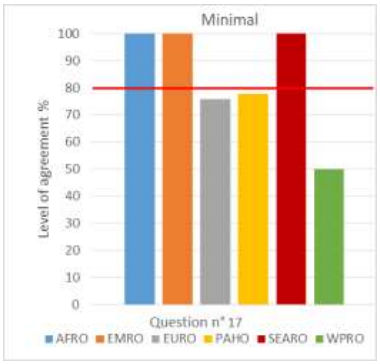
17) Limit of detection of minor variants

Minimal: ≤20% (that is 20 resistant bacteria out of 100)

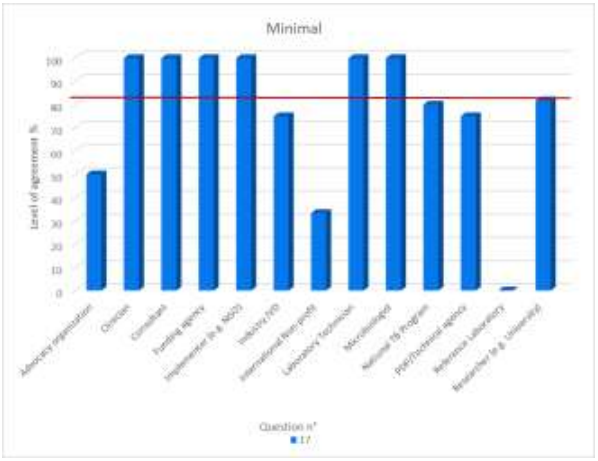
Level of agreement 78%

Stratified by WHO region





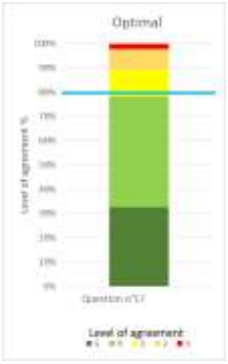
Stratified by professional role

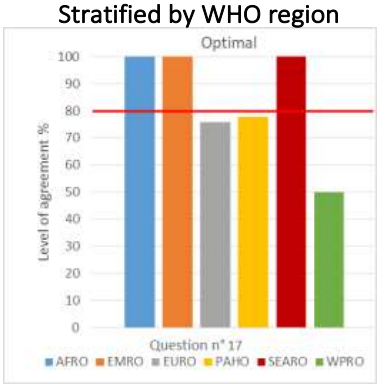


Optimal: ≤10% (that is 10 resistant bacteria out of 100)

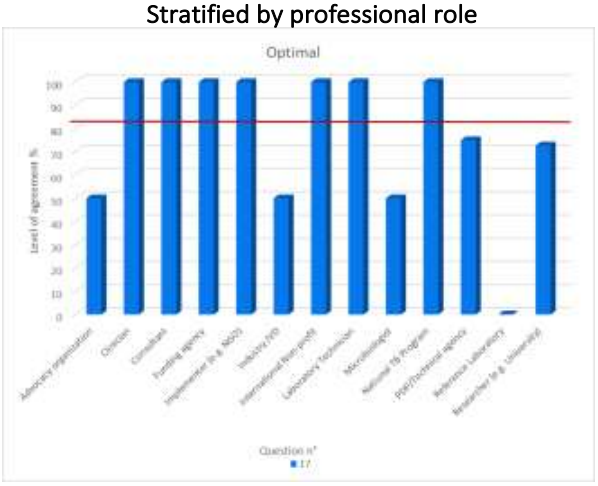
Level of agreement 78%

Stratified by WHO region





Stratified by professional role



Comments:

- Minimal of </= 30% would be equivalent to Sanger and still useful.
- ≤10% and ≤5%
- This really depends if this is targeted or WGS. Answers are for a targeted approach.
- This should be lower
- Again, not informed about this to give opinion in either direction
- I would like 1%. Since now I see that there are newer seq technologies coming out that can sequence a single cell!
- Both should be lower

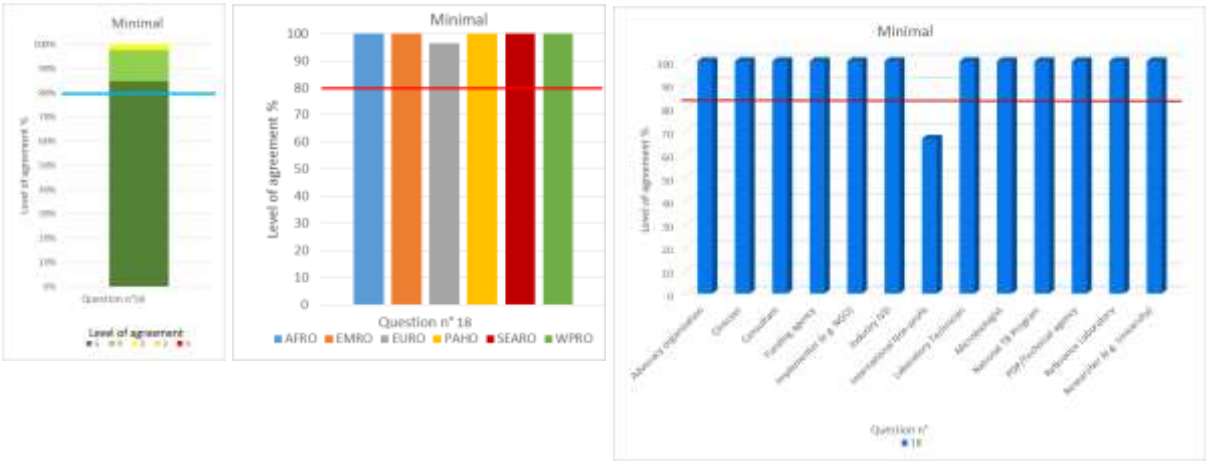
- Further clarification required, what concentration of total bugs, Is this at the LoD for the minor variant? Do I have to do LoD for ALL mutations claimed
- 10% is already pathologic and on going for resistance selection
- optimal <5%
- Optimal: ≤1%. If we conventionally take 1% as the critical proportion for phenotypic DST

18) Analytical specificity for TB detection

Minimal: No cross-reactivity with other organisms including nontuberculous mycobacteria

Level of agreement 98%

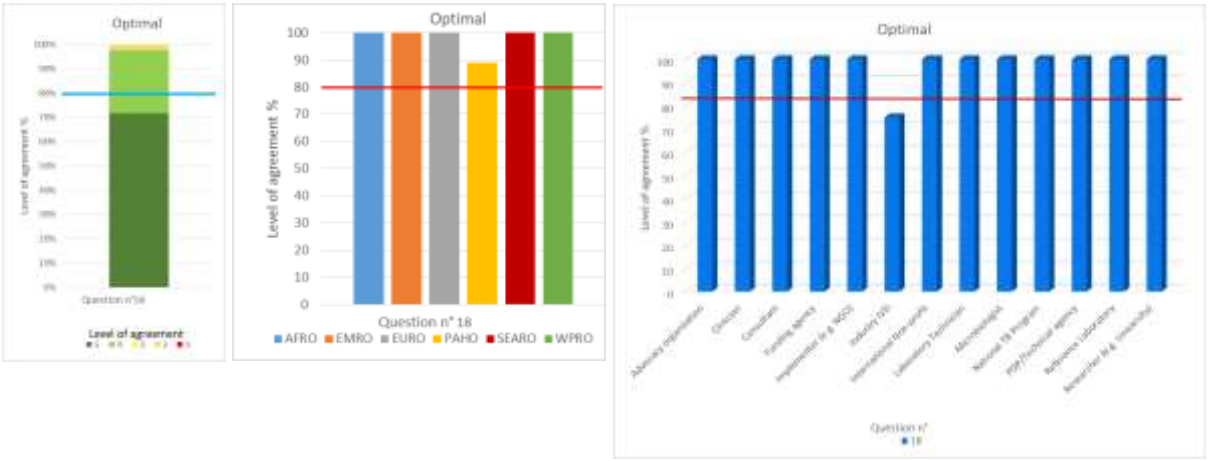
Stratified by WHO region **Stratified by professional role**



Optimal: No cross-reactivity with other organisms. NTM identification should be also available

Level of agreement 98%

Stratified by WHO region **Stratified by professional role**



Comments:

- With the advancements being made in NGS, I think OPT should be able to discriminate MTBC vs NTM vs other lung infections, or even maybe discriminate bacterial vs viral infections.
- The Optimal should be the minimal

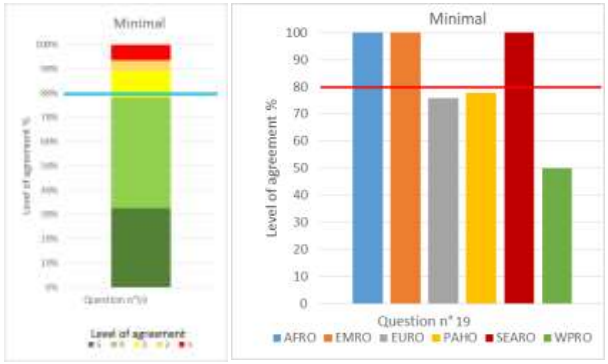
19) Indeterminate results during detection

Indeterminate results may be caused by a lower sensitivity for detecting TB during the second reaction.

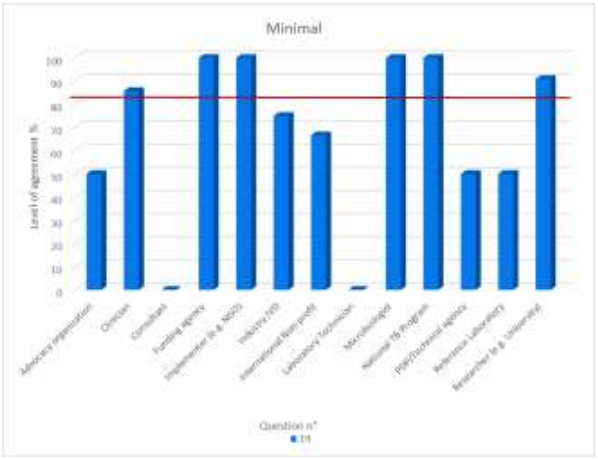
Minimal: <10%

Level of agreement 78%

Stratified by WHO region



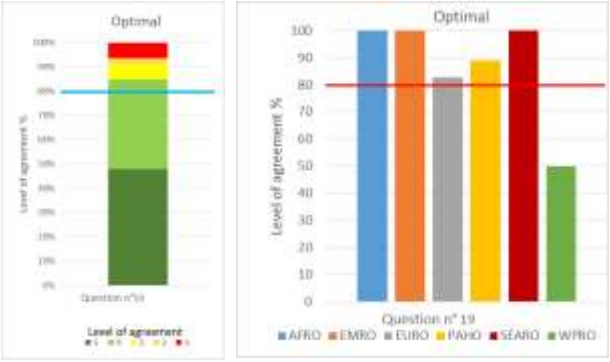
Stratified by professional role



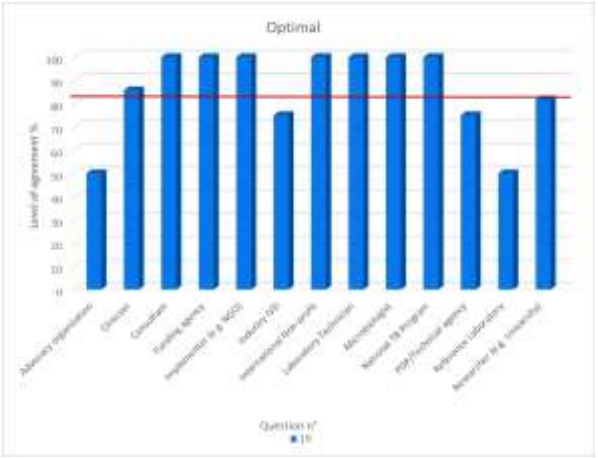
Optimal: <5%

Level of agreement 85%

Stratified by WHO region



Stratified by professional role



Comments:

- <5% and <1%
- 10% of indeterminate results is too high
- Vague question
- Would encourage lower limits for both, 3% and 7%
- <10% seems relatively high, even for the second reaction. is a lower threshold feasible?

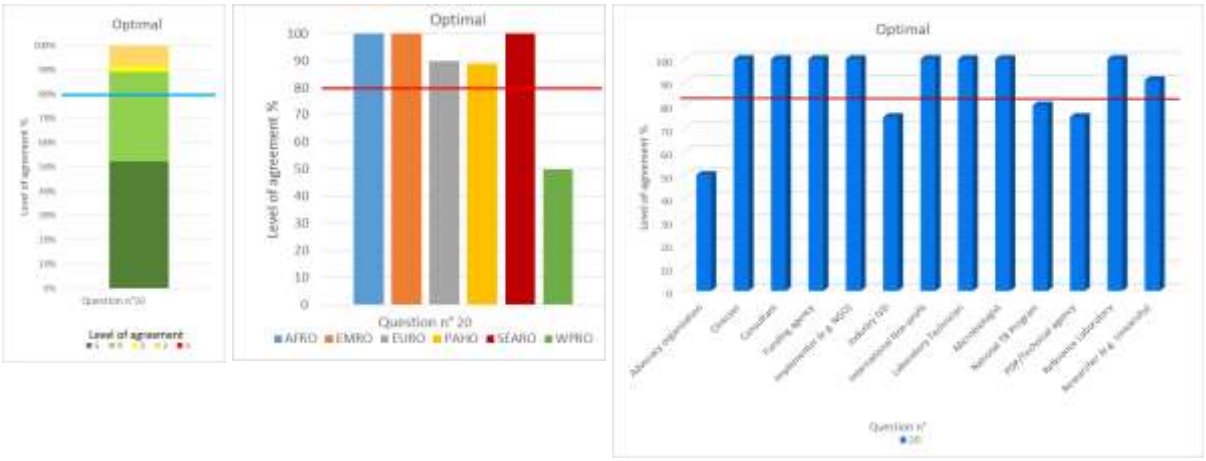
- I do not understand the question
- This is too high to be acceptable. Thin of the cost to the lab if there are 10% indeterminates' who would use the assay
- too high

20) Reproducibility

Optimal/Minimal: Interassay coefficients of variance should be ≤10.0% at the high and low extremes of the assay

Level of agreement 89%

Stratified by WHO region **Stratified by professional role**



Comments:

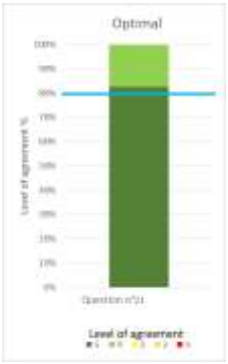
- ≤5%
- Again, not informed to offer meaningful opinion on this.
- For all aspects of the assay, what about mixtures? Is acceptable but needs to be further defined
- too high variance if 10%

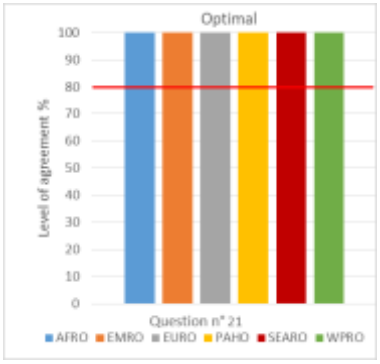
21) Interfering substances

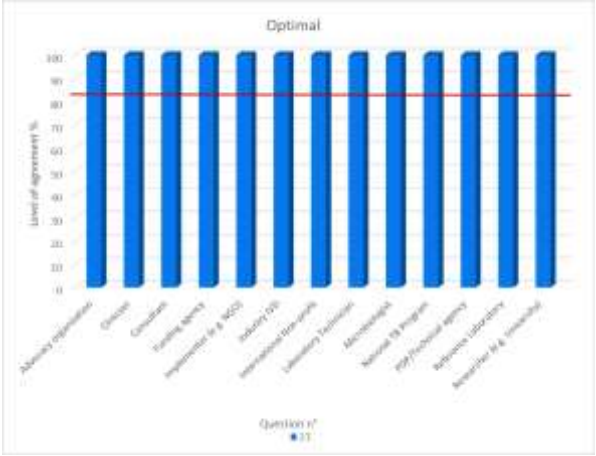
Optimal/Minimal: No interference should be caused by those substances known to occur in the human respiratory and pulmonary tracts, including blood that could potentially inhibit a PCR reaction, and substances used to treat or alleviate respiratory disease orsymptoms

Level of agreement 100%

Stratified by WHO region | **Stratified by professional role**







Comments:

- particularly anti-TB agents
- Should provide an appendix with the list

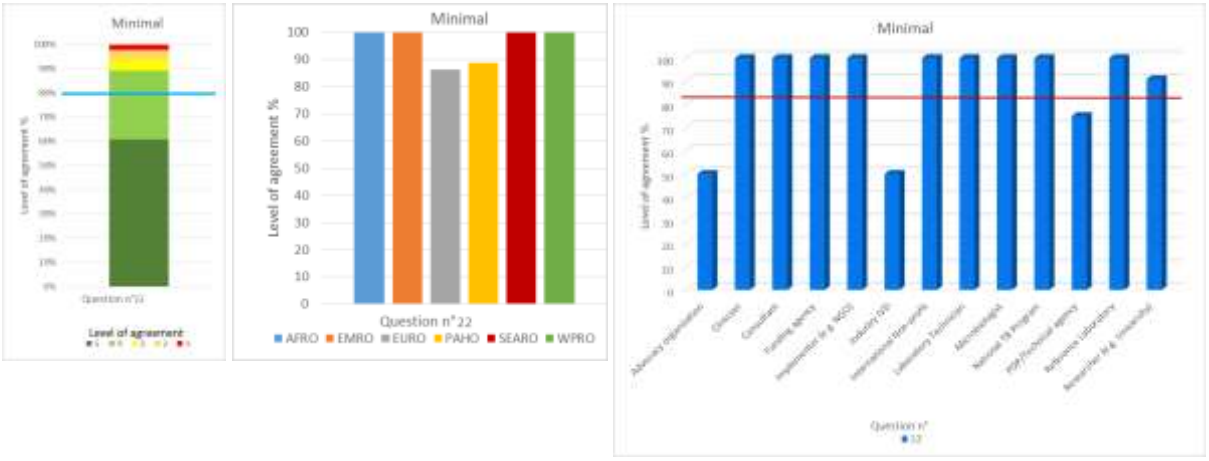
22) Treatment monitoring capability

A test that is able to replace smear microscopy for treatment monitoring (for example, by detecting viable bacteria) is more likely to be adopted and to completely replace smear microscopy; thus, it would have a larger market as well.

