Tuberculosis 115 (2019) 140–145

Contents lists available at ScienceDirect

Tuberculosis

journal homepage: www.elsevier.com/locate/tube

Diagnostics

Quantification of viable bacterial load in artificial sputum spiked with *Mycobacterium tuberculosis*

Sven O. Friedrich[a,b], Eva Kolwijck[c], Miriam N. Karinja[d], Lize van der Merwe[b], Andreas H. Diacon[a,b,*]

[a] Division of Medical Physiology, MRC Centre for Tuberculosis Research, DST/NRF Centre of Excellence for Biomedical Tuberculosis Research, Faculty of Medicine and Health Sciences, Stellenbosch University, Tygerberg, South Africa
[b] Task Applied Science, Bellville, Cape Town, South Africa
[c] Department of Medical Microbiology, Radboud University Nijmegen Medical Centre, Nijmegen, the Netherlands
[d] Swiss Tropical and Public Health Institute, University of Basel Switzerland & Center for Research in Therapeutic Sciences, Strathmore University, Nairobi, Kenya

ARTICLE INFO

Keywords:
Mycobacterium tuberculosis
Artificial sputum
Storage
Viability
Validation
Proficiency

ABSTRACT

Objective: Artificial sputum spiked with *Mycobacterium tuberculosis* could serve for validation of procedures that determine viable mycobacterial load.
Design: Artificial sputum specimens prepared in-house were spiked with low, medium or high concentrations of *Mycobacterium tuberculosis* H37Rv stock solution. In a first series, a single technologist processed two batches of specimens daily with high load that were stored refrigerated or at room temperature for up to 8 days. In a second series, nine different technologists processed freshly made batches of specimens with low, medium or high loads. We recorded time to positivity (TTP) in duplicate liquid cultures made from each specimen.
Results: Specimens were well grouped around the mean TTP (hours; standard deviation) of low: 271.7 (25.9), medium: 233.5 (16.3), and two batches of high load: 186.9 (12.3) and 191.8 (9.0), respectively. A variance component model that included load, storage temperature, days of storage until processing, batch of specimens made, sample ID and technologist ID as random effects in a linear mixed-effects model identified only load, technologist and residual as significant contributors to overall TTP variance.
Conclusion: Artificial sputum specimens with reproducible and stable viable mycobacterial loads can be made that could serve for training and validation purposes.

1. Introduction

A standard artificial sputum spiked with *Mycobacterium tuberculosis* could be useful to calibrate laboratory equipment, assess processing proficiency, perform external quality assessment and confirm inter-laboratory agreement of quantitative results of microscopy, cell culture or molecular assays performed on sputum. Such a tool would be of great benefit in preparation for multicenter clinical trials, however, a recognized protocol to create consistent samples is still not in existence [1].

Artificial sputum is a convenient matrix to generate specimens spiked with bacteria as its composition can be exactly defined and it is free of any contaminating cells which might be present if samples from patients with lung disease were collected [2]. Yamada *et al.* showed that artificial sputum samples can be spiked with various types of cells and smears made from these specimens examined by microscopy can appear nearly identical to smears made from real clinical samples [3–5].

Short term storage of sputum samples from patients with pulmonary tuberculosis (TB) is sometimes necessary as analytical laboratories can be some distance away from the facilities where samples are collected. Paramasivan et al. showed that by using microscopy, *M. tuberculosis* presence can still be shown after 28 days of storage at room temperature, but recommended a maximum of 3 days storage for culture assays as mycobacterial viability declines rapidly [6]. Banda *et al.* found that after 4 weeks of storage, mycobacteria from samples refrigerated at 4 °C retained 67% viability compared to 37% when kept at room temperature [7]. Kolwijck *et al.* reported that sputum specimens collected for early bactericidal activity studies can be stored refrigerated for at least

Abbreviations: AFB, acid-fast bacilli; MGIT, Mycobacterial Growth Indicator Tube; TTP, time to positivity

* Corresponding author. Division of Medical Physiology, MRC Centre for Tuberculosis Research, DST/NRF Centre of Excellence for Biomedical Tuberculosis Research, Faculty of Medicine and Health Sciences, Stellenbosch University, Tygerberg, South Africa.
E-mail address: ahd@sun.ac.za (A.H. Diacon).

