

https://dspace.org/

Estimation of therapeutic threshold for tuberculosis using adapted nominal group technique and clinical vignettes in clinical and community settings in Southern Africa

Alfred Kipyegon Keter; Alastair Van Heerden; Tom Decroo; Tom Boyles; Shannon Bosman ...

https://hdl.handle.net/10539/48677
WOS

Downloaded from a DSpace repository. https://dspace.org/

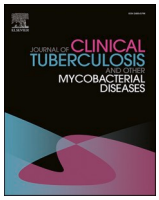
Journal of Clinical Tuberculosis and Other Mycobacterial Diseases 40 (2025) 100529



Contents lists available at ScienceDirect

Journal of Clinical Tuberculosis and Other Mycobacterial Diseases

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



Estimation of therapeutic threshold for tuberculosis using adapted nominal group technique and clinical vignettes in clinical and community settings in Southern Africa

Alfred Kipyegon Keter [a,b,c,d,*], Alastair Van Heerden [b,e], Tom Decroo [a], Tom Boyles [f,g,h], Shannon Bosman [b], Thandanani Madonsela [b], Lindani Innocent Msimango [b], Lenika Naiken [b], Carlos Kiyan [a], Mashaete Kamele [i], Irene Ayakaka [i], Klaus Reither [j,k], Bart Karl Mario Jacobs [a], Lutgarde Lynen [a]

[a] Institute of Tropical Medicine, Nationalestraat 155, 2000 Antwerp, Belgium
[b] Center for Community Based Research, Human Sciences Research Council, Pietermaritzburg, South Africa
[c] Ghent University, Krijgslaan 281, Building S9,9000 Ghent, Belgium
[d] The Aga Khan University, 3rd Parklands Avenue, Off Limuru Road, Nairobi, Kenya
[e] SAMRC/WITS Developmental Pathways for Health Research Unit, Department of Paediatrics, School of Clinical Medicine, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, Gauteng, South Africa
[f] Clinical HIV Research Unit, Helen Joseph Hospital, Johannesburg 2092, South Africa
[g] Right to Care (NPC) Centurion, South Africa
[h] London School of Hygiene and Tropical Medicine London, London WC1E 7HT, UK
[i] SolidarMed Lesotho, Premium House, 224 Kingsway Street, Maseru, Lesotho
[j] Swiss Tropical and Public Health Institute, Kreuzstrasse 2, 4123 Allschwil, Switzerland
[k] University of Basel, Basel, Switzerland

ARTICLE INFO

Keywords:
Adapted nominal group technique
Therapeutic threshold
Clinical decision-making
Clinical vignettes
Tuberculosis
Diagnostic uncertainty

ABSTRACT

Background: When confronted with diagnostic uncertainty and a decision on whether to start treatment or not, clinicians consider the potential harm and benefit of offering versus withholding treatment. Treatment can be offered if the probability of tuberculosis (TB) in the patient is above the "therapeutic threshold" (ThT): the probability of disease at which the expected utility of treating and not treating is the same. We estimated ThT for TB in clinical and community settings in Southern Africa using two methods: an adapted nominal group technique (aNGT), and decisions made based on clinical vignettes (CVs).

Methods: We enrolled health professionals involved in the routine management of TB patients in South Africa and Lesotho. The participants elicited, discussed and refined the harms of false positive (FP) and false negative (FN) treatment decisions for stable ambulatory patients in the clinical and community settings. They weighed all harms according to their importance in treatment decisions by distributing 100 points. ThT, calculated as the sum of the weights of the harms of the FP decision divided by the total weight, was estimated using a hierarchical Beta regression model. For the CVs, participants were presented with ten hypothetical TB cases in each setting and asked to indicate whether they would offer TB treatment or not. ThT was estimated using the generalized linear model for binary outcomes.

Results: We enrolled 138 health professionals (aNGT: 123, CVs: 130 and 115 in both). Using aNGT, the overall ThT was 37.7% (95% credible intervals (95% CrI): 35.8–39.8) and 38.2% (95% CrI: 35.9–40.6) in the clinical and community settings, respectively. Compared to aNGT, CVs produced a significantly lower estimate in the clinical setting (27.7%; 95% CrI: 23.8–31.3) but similar in the community setting (37.7%; 95% CrI: 33.1–41.7). We did not find significant differences across the subgroups defined by the measured covariates.

Conclusion: The aNGT produced a reliable estimate of ThT. The difference in the estimates of ThT between the aNGT and CVs may have a limited impact on clinical decisions. Factors influencing ThT and the acceptability of results by healthcare workers will be explored in focus group discussions and in-depth interviews.

