LabCoP EXTENDED ECHO SESSION

16 April 2026

Pooled sputum testing for molecular TB diagnosis – Lessons from the Start4All multi-country experience

Presenters: Dr. Tom Wingfield (Liverpool School of Tropical Medicine), Dr. Jérémie Piton (UNITAID), Dr. Alexei Korobitsyn (WHO), Dr. Vibol Iem (Liverpool School of Tropical Medicine), Dr. Valérie Donkeng (Centre Pasteur du Cameroun), Dr. Tushar Garg (Stop TB Partnership), Dr. Jacob Creswell (Stop TB Partnership), Dr. Rachel Byrne (Liverpool School of Tropical Medicine)

What is the diagnostic accuracy of pooled sputum testing?

1. What is the sensitivity of individual test and pool test compared with MGIT?

Below is a screenshot from Dr Tushar Garg's presentation on the performance of pooled Xpert Ultra vs Individual Xpert Ultra against MGIT culture from our Start4All study. This was calculated from data from 13,531 participants with complete data (pooled, individual Xpert Ultra, culture), across seven high TB burden countries.

	Pooled testing	Individual Xpert Ultra	Difference
Sensitivity vs culture MRS	**84.9%** 95% CI 82.5–87.0	**88.0%** 95% CI 85.9–90.0	**−3.2 %** 95% CI −4·4 to −2·1 , p < 0.01
Specificity vs culture MRS	**98.3%** 95% CI 98.0–98.5	**97.7%** 95% CI 97.4–97.9	**+0.6 %** 95% CI 0·5 – 0·7, p<0·01

2. Diagnostic accuracy is largely influenced by disease prevalence; under what prevalence conditions was the study conducted, are there studies conducted at different settings?

In Start4All diagnostic accuracy study across 7 countries, 5,135 (36·6%) participants were recruited in communities, 4,249 (30·3%) in Primary Health Care (PHC) facilities, 4,106 (29·2%) in District Health (DH) facilities and 550 (3·9%) were children. On individual Xpert Ultra, 1,248 (8·9%) tests were positive (693 [16·9%] in DH, 334 [7·9%] in PHC, and 210 [4·1%] in communities.

3. Pooled sputum molecular TB testing may be more efficient, yes, but to what extent is its accuracy compared to index test of TB and cost analysis to use in low resource countries?

In terms of performance, sensitivity was 3 percentage points lower with pooled vs individual Xpert Ultra, which is traded off for slightly better specificity, against culture as reference.

Start4All

What are some key considerations for pooled sputum testing?

4. Can we conduct a pooled sputum testing with existing Xpert machine?

Yes, pooled testing can be performed on existing GeneXpert systems without modifying the instrument. However, it requires changes to laboratory workflows, including sample handling, pooled testing procedures, and data recording, supported by SOPs and staff training.

5. Is pooling possible for testing other samples or it's only for sputum alone?

Data presented here is only for sputum. Evidence for other sample types remains limited. But Start4All Phase 2 will explore pooled tongue swabs and pooled sputum swabs on near point-of-care (NPOC) platforms (results expected next year in 2027).

6. How to combine 4 sputa in the 1 pooling test?

See the video in the Start4All "Pool-kit", a toolkit with practical resources to support implementation of pooled testing.

7. Should pooling be done under BSC?

Pooled testing should be conducted under the same laboratory safety procedures as individual testing for that test site.

8. Is there any limit to the number of samples that could be pooled?

WHO currently recommends pools up to four samples. Pools of four specimens are commonly used in operational settings, although smaller pools may be selected depending on specimen availability or laboratory workflow. Larger pool sizes may offer additional gains at very low positivity levels, but increase dilution and risk of reduced sensitivity, particularly in samples with low bacterial load. Larger pool sizes were also not part of the 2026 WHO recommendations on pooled testing.

9. In your study you pooled 4 samples, now what should be the maximum number of samples for a pool that does not affect the limit of detection and compromise of sensitivity?

Dorfman method predicts the optimum number of specimens per pool based on prevalence. This shiny app can do it. https://bilder.shinyapps.io/PooledTesting. Look for Hierarchical Testing —> Calculate for one testing configuration.