https://doi.org/10.1016/j.tube.2019.03.001
Received 14 November 2018; Received in revised form 28 February 2019; Accepted 4 March 2019

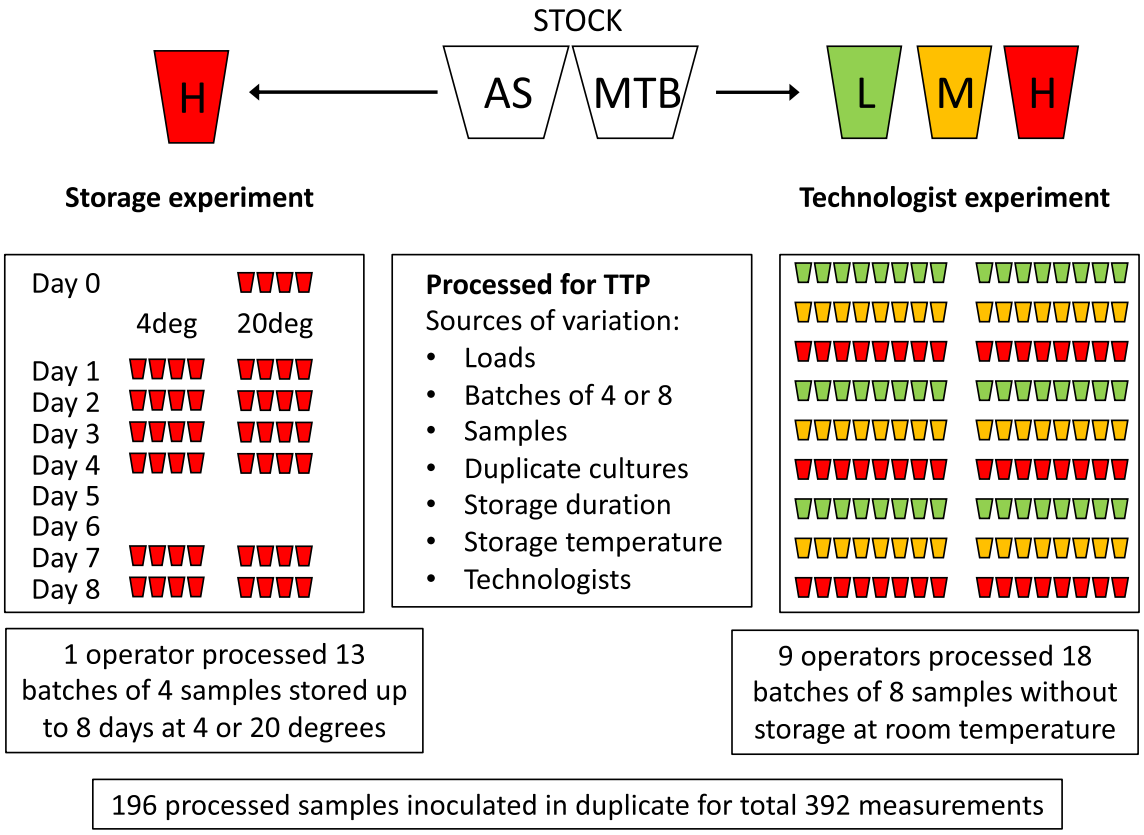

Fig. 1. Experimental setup of artificial sputum specimens spiked with different loads of *M tuberculosis*. The mycobacterial stock solution (MTB) was used undiluted (H = high) or diluted as described (M = medium or L = low) to spike vials with artificial sputum. Batches of 4 or 8 samples were processed immediately after preparation by nine different technologist or kept for up to 8 days at room temperature (20 °C) or refrigerated (4 °C) at 4–8 °C before being processed by one technologist. TTP = Time to positivity.

3 days without a significant change in colony forming unit counts on agar plates or time to culture positivity in liquid culture [8].

The aim of this study was to assess the feasibility of producing and storing different batches of artificial sputum specimens spiked with a range of concentrations of *M. tuberculosis*. Predictable and reproducible quantification of the mycobacterial load of such samples could prove beneficial for staff training and method validation.