* Corresponding author.
E-mail addresses: alfred.keter@aku.edu (A.K. Keter), bkjacobs@itg.be (B.K. Mario Jacobs).

https://doi.org/10.1016/j.jctube.2025.100529

Available online 24 April 2025
2405-5794/© 2025 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

1. Introduction

Globally, tuberculosis (TB) is the major cause of morbidity and mortality.[1] Approximately 10 million new TB infections and more than 1.5 million deaths due to TB occurred in 2020.[22] Due to the inadequate diagnostic accuracy of the available tests and the regular disagreement between what is suggested by clinical presentation and implied by the tests, the decision to treat patients for tuberculosis is often taken under conditions of uncertainty.[3,4] Before the use of Xpert® MTB/RIF, a rapid molecular test, half of the patients in high-burden settings started TB treatment empirically, i.e., without bacteriological confirmation.[5] With the roll-out of Xpert® MTB/RIF, this proportion has declined. Nonetheless, in 2021 only 63 % of pulmonary TB cases were bacteriologically confirmed globally.[2].

When confronted with diagnostic uncertainty and a decision on whether or not to start treatment, the clinician should consider the potential harm and benefit of offering versus withholding treatment. Treatment decisions can be based on "therapeutic threshold" (ThT): the probability of disease at which the expected utility (foregone harm) of treating and not treating is the same.[6] It considers the net harm of false-positive (FP) and false-negative (FN), and the net benefit of true-positive (TP) and true-negative (TN) treatment decisions.[6,7] ThT is lower when the patient is in a worse or unstable clinical condition because the consequences of withholding treatment are greater, and higher when the treatment to be offered has greater side effects, is expensive or accompanied by psychosocial harm e.g., stigma or emotional distress.[8] Clinicians usually implicitly weigh these harms against their perceived probability that the patient has TB and offer treatment only when the expected harm of not treating is higher than the expected harm of treating. In other words, empiric treatment initiation is considered only when the perceived probability of TB in an individual under observation exceeds the ThT.

The available methods for estimating the ThT are broadly classified into prescriptive and descriptive.[9] Prescriptive methods quantify harms and benefits based on theoretical models.[3,6,7,9] These are typically extensions of *Expected Utility Theory* (EUT) which focuses on objective quantifiable harms (mortality, cost, morbidity, etc.).[10] Other prescriptive methods also try to quantify subjective sources of harm (regret, fear of retaliation, etc.).[4,11] Descriptive methods, on the other hand, derive the threshold from observing clinical practice or a surrogate thereof.[8,12–14] Thus, the ThT is impacted by unquantifiable cognitive and emotional factors. [9].

However, no gold standard method currently exists. Such a method should be easy and rapid to perform and give a reliable, reproducible, and accurate estimate of the ThT. Ideally, it should also allow room for reflection and feedback by a representative group of local physicians and healthcare workers to enhance the acceptability of and compliance with the guidelines during implementation.

Previous studies that use harm and/or regret-based approaches to estimate the ThT for TB produce different thresholds in different countries, varying from 20 % to over 50 %, and it is difficult to set one universal threshold.[3,4,8] In fact, local factors are expected to influence the threshold, meaning any method would need to be applied in a local setting and repeated when associated conditions change over time. Additionally, the current estimates and discussions of ThT apply to the clinical setting only. However, we cannot assume that physicians will give similar weights to harm in a setting of community-based active case finding.

We estimate the ThT using two methods: an adapted Nominal Group Technique (aNGT) and decisions based on clinical vignettes (CVs), in the clinical (passive TB case-finding) and community (active TB case-finding) settings in Lesotho and South Africa. Clinical vignettes (descriptive method) are hypothetical clinical scenarios representing TB patients with different probabilities of TB. Participants are asked whether they would want to treat a patient presented in a vignette for pulmonary TB or not.[8,12,13] Nevertheless, CVs are not trivial to analyse and in their current form and application give no room for clarification or discussion of individual differences in decisions. The aNGT (prescriptive method) is based on a consensus-building exercise that allows clinicians to share, discuss and reflect upon the harms and benefits of treating versus not treating a patient.[12,13,15–17] This method has been piloted for a variety of infectious diseases in low-resource settings but has not been applied yet to the concept of therapeutic threshold in TB. We compared the estimates derived using the aNGT to those based on CVs.

2. Methods

2.1. Study design

This was a mixed-methods study involving both the qualitative and quantitative components: aNGT (mixed-methods) and CVs (quantitative). Participant observation was conducted during the aNGT sessions. In-depth interviews and focus group discussions were also completed. In this paper, we only present the quantitative findings.