Start4All

10. What do you do if the pool test result is positive, and when you did the test to individual sample all 4 showed negative result? Would you redo the test for all 4 samples?

This sometimes happens. It is a good question. There is some variability of Ultra tests, and even with individual tests, sometimes there are discrepant results, especially with low/trace semi-quantitative results. There is a real possibility that someone in that pool might test positive for TB if you retested everyone, but the individual Xpert result is taken as the result (just as in a standard lab approach). We recommend following up each person in the pool until they get better, including follow-up TB tests as indicated.

11. What happens if the pooled results are TB detected Trace and RIF indeterminate?

If the pooled test result is MTB detected trace, all individual specimens included in the pool must undergo deconvolution individual testing. Because pooled testing dilutes the original specimens, the semi-quantitative category observed in the pooled result may not reflect the bacillary load of the individual specimen responsible for the signal. For example, a pooled result reported as trace may originate from an individual specimen that would yield a higher semi-quantitative category when tested individually.

12. In a pooled test containing three samples, if one sample has a very low MTB detection, can the two negative samples affect the result and lead to a false-negative outcome?

Pooled testing introduces some dilution of bacterial genetic material, which may slightly reduce analytical sensitivity. Evidence suggests this effect is generally small, although detection may be lower in specimens with very low bacillary loads. Diagnostic accuracy may also be lower when sputum is collected through community-based case finding compared with facility-based collection, potentially reflecting differences in bacterial load or sample quality. Programmes implementing pooled testing should therefore balance efficiency gains with potential reductions in sensitivity, particularly in populations where many specimens contain very low bacterial concentrations or where samples are collected in community settings.

13. Will a too mucopurulent sample affect the result of other sample result?

The sample preparation is the same as with an individual test. In our study, pooled testing involves mixing processed sputum samples after the addition of sample reagent, which liquefies and homogenises the specimen. This reduces variability related to sputum consistency, including mucopurulent samples. There is no evidence that a mucopurulent sample adversely affects the results of other samples in the pool, provided standard processing procedures are followed.

14. Is there a difference in the amount of sample for pooled vs individual testing?

This is explained in detail in the Start4All toolkit. National guidelines typically recommend collecting 1–4 mL of sputum, and pooled testing fits within these existing requirements for individual molecular testing. If a specimen is acceptable for individual testing, it can also be included in pooled testing. Specimens with less than 1 mL of raw sputum may not provide enough volume for pooled testing. In such cases, pooled testing is not recommended, and laboratories should follow standard procedures, such as individual testing, or requesting a new sample, in line with NTP guidance and local protocols.

15. While pooling, do we need to mix? if yes how do we reduce the risk of aerosols?

Yes, mixing is required. As with individual testing, sputum is first mixed with sample reagent, which liquefies the specimen and reduces infectiousness. Mixing should be performed using standard laboratory procedures, such as gentle inversion or swirling in a closed container, to avoid aerosol generation. Laboratories should follow routine biosafety practices, including appropriate PPE, working in a well-ventilated area or designated workspace, and minimizing open handling steps.

16. What is the recommended sample volume when performing pooling? Is this amount sufficient in cases of low bacteraemia? Or could the high amount of DNA inhibit the PCR? Which cutoff is used?

A minimum of 1 mL of raw sputum is recommended for pooled testing. After processing with sample reagent at the standard 2:1 ratio, this yields approximately 3 mL of processed specimen. For pools of four, around 1 mL of processed specimen from each sample is combined to form the pooled sample (about 4 mL total), with ~2 mL used for testing and the remainder available for repeat testing if needed. The study did not identify issues related to PCR inhibition from higher DNA concentrations. However, pooled testing introduces dilution, which may reduce sensitivity in specimens with very low bacterial load.

17. Is there any comparison of pooling of only presumptive TB cases vs mixed presumptive TB and Presumptive DR-TB samples?

This study evaluated pooled testing among people with presumptive TB and did not specifically assess pooled testing strategies for presumptive DR-TB populations. Further evidence would be needed to assess the performance and appropriateness of pooled testing in DR-TB populations.