2. Materials and methods

2.1. Mycobacterial stock cultures

All experiments were conducted using *M. tuberculosis* strain H37Rv purchased from the American Type Culture Collection (ATCC 27294; Manassas, VA) and first described in 1972 [9]. A subculture of this strain was incubated in a Mycobacteria Growth Indicator Tube (MGIT; Becton Dickinson, Sparks, MD) enriched with growth supplement (OADC; Becton Dickinson) and antibiotics (PANTA; Becton Dickinson) according to manufacturer recommendations. In the grown culture, the presence of acid-fast bacilli (AFB) was confirmed by microscopy of Ziehl-Neelsen stained smears and contamination was excluded by lack of growth on blood agar plates (National Health Laboratory Service, Cape Town, South Africa) after incubating for 48 h at 37 °C. Stock cultures were made by growing 20 µl of culture suspension in 3.98 ml MGIT medium containing OADC/PANTA. An aliquot of 500 µl from this dilution was used to inoculate six fresh and supplemented MGITs. These tubes were incubated for two additional days after being flagged positive by the instrument, removed, inverted and left standing. The upper volume of each tube was harvested, leaving approximately 1 ml at the bottom and united in one 50 ml sterile centrifuge tube and transferred as aliquots of 200 µl into 1 ml screw cap cryogenic vials (Corning, Oneonta, NY) with regular vortexing in-between to assure an even distribution of bacteria. Thereafter, these vials were stored frozen at −80 °C for one month.

2.2. Preparation of artificial sputum samples and storage

Artificial sputum was created using a modified protocol published elsewhere [2]. An amount of 10 g methyl cellulose (Sigma-Aldrich, Cape Town, South Africa) was carefully added to 1.0 l sterile water at 50 °C with continuous stirring to avoid clumping. One emulsified egg was then added and the solution stirred until it was evenly suspended. The mixture was divided into bottles, which were autoclaved for 20 min at 120 °C and stored at 4 °C for one month. For each experiment, the necessary frozen mycobacterial stock was thawed, united and vortexed before every pipetting step to disperse the cells homogenously. Specimens were prepared by placing 3 ml of artificial sputum in a sterile sputum collection jar, spiking with 0.5 ml of mycobacterial stock and mixing by shaking. Neat stock was used for high load and stock diluted 1:5 or 1:25 for medium and low load, respectively.

2.3. Sample processing

Technologists transferred the complete specimen into a 50 ml centrifuge tube and decontaminated it by adding 5 ml of 2% NaOH-NALC (BBL Mycoprep, Becton Dickinson). The mixture was incubated for 20 min at room temperature with vortexing every 5 min and the reaction was stopped by adding phosphate buffered saline (PBS; pH 6.8; Becton Dickinson) to a final volume of 50 ml. The tube was then centrifuged for 15 min at 4 °C and 3000 × *g*, the supernatant discarded and the remaining pellet re-suspended in 2 ml PBS. Of this, 0.5 ml was used to inoculate each of two sterile MGITs supplemented with OADC/