Minimal: Yes (preferable)

Level of agreement 89%

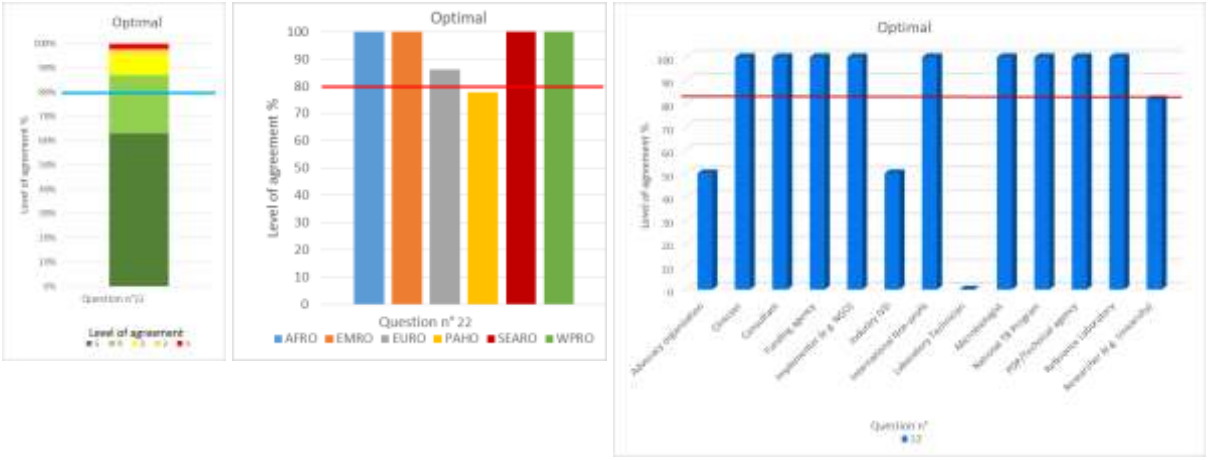
Stratified by WHO region | Stratified by professional role



Optimal: Yes (mandatory)

Level of agreement 87%

Stratified by WHO region | Stratified by professional role



Comments:

- Microscopy will never be fully replaced - in many settings, it will remain the best / cheapest method to follow-up a patient under treatment or the infectiousness of a patient beginning treatment
- WOuld suggest No (minimal) and Yes (preferably; optimal) as the tests will likely be a molecular test and unlikely to be capable of monitoring

- Would require RT-PCR which we are very good at
- MIN: may be able to build capacity for this OPT:Preferable (If we are testing Nucleic acids, we may be able to assess RNA from a specimen simultaneously in a test that would allow for us to see that transcription is occuring, which means live cells) But, I would NOT require this as it may limit technologies from coming forward by asking too much. Another key would be to have a more definitive value on genomic units for bacterial load, to eliminate microscopy all together from both Dx and Tx monitoring.
- Wording is ambiguous and may cause confusion - 'yes' for minimal level may limit the types of technologies/developers to build such a test.
- This requires two different performance so it can't be achieved in most cases and will complicate the assay design

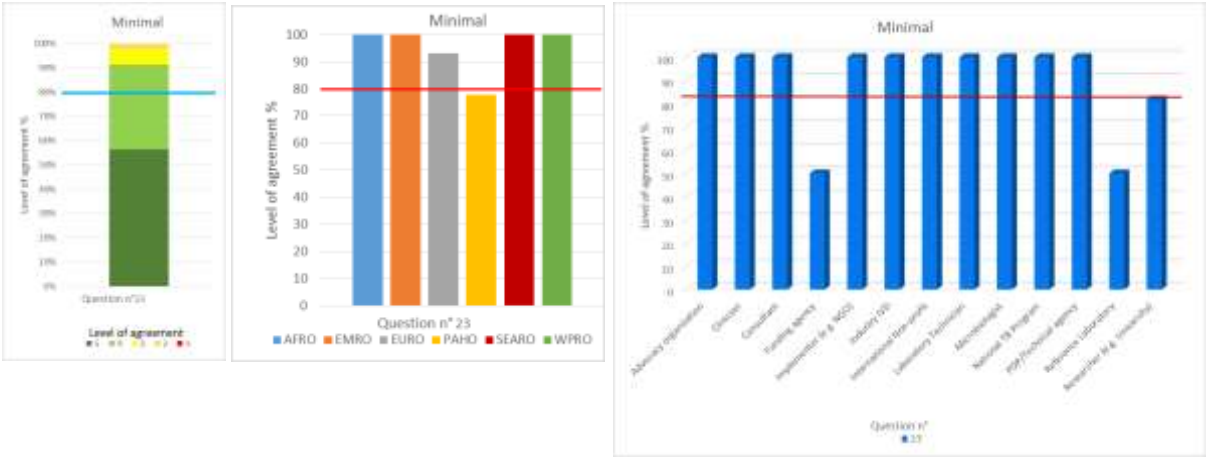
23) Multiuse platform

Any technology entering this market should be useful for diagnosing also relevant diseases other than TB. The diseases to be targeted should be those among the WHO list of poverty-related diseases, communicable diseases and AMR priorities. Multiplex testing or the ability to use a platform to perform different tests will likely increase the acceptability of the new test, especially in the private sector.

Minimal: Yes (achievable)

Level of agreement 91%

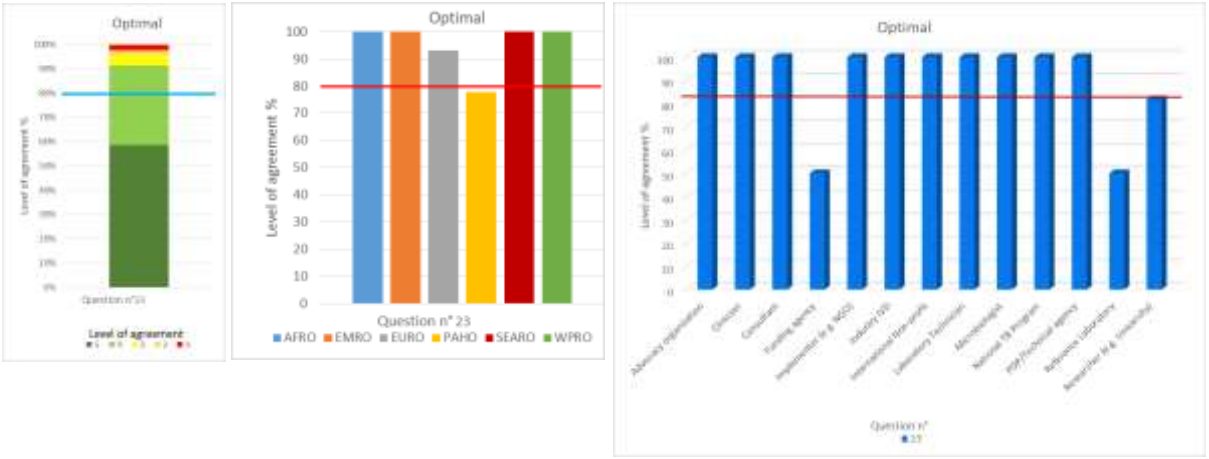
Stratified by WHO region **Stratified by professional role**



Optimal: Yes (demonstrated)

Level of agreement 91%

Stratified by WHO region **Stratified by professional role**



Comments:

- This would raise training and equipment costs
- Too complicated target
- Would also encourage open source platform at least as optimal

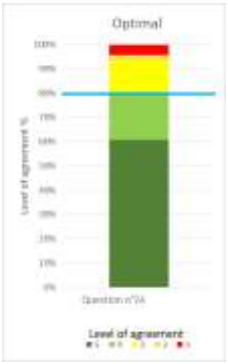
- You should not be dictating the menu selection for the manufacturer as this is interference in their business and a law suit waiting
- Development of assay for other diseases should not at all impact TB diagnosis.
- not necessary

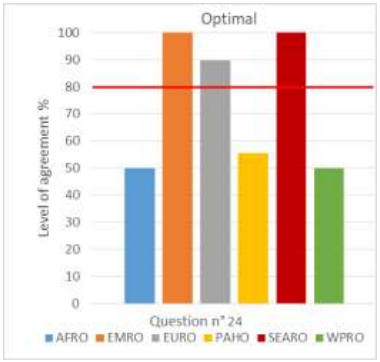
24) Sample type

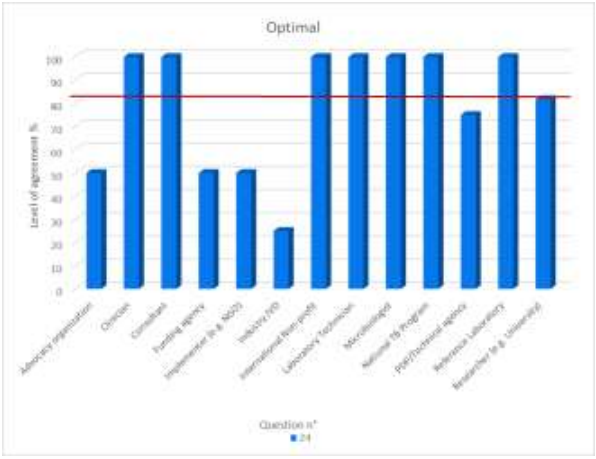
Optimal/Minimal: Unprocessed sputum

Level of agreement 80%

Stratified by WHO region | **Stratified by professional role**







Comments:

- What you mean, unprocessed sputum??
- Include other types of samples is necessary
- agree - time / temp storage - may be covered later
- Possible, but we have no direct experience yet
- This is limiting. What about EPTB samples. And as we know cfDNA using capture technologies for sample prep may allow other fluids like blood and urine. OPT should cover more sample types, while SPUTUM is minimal either unprocessed or processed.
- difficult to obtain in pregnant and pediatric populations. we need easily accessible specimens
- Why is this Optimal, Optimal should include more than Minimal in this case. Also clarification is required ALL assays currently require processed sputum. This means that unprocessed sputum is placed in the disposable
- Should be capable of other more easily collected sample types
- Minimal: unprocessed sputum, Ideal: unprocessed sputum and additional clinically relevant specimens for TB or other diseases of importance (WHO list of poverty-related diseases, communicable diseases and AMR priorities)

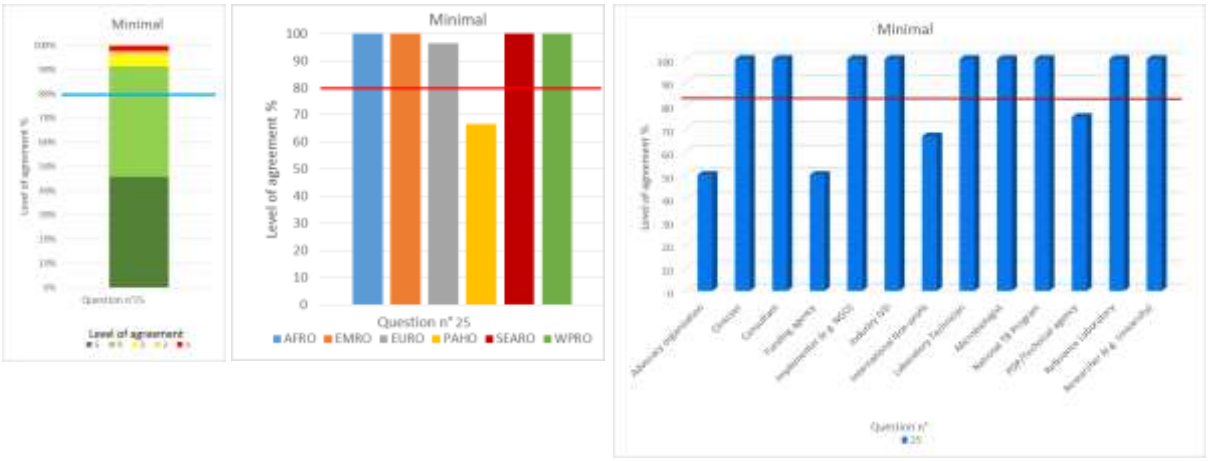
25) Sample volume

The lowest volume possible for all types of samples should be 0.1 ml, especially since HIV-positive patients may have difficulty providing a sample; however, this should not come at the expense of decreased sensitivity.

Minimal: <0.5-2 mL

Level of agreement 91%

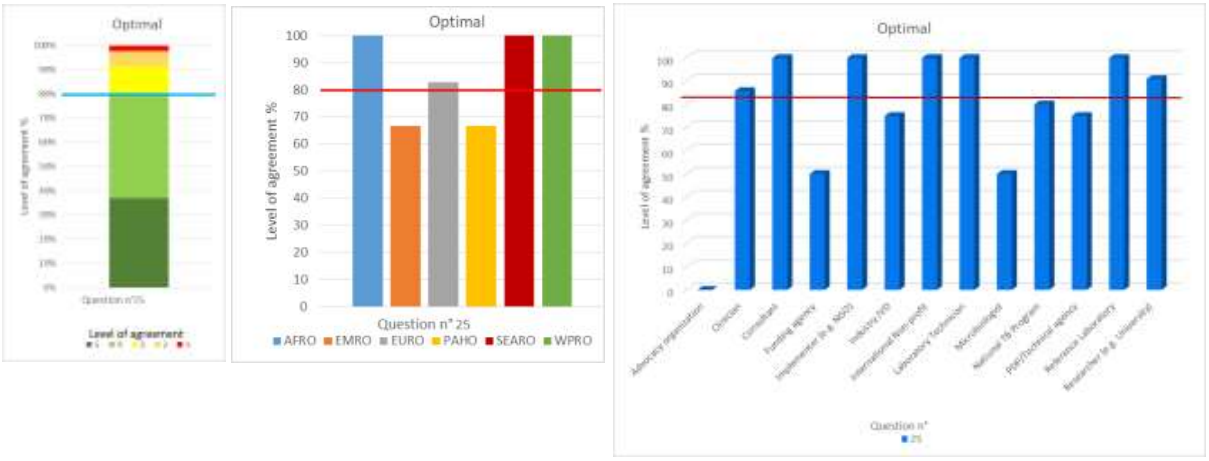
Stratified by WHO region **Stratified by professional role**



Optimal: Up to 10 mL

Level of agreement 80%

Stratified by WHO region **Stratified by professional role**



Comments:

- Reduced sensitivity likely due to low sample volume would negate the advantage unless the patient is known to have a high bacterial burden
- up to 2 ml
- Our goal is to achieve extreme sensitivity therefore need only small samples

- Again, not informed enough to offer meaningful opinion
- For sputum: this depends on processed or unprocessed but <2ml for MIN Raw. OPT- again we should address other sample types and possibly list for each possible sample type and NOT limit to sputum. We need to try to get researchers to use capture technologies to process or concentrate cells or DNA either externally or within the technology…so maybe this is too simplified as it is stated.
- For lowest possible volume, the optimal criteria should be < 0.5 - 2 mL and minimal should be < 10 mL.
- Is it possible the minimal and optimal conditions are swapped?
- Needs to be reworked as the only sample type is sputum. Should be restated as this is incorrect for the Minimal. What happens if a reflex solution is provided?
- Minimal and optimal should be switched.
- It would be great to have a test not affected by sample volume, mantaing same high sensitiviy, specificity and accuracy

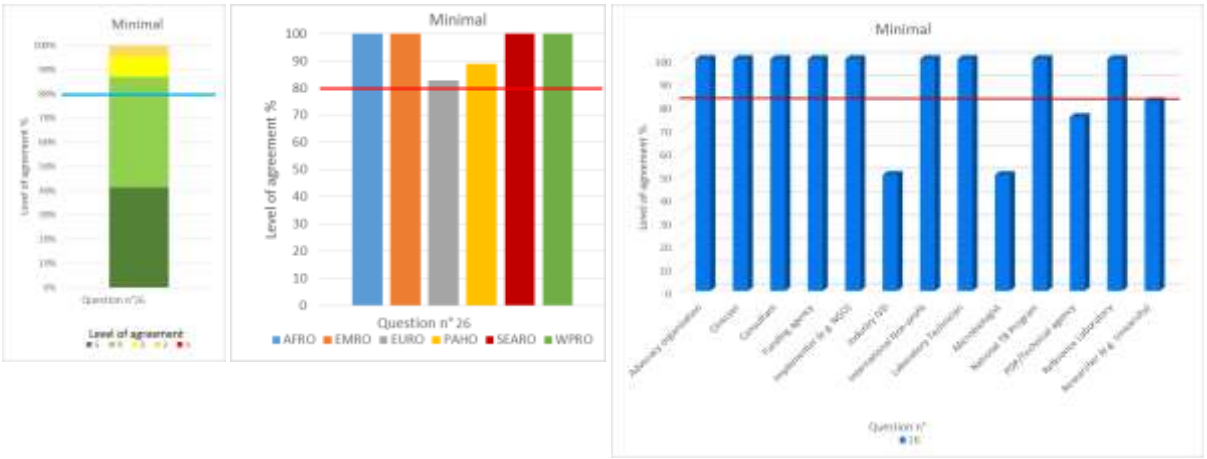
26) Manual preparation of samples (steps needed after obtaining sample)

Precise volume control and precise timing should not be required. Only general/cross-cutting laboratory skills required; no specific analytical procedures based on additional instruments should be required (e.g. DNA quantification, gel electrophoresis, serial dilutions…).

Minimal: Less than 5 steps

Level of agreement 87%

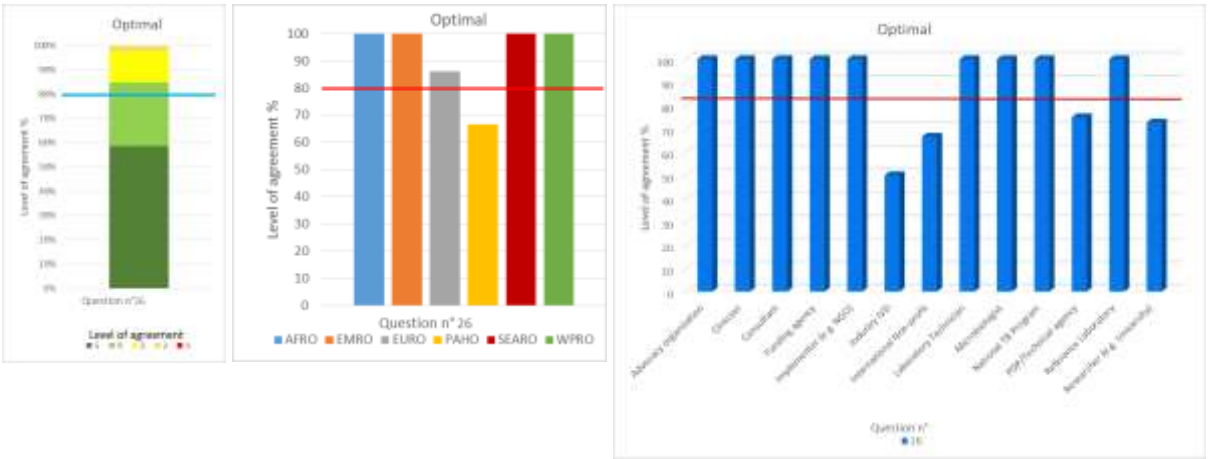
Stratified by WHO region **Stratified by professional role**



Optimal: No steps or 1 step

Level of agreement 85%

Stratified by WHO region **Stratified by professional role**



Comments:

- Control points are requiered
- I think 5 steps might be too difficult for a microscopy center level. Would say 3 or less
- Sample prep will be critical to ensure an accurate and reproducible result
- Spum + KOH + Neutralization may be sufficient, but need to test

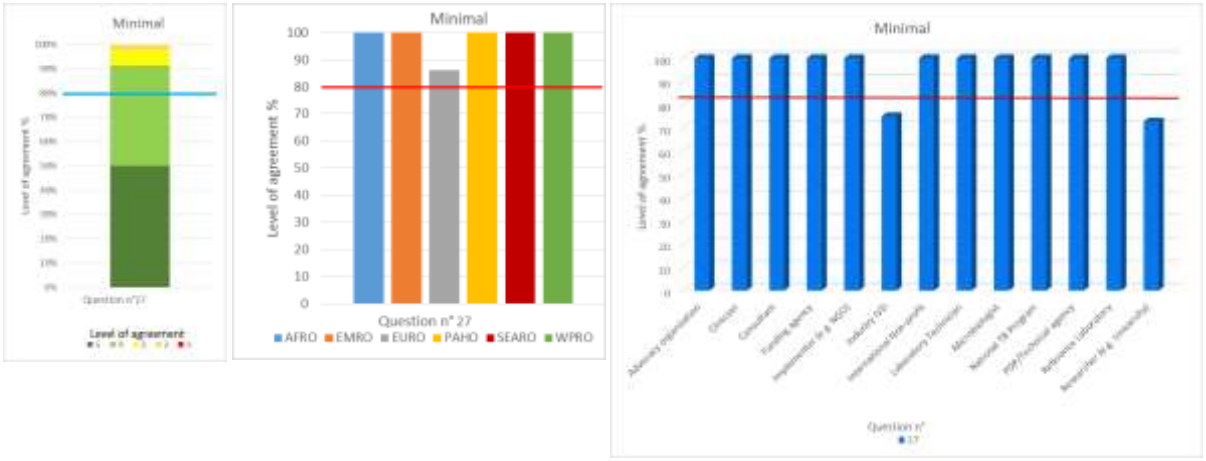
- Min 3 steps
- How are steps defined? There will always be a step that someone must do. Does this include entry of information in tot he platform?
- 5 steps are too numerous
- Minimal: Volume control (>50microL) or 'drops' with dedicated sample transfer device. Precise timing (+/- 5min per step). Similar to serology Lateral Flow Rapid Test requirements; 5 steps.
- No steps is unlikely
- Much depends on the nature of the steps.