Adapted Nominal Group Technique: The aNGT is a highly structured interactive group activity comprising five stages, which can be conducted face-to-face or online. The first stage introduces clinical decision-making and the relevant problem. The second stage involves the generation of ideas resulting in a list of statements on the harms of FP and FN clinical decisions. This is transmitted electronically through Mentimeter (https://www.mentimeter.com/), a web-based platform for transmitting votes and comments in real-time. The elicited harms are obscured until everyone has submitted their statements. The transmitted harms are then revealed and all harms are discussed and refined to ensure a non-overlapping complete list of harms. In the third stage, the participants are invited to rate the harms on a six-point Likert scale of 0–5 based on their importance in influencing treatment decisions, with 0 suggesting the harm is not a factor to consider during treatment decisions and 5 suggesting the factor is an extremely important factor driving treatment decisions (Appendix A, Table A.1). At this stage, the participants are allowed to discuss and reflect on the importance of the harms. The discussion is important to ensure that everyone understands each of the harms and agrees on their importance. In classic NGT, options would then be eliminated. However, in the aNGT all harms were retained unless everyone agreed that a given harm should not be considered in treatment decisions. In the fourth stage, the participants were asked to weigh each harm according to their importance in treatment decisions by distributing a total of 100 points (equivalent to 100 percentage points) to all the harms.[15,16] The relative weights of the harms were revealed to the group followed by a brief discussion to establish if the weights assigned to the harms were realistic. After this, the participants were allowed to revise their weightings. The process was performed twice, the first assumed the clinical setting involving stable ambulatory patients. The second assumed a community setting. A detailed description of these stages and the activities, including sample captions of the Mentimeter output, has been described by Jacobs et al. [17].

Clinical vignettes: We developed different web-based vignettes of adult patients being evaluated for pulmonary tuberculosis in clinical and community settings. For the clinical setting, we considered stable ambulatory patients. For the community setting, the vignettes considered hypothetical stable participants being screened for TB. A set of ten hypothetical scenarios for the clinical and community settings with varying probabilities of TB were provided (Supplementary material, TheraTUB Clinical vignettes). The participants were asked to indicate whether they would treat or not, the rationale, and what they think is the probability of TB in each of the cases. The probability of TB for each of the CVs presented to the participants was estimated by a separate group of TB experts based on expert knowledge. They were also estimated using a prediction model based on latent class analysis.[18] The two estimates were harmonised following a discussion with a separate

group of experts in TB to obtain one estimate.

2.2. Study population, setting, study size and procedures

Study population: Physicians, nurses and program managers involved in the routine management of TB patients in South Africa and Lesotho.

Sample size: No formal sample size calculation was conducted. An estimate of 200 participants was proposed for the CVs based on two studies that produced data with good precision with a sample size of 160–200.[8,12,13] A sample size of 5–10 participants was proposed for each aNGT session based on a pilot study which showed that the sessions with such a number work well especially when online discussions are involved.[17].

Study procedure: Emails were sent to key individuals in the field of infectious diseases, and through relevant listserv mailing lists (in compliance with the General Data Protection Regulation (GDPR) and Protection of Personal Information Act (POPIA) through a third party) in South Africa and Lesotho with the request to forward it to potentially interested individuals. A link with further information was provided for interested subjects to read. They could click the link to agree to participate. Recipients of the emails were encouraged to disseminate further the messages through their networks and closed professional WhatsApp groups. Up to three reminders were sent. An announcement was also made at a national TB conference in South Africa and a link was shared.

Participants in the aNGT sessions had their responses captured using Mentimeter, a web-based survey platform. Upon providing informed consent, the participants in the CVs were invited to provide their responses to the vignettes using Formsite®, a web-based survey platform.

A participation fee of ZAR800 (≈US $50) for CVs and ZAR1400 ((≈US $80) for aNGT was given to each participant for the time spent in the study.

2.3. Inclusion and exclusion criteria

Participants were eligible if their work involved making clinical decisions on the initiation of TB treatment, including indirectly (e.g., through policy decisions as program managers) in Lesotho or South Africa, had a (para-)medical background, and provided informed consent.

Participants who were not allowed to make a diagnosis of pulmonary tuberculosis without bacteriological confirmation (e.g., without smear microscopy or Xpert) were excluded.

2.4. Ethical and regulatory review

This study was approved by the Institutional Review Board of the Institute of Tropical Medicine (ITM), the National Ethics Committee (NH-REC) of Lesotho, the Human Sciences Research Council Research Ethics Committee, and the Provincial Department of Health in South Africa. The study was carried out according to the principles stated in the Declaration of Helsinki, all applicable regulations and according to the most recent Good Clinical Practice (GCP) guidelines.