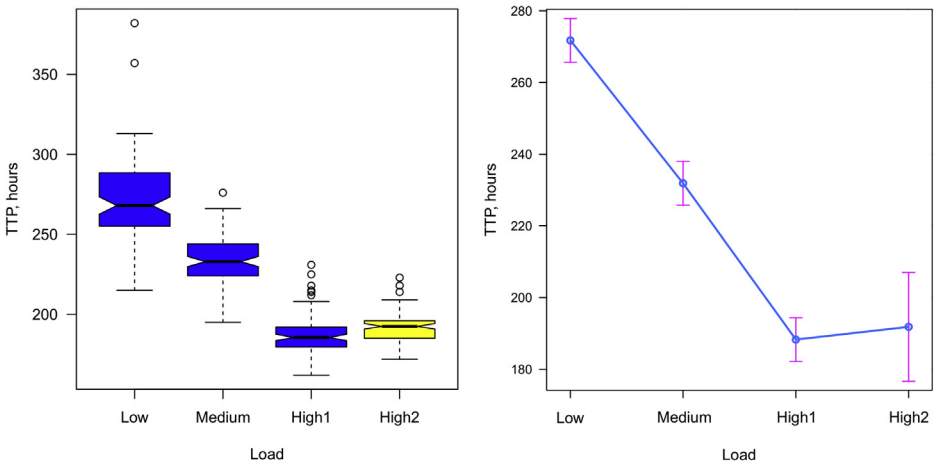

Fig. 2. Observed TTP (left panel) and estimated mean (95% CI) TTP (right panel) from artificial sputum samples spiked with low, medium and two high loads of *M. tuberculosis*. The notched boxplots on the left panel show the distribution of the 392 observations. The right panel shows the estimated means of the loads with 95% confidence intervals. In notched box plots the box covers the middle 50% of values and the line inside the box shows the median. The notches in the boxes are approximate 95% confidence intervals for the median. The whiskers extend from the quartiles to the smallest and largest values, respectively, unless they are extreme and indicated with a circle. (TTP = Time to positivity).

Table 1
Mean and standard deviation (SD) of time to culture positivity returned from artificial sputum samples spiked with different loads of *M tuberculosis*.

Load	Observations (n)	Mean TTP (h)	SD
Storage experiment			
High, batch 1	104	191.8	9.0
Load and technologist experiment			
Low	96	271.7	25.9
Medium	96	233.5	16.3
High, batch 2	96	186.9	12.3
Total	392		

TTP: Time to culture positivity in hours; SD: standard deviation.

PANTA These tubes were incubated and tested for AFB and contamination as previously described. Time to culture positivity (TTP) was reported. All technologists used in the following experiments had undergone at least 3 months of training and were proficient in this procedure.

2.4. Experiments

Two experiments were designed, namely a “storage” and a “technologist” experiment. The former was performed on high mycobacterial load specimens by a single technologist who processed batches of 4 specimens immediately or after being stored for 1, 2, 3, 4, 7 or 8 days at room temperature or refrigerated at 4–8 °C (Fig. 1, left-hand side). This allowed determination of variation between batches stored over different times at different conditions. The technologist experiment used batches of 8 specimens with low, medium or high loads. On the same day they were made, 9 different technologists randomly picked

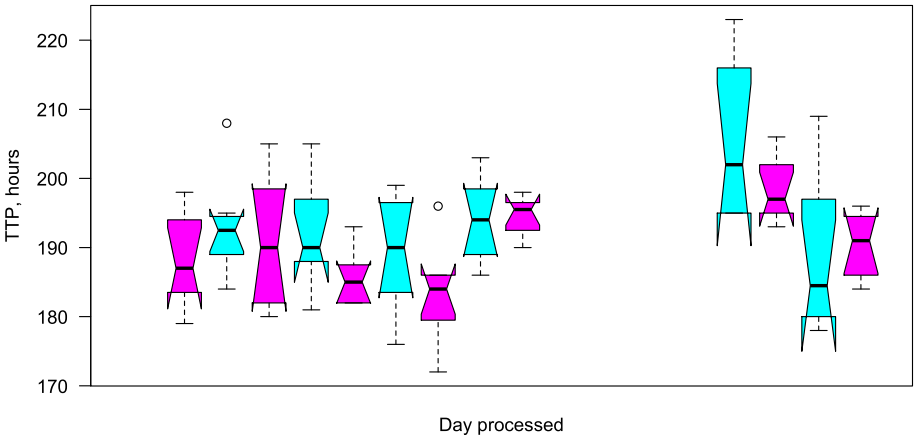

Fig. 3. TTP observed in storage and temperature experiment. Boxplots of observed TTP (n = 104) from batches of 4 high-load specimens stored at 4 °C (blue) or at room temperature (purple). Samples were processed by the same operator on the days indicated. Neither temperature nor day was a significant predictor of TTP. The within-day variation is fairly consistent and seems to be smaller than the between-day variation. The difference between the estimated day effects was not significant. A slight deviation seems to be noted for the cold-stored TTP on the second Monday. (TTP = Time to positivity). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