27) Reagent integration

Minimal: No specific indications, but refer to reagent kit storage and stability for restrictions

Level of agreement 91%

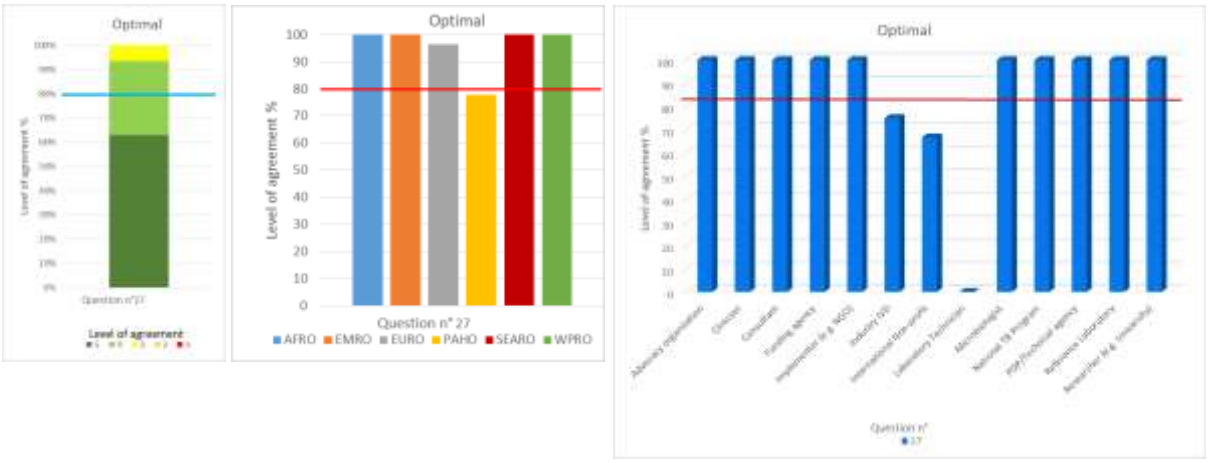
Stratified by WHO region **Stratified by professional role**



Optimal: All reagents should be contained in a single device

Level of agreement 93%

Stratified by WHO region **Stratified by professional role**



Comments:

- I do not think that it is possible to contain all reagents in a single device
- Should specify number of external reagents allowed in line with steps in prep allowed. Also would specify that additional reagents need to be part of price considered (and not separately charged)
- Integration into a single device will raise costs so keeping reagents separate and centralizing with QC will keep costs down
- Tubes with dried reagents
- MIN- All reagents required should be provided. NO external procurement necessary

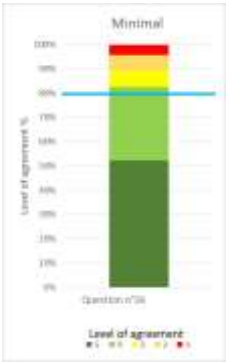
- Minimal needs clarification as this means nothing as written and currently in all cases there will need to be a liquefaction reagent which is not part of the test cassette
- Minimal may include some reagent integration anyway

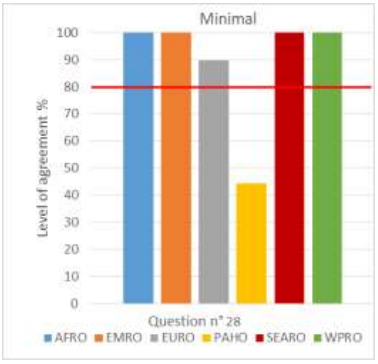
28) Time to result

Minimal: <24 hours for detection and DST

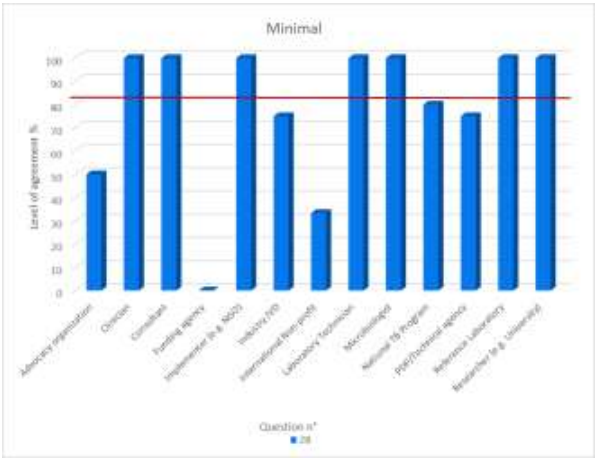
Level of agreement 83%

Stratified by WHO region





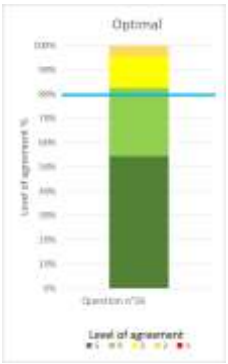
Stratified by professional role

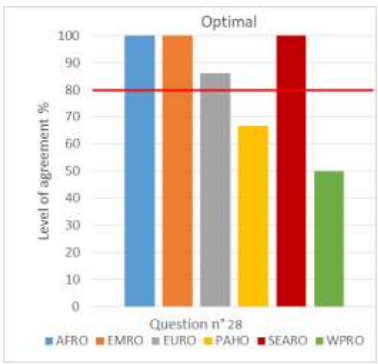


Optimal: <30 minutes for detection and DST

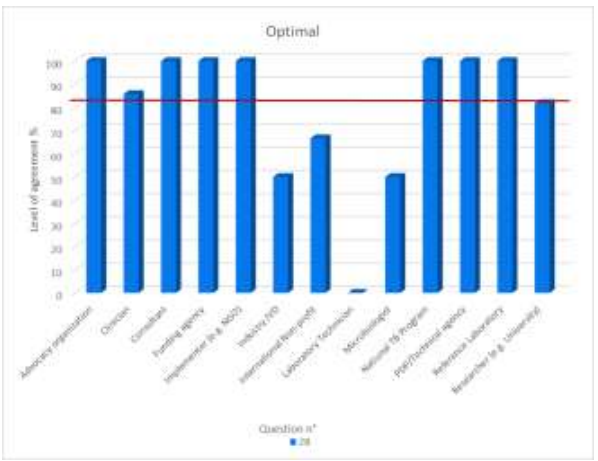
Level of agreement 83%

Stratified by WHO region





Stratified by professional role



Comments:

- To provide results during same visit the results should be available in< 8hr
- Quite impossible to produce a method doing all this in < 30 minutes...
- Less than 24h is OK, but more than 2 or 3 hours requires a second visit to health facility, which is not desirable.
- I suspect it will require 2-3 days
- Depends on conc. of TB and complexity of test
- This is a huge gap between the two, and many patients will be LTFU if it's not completed within a day. I would instead recommend 6 hours for minimal

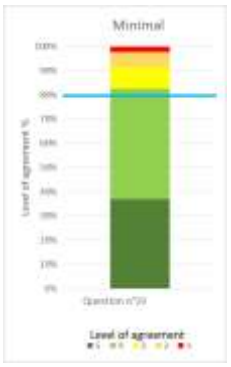
- I think Minimal should be closer to Xpert (maybe we expand write preferable <2hrs, but <24hrs acceptable) OPT: preferable 30min, but acceptable < 1hr
- <12 hours should be the minimal to avoid significant LTFU
- 2 hours or longer (<12 hours) would be well acceptable
- Times are very unrealistic and provide no guidance to the manufacturer. We know that this will cause problems if it is 23 hours and 59 minutes
- Minimal: <8 hrs. Same day results need to be less than 2hrs
- Range from 15 min to 2 hours would be as good.

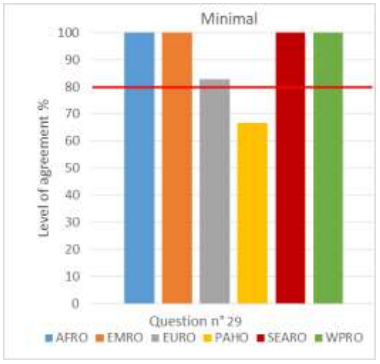
29) Daily throughput

Minimal: >10tests

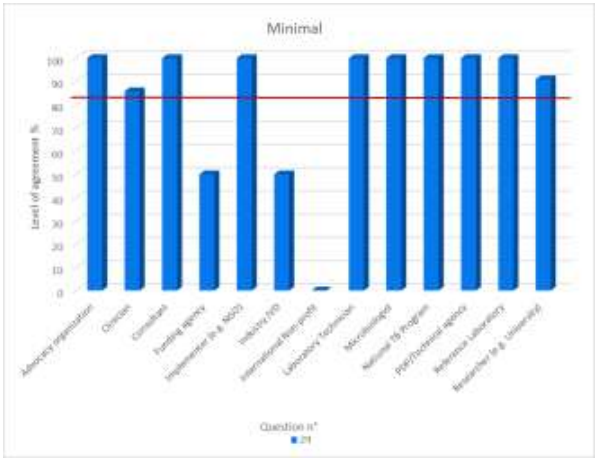
Level of agreement 83%

Stratified by WHO region





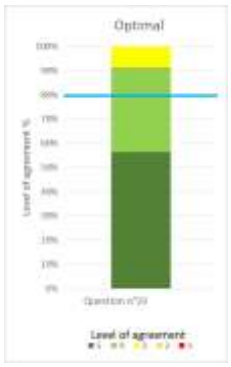
Stratified by professional role

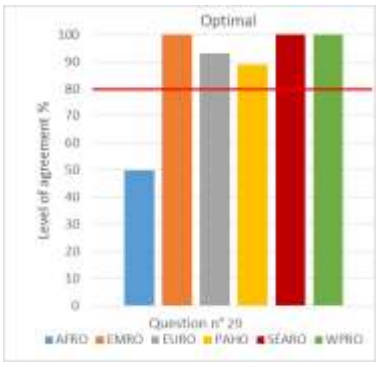


Optimal: >25tests

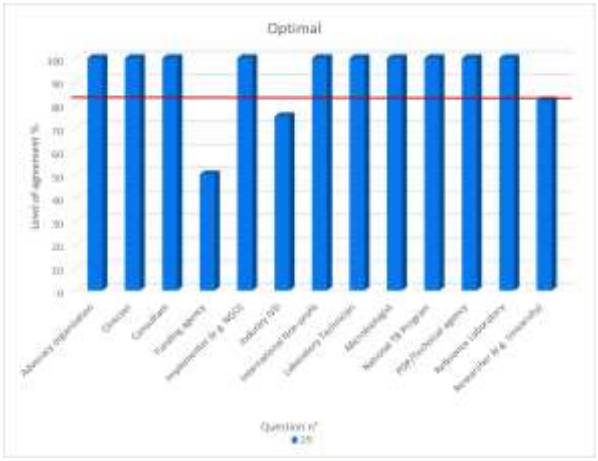
Level of agreement 91%

Stratified by WHO region





Stratified by professional role



Comments:

- Peripheral centers may only see a few patients a day. Even a single test would be useful if within price point. Otherwise batching will be necessary.
- Batching samples may be one way in which to reduce costs but will also extend the turnaround time depending on need. Having flexibility will be able to address different needs
- Minimal: at least 25
- For OPT: may not have any restriction but a platform that can be flexible to have high capacity (100) or with limited lower capacity of 25. Maybe this is already clear as it is stated.
- Seems high for minimal

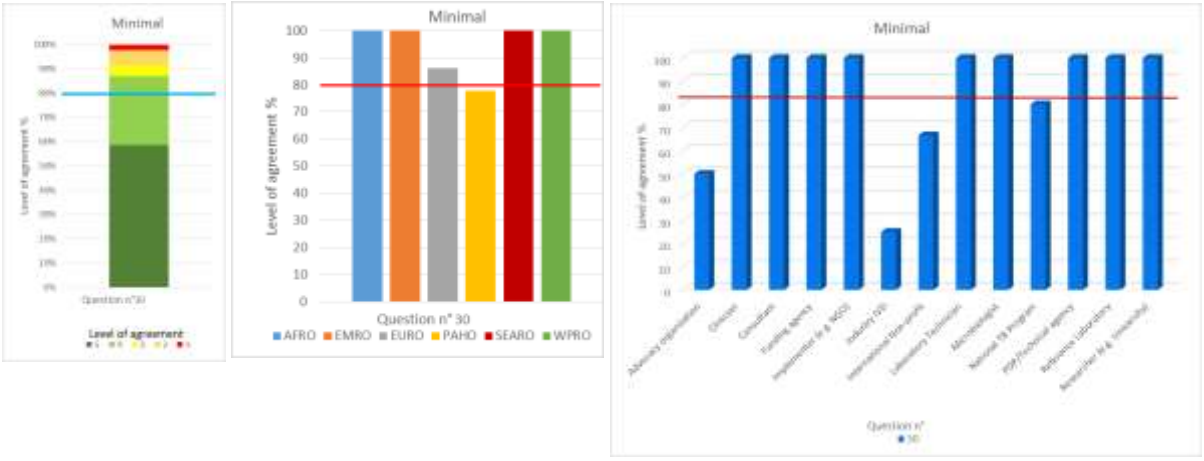
- we need rapid assays and atleast 50 a day
- How is this defined, one unit, one rack what? If this is at the lowest level of the system then why greater than 10 per day? This si also deendent upon the turn around time so this needs to be defined
- Optimal would be more than 100. Minimal needs to be greater than 20.
- should be minimal for one or two tests
- optimal may be a bit lower in a routine setting, e.g. >15-20

30) Sample capacity and throughput

Minimal: Batching should be possible

Level of agreement 87%

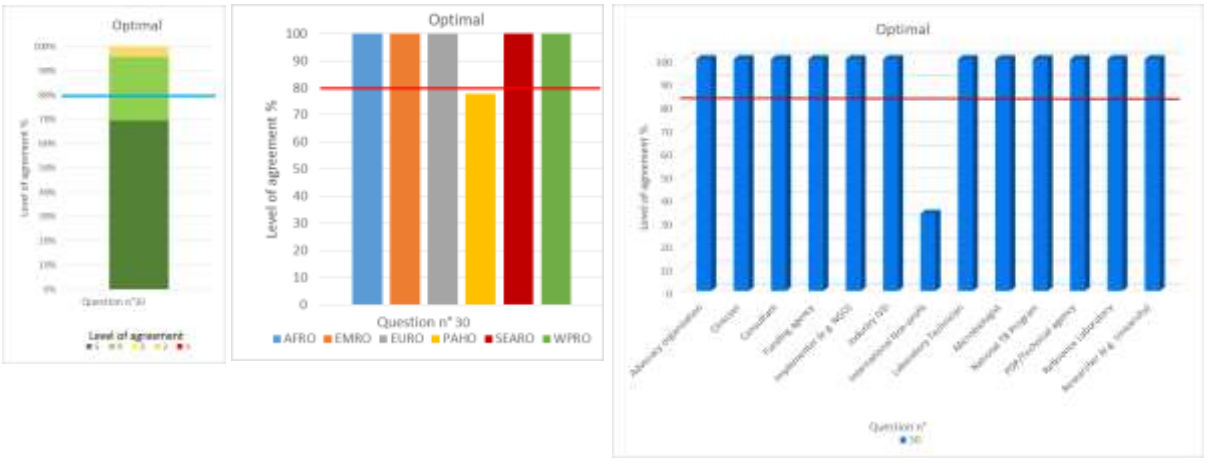
Stratified by WHO region **Stratified by professional role**



Optimal: Multiple samples should be able to be tested at the same time; random access should be possible

Level of agreement 96%

Stratified by WHO region **Stratified by professional role**



Comments:

- I am not sure why batching is a requirement- at peripheral centers it may be helpful just to test 1 sample to have results during the same patient visit.
- Not possible for NGS random access
- I think batching limits TAT and should not be considered at minimal these days. I would like to push for capacity to test multiple samples at once for MIN. and OPT include the the random access, so that you can add tests to run at anytime.
- No batching

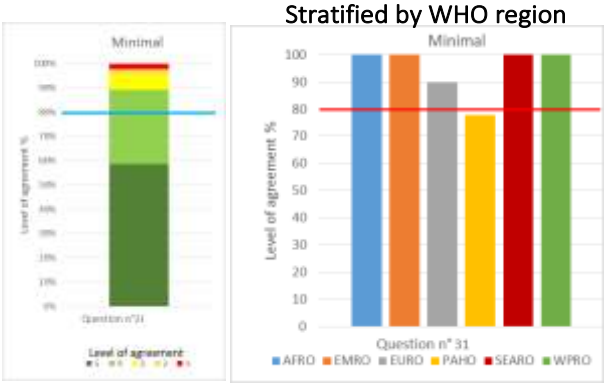
- Batching defeats the purpose of the assay

31) Walk-away operation

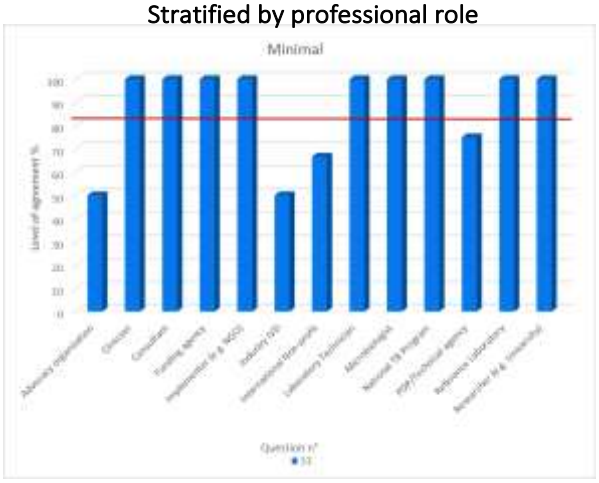
Minimal: No more than 1 step of operator intervention should be needed once the sample has been placed into or on the system