What are some programmatic considerations for implementing pooled sputum testing?

18. Pooling seems to be beneficial in settings with low prevalence of TB. Why was the diagnostic accuracy of the pooling strategy reduced when sputum was collected in community settings (sensitivity range of 25-96%, 4 studies) compared with facility-based sputum collection (sensitivity range of 84-100%, 7 studies). One would imagine pooling to be beneficial for example in community drives, where majority of samples would be negative and pooling would be beneficial compared to facility based. The population at facility is usually already pre-selected.

While pooled testing is more efficient in low-prevalence settings such as community screening, diagnostic accuracy depends largely on bacterial load rather than prevalence. In community-based settings, individuals often have earlier or less advanced disease, leading to a higher proportion of specimens with very low bacillary load. Because pooled testing introduces a dilution effect, detection may be reduced in these low-burden samples, which can lower sensitivity. In contrast, facility-based populations are more likely to include individuals with higher bacterial loads, resulting in higher sensitivity even with pooled testing. As such, pooled testing can offer important efficiency gains in community settings, but programmes should balance these benefits against the potential for reduced sensitivity in populations with predominantly low bacillary load.

19. Is there a calculation to guide the number of tests per pool varied by estimated prevalence in the population you are testing?

Yes, the optimal pool size can be guided by the expected prevalence in the population. Dorfman method predicts this. Practical tools are also available, such as this Shiny app: https://bilder.shinyapps.io/PooledTesting. (see “Hierarchical Testing” → “Calculate for one testing configuration”), which allows users to estimate the optimal number of samples per pool under different prevalence scenarios.

20. Pooled testing is a great especially for resource-limited settings, but with low prevalence.

Pooled testing is most beneficial in low-prevalence settings. Based on the classical Dorfman framework, efficiency gains decline rapidly as prevalence increases. Around 25% positivity, most of the efficiency gains are lost, and by ~30%, pooled testing offers little to no benefit compared to individual testing. In practice, ~25% can be considered the practical cut-off for pooled testing, while ~30% represents a theoretical ceiling where pooled testing may still be feasible but is essentially not useful.

21. Why is pooled testing used despite the need for subsequent individual testing when a pooled sample tests positive?

You are right that positive pools need retesting of individual samples. However, negative pools do not require retesting, and this is where you get much of the cartridge and costs savings. In low-prevalence settings, the majority of pools are negative, allowing multiple individuals to be cleared with a single test. This leads to substantial savings in cartridges, time, and laboratory workload despite the need for follow-on testing in positive pools.

22. Is pooling suitable for high endemic areas?

Pooled testing can be suitable in high-burden settings, but its efficiency depends on the local positivity rate. It works best when laboratory positivity is low to moderate. In practice, pooled testing remains efficient up to around 25% positivity; beyond this, efficiency gains decline rapidly, and by ~30% there is little to no benefit.

23. What is the limit of detection of pooled samples compared to individual sample? Also, did you compared different number of samples used for pooling and what were their sensitivity?

Pooled testing introduces a dilution effect, which may slightly reduce analytical sensitivity, particularly in specimens with very low or Trace bacillary load. While an exact limit of detection for pooled samples was not formally established in this study, we observed an overall reduction in sensitivity of about 3% compared to individual testing, with the largest impact seen in very low and trace categories.

We also evaluated different pool sizes (2, 3, and 4 samples). Most pools contained four samples, and sensitivity remained high across pool sizes, with only small differences observed. The modest reduction in sensitivity was primarily driven by low bacterial load rather than the number of samples per pool.

24. Did you take all the samples received in the programmatic setting? or calculated the sample size for the study?

This was conducted within a study setting. We included all individuals meeting eligibility criteria that were able to give sufficient sputum for both individual and pooled testing during the study period.

25. Did you identify in your studies groups at risk of TB or DR-TB (e.g. household contacts, retreatments) that would benefit directly from individual testing rather than pooling approach in a supposed diagnostic algorithm starting with pool testing? Do you have trends of correlation of specific groups with positive testing? Thanks, and congratulations!