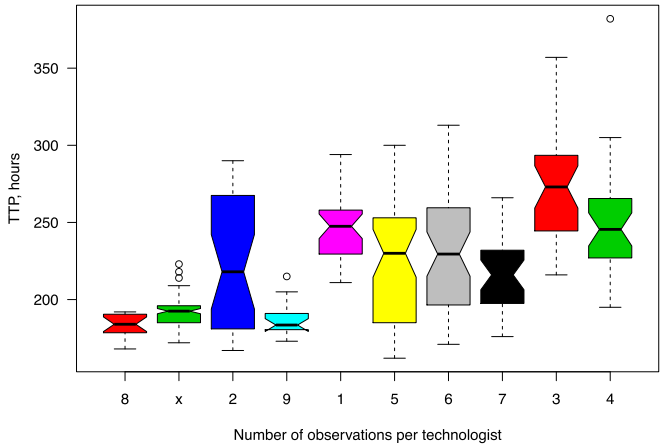

Fig. 4. TTP observed in technologist and load experiment. Notched boxplots of TTP observed from each technologist with the number of observations made. The variance in TTP between technologists is significant after taking the different loads into account. The third operator from the left (blue) seems to deviate not by a wider spread but by a larger median than his peers, indicating good precision but suboptimal accuracy. The very last operator (n = 104 observations) is from the storage experiment. The approximate 95% confidence interval for the median does not overlap any of the others. This makes the variance component for technologists significant. (TTP = Time to positivity). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 2
Estimated components of variance including all possible sources of variation.

Source of variation	n	Variance	Standard Deviation	
			Estimate	95% CI
Load	4	1545.00	39.30	19.1 to 84.1
Batch	31	7.14	2.67	0.0 to 6.3
Sample ID	196	8.99	3.00	0.0 to 6.9
Days stored	7	8.06	2.84	0.0 to 9.6
Temperature	2	0.00	0.00	0.0 to 20.8
Technologist	10	53.91	7.34	3.6 to 14.2
Residual	392	237.70	15.42	14.0 to 16.9

between 2 and 6 batches to process but remained blinded to the loads of the batch of samples they were processing (Fig. 1, right-hand side). This experiment tested the variation between the mycobacterial loads, batches of samples processed together and between technologists. Different batches of high-load samples were made for the experiments.

2.5. Statistical analysis

Data from the two experiments were combined for analysis, including all data for a full model. We first described the observed means of TTP obtained from batches of low, medium and two batches of high mycobacterial loads. We then constructed a single mixed-effects linear model. Variables for load (low, medium, high), number of days of delay to processing, storage temperature (refrigerated or room temperature) and the statistical interaction between days of delay and temperature were included as fixed effects. Technologist ID, batch ID and sample ID were specified as random effects. The variables were included as fixed or random effects or interactions as needed to estimate their effects or variances and adjust for possible correlations. The residual variation not explained by any other factor considered above is equal to the variation between duplicate cultures inoculated from the same re-suspended pellet. We used a stepwise procedure to, one-by-one, eliminate the factors that did not contribute significantly to the model, until the final linear mixed-effects model, containing only significant predictors, remained. We also created a variance component model, including all factors as random effects, in a linear mixed-effects model, to illustrate the significance as well as the other variances. All statistical analyses were done using functions from R and from R packages lme4 and lmerTest [10,11].

3. Results

3.1. Observed TTP

A total of 392 TTPs were determined from 196 specimens divided into 31 batches from 4 batches of artificial sputum with different loads. A total of 10 operators were involved in the study. None of the cultures was contaminated. In the storage experiment, 104 observations were made from one batch of high load sputum. In the technologist experiment, 96 observations were made from batches with low or medium and a second batch of high load. The distribution of the loads and the mean TTP found with high loads over time are illustrated in Fig. 2. The observed mean TTP values for all batches of loads are shown in Table 1. The loads were significantly different from each other and well spread with a very similar increase in mean TTP from high to medium and from medium to low load.

3.2. Storage and temperature experiment

The statistical interaction between number of days stored and temperature was not significant ($p = 0.867$), neither was temperature alone ($p = 0.540$) nor day of storage alone ($p = 0.537$) when the interaction term was removed. Inspection of the observed and the modelled daily mean TTP with confidence intervals confirmed that storage conditions had little impact on TTP and that there was no linear trend over time (Fig. 3). Mean TTP did not vary significantly from day to day.

3.3. Load and technologist experiment

Nine operators produced 288 observations (96 from each load) contributing 48 (2 operators), 32 (5 operators) or 16 (2 operators) each. We modelled TTP as a function of load and operator (Fig. 4). The operators were comparable bar one whose confidence interval was narrow indicating good precision but poor accuracy.