Level of agreement 89%

Stratified by WHO region



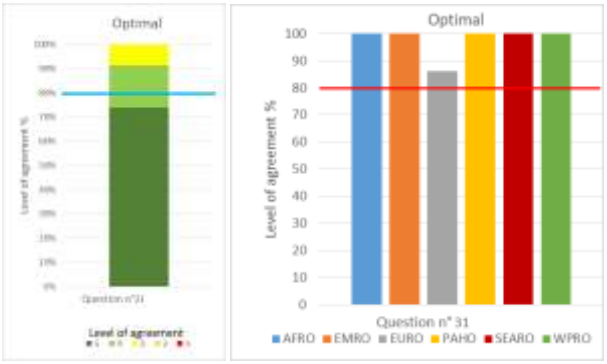
Stratified by professional role



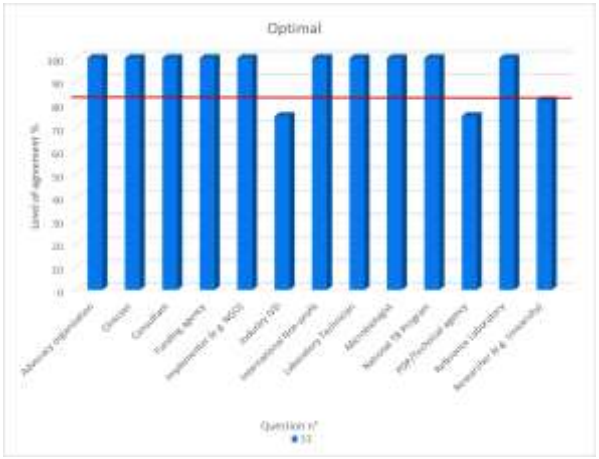
Optimal: These features are required; there should not be a need for operator intervention once the sample has been placed into or on the instrument

Level of agreement 91%

Stratified by WHO region



Stratified by professional role



Comments:

- Once sample is loaded, there should not require any intervention
- Both should be walk away.
- Needs to be better defined. What happens when the assay is done running
- Minimal: no more than 3 steps

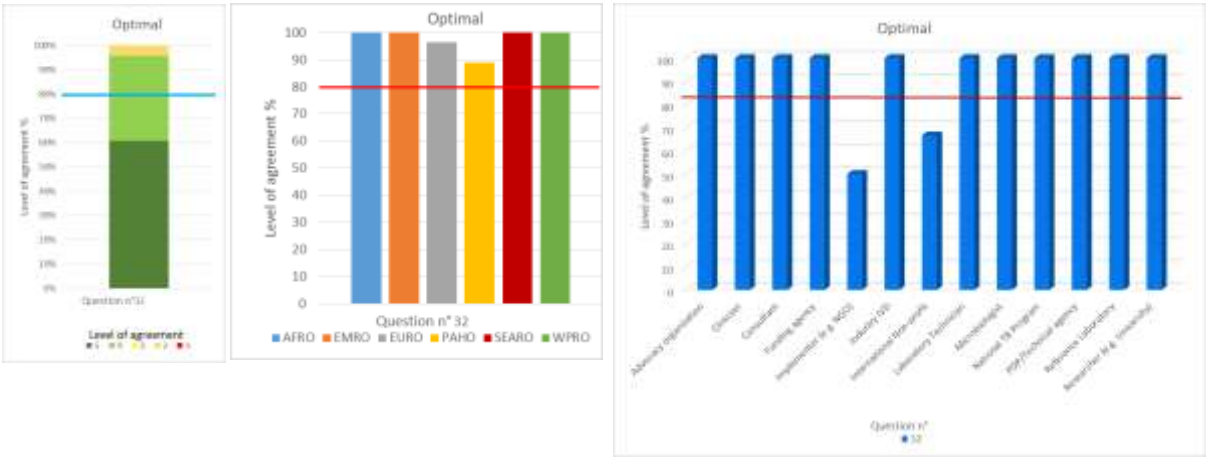
32) Biosafety

A biosafety cabinet is not commonly available at the level of a microscopy centre; low-risk TB laboratories follow the minimum biosafety requirements as described in WHO's Tuberculosis Laboratory Biosafety Manual (World Health Organization, Tuberculosis laboratory biosafety manual. 2012: Geneva, Switzerland).

Optimal/Minimal: Similar to those for smear microscopy (low-risk TB laboratories)

Level of agreement 96%

Stratified by WHO region | **Stratified by professional role**



Comments:

- should be usable in the community
- I think it would be better to imliment first at the hospital or district level rather than microscopy centers first
- OPT- BSC not required. However, for some other specimen types EPTB this may need consideration.
- But we also know that what is currently done at the microscopy level places undo risk on the users

33) Waste disposal – solid

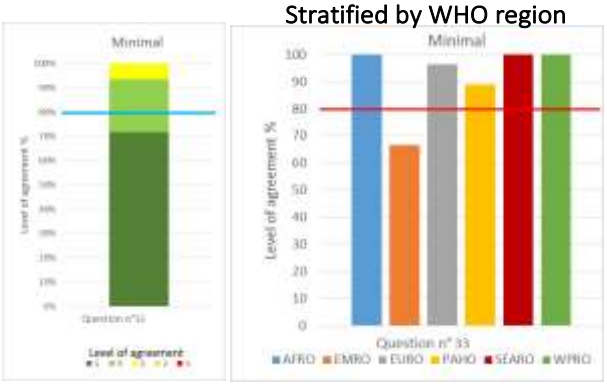
Increasing the amount of waste generated compared to a smear microscopy laboratory should be avoided.

Green friendly, sustainable packaging minimizing the environmental impact of packaging should be considered for the product's entire lifecycle.

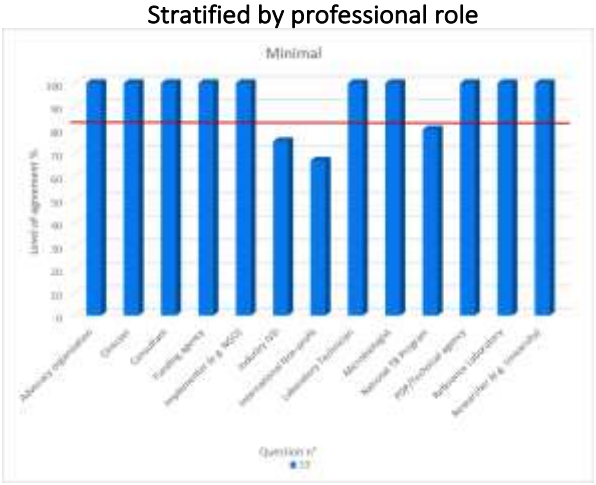
Minimal: Should require no more than current WHO-endorsed assays at peripheral level

Level of agreement 93%

Stratified by WHO region



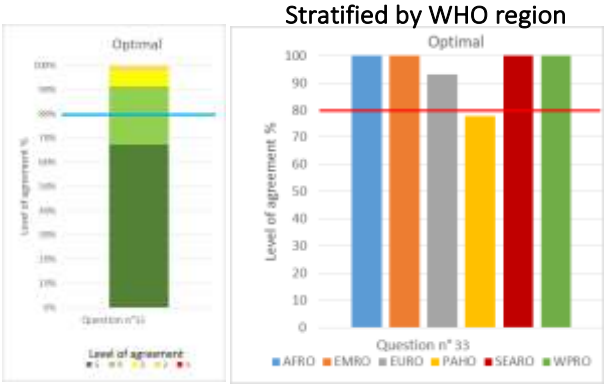
Stratified by professional role



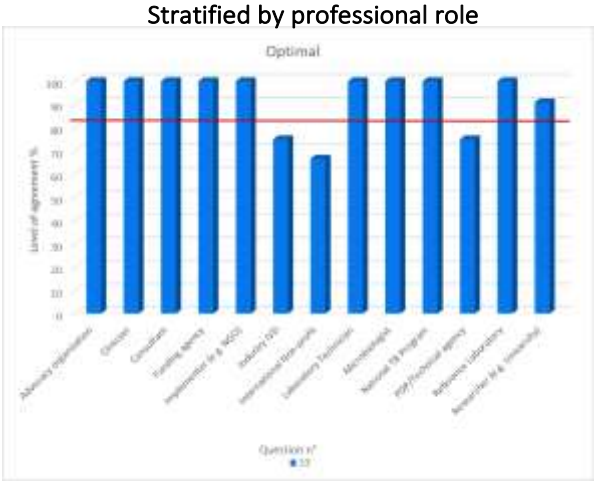
Optimal: Should require no more than smear microscopy

Level of agreement 91%

Stratified by WHO region



Stratified by professional role



Comments:

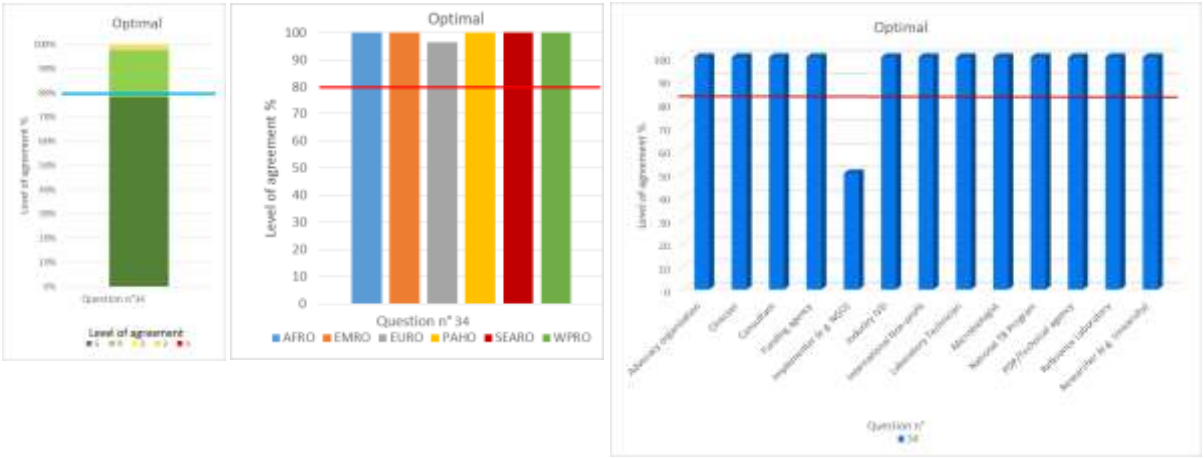
- not recommended at peripheral level at first roll out
- Need more info
- This is NOT the case for Xpert, so for OPT, we need to have companies include possibly a recycling strategy for cartridges, chips, or compartmentalized test which are made of plastics.
- We should be pushing to minimize waste

34) Waste disposal – infectious

Optimal/Minimal: Similar to those for smear microscopy (low-risk TB laboratories)

Level of agreement 98%

Stratified by WHO region **Stratified by professional role**



Comments:

- should be a killed sample - disposable through usual community routes
- OPT: I think the technology should include a decon step at the end, either via UV or Heat to eliminate risk.

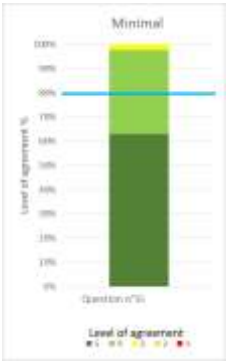
35) Instrument

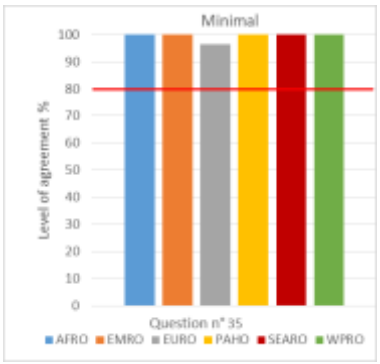
Ideally, a single device is preferred but modular solutions would be acceptable.

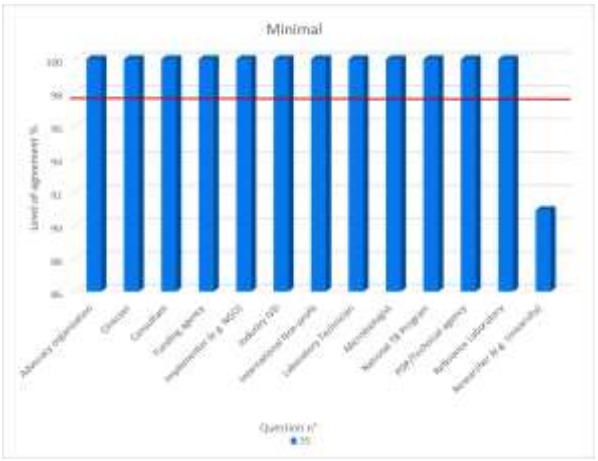
Minimal: Build on a modular concept allowing to tailor needs and upgrade additional functionalities at any time

Level of agreement 98%

Stratified by WHO region | Stratified by professional role



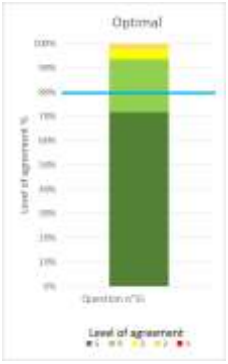


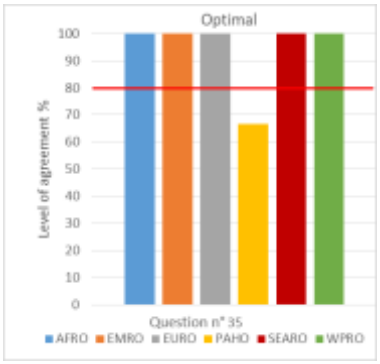


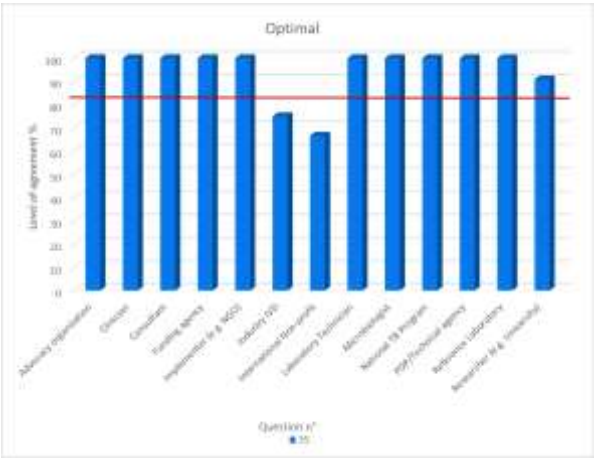
Optimal: Ideally, would be a single integrated system that is modular to allow throughput to be increased if needed

Level of agreement 93%

Stratified by WHO region | Stratified by professional role







Comments:

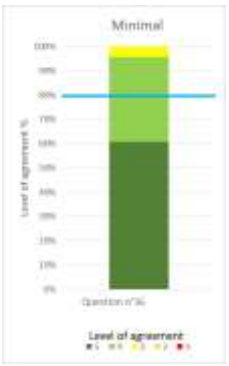
- difficult and more complicated for a totally integrated system. Suggest separate for now
- Varying the # of samples per run is possible in our system
- However, the statements for Optimal and Minimal are similar what is the basic concept here?

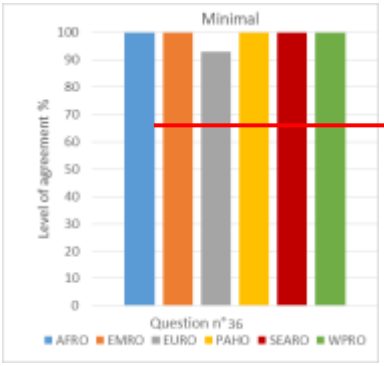
36) Power requirements

Minimal: Capable of running on standard electricity plus an ad hoc certified uninterrupted power supply unit delivered with the system to enable a cycle to be completed in case of a power outage; a circuit protector must be integrated within the system; the uninterrupted power supply should be preferably integrated within the system

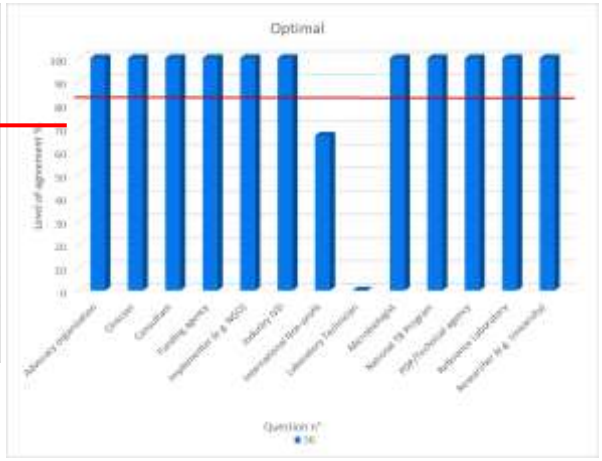
Level of agreement 96%

Stratified by WHO region





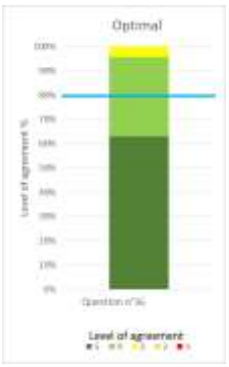
Stratified by professional role

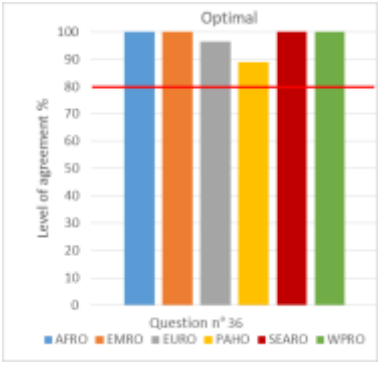


Optimal: Capable of running on standard electricity plus an ad hoc certified uninterrupted power supply unit delivered with the system to enable a cycle to be completed in case of a power outage; the uninterrupted power supply and circuit protector must be integrated within the system. The system should be also compatible for switching it in a battery operated device with the ability to run for 1 day on the battery, and with recharging capability (which could be solar powered)

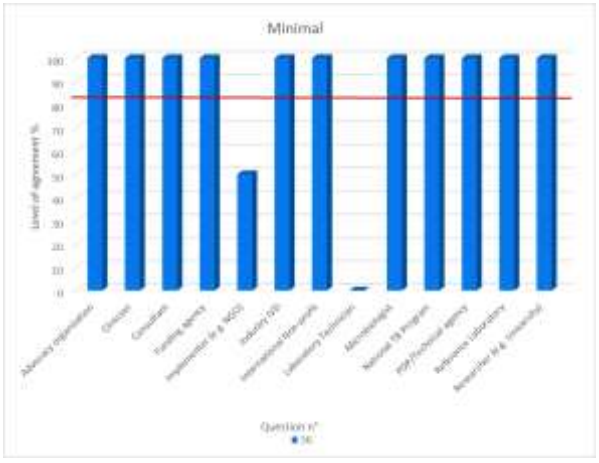
Level of agreement 96%

Stratified by WHO region





Stratified by professional role



Comments:

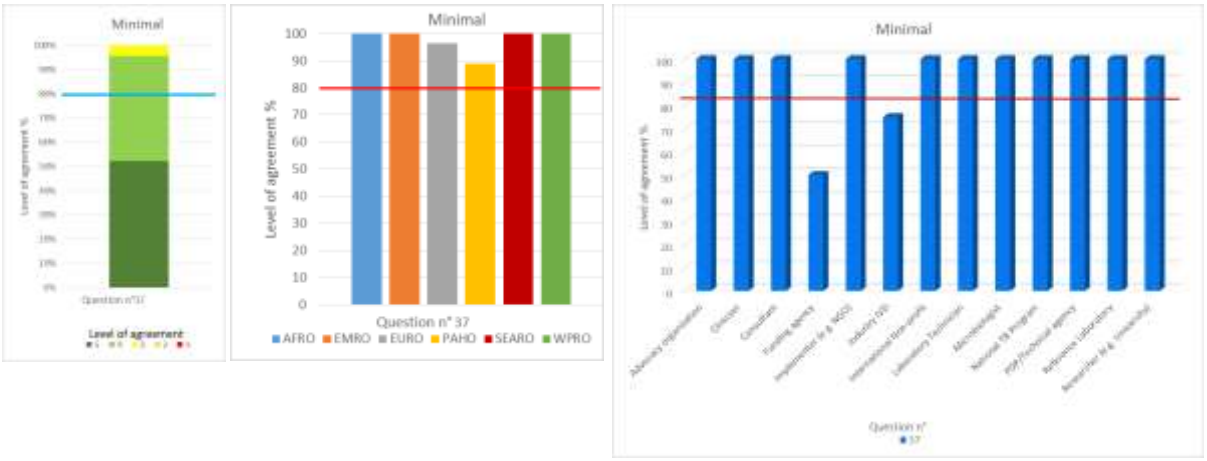
- In both cases, the size of the battery pack will depend on local conditions. Integrating this in the system will either lead to extra costs (maybe there is no need for a huge battery pack) or insuffcient power (where there are long electricity cuts). What could be integrated is a current-correcting device (in many settings voltage and frequency are irregular, even if there are no electricity cuts)
- portability is key
- Interoperable with conventional power or alternative energy options. Battery should last 2-3 days.
- Current regulations around batteries will affect this are force less robust chemistries to be utilized

37) Maintenance and calibration

Minimal: Preventative maintenance should not be needed more than once a year. Users should be able to monitor the machine status independently from manufacturers' intervention by the use of appropriate internal/external controls; results for such controls can be shared with manufacturers or appropriate control bodies to schedule appropriate on demand intervention (maintenance/calibration). An alert should be included to indicate when maintenance is needed according to manufacturer's indications. Software updates should be provided remotely

Level of agreement 96%

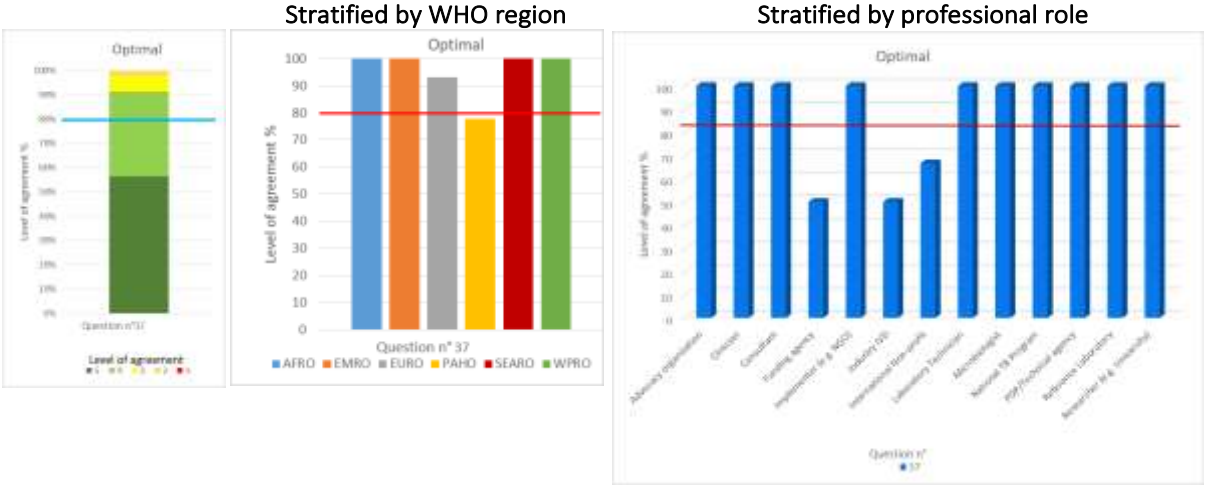
Stratified by WHO region | **Stratified by professional role**



Optimal: Preventative maintenance should not be needed more than once every two years. Users should be able to monitor the machine status independently from manufacturers' intervention by the use of appropriate internal/external controls; results for such controls can be shared with manufacturers or appropriate control bodies to schedule appropriate on demand intervention (maintenance/calibration). An alert should be included to indicate when maintenance is needed according to manufacturer's indications. Software updates should be provided remotely

Level of agreement 91%

Stratified by WHO region | **Stratified by professional role**



Comments:

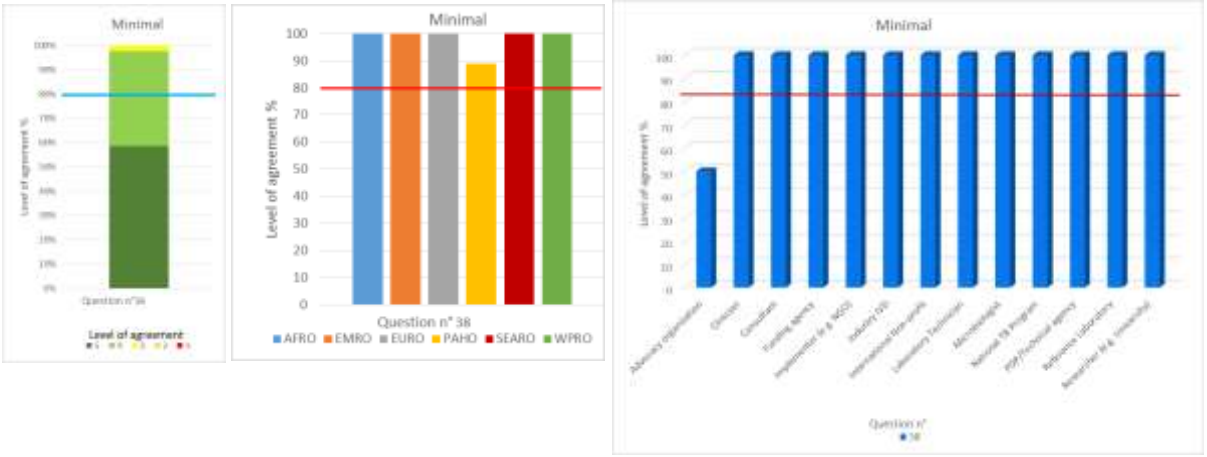
- Preventative maintenance should NOT be required (preferred), but will accept at >1 year intervals.
- Annual check necessary
- Ignoring the aspect of tracking and trending performance so as to be proactive in maintenance
- Remote access to machine status should be available for both minimal and optimal to ensure fleet management from central location

38) Data analysis

Minimal: Exported data should be capable of being analysed on a separate or networked PC

Level of agreement 98%

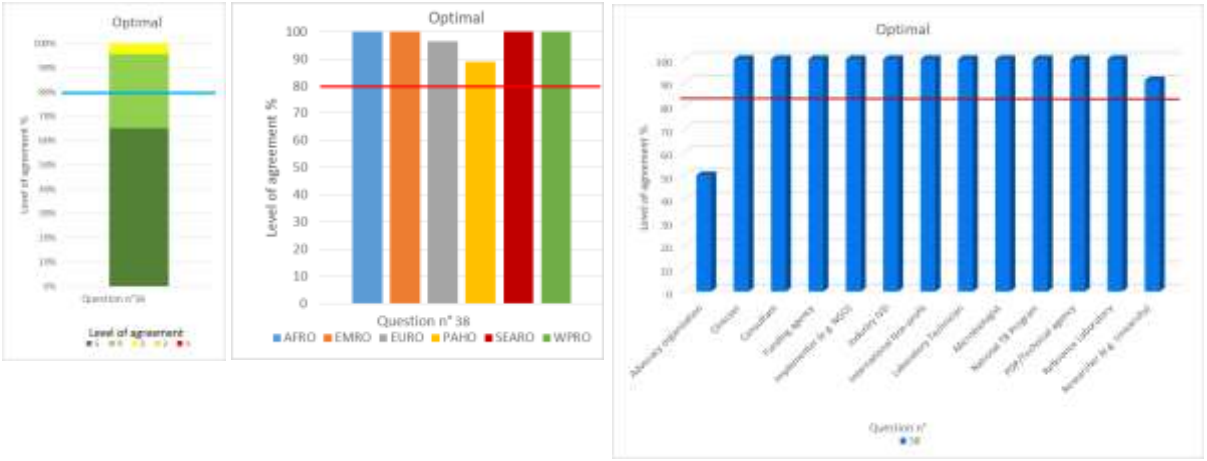
Stratified by WHO region **Stratified by professional role**



Optimal: Data analysis should be integrated into the device; a PC should not be required; exported data should be capable of being analysed on a separate or networked PC

Level of agreement 96%

Stratified by WHO region **Stratified by professional role**



Comments:

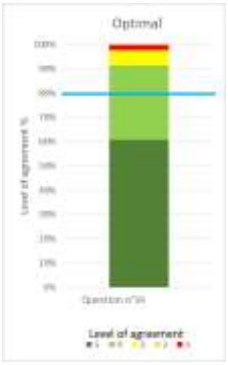
- Again, not informed enough to offer meaningful opinion
- MIN - USB port required
- BUT this gets to a route issue of data interpretation for resistance
- Data analysis should be integrated into the device but with possibility to customize the analysis depending on the specific indicators used at country level.

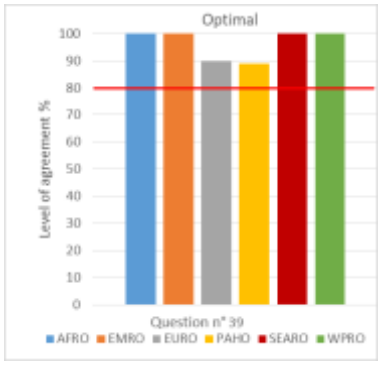
39) Regulatory requirements

Optimal/Minimal: Manufacturing of the assay and system should comply with ISO15189 or higher standards or regulations, and comply with ISO IEC 62304 Medical Device Data Systems; the manufacturing facility should be certified and authorized for use by a regulatory authority that is a member of the International Medical Device Regulators Forum, formerly known as Global Harmonization Task Force; the assay must be registered for in vitro diagnostic use

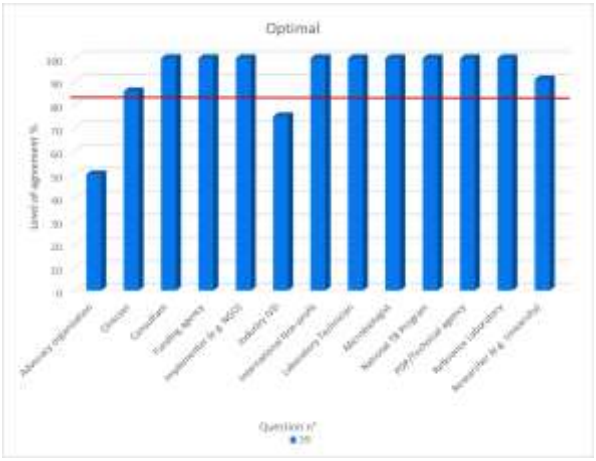
Level of agreement 91%

Stratified by WHO region





Stratified by professional role



Comments:

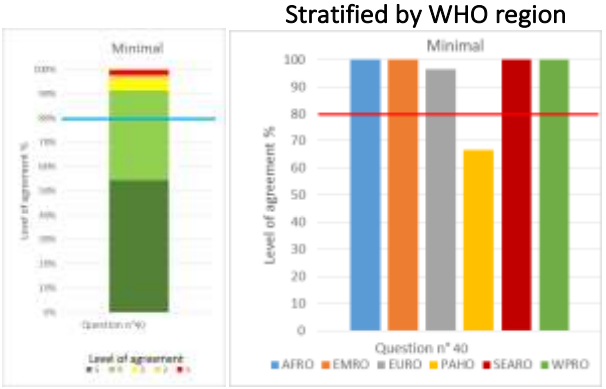
- Again, not informed enough to offer meaningful opinion
- I am not expert in this question
- Not clear how IVD use registration can be interpreted, as CE-IVDR may not be relevant nor feasible given the target markets of high-burden countries. Would recommend removing, or clarifying it with language of one or more SRA (stringent regulatory authority approval). Assume WHO positive recommendation is handled separately in the later sections.
- ISO13485 is what you mean as stated the manufacturer needs to be compliant to a clinical laboratory requirement

40) Data export (connectivity and interoperability)

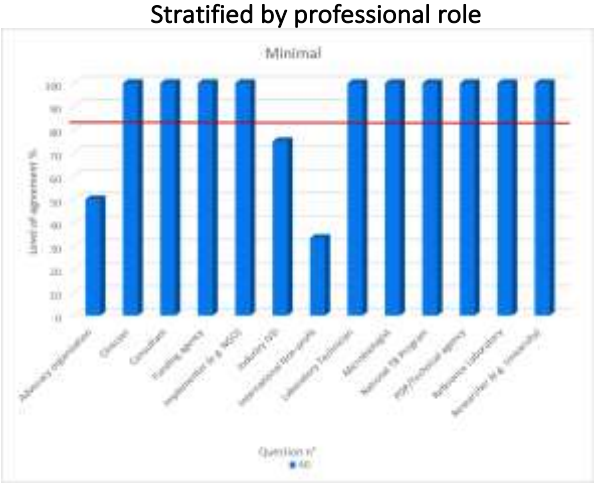
Minimal: Integrated ability for all data to be exported from the device in a user-friendly format (including data on use of the device, error rates or rates of invalid tests, and non-personalized results) over a USB port. Bluetooth connectivity should also be available. It should also be possible to import data (e.g. for updating interpretation rules or databases)

Level of agreement 91%

Stratified by WHO region



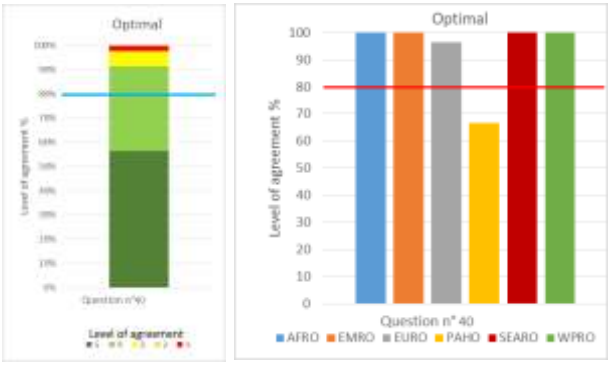
Stratified by professional role



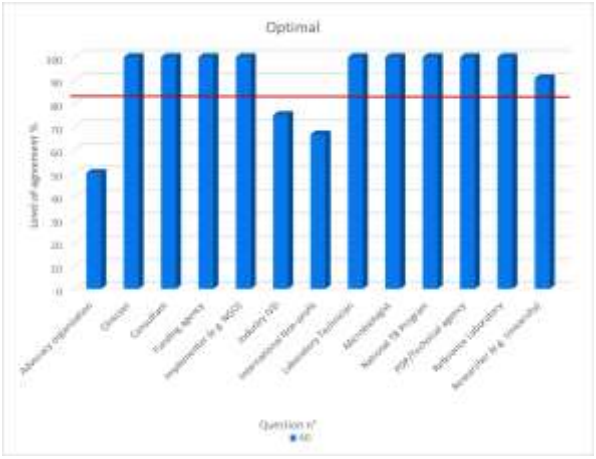
Optimal: All data should be able to be exported (including data on use of the device, error rates and rates of invalid tests, and personalized, protected results) over a USB port and network; network connectivity should be available through an Ethernet, Wi-Fi, and GSM/UMTS mobile broadband modem, or a combination of these; results should be encoded using a documented standard (such as HL7) and be formatted as JSON text; JSON data should be transmitted through HTTP(S) to a local or remote server as results are generated; results should be stored locally and queued during network interruptions to be sent as a batch when connectivity is restored. Bluetooth connectivity should also be available. It should also be possible to import data (e.g. for updating interpretation rules databases)

Level of agreement 91%

Stratified by WHO region



Stratified by professional role



Comments:

- Bluetooth should not be necessary- USB should be sufficient.
- needs better connectivity to address need for interpretation
- Again, not informed enough to offer meaningful opinion
- Define user friendly format, the format should be complaint with current practices and should be defined to the end user. Does NOT address the critical aspects of control of the data which is critical. When compliance to data standards is included then I could agree with the statements but currently this is ignored bwhich doesn't put the security of the data first

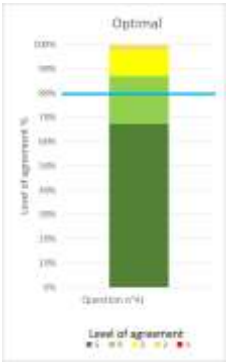
41) Electronics and software

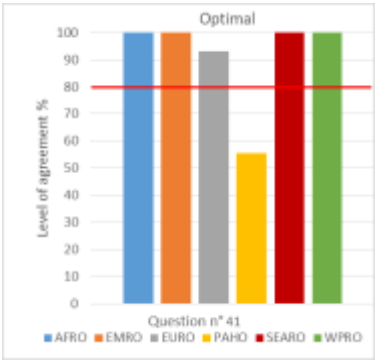
If an external device (separate PC, tablet, mobile…) is needed, it will likely limit the ability to update software, since staff with the skills needed to operate a PC are not present in all microscopy centres.