This study did not specifically stratify or prioritise high-risk groups (e.g. household contacts or people with previous TB treatment) for individual versus pooled testing. Identifying which populations may benefit more from individual testing vs pooled testing is an important question, and one we aim to address in Start4All Phase 2. This will include exploring adaptive strategies, such as using CAD scores to guide whether samples are tested individually or pooled.

26. How is a sensitivity drop considered acceptable during standardization of pool size?

This will be an operational question for NTPs, implementers, and TB-affected communities in your setting to decide. The drop we found (overall) was 3 percentage points but the confidence intervals for pooled and individual testing overlap, which means that, given the uncertainty in the estimates, the true sensitivity of the two approaches could be very similar. In simple terms, this suggests that the difference observed is small and that overall sensitivity is broadly comparable.

27. How did you batch the pooled sputum specimen at large sputum collection centres?

During our study in Start4All, pools were created by combining consecutive samples as part of the research study design without selection based on clinical or demographic factors, provided there was sufficient volume. **WHO currently recommends pools up to four samples**. Pools of four specimens are commonly used in operational settings, although smaller pools may be selected depending on specimen availability or laboratory workflow. Larger pool sizes may offer additional gains at very low positivity levels, but increase dilution and risk of reduced sensitivity, particularly in samples with low bacterial load. Larger pool sizes were also not part of the 2026 WHO recommendations on pooled testing.

28. Which one would be more acceptable (for NTP and clients) - drop in sensitivity of 3% or saving (cartridges) of about 50%?

There is no single answer, as this depends on context and programme priorities. It is a trade-off between a small reduction in sensitivity (~3%) and substantial efficiency gains (~50% fewer cartridges).
In low-resource or low-coverage settings, testing more people may outweigh the small loss in sensitivity, while higher-risk settings may prioritise maximising sensitivity. Ultimately, acceptability will depend on how programmes balance access, cost, and diagnostic performance within their specific context.

29. Is there any data on use of pooled sputum testing in prison settings?

Yes, a few studies, including from Brazil, have shown that pooled sputum testing is feasible, sensitive, and efficient for TB screening in prisons. However, evidence remains limited.

30. For the Cameroon experience,

   a. How did you standardize the pool size?

   Labs typically use a single pools size based on their historical TB test positivity rate. So, a lab with an historical TB test positivity rate of 12% may typically test in pools of 2, and they might just use occasional pools of 3 depending on the workflow and cartridge availability. And a lab with an historical positivity rate of 7% might use pools of 4 typically but then might prepare a pool of 4 and a pool of 3 on a day when they have 7 specimens to test.

   b. How did you identify and confirm false positive results on pool testing?

   We might better call these 'discordant positive results.' This happens when the pool test has a positive result and none of the subsequent individual tests are positive. We do not' consider these 'false' positives because there is often a specimen with TB in the pool, but this is missed on the subsequent individual test because of assay variability. This nearly always happens when the pool result has a small concentration of TB, either a Trace or Very Low result on Ultra. In these cases, we follow up each person that had a negative result from the positive pool until the person gets better, including with additional TB testing as indicated. Often, someone from one of these discordant pools will have a subsequent positive test after re-testing on Ultra.

c. What were the criteria for choosing the number of samples per pool?

For the pilot phase, we had a fixed pool size of 3 because we were still doing an evaluation. With an evidence base, we move to a scale phase with a flexible pool size based on lab positivity rate.

d. What did you mean by "tracking false positive pools" and how common are those?

It might be better to say, "tracking discordant positive pools." We track the rate and type of discordant positive pools. These are pools that are positive on the pool test but then have no individual positive results when each specimen is tested individually. This can happen due to the variability of the PCR assay, and it has been reported at similar rates on other types of specimen pooled testing (eg for COVID pooled testing and STI pooled testing). We typically only observe these discordant positive pools when the pool result has a low concentration of TB, either a Trace or a Very Low result on Ultra. An acceptable rate of discordant positive pools is 3-10%. If a lab has a higher or lower rate, this might indicate issues with test procedures. In addition, if there are multiple events where a pool with a higher positive result (either Low or Medium or High on Ultra) is discordant, with no individual Ultra positives in the pool, then this might indicate issues with specimen handling and/or pooled testing procedures.

e. Is Pooled sputum testing currently being done at CPC so that we can refer to CPC in a study that need to diagnose TB in patients?