3.4. Variance components of the final model

The final model contained only load as fixed effect. After stepwise elimination, the only remaining random effects with significant variation (lower confidence limit of SD above zero) were technologist ID and residual variation (Table 2 and Fig. 5). The total variance is the variance attributable to all TTPs. Load is, by design, the largest source of

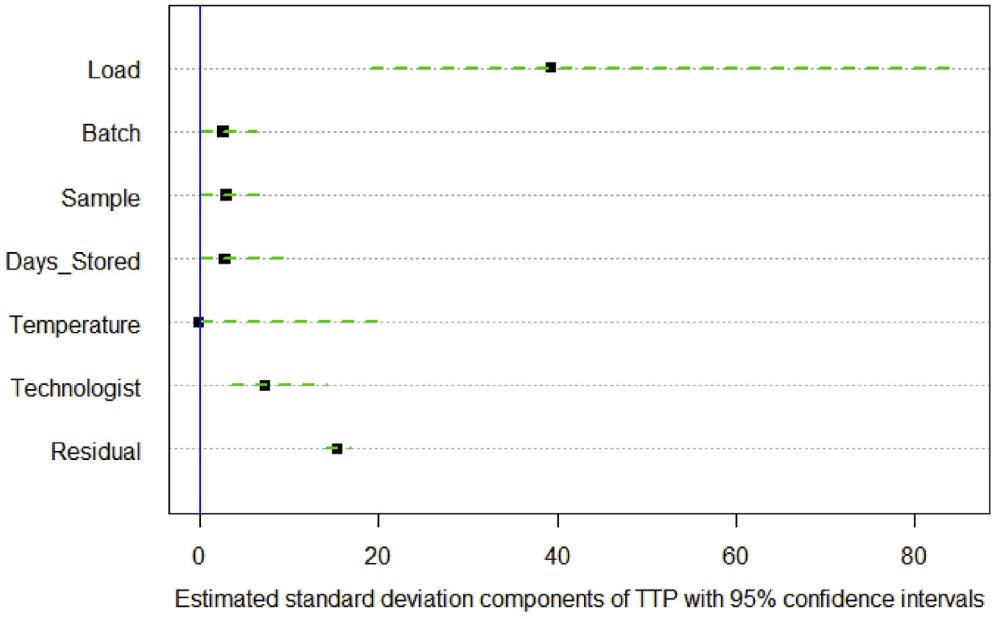

Fig. 5. Sources of variance of TTP from full model. The dots indicate the estimated standard deviations as indicated in Table 2. They graphically illustrate the contribution of each factor to the overall variation of TTP, relative to all the other factors in the model. Only factors whose green bars (95% confidence intervals of the SD) do not overlap with zero contribute significantly to the overall variation. Apart from load, which is the largest source of variation by design, significant variance comes only from technologists and the residual, within sample variance. (TTP = Time to positivity). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

variation. The manufacture of batches with the same loads and samples made from these batches do not significantly contribute to the overall variation. Neither do storage temperature nor time of storage up to 7 days. Apart from the residual, process-inherent variation the only remaining significant factor is the technologist processing the samples. The residual is defined as the variance not explained by any other predictor and can be termed sample variance that can no longer be separated into components. Interestingly, there was more variation within samples than between samples after all known causes of variation were removed or adjusted for.

4. Discussion

We found that artificial sputum specimens spiked with live *M. tuberculosis* were stable over at least 8 days, even without refrigeration, and returned a clinically relevant range of TTP results in liquid culture. The largest variance, apart from that introduced by spiking with different mycobacterial stock concentrations, was that of processing staff and the residual within-sample variation. Culture contamination was not observed. This indicates that spiked artificial sputum could serve as a tool for inter-laboratory comparison, method validation and proficiency assessments.

The ingredients for the artificial sputum are easily obtained and H37Rv, first described in 1972, is used in many laboratories as the reference strain of *M. tuberculosis* [9,12]. TTPs from batches made to contain low, medium and high loads of *M. tuberculosis* were well grouped together and had noticeably different mean TTPs. Two batches made to contain high load for separate experiments had very similar mean TTPs. Batches made and specimens within batches varied very little. There was also no significant variation between specimens processed immediately or stored for up to 8 days, either refrigerated or at room temperature. These findings indicate that consistent readings can be obtained from artificial sputum samples with different mycobacterial loads that are produced in-house using inexpensive standard ingredients. Artificial sputum might be a convenient matrix to house live mycobacteria before, during and after transport, and artificial sputum spiked with different loads of *M. tuberculosis* could serve as validation experiments within and between laboratories, as well as for staff proficiency testing.