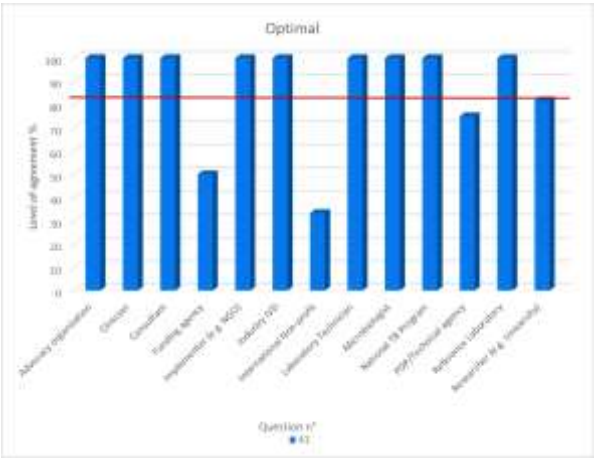
Optimal/Minimal: Should be integrated into the instrument

Level of agreement 87%

Stratified by WHO region **Stratified by professional role**







Comments:

- Cell phones are ubiquitous and I can't see a technician being unfamiliar with phone operation or updates.
- BUT this is not completely in alignment with previous attributes for analysis
- Should be integrated and requires minimal training.
- every one knows to use a mobile

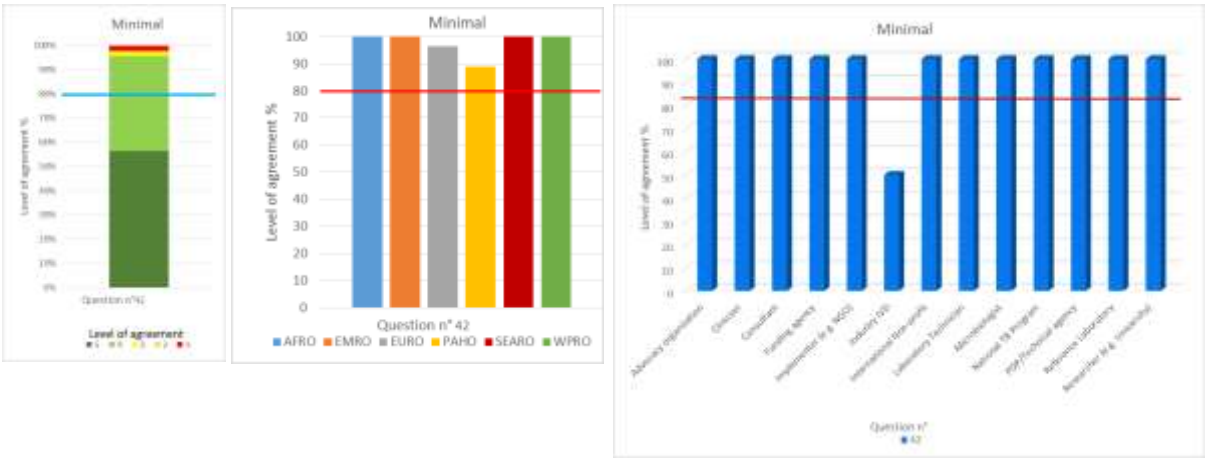
42) Operating temperature and humidity level

High environmental temperatures and high humidity are often present in countries where TB is endemic. Dust also is a problem in these settings, and the need to adequately protect optics should be considered.

Minimal: Between +5 °C and +40 °C with 70% humidity

Level of agreement 96%

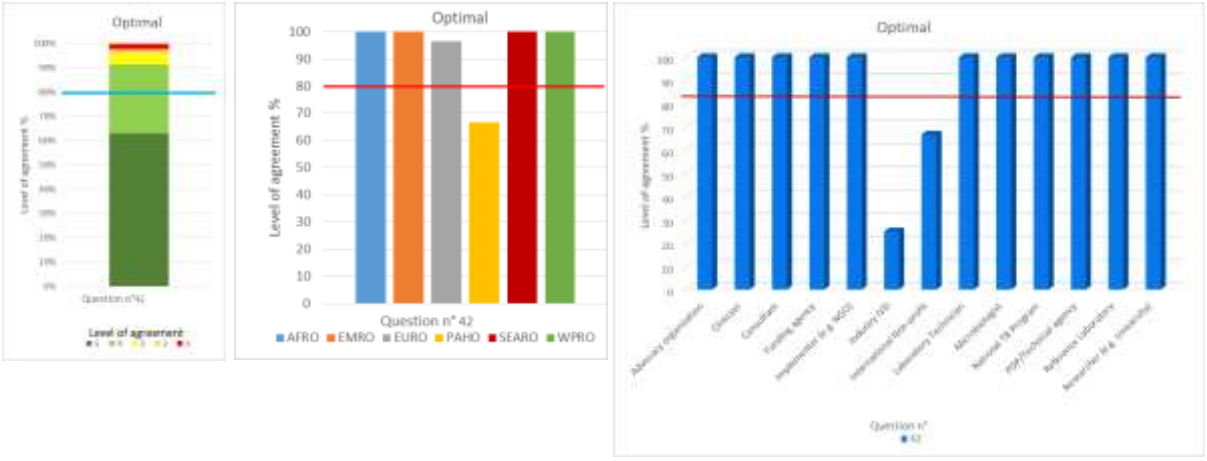
Stratified by WHO region **Stratified by professional role**



Optimal: Between +5 °C to +50 °C with 90% humidity

Level of agreement 91%

Stratified by WHO region **Stratified by professional role**



Comments:

- I don't think that this extreme will be necessary and a more modest environmental control needs should be required
- 50 degrees might be challenging
- is 40-50C realistic????

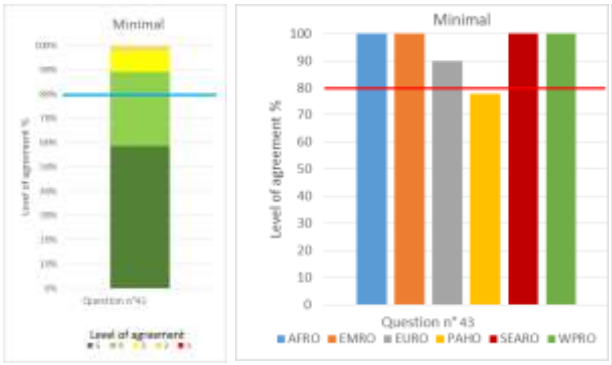
- dust is mentioned but not included in minimal/optimal wording
- Temperature and humidity are two separate variables which need to be addressed separately. As stated this combines the two variables
- Minimal: +5 to +35C; Optimal: +5 to +40C, 85% humidity

43) Reagent kit – transport

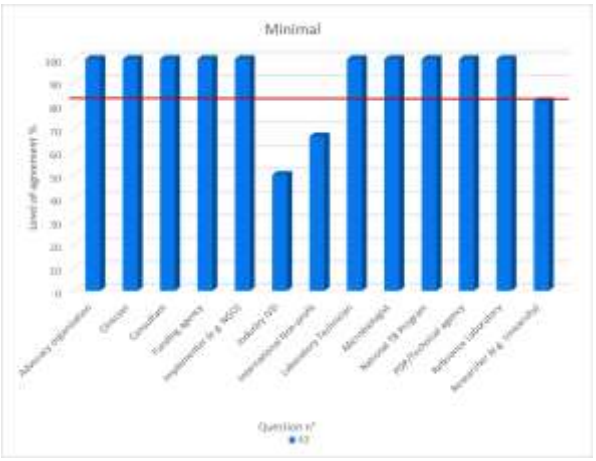
Minimal: No cold chain required; should be able to tolerate stress during transport for a minimum of 72 hours at -15 °C to +40 °C

Level of agreement 89%

Stratified by WHO region



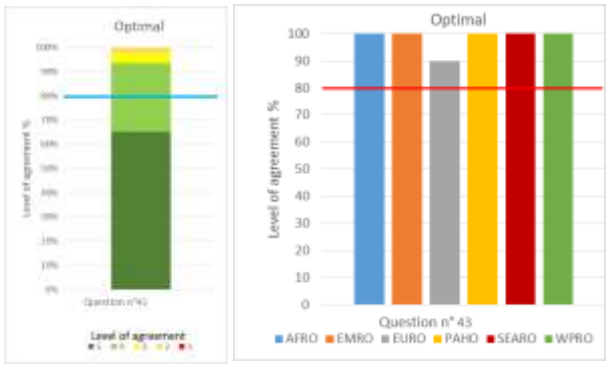
Stratified by professional role



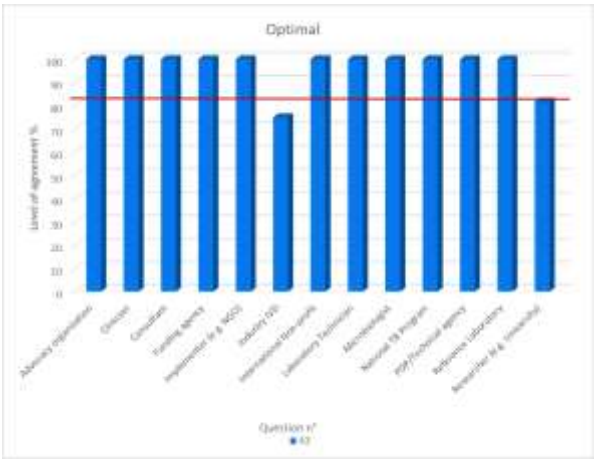
Optimal: No cold chain should be required; should be able to tolerate stress during transport for a minimum of 72 hours at -15 °C to +50 °C

Level of agreement 93%

Stratified by WHO region



Stratified by professional role



Comments:

- Would consider high temp for 72h
- Addressing temp extremes should be a priority during transportation of reagents
- Are these temps realistic?
- Need to address humidity also
- Minimal: cold chain

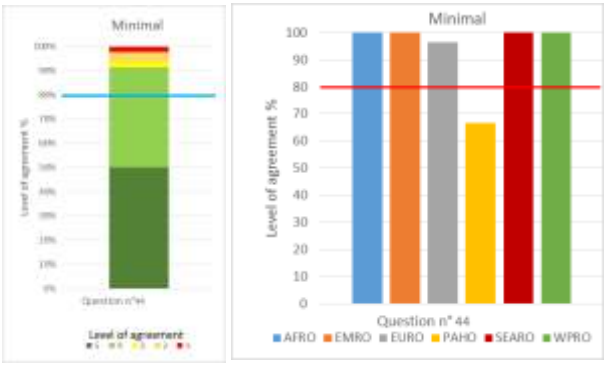
- Would extend transport time, as in some settings it may be needed more than 3 days to deliver the reagent kits
- Optimal: stress tolerance to be increase as "for a minumum of 1 week"

44) Reagent kit – storage and stability

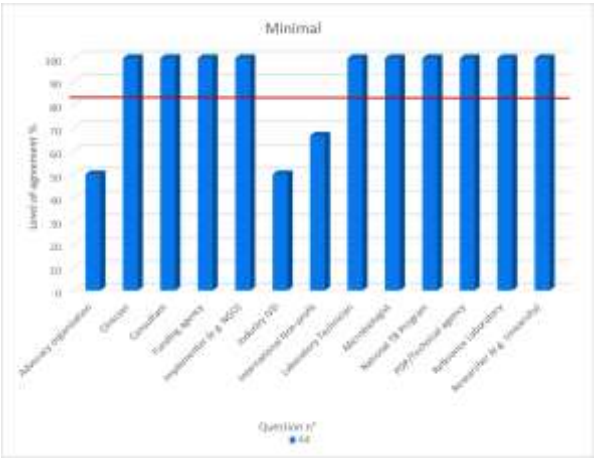
Minimal: 12 months at +5 °C to +35 °C with 70% humidity; should be able to tolerate stress during transport for a minimum of 72 hours at +40 °C; no cold chain should be required

Level of agreement 91%

Stratified by WHO region



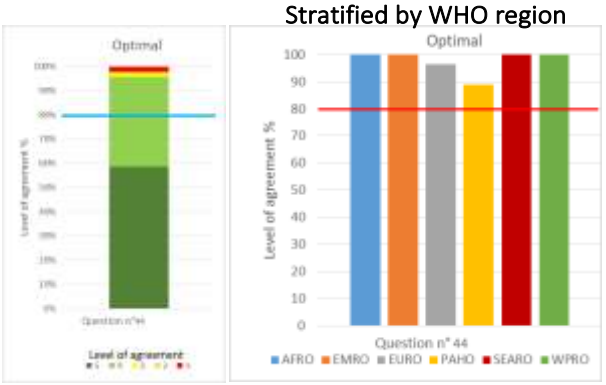
Stratified by professional role



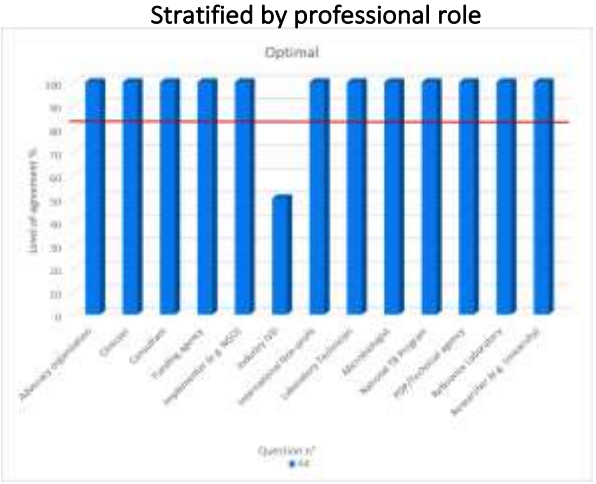
Optimal: 2 years at +5 °C to +40 °C with 90% humidity; should be able to tolerate stress during transport for a minimum of 72 hours at +50 °C; no cold chain should be required

Level of agreement 96%

Stratified by WHO region



Stratified by professional role



Comments:

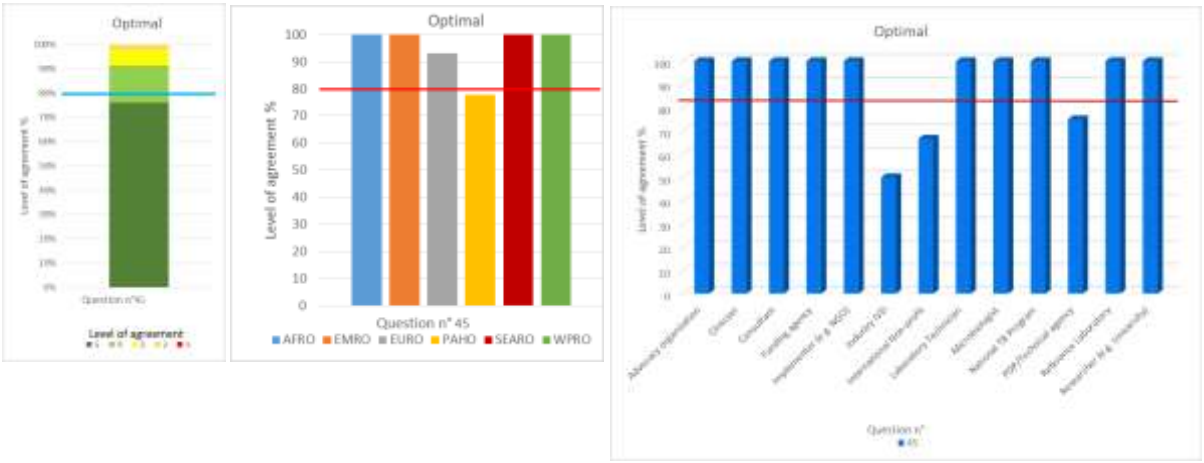
- 12 months is too short due to the length of time between manufacture and shipping.
- Unknown
- If possible would increase shelf life to 18 months minimal and 3 years optimal given supply chain management issues among programs
- Minimal: 18 months OPT: 2 years or longer
- Not in alignment with storage and humidity is a separate variable
- Minimal: frozen or 2-8C; 12 months

45) Additional supplies (not included in kit)

Optimal/Minimal: None

Level of agreement 91%

Stratified by WHO region **Stratified by professional role**



Comments:

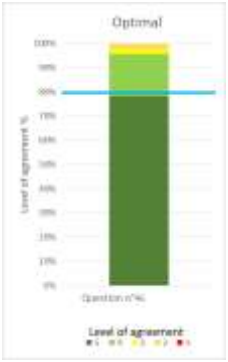
- Clarify that none outside of things necessary for smear.
- When minimal level is not indicated, assume that it means 1 or more additional supplies may be needed. i.e. this should not mean minimal and optimal are the same level for this criterion. Recommend checking language to ensure this is consistent throughout TPP.
- So does the sputum cup need to be part of the kit?
- Sample transfer devices, timer and other consumables, in alignment with needs in serology Lateral Flow Rapid Test procedures

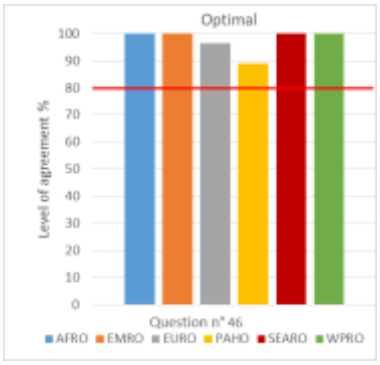
46) Internal quality control

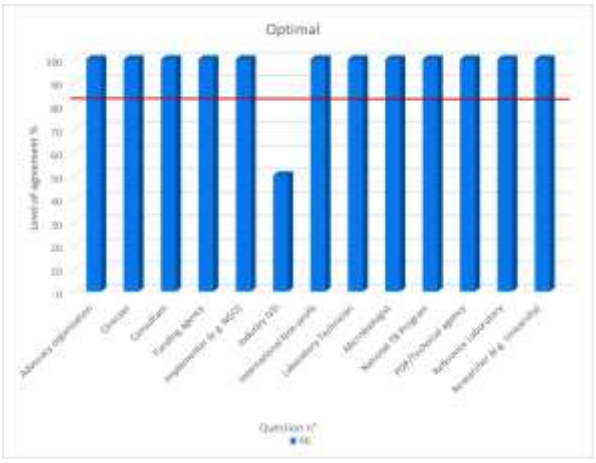
Optimal/Minimal: Full controls for sample processing, amplification and detection of TB and any target for DR should be included. Internal controls for analysis and reporting (e.g. software version) should be included. A monitor (remote) system for checking results on the controls should be also considered

Level of agreement 96%

Stratified by WHO region **Stratified by professional role**







Comments:

- Should be for minimal as well
- ELIMINATING the need for EQA
- How do you control for ALL DR mutations which is what is stated?
- External run controls and internal spike in control (optional) for entire workflow and representative targets for DR and drug susceptibility, as well as TB vs relevant negative controls

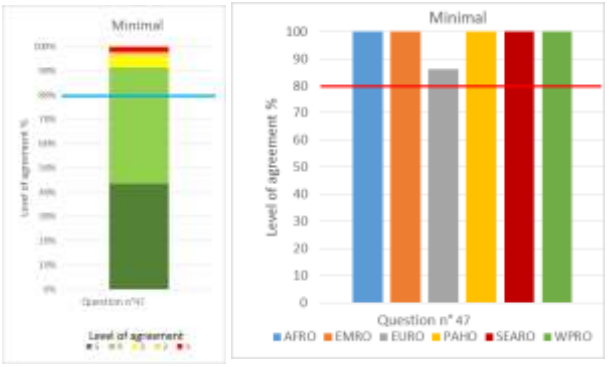
47) Training and education

Trainings should be developed according to Continuing Education and Training (CET) models and Individualized Training Programs (ITP) to ensure that only properly trained, accredited people can perform the assay. Online and remote support systems should be available for retraining, monitoring/evaluating and updating ("refresher") training.