Yes, you can contact Dr. Valerie Donfankeng.

31. What is the experience with pooling in high-throughput Xpert laboratories? i.e., labs operating GX80 or multiple platforms? There are concerns that workflows will become too complex in the ability to track and manage so many specimens?

One of the key considerations we have highlighted is to have a reliable system to track and map individual samples to their respective pools. As outlined in the Start4All toolkit, this includes having adequate space to organise samples while awaiting pooled results, and a data system (e.g. registers, log sheets, or digital tools) to ensure that samples from positive pools can be easily retrieved for individual testing. The SOP also recommends incorporating a clear linkage between pool IDs and individual samples to simplify tracking and minimise errors.

32. Many electronic systems have been designed around a simple assumption in that one sample equates to one test, one patient and one result. Is adding a Pool ID field the only suggested change? If we are to continue linking tests with cartridge use, should systems not be updated to support parent child relationships between pools and samples, many to many mappings between tests and patients, conditional workflows triggered by results, multi-stage turnaround times etc?

While adding a Pool ID field is a useful first step, more robust adaptations are often needed. Ideally, systems should support parent–child relationships between pools and individual samples, allow linkage of multiple samples to a single test, and enable retrieval of individual samples for follow-on testing when a pool is positive. Conditional workflows and tracking of multi-stage turnaround times can further strengthen implementation, particularly in high-throughput settings. In the absence of such system-level adaptations, these processes can be managed using structured paper registers or simple digital tools, as outlined in the Start4All toolkit.

33. Any suggestion on pooling for Contacts (per index cases) as part of ACF, rather than simply in the community in general

Pooled testing is an approach to support expansion of molecular TB testing, particularly in community-based case finding (CBCF, formerly called ACF) and mass screening strategies. It can be considered for contacts, but its suitability depends on the expected positivity within the group. If contacts are tested around an index case, positivity may be higher than in general community screening, which can reduce the efficiency of pooled testing. The key point is that pooled testing is most valuable when expected positivity is low. In practice, pooled testing may be useful for larger or lower risk contact groups, while individual testing may be preferred for high-risk contacts or when clinical presumption of TB is high.

34. How do you decide which specimens to pool if, for example, you have 20 specimens to be tested in the lab.

During our study in Start4All, pools were created by combining consecutive samples as part of the research study design without selection based on clinical or demographic factors, provided there was sufficient volume. We preferentially included pools of 4 sputum samples. If the number was not divisible by 4, we included all left-over sputum in pools of 3 or 2. These were much smaller numbers.

35. Could we apply similar pooling method for TrueNat and potentially NPOC?

The product for which eligible data were assessed and met the class-based performance criteria for LC-aNAATs for the WHO recommendation on pooled testing was Xpert MTB/RIF Ultra.

While we would anticipate that other molecular tests using sputum, such as TrueNat platforms, may show similar performance, evidence is still needed. Start4All Phase 2 will explore pooled tongue swabs and pooled sputum swabs on near point-of-care (NPOC) platforms (results expected next year in 2027).

36. What are the challenges for non-hierarchical pool testing?

Non-hierarchical pooled testing (e.g. matrix or combinatorial designs) can improve efficiency but is more complex to implement. Samples are included in multiple pools, which requires more sophisticated tracking systems, data management, and result interpretation.

In these approaches, results are often inferred rather than directly confirmed through individual retesting, which can make them harder to explain and less intuitive for end users. This may affect acceptability and trust, particularly in settings where systems are already under pressure or where clear, transparent workflows are essential.

This added complexity makes workflows harder to manage, especially in high-throughput or resource-limited settings. For these reasons, hierarchical (Dorfman) pooled testing is generally preferred for routine programmatic use due to its simplicity and ease of integration into existing laboratory systems.