From the perspective of those involved in processing, artificial specimens look very similar to routine specimens and are processed using the same standard protocols. The variance component analysis demonstrated that the largest variance, apart from that intended by spiking for different loads, is that between technologists processing and the residual, within-sample variance. Outliers due to individual processing techniques were seen in the experiment comparing loads and operators (Fig. 4), where one operator had similar precision but a different mean than his peers, indicating suboptimal accuracy. This points towards an anomaly in that operator's processing technique can be remedied by re-training. The other outlier was seen on the 7th day of the same person processing batches of the same high-load samples (Fig. 3). This was investigated and was eventually put down to “Monday Blues” which is a self-limiting condition that can affect even otherwise stable, mostly younger human beings.

The residual variance, i.e. that found between duplicate MGIT tubes incubated from the same re-suspended pellet by the same person at the same time, was larger than that between operators. In other words, for each operator, single TTPs appear not to be very precise yet the operators' averages are very similar suggesting accurate mean TTPs (Fig. 4). This within-sample variance is likely due to heterogeneous distribution of bacteria or clumps in the re-suspended pellet following decontamination and centrifugation. The magnitude of this residual variance might be specific to artificial sputum. Whether a mean TTP reading from two or more MGIT tubes incubated from the same clinical sample can increase accuracy needs to be investigated.

This report is, to our knowledge, the first describing the use of artificial sputum specimens spiked with the H37Rv strain of *M. tuberculosis* to determine the precision of multiple preparations of liquid cultures. Clinical sputum samples that also contain bacterial contaminants and host cells have been shown to do better when refrigerated [7,8,13], and storage of clinical samples frozen at −20 °C for > 100 days does not seem to critically affect the viability of *M. tuberculosis* [14]. This study, however, cannot make recommendations for clinical samples, and it was not within the scope to test other methods of quantifying mycobacterial sputum load such as sputum smear microscopy, colony-forming unit counting, Xpert MTB/RIF [15,16], the Mycobacterial load (MBL) assay [17] or the LAM detection assay [18] that have recently been proposed.

We conclude that artificial sputum samples spiked with *M. tuberculosis* are promising for validation experiments and staff proficiency testing as they are stable when stored for short durations, can be processed equivalent to routine clinical samples and, for the same mycobacterial load, the variation is a function of sample processing. The effect of transporting samples must be investigated to determine this method's practical suitability for between-laboratory comparisons or for making proficiency panels and modifications that increase standardization, reduce variation and simplify the analysis of results which could increase its practical value.

Conflicts of interest

The authors declare that they have no competing interests.

Author contributions

AHD and EK were responsible for the study design, data collection and interpretation, drafting and revision of the manuscript. SOF drafted the manuscript and assisted with data collection and interpretation. MNK was responsible for the study design and data collection. LvdM performed the statistical analyses.

Editorial assistance

The authors would like to thank Lauren Philips for editorial assistance during the drafting of this manuscript.

Funding

The study was funded by TASK Foundation.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.tube.2019.03.001.

References

[1] Wallis RS, Pai M, Menzies D, et al. Biomarkers and diagnostics for tuberculosis: progress, needs, and translation into practice. Lancet 2010;375:1920–37.
[2] Aber VR, Allen BW, Mitchison DA, Ayuma P, Edwards EA, Keyes AB. Quality control in tuberculosis bacteriology. 1. Laboratory studies on isolated positive cultures and the efficiency of direct smear examination. Tubercle 1980;61:123–33.
[3] Yamada H, Mitarai S, Aguiman L, Matsumoto H, Fujiki A. Preparation of mycobacteria-containing artificial sputum for TB panel testing and microscopy of sputum smears. Int J Tuberc Lung Dis Off J Int Union Tuberc Lung Dis 2006;10:899–905.
[4] Yamada H, Matsumoto H, Mitarai S, Fujiki A. [Stability for long-term storage and reproducibility of positivity in the panel test slide prepared with the polyacrylamide-based artificial sputum]. Kekkaku 2008;83:65–71.
[5] Yamada H, Mitarai S, Wahyunitisari MR, et al. Improved polyacrylamide-based artificial sputum with formalin-fixed tubercle bacilli for training of tuberculosis microscopists. J Clin Microbiol 2011;49:3604–9. https://doi.org/10.1128/JCM.00370-11.
[6] Paramasivan CN, Narayana AS, Prabhakar R, Rajagopal MS, Somasundaram PR, Tripathy SP. Effect of storage of sputum specimens at room temperature on smear and culture results. Tubercle 1983;64:119–24.
[7] Banda HT, Harries AD, Boeree MJ, Nyirenda TE, Banerjee A, Salaniponi FM.