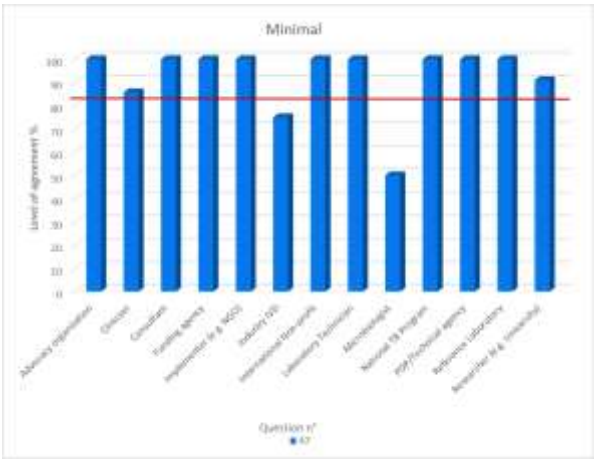
Minimal: 3 days (or 24 work hours) for staff at the level of a laboratory technician

Level of agreement 91%

Stratified by WHO region



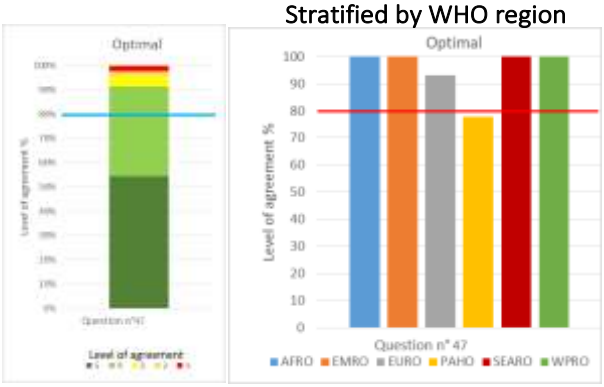
Stratified by professional role



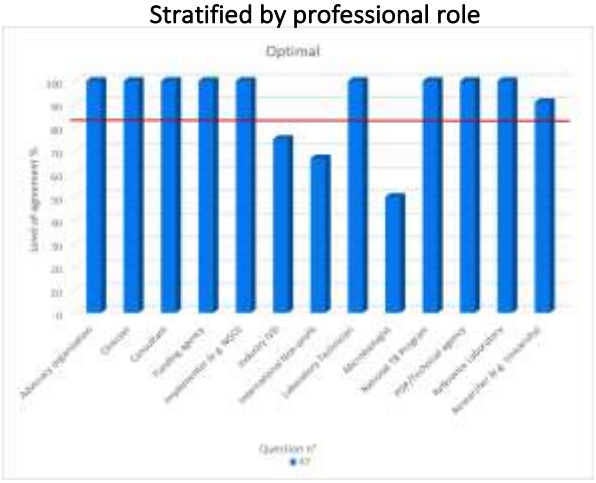
Optimal: 6 work hours for staff at the level of a microscopy technician

Level of agreement 91%

Stratified by WHO region



Stratified by professional role



Comments:

- I think its a mistake to roll this out to peripheral microscopy labs first.
- Minimal: one day
- We have some experience with teach over long distances

- Min 2 days (12 hrs) lab staff, Other staff maybe longer 3 days. OPT: <4 hrs LAB staff/ HCW at lower levels for testing 1 day (6-8 hrs)
- This is ridiculous for a Minimal
- it should be the same due to potential easy-to-use device.
- timing of training sessions for both optimal and minimal should be increased

ANNEX 2 – Target Product Profile draft document submitted together with the survey

Target Product Profile:
Next-Generation Drug-Susceptibility Testing at Peripheral Centres (ver. 18Mar2019)

This is a working draft intended to promote discussion between the different groups of stakeholders that are involved in ongoing projects that may inform this TPP. This document was drafted by NDWG NGS and next-generation DST Task Force (Emmanuel André, Martina Casenghi, Paolo Miotto, Camilla Rodrigues, Timothy Rodwell, Philip Supply, Timothy Walker).
This document aims at providing TPP guidance for the development of rapid drug-susceptibility tests that can be used at the microscopy-centre level of the health-care system (the rapid DST test). This TPP is updating the previous existing document [1], as i) the key assumptions for the development timeframe (<5 years) are expired, ii) new drugs and regimens are under development (https://www.newtbdrugs.org/, latest accessed 4 February 2019), iii) new treatment policies are available [2], and iv) new technologies are entering the TB diagnostic landscape [3].
The diversity of resources and needs in different countries, the geographical variation in the epidemiology of TB and related co-morbidities, and drug-resistant TB, together with the specialized nature of the different technical procedures make mandatory the adoption of proper implementation strategies specific for any given new assay developed. Providing guidance for implementation strategies is beyond the scope of this document, however the characteristics defined in the TPP should be regarded in the perspective of an implementation framework [4].

DEFINITIONS and CLARIFICATIONS:

- Characteristic – refers to a specific requirement or specification that is measurable.
- Minimal – for a specific characteristic, refers to the lowest acceptable output for that characteristic. For clarification, solutions must meet the 'Minimal' characteristic in order to be acceptable (CAVEAT: a test may still be acceptable if shortcomings pertain to the soft targets and if specific hard targets (marked with an asterisk) are only missed marginally).
- Optimal – for a specific characteristic, provides the best output for that characteristic that is believed to be realistically achievable. Meeting the 'Optimal' characteristics provides the greatest differentiation and the greatest impact for the end users, clinicians and patients. Developers would ideally design and develop their solutions to meet the 'Optimal' requirements for all characteristics.

NOTE: The optimal and minimal requirements/characteristics define a range.

Abbreviations: AMK – amikacin; BDQ – bedaquiline; CAP – capreomycin; CLO – clofazimine; DLM – delamanid; DST – drug susceptibility testing; FQ – fluoroquinolones; INH – isoniazid; KM – kanamycin; LEV – levofloxacin; LZD – linezolid; MDR – multidrug resistant TB (RIF, INH); MOX –moxifloxacin; PZA – pyrazinamide; RIF – rifampin; WHO – World Health Organization; XDR – extensively drug resistant TB (INH, RIF, FQ, AMK, KAN, CAP).

Characteristic	Optimal (ideal) requirements	Minimal requirements	Explanations	References
Scope				
Key assumptions	The development **time is < 5 years**; this approach would use **1 solution** for TB detection and DST; this TPP has taken the developers' perspective by assuming that new regimens will be implemented and available in parallel with current standard-of-care regimens, at least initially			
Rationale	To provide support for **effective anti-TB therapy** in the context of the roll-out of new regimens; to provide the characteristics and qualities of a test that would have a sufficiently rapid turnaround time (that is, results can be provided during the same visit) for TB detection and would provide data about DST that can be used to inform treatment decisions			
Goal	Diagnosis of TB disease and detection of drug resistance to inform decision making about the optimal (individualized) regimen	Diagnosis of TB disease and detection of drug resistance to provide rapid triage of patients and identification of adequate treatment regimen (1st line treatment *vs* 2nd line treatment)	The market for a test that includes DST and detection is all patients tests for TB, which is approximately 10 times the number of detected cases, or about 60 million patients. If DST were performed in a second step, the market would be all patients in whom TB had been detected (or about 9 million). The market for a test to detect PZA resistance is different because the current achievable performance characteristics of a molecular test for PZA resistance is a maximum of 95% for both sensitivity and specificity; therefore, a test for PZA resistance could be used as a follow-on test only if RIF resistance has been confirmed (a higher prevalence of resistance leads to a higher PPV for the detection of resistance to a particular anti-TB agent). This means that the market for testing for PZA resistance is only as large as the number of patients confirmed to have MDR-TB, which is about 450 000, although the number is likely to increase as testing for MDR-TB increases.	
Priority of anti-TB agents for testing[a]	In order of decreasing importance: 4. RIF + INH + FQ 5. BDQ + LZD 6. CLO + DLM + pretomanid + AMK + PZA	In order of decreasing importance: 4. RIF + INH 5. FQ + KM* 6. AMK + CAP*	Drug prioritization considers universal DST access (END-TB Strategy) and that effective administration of anti-TB drugs can be achieved only by knowing susceptibility testing results. This is a general principle becoming crucial especially for MDR/XDR treatment. The proposed prioritization keep into account that FQs are relevant for both MDR and INH-R/RIF-S cases. In addition, new tests should be useful for triaging patients for short MDR regimen. The minimal requirements keep into account the transition time required for the complete drop-off from the regimens of the two injectable drugs no longer recommended (*), estimated in the range	[2, 5-12]

	(FQ always include LEV, MOX; any additional drug listed in the WHO treatment guidelines)	(FQ always includes LEV, MOX)	of 1-2 years depending upon the setting considered (especially where access to late generation drugs such as BDQ and LZD is strictly regulated and likely less available at peripheral level). The optimal requirements keep into account the full compliancy with the new WHO guidelines. To be also noted that treatment guidelines are moving forward to an "all oral" regimen; in this context, the relevance of AMK is expected to further decrease in the next 3-5 years. The differentiation of resistance among FQs is more a function of interpreting mutations (that is, evaluating the hierarchical structure of mutations) rather than detecting different mutations. **Sequence of detection and drug-resistance testing:** The proportion of patients with a diagnosis of TB who experienced pre-treatment loss to follow-up was found in the range of 4 to 38%. This scenario might vary substantially among countries. Initially testing for TB and DST might come at the expense of the sensitivity for TB detection, depending on the platform used and cost. However, a delay in DST might result in patients receiving inappropriate treatment until they return, assuming that the DST result will not be known in time to inform initial decisions about treatment. The acceptability of a longer wait time might vary among countries, and informing the patient of results on the same day if the result is not available during the first visit, might be associated with substantial costs.
Assay design	The assay should be designed in such a manner that the addition of or removal of analytes does not require extensive analytical and clinical re-verification and revalidation of the assay		The assay should be designed in such a manner that it is capable of being updated as needed, with minimal redevelopment required. For whole genome sequencing-based assays, this include the possibility to adjust sequence interpretation for new drugs.
Target population	Target groups are all patients suspected of having TB, with a special focus on those at high risk of morbidity and mortality from drug-resistant TB, such as people living with HIV and those at high risk of having MDR-TB (for example, household contacts of patients diagnosed with MDR-TB, and persons with a history of TB, especially those for whom first-line therapy has		The optimal target population should be all adults with signs and symptoms of, although the resource implications need to be considered. Children aged <11 years have limited ability to produce sputum for testing. Therefore, initial validation studies should focus on adults. **WHO's categories:** High-incidence countries are those with > 100 cases per 100 000 population; medium-incidence countries are

	failed) in countries with a medium incidence to a high incidence of TB as defined by WHO		those with 20–40 cases per 100 000 population; and low-incidence countries are those with < 10 cases per 100 000 population [5].	
Target user of test[a]	Health-care workers with minimal training necessary	Health-care workers with minimal/moderate training	Minimal training: users are health-care workers with limited or no competency in general laboratory practice (beginner users). Moderate training: users are health-care workers with minimal/moderate competency in general laboratory practice (competent/proficient users) The "Competency Guidelines for Public Health Laboratory Professionals" was used for providing a term of reference [13].	[5, 14]
Setting (level of the health-care system)	Peripheral and/or microscopy-centre level of the health-care system		Implementation at the microscopy-centre level should be feasible using the specifications as outlined. This would embed the test in an infrastructure that is based around smear microscopy. However, the test could be implemented at higher levels of care as well. Testing for resistance to the anti-TB agents included in second-line therapy could be incorporated into separate reactions, but ideally it would be feasible to test the same specimen.	[14-17]
Pricing				
Price of individual test (cost of reagents and consumables only; after scale-up; ex-works; excluding shipping and subsidiary factors. Non-negotiated prices)[a]	Detection of RIF + INH: 15-20 (±15%) USD; Detection of RIF +INH + FQ + AMK: 40-50 (±15%) USD; Detection of RIF + INH + FQ + AMK + KM* + CAP*: 50-60 (±15%) USD (FQ always includes LEV, MOX)		Meeting participants emphasized the critical need for the price to be kept within an affordable range. A price that is higher than available technologies would be justified only if the new test brings substantial added value in terms of vastly improved performance, greater suitability for decentralization, and the number of anti-TB agents for which resistance can be detected. Estimated ranges are based on current costs (expressed as min/max range) of the gold standard phenotypic DST that such new assays are expected to replace. We are currently unable to provide a price range for tests covering the "optimal" list of prioritized drugs because (i) the price might vary depending on the number of drugs considered, and (ii) there are no evidence data for foreseen what would be the cost of assays testing for new drugs such as DLM, LZD, and CLO. However, considering that phenotypic DST for first- and second-line drugs is estimated in the range of USD 50-100 (±30%), a new assay providing DST results for all the drugs listed among the "optimal" prioritization list should be within the same range, and any additional costs should be related to (i) the reduction of the turn-around-time, (ii) the increase of	[18-23]

			easiness of use and reduction of hands-on time, (iii) the reporting system, (iv) the reduction of need of extra reagents and/or equipment, and (v) the inclusion of additional drugs and/or additional testing (i.e. epidemiological details and/or comorbidities). Assay developers should consider cost-effectiveness in the context of willingness to-pay for disability-adjusted life year (DALY): screening tests should be cheaper enough for being cost-effective compared to current diagnostic algorithms, especially in high-burden, low-middle income settings. Finally, assay developers should consider that Global Drug Facility (GDF) negotiated prices for current assays for rapid DST endorsed by the WHO are <10 USD/test. (*) KM and CAP are no longer recommended, and are estimated to last in the clinical practice no longer than 1-2 years, depending upon the setting considered. Price estimates are considering that AMK, KM, and CAP share most relevant targets; thus, further implementation of an assay for detecting KM and CAP is not expected to be an expensive process.	
Capital costs for the instrument (non-negotiated prices)	<5,000 USD (including warranties, service contracts and technical support)	<15,000 USD (including warranties, service contracts and technical support)	The lower the capital costs of the instrument are, the lower the initial cost would be, and thus the barrier to implementation would also be lower, particularly since the volume of instruments that would be distributed to microscopy centres is sizeable. The cost of the instrument should also include warranties, service contracts and technical support. Cost-effectiveness should be then evaluated during implementation according to the number of drugs/targets that a given technology can cover, the assay multiplexing, and the multipurpose options offered.	
Performance				
Limit of detection – TB detection after first reaction	< 4.5 genome equivalents/reaction and < 10e2 CFU/assay using one sample	between 10e2 CFU/assay and 10e5 CFU/assay using one sample	Limit of detection testing should be performed as outlined in the United States Food and Drug Administration's guidance document.	[24-26]
Limit of detection – TB detection	≥4.5 genome equivalents/reaction	between 10e2 CFU/assay and 10e5	A slightly decreased analytical sensitivity for TB detection in the second reaction for resistance testing (in comparison with the first reaction) both for the optimal and minimal requirements will avoid	[27]

after second reaction for DST	and 131 CFU/mL of sputum	CFU/assay using one sample	resistance calls (for example, no *M. tuberculosis* but resistance present) but will come at the expense of a slightly lower sensitivity for DST.	
Diagnostic sensitivity for TB detection[a]	Sensitivity for detecting TB should be > 95% for a single test when compared with 2 liquid cultures; for smear-negative TB it should be > 68%;~~%;~~ for smear-positive TB it should be 99%	Sensitivity should be > 80% for a single test when compared with 2 liquid cultures; for smear-negative TB it should be > 60%; for smear-positive TB it should be 99%	The sensitivity specified is considering currently available technologies as baseline.	[28]
Diagnostic specificity for TB detection[a]	Specificity should be > 98% for a single test when compared with culture	Specificity should be > 98% for a single test when compared with culture		[24, 29, 30]
Diagnostic sensitivity for DST compared against genetic sequencing as the reference standard[a]	Sensitivity should be > 98% for detecting targeted SNPs for resistance when compared with genetic sequencing	Sensitivity should be > 98% for detecting targeted SNPs for resistance when compared with genetic sequencing	For Next Generation Sequencing technology-based assays: currently, there are no clear guidelines on what is a reference for a NGS-based diagnostic assay. In general, validating NGS results using different platforms plus different analysis pipelines is considered appropriate.	[24, 29, 30]
Diagnostic sensitivity for DST compared against phenotypic DST as a reference standard[a]	RIF, INH, FQ, BDQ, LZD, CLO, DLM, pretomanid, AMK, PZA, KM*, CAP*: >95% sensitivity for detection of phenotypic resistance	RIF: >95% sensitivity for detection of phenotypic resistance. INH, FQ: >90% sensitivity for detection of phenotypic resistance. BDQ, LZD, CLO, DLM, pretomanid, AMK, PZA, KM*, CAP*: ≥80% sensitivity for detection of phenotypic resistance	Modelling data suggest that for rapid DST to be more cost effective than culture, on a currently available platform it must attain an aggregated sensitivity of 88% for all clinically relevant mutations. A lower sensitivity could be tolerated for a test with high specificity, particularly if the prevalence, and thus the pretest probability, are high. The sensitivity achieved against a phenotypic internationally recognized reference standard (e.g. World Health Organization, Clinical Laboratory Improvement Amendment) will be only as good as the mutations that are targeted (that is, even if all known mutations conferring INH resistance are detected with 100% sensitivity when compared against a sequencing reference standard, 100% sensitivity cannot be achieved against a phenotypic reference	[24, 29, 30]

		standard because the knowledge of all molecular targets that confer resistance is not complete). Frequency of mutations at different drug resistant loci may vary depending upon several factors including (but not limited to) geographical region, local epidemiology and outbreaks; thus, implementation of molecular assays should carefully take into account the local epidemiology in order to achieve proper sensitivity. The requirements keep into account the transition time required for the complete drop-off from the regimens of the two injectable drugs no longer recommended (*), estimated in the range of 1-2 years depending upon the setting considered (especially where access to late generation drugs such as BDQ and LZD is strictly regulated and likely less available at peripheral level).	
Diagnostic specificity for DST compared against genetic sequencing as the reference standard[a]	Specificity should be ≥98% for any anti-TB agent for which the test is able to identify resistance when compared against genetic sequencing as the reference standard	If alternative regimens are available, effective, safe and not too cumbersome, then a lower PPV might be tolerated. Because the pretest probability is low when all-comers without any additional risk factors are tested in settings with a low prevalence of resistance, the specificity has to be very high: if the prevalence of resistance is about 3% according to surveillance data, then a specificity of 99% results in a PPV of only 74%. A very high specificity (for example, ≥99.7%) is necessary in order to reach a PPV of >90%; if the prevalence of resistance is ≥20% (for example, when resistance to RIF is used as an indicator or when testing is only done in high-risk patients), a specificity of >97% is sufficient to achieve a PPV of 90%. To be noted that mutations conferring resistance systematically missed by current phenotypic reference standard methods, as well as mutations not associated with phenotypic resistance exist [31].	[32, 33]
Diagnostic specificity for DST compared against phenotypic DST as a reference standard[a]	The specificity of targeted sequencing for the mutations included for any anti-TB agent for which the test is able to identify resistance should be ≥ 98% when compared against the phenotypic reference standard recommended for each anti-TB agent	The estimates of specificity for molecular tests in comparison with phenotypic testing as a reference-standard might be falsely low as the reference-standard has limited sensitivity. Therefore it is important to use the optimized phenotypic reference standard for a drug in comparison.	[24, 29, 30]

			To be noted that mutations conferring resistance systematically missed by current phenotypic reference standard methods, as well as mutations not associated with phenotypic resistance exist [31].
Limit of detection of minor variants	≤10% (that is 10 resistant bacteria out of 100)	≤20% (that is 20 resistant bacteria out of 100)	
Analytical specificity for TB detection	No cross-reactivity with other organisms. NTM identification should be also available.	No cross-reactivity with other organisms including nontuberculous mycobacteria	
Indeterminate results during detection[a]	< 5%	< 10%	Indeterminate results may be caused by a lower sensitivity for detecting TB during the second reaction.
Reproducibility	Interassay coefficients of variance should be ≤ 10.0% at the high and low extremes of the assay		This applies if the quantitative outcomes of a test are measurable (for example, for the limit of detection, and cycle threshold values).
Interfering substances	No interference should be caused by those substances known to occur in the human respiratory and pulmonary tracts, including blood that could potentially inhibit a PCR reaction, and substances used to treat or alleviate respiratory disease or symptoms		
Treatment monitoring capability	Yes (mandatory)	Yes preferable)	A test that is able to replace smear microscopy for treatment monitoring (for example, by detecting viable bacteria) is more likely to be adopted and to completely replace smear microscopy; thus, it would have a larger market as well.
Multiuse platform	Yes (demonstrated)	Yes (achievable)	Any technology entering this market should be useful for diagnosing also relevant diseases other than TB. The diseases to be targeted should be those among the WHO list of poverty-related diseases, communicable diseases and AMR priorities. Of course, proper implementation strategies should be in place to select which additional diseases should be targeted along with TB in a given setting. Multiplex testing or the ability to use a platform to perform different tests will likely increase the acceptability of the new test, especially in the private sector.

Operational characteristics				
Sample type	Unprocessed sputum	Unprocessed sputum		
Sample volume	up to 10 mL	< 0.5-2 mL	The lowest volume possible for all types of samples should be 0.1 ml, especially since HIV-positive patients may have difficulty providing a sample; **however, this should not come at the expense of decreased sensitivity**. If a higher volume is available, the test should be able to use it if doing so would increase sensitivity. Additionally, the ideal test would need only 1 sample even if requires 2 or more steps or reactions.	
Manual preparation of samples (steps needed after obtaining sample)[a]	No steps or 1 step	Less than 5 steps	Precise volume control and precise timing should not be required. Only general/cross-cutting laboratory skills required; no specific analytical procedures based on additional instruments should be required (e.g. DNA quantification, gel electrophoresis, serial dilutions…). Devices such as a centrifuge or heat block are available only infrequently at the level of microscopy centres; therefore, these should not be required for novel assays. The procedure should take advantage of automation as much as possible.	[14, 15]
Reagent integration	All reagents should be contained in a single device	no specific indications, but refer to reagent kit storage and stability for restrictions		
Time to result[a]	< 30 minutes for detection and DST	< 24 hours for detection and DST	The need for rapid turnaround, the possibility of batching or using random access for testing, and the ability to test multiple samples at the same time are interrelated. The time to result is probably the most important parameter since extending the wait time for patients may result in loss to follow-up.	[34, 35]
Daily throughput	> 25 tests	> 10 tests	The daily throughput needed in most microscopy centres is <10 tests per day. Daily throughput requirements are considering currently available technologies as baseline.	
Sample capacity and throughput	Multiple samples should be able to be tested at the same time; random access should be possible	Batching should be possible	Ideally, 1 sample should not occupy the instrument without it still being able to process other samples (that is, random access or parallel analyses should be possible). If the platform is multiplexed, then running different assays at the same time should be feasible.	

Walk-away operation	These features are required; there should not be a need for operator intervention once the sample has been placed into or on the instrument	No more than 1 step of operator intervention should be needed once the sample has been placed into or on the system	Once the sample has been loaded into an instrument, then further operator intervention should not be required until detection has occurred. This characteristic is related to the characteristics for sample preparation and assay processing (that is, the steps needing to be completed after a sample has been obtained).	
Biosafety	similar to those for smear microscopy (low-risk TB laboratories)	similar to those for smear microscopy (low-risk TB laboratories)	A biosafety cabinet is not commonly available at the level of a microscopy centre; low-risk TB laboratories follow the minimum biosafety requirements as described in WHO's Tuberculosis Laboratory Biosafety Manual.	[36, 37]
Waste disposal – solid	Should require no more than smear microscopy	Should require no more than current WHO-endorsed assays at peripheral level	Further information is provided in WHO's Tuberculosis Laboratory Biosafety Manual. Increasing the amount of waste generated compared to a smear microscopy laboratory should be avoided. Green friendly, sustainable packaging minimizing the environmental impact of packaging should be considered for the product's entire lifecycle.	[37]
Waste disposal – infectious	similar to those for smear microscopy (low-risk TB laboratories)	similar to those for smear microscopy (low-risk TB laboratories)	Low-risk TB laboratories as described in WHO's Tuberculosis Laboratory Biosafety Manual.	[37]
Instrument	Ideally, would be a single integrated system that is modular to allow throughput to be increased if needed	build on a modular concept allowing to tailor needs and upgrade additional functionalities at any time	Ideally, a single device is preferred but modular solutions would be acceptable (for example, for separate sample processing and detection)	
Power requirements[a]	Capable of running on standard electricity plus an ad hoc certified uninterrupted power supply unit delivered with the system to enable a cycle to be completed in case of a power outage; the	Capable of running on standard electricity plus an ad hoc certified uninterrupted power supply unit delivered with the system to enable a cycle to be completed in case of a power outage; a circuit	Continuous power is not always available at the level of a microscopy centre, and current experience with the use of electrical devices in settings where power supply can be intermittent showed challenges in finding appropriate uninterrupted power supply (UPS) solutions suitable for a given instrument. UPS should come together with the instrument, and manufacturers must provide UPS certified to meet the goal of ensuring enough power for enabling a cycle to be completed. Also, in the optimal situation, it should be possible to	[14, 15]

	uninterrupted power supply and circuit protector must be integrated within the system. The system should be also compatible for switching it in a battery operated device with the ability to run for 1 day on the battery, and with recharging capability (which could be solar powered)	protector must be integrated within the system; the uninterrupted power supply should be preferably integrated within the system	switch the system into a battery operated device that can be recharged, possibly using solar power.
Maintenance and calibration[a]	Preventative maintenance should not be needed more than once every two years. Users should be able to monitor the machine status independently from manufacturers' intervention by the use of appropriate internal/external controls; results for such controls can be shared with manufacturers or appropriate control bodies to schedule appropriate on demand intervention (maintenance/calibration). An alert should be	Preventative maintenance should not be needed more than once a year. Users should be able to monitor the machine status independently from manufacturers' intervention by the use of appropriate internal/external controls; results for such controls can be shared with manufacturers or appropriate control bodies to schedule appropriate on demand intervention (maintenance/calibration). An alert should be included to indicate	Maintenance and calibration represent two challenging points for any device to be placed at peripheral level. A maintenance alert is necessary to ensure proper functioning in settings where it is unlikely that the same person will always handle the device and that records will be kept about the duration of use. Furthermore, it will be essential that only simple tools and minimal expertise are necessary to perform maintenance, given the quantity of devices that are likely to be used; additionally, service visits are unlikely to be feasible outside of urban settings.

	included to indicate when maintenance is needed according to manufacturer's indications. Software updates should be provided remotely	when maintenance is needed according to manufacturer's indications. Software updates should be provided remotely		
Data analysis	Data analysis should be integrated into the device; a PC should not be required; exported data should be capable of being analysed on a separate or networked PC	exported data should be capable of being analysed on a separate or networked PC		
Result documentation, data display	An integrated results screen and the ability to save and print results should be included; the device should have a USB port	An integrated results screen and the ability to save results should be included; the device should have a USB port	Results should be simple to interpret (for example, positive versus negative for TB detection, or present versus absent for drug resistance). Information that would allow a more detailed interpretation of results should be available (for example, information on the mutations detected) for surveillance purposes or more differentiated clinical decision-making; however, it should be possible to hide this information if necessary.	
Regulatory requirements	Manufacturing of the assay and system should comply with ISO15189 or higher standards or regulations, and comply with ISO IEC 62304 Medical Device Data Systems; the manufacturing facility should be certified and authorized for use by a regulatory authority that is a member of the International Medical Device Regulators Forum, formerly known as Global Harmonization Task Force; the assay must be registered for in vitro diagnostic use			
Data export (connectivity and interoperability)	All data should be able to be exported (including data on use of the device, error	Integrated ability for all data to be exported from the device in a user-friendly format	Mobile phone capacity is frequently available even at the level of microscopy centres. This could be leveraged for data export, quality control, supply-chain management and surveillance. As the systems will be implemented in peripheral microscopy centers, connectivity	[36, 38]

	rates and rates of invalid tests, and personalized, protected results) over a USB port and network; network connectivity should be available through an Ethernet, Wi-Fi, and GSM/UMTS mobile broadband modem, or a combination of these; results should be encoded using a documented standard (such as HL7) and be formatted as JSON text; JSON data should be transmitted through HTTP(S) to a local or remote server as results are generated; results should be stored locally and queued during network interruptions to be sent as a batch when connectivity is restored. Bluetooth connectivity should also be available. It should also be possible to import data (e.g. for updating interpretation rules databases)	(including data on use of the device, error rates or rates of invalid tests, and non-personalized results) over a USB port. Bluetooth connectivity should also be available. It should also be possible to import data (e.g. for updating interpretation rules or databases)	should be adapted to the actual situation (data transfer cannot rely on high-speed internet connectivity, and the format of the data should be adapted accordingly). Data export must include raw data and interpreted results, allowing further re-analysis in case of updated interpretation guidelines.

Electronics and software	Should be integrated into the instrument	Should be integrated into the instrument	If an external device (separate PC, tablet, mobile…) is needed, it will likely limit the ability to update software, since staff with the skills needed to operate a PC are not present in all microscopy centres. Furthermore, security will be an issue, and separate PCs might be stolen.	
Operating temperature and humidity level	Between +5 °C to +50 °C with 90% humidity	Between +5 °C and +40 °C with 70% humidity	High environmental temperatures and high humidity are often present in countries where TB is endemic. Dust also is a problem in these settings, and the need to adequately protect optics should be considered. Tropicalized instruments/devices should be available for implementation in such settings.	[36, 39]
Reagent kit – transport	No cold chain should be required; should be able to tolerate stress during transport for a minimum of 72 hours at -15 °C to +50 °C	No cold chain required; should be able to tolerate stress during transport for a minimum of 72 hours at -15 °C to +40 °C	Refrigerated transport is costly and often cannot be guaranteed for the entire transportation process. Frequent delays in transport are commonplace.	[14, 15, 40]
Reagent kit – storage and stability	2 years at +5 °C to +40 °C with 90% humidity; should be able to tolerate stress during transport for a minimum of 72 hours at +50 °C; no cold chain should be required	12 months at +5 °C to +35 °C with 70% humidity; should be able to tolerate stress during transport for a minimum of 72 hours at +40 °C; no cold chain should be required	High environmental temperatures and high humidity are often present in countries where TB is endemic; they are especially problematic during the transport of reagents and systems.	[36, 39]
Additional supplies (not included in kit)	None	None		
Internal quality control	Full controls for sample processing, amplification and detection of TB and any target for DR should be included. Internal controls for analysis and reporting (e.g. software version) should be included. A monitor (remote) system for checking results on the controls should be also considered.			[39, 41]
Training and education	6 work hours for staff at the level of a microscopy technician	3 days (or 24 work hours) for staff at the	Trainings should be developed according to Continuing Education and Training (CET) models and Individualized Training Programs (ITP) to ensure that only properly trained, accredited people can perform	

	level of a laboratory technician	the assay. Online and remote support systems should be available for retraining, monitoring/evaluating and updating ("refresher") training. All the phases of the training should be properly documented.

[a] These characteristics were considered to be the most important.

REFERENCES

1. World Health Organization, *High-priority target product profiles for new TB diagnostics: report of a consensus meeting*. 2014, World Health Organization.
2. World Health Organization, *WHO treatment guidelines for multidrug- and rifampicin-resistant tuberculosis, 2018 update*. 2018: Geneva, Switzerland.
3. UNITAID, *Tuberculosis diagnostics technology landscape*. 2017, World Health Organization.
4. World Health Organization, *Implementing tuberculosis diagnostics: a policy framework*. 2015: Geneva, Switzerland.
5. World Health Organization, *Global tuberculosis report 2018*. 2018: Geneva, Switzerland.
6. World Health Organization, *Guidelines for treatment of drug-susceptible tuberculosis and patient care, 2017 update*. 2017: Geneva, Switzerland.
7. World Health Organization, *WHO treatment guidelines for isoniazid-resistant tuberculosis: Supplement to the WHO treatment guidelines for drug-resistant tuberculosis*. 2018: Geneva, Switzerland.
8. World Health Organization, *Rapid communication: key changes to treatment of multidrug- and rifampicin-resistant tuberculosis (MDR/RR-TB)*. 2018: Geneva, Switzerland.
9. Ahuja, S.D., et al., *Multidrug resistant pulmonary tuberculosis treatment regimens and patient outcomes: an individual patient data meta-analysis of 9,153 patients.* PLoS Med, 2012. **9**(8): p. e1001300.
10. Heyckendorf, J., et al., *What Is Resistance? Impact of Phenotypic versus Molecular Drug Resistance Testing on Therapy for Multi- and Extensively Drug-Resistant Tuberculosis.* Antimicrob Agents Chemother, 2018. **62**(2).
11. World Health Organization, *Implementing the END TB strategy: the essentials*. 2015: Geneva, Switzerland.
12. MacPherson, P., et al., *Pre-treatment loss to follow-up in tuberculosis patients in low- and lower-middle-income countries and high-burden countries: a systematic review and meta-analysis.* Bull World Health Organ, 2014. **92**(2): p. 126-38.
13. Ned-Sykes, R., et al., *Competency Guidelines for Public Health Laboratory Professionals: CDC and the Association of Public Health Laboratories.* MMWR Suppl, 2015. **64**(1): p. 1-81.
14. Pai, M. and M. Schito, *Tuberculosis diagnostics in 2015: landscape, priorities, needs, and prospects.* J Infect Dis, 2015. **211 Suppl 2**: p. S21-8.
15. Pai, M., *Innovations in Tuberculosis Diagnostics: Progress and Translational Challenges.* EBioMedicine, 2015. **2**(3): p. 182-3.

16. Cobelens, F., et al., *Which new diagnostics for tuberculosis, and when?* J Infect Dis, 2012. **205 Suppl 2**: p. S191-8.
17. Keeler, E., et al., *Reducing the global burden of tuberculosis: the contribution of improved diagnostics.* Nature, 2006. **444 Suppl 1**: p. 49-57.
18. Dowdy, D.W., et al., *Cost-effectiveness of rapid susceptibility testing against second-line drugs for tuberculosis.* Int J Tuberc Lung Dis, 2014. **18**(6): p. 647-54.
19. Adepoyibi, T., et al., *Which attributes within target product profiles for tuberculosis diagnostics are the most important to focus on?* Int J Tuberc Lung Dis, 2018. **22**(4): p. 425-428.
20. Pantoja, A., S.V. Kik, and C.M. Denkinger, *Costs of novel tuberculosis diagnostics--will countries be able to afford it?* J Infect Dis, 2015. **211 Suppl 2**: p. S67-77.
21. Murray, M., et al., *Cost-effectiveness of triage testing for facility-based systematic screening of tuberculosis among Ugandan adults.* BMJ Glob Health, 2016. **1**(2): p. e000064.
22. Padmasawitri, T.I.A., et al., *Disparities in model-based cost-effectiveness analyses of tuberculosis diagnosis: A systematic review.* PLoS One, 2018. **13**(5): p. e0193293.
23. Van't Hoog, A.H., et al., *The potential of a multiplex high-throughput molecular assay for early detection of first and second line tuberculosis drug resistance mutations to improve infection control and reduce costs: a decision analytical modeling study.* BMC Infect Dis, 2015. **15**: p. 473.
24. World Health Organization, *WHO Meeting Report of a Technical Expert Consultation: Non-inferiority analysis of Xpert MTB/RIF Ultra compared to Xpert MTB/RIF*. 2017, World Health Organization.
25. Ninan, M.M., et al., *The diagnostic utility of line probe assays for multidrug-resistant tuberculosis.* Pathog Glob Health, 2016. **110**(4-5): p. 194-9.
26. Administration, F.a.D., *Class II Special Controls Guideline: Nucleic Acid-Based In Vitro Diagnostic Devices for the Detection of Mycobacterium tuberculosis Complex in Respiratory Specimens - Guideline for Industry and Food and Drug Administration Staff*. 2014.
27. Helb, D., et al., *Rapid detection of Mycobacterium tuberculosis and rifampin resistance by use of on-demand, near-patient technology.* J Clin Microbiol, 2010. **48**(1): p. 229-37.
28. Sun, A.Y., et al., *Modeling the impact of alternative strategies for rapid molecular diagnosis of tuberculosis in Southeast Asia.* Am J Epidemiol, 2013. **178**(12): p. 1740-9.
29. Nathavitharana, R.R., et al., *Multicenter Noninferiority Evaluation of Hain GenoType MTBDRplus Version 2 and Nipro NTM+MDRTB Line Probe Assays for Detection of Rifampin and Isoniazid Resistance.* J Clin Microbiol, 2016. **54**(6): p. 1624-1630.
30. World Health Organization, *The use of molecular line probe assays for the detection of resistance to second-line antituberculosis drugs. Policy guidance.* 2016, World Health Organization: Geneva, Switzerland.
31. World Health Organization, *The use of next-generation sequencing technologies for the detection of mutations associated with drug resistance in Mycobacterium tuberculosis complex: technical guide*. 2018: Geneva, Switzerland.
32. Dalton, T., et al., *Prevalence of and risk factors for resistance to second-line drugs in people with multidrug-resistant tuberculosis in eight countries: a prospective cohort study.* Lancet, 2012. **380**(9851): p. 1406-17.
33. Horne, D.J., et al., *Diagnostic accuracy and reproducibility of WHO-endorsed phenotypic drug susceptibility testing methods for first-line and second-line antituberculosis drugs.* J Clin Microbiol, 2013. **51**(2): p. 393-401.

34. Claassens, M.M., et al., *Tuberculosis patients in primary care do not start treatment. What role do health system delays play?* Int J Tuberc Lung Dis, 2013. **17**(5): p. 603-7.
35. Sreeramareddy, C.T., et al., *Time delays in diagnosis of pulmonary tuberculosis: a systematic review of literature.* BMC Infect Dis, 2009. **9**: p. 91.
36. Denkinger, C.M., et al., *Are peripheral microscopy centres ready for next generation molecular tuberculosis diagnostics?* Eur Respir J, 2013. **42**(2): p. 544-7.
37. World Health Organization, *Tuberculosis laboratory biosafety manual*. 2012: Geneva, Switzerland.
38. Andre, E., et al., *Connectivity of diagnostic technologies: improving surveillance and accelerating tuberculosis elimination.* Int J Tuberc Lung Dis, 2016. **20**(8): p. 999-1003.
39. Banoo, S., et al., *Evaluation of diagnostic tests for infectious diseases: general principles.* Nat Rev Microbiol, 2010. **8**(12 Suppl): p. S17-29.
40. Kuupiel, D., V. Bawontuo, and T.P. Mashamba-Thompson, *Improving the Accessibility and Efficiency of Point-of-Care Diagnostics Services in Low- and Middle-Income Countries: Lean and Agile Supply Chain Management.* Diagnostics (Basel), 2017. **7**(4).
41. Parsons, L.M., et al., *Laboratory diagnosis of tuberculosis in resource-poor countries: challenges and opportunities.* Clin Microbiol Rev, 2011. **24**(2): p. 314-50.