Viability of stored sputum specimens for smear microscopy and culture. Int J Tuberc Lung Dis Off J Int Union Tuberc Lung Dis 2000;4:272–4.
[8] Kolwijck E, Mitchell M, Venter A, Friedrich SO, Dawson R, Diacon AH. Short-term storage does not affect the quantitative yield of Mycobacterium tuberculosis in sputum in early bactericidal activity studies. J Clin Microbiol 2013;51:1094–8. https://doi.org/10.1128/JCM.02751-12.
[9] Kubica GP, Kim TH, Dunbar FP. Designation of strain H37Rv as the neotype of Mycobacterium tuberculosis. Int J Syst Bacteriol 1972;22:99–106. https://doi.org/10.1099/00207713-22-2-99.
[10] Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Softw 2015;67:1–48. https://doi.org/10.18637/jss.v067.i01.
[11] Kuznetsova A, Brockhoff PB, Christensen RHB. lmerTest package: tests in linear mixed effects models. J Stat Softw 2017;82(13):1–26. https://doi.org/10.18637/jss.v082.i13http://doi.org/10.18637/jss.v082.i13.
[12] Köser CU, Niemann S, Summers DK, Archer JAC. Overview of errors in the reference sequence and annotation of Mycobacterium tuberculosis H37Rv, and variation amongst its isolates. Infect Genet Evol J Mol Epidemiol Evol Genet Infect Dis 2012;12:807–10. https://doi.org/10.1016/j.meegid.2011.06.011.
[13] Bhat J, Selvakumar N, Rao VG, Gopi PG, Yadav R, Wares DF. Yield of culture of Mycobacterium tuberculosis complex in sputum samples transported from tribal areas. Int J Tuberc Lung Dis Off J Int Union Tuberc Lung Dis 2011;15:478–82. https://doi.org/10.5588/ijtld.10.0234.
[14] Tessema B, Beer J, Emmrich F, Sack U, Rodloff AC. Rate of recovery of Mycobacterium tuberculosis from frozen acid-fast-bacillus smear-positive sputum samples subjected to long-term storage in Northwest Ethiopia. J Clin Microbiol 2011;49:2557–61. https://doi.org/10.1128/JCM.00059-11.
[15] Kayigire XA, Friedrich SO, Venter A, et al. Direct comparison of Xpert MTB/RIF assay with liquid and solid mycobacterial culture for quantification of early bactericidal activity. J Clin Microbiol 2013;51(6):1894–8. https://doi.org/10.1128/JCM.03290-12.
[16] Friedrich SO, Rachow A, Saathoff E, et al. Assessment of the sensitivity and specificity of Xpert MTB/RIF assay as an early sputum biomarker of response to tuberculosis treatment. Lancet Respir Med 2013;1:462–70.
[17] Honeyborne I, Mtafya B, Phillips PPJ, et al. The molecular bacterial load assay replaces solid culture for measuring early bactericidal response to antituberculosis treatment. Land GA. J Clin Microbiol 2014;52(8):3064–7. https://doi.org/10.1128/JCM.01128-14.
[18] Kawasaki M, Echiverri C, Raymond L, Cadena E, Reside E, Gler MT, Oda T, Ito R, Higashiyama R, Katsuragi K, Liu Y. Lipoarabinomannan (LAM) in sputum to detect bacterial load and treatment response in pulmonary TB patients: Analytic validation and evaluation in two cohorts. PLOS Medicine 2019. accepted for publication.