All participants provided signed informed consent through Formsite® before entering the study.

Participants could separately agree or disagree to participate in each of the study components.

2.5. Statistical analysis

The participant and session characteristics were summarized using mean and standard deviation (SD) for continuous variables, and frequencies and percentages for categorical variables.

Estimation of the ThT using aNGT was based on the weighted harms. The sum of the weights of the harms of FP clinical decision divided by the total weights yields an estimate of the ThT for each participant. We analysed this outcome using the random intercept Bayesian Beta regression model with demographic characteristics (age and sex), study site, type of aNGT session (online vs. in-person), order of discussion of the harms (FP harms first or FN harms), order of participation in aNGT and CVs, profession and years of work experience included as fixed effects. The ThT based on the CVs was estimated by adapting the method of Plasencia et al. 1992.[19] Estimation of ThT was done by inverting the probit link of the random intercept Bayesian probit regression model when the modelled probability of treating TB was 50 %. The expert elicited probability of TB was included as the covariate of interest. In this analysis, we included demographic characteristics (age and sex), study site, order of participation in aNGT and CVs, profession and years of work experience as fixed effects. In both analyses, we standardized the estimates to the total population. The random intercept was motivated by the differences between the different sessions in aNGT and between different participants in CVs which may explain a larger portion of the observed variability. The unknown model parameters were assigned priors from the Gaussian distribution. The estimates of ThT were compared by calculating the differences between the posterior distributions of the ThT and assessing whether the 95 % credible intervals (95 % CrI) of the distribution of the differences covered zero.

Based on CVs, we conducted a sensitivity analysis whereby we included the participant-elicited probability as the covariate of interest. We adjusted for the same covariates and estimated the ThT by inverting the probit link of the random intercept Bayesian probit regression model as above. This was done for all the participants and among the doctors only.

Convergence in the models was assessed by running three chains, each with 50,000 Markov Monte Carlo (MCMC) iterations. The first 25,000 iterations were discarded as burn-in. Gelman-Rubin statistic and trace plots were used to evaluate model convergence. Inferences were based on the median of the posterior distribution with the corresponding 95 % CrI. Bayesian inference was done using the Just Another Gibbs Sampler (JAGS) function implemented in the *R2jags* package in R version 4.2.2.[20,21].

3. Results

Overall, 138 participants were enrolled. We enrolled 123 participants in 17 aNGT sessions, 130 in the CVs, and 115 in both the aNGT sessions and CVs (Table A.3 in the Appendix).

The inclusion criteria effectively restricted the study population to doctors in Lesotho, resulting in a considerably different participant population by design. Overall, 42.6 % of all the participants were doctors. Three-quarters of them had a degree in medicine as their highest qualification. Overall, one-third of the participants were male with an imbalance between the two sites. Overall, the mean age was 40.5 (SD: 9.5) years with participants from South Africa having older age compared to those from Lesotho. (Table 1).

A higher proportion of participants from South Africa had a longer work experience compared to their counterparts in Lesotho. At the time of the study, two-thirds of the participants were working in a district, secondary or tertiary hospital. In the month preceding the study, the majority of the participants from South Africa had initiated treatment in at least five patients with TB compared to Lesotho (Table 1).

The subset of participants that completed both the aNGT and CVs was similar to the whole sample of participants that completed CVs (Table A.4 in the Appendix).

4. Estimation of therapeutic threshold

4.1. Adapted nominal group technique

Overall, the estimated ThT was 37.7 % (95 % CL: 35.8, 39.8) in the clinical setting and 38.2 % (95 % CL: 35.9, 40.6) in the community setting.

Table 1
Demographic characteristics of the participants in both the aNGT and clinical vignettes.

		Study site		
		Lesotho	South Africa	Total
Characteristics	**n**	**n = 38**	**n = 77**	**n = 115**
Age (years)				
Mean (SD)	114	33.6 (4.1)	44.0 (9.4)	40.5 (9.5)
Range (Min. – Max.)		26–42	27–66	26–66
Male, n (%)	115	19 (50.0 %)	18 (23.4 %)	37 (32.2 %)
Years of experience in TB treatment, n (%)				
< 1 Year		2 (5.3 %)	3 (3.9 %)	5 (4.3 %)
1–5 Years	115	24 (63.2 %)	20 (26.0 %)	44 (38.3 %)
6–10 Years		9 (23.7 %)	21 (27.3 %)	30 (26.1 %)
> 10 Years		3 (7.9 %)	33 (42.9 %)	36 (31.3 %)
Profile, n (%)				
Doctors		38 (100 %)	11 (14.3 %)	49 (42.6 %)
Nurses		0 (0.0 %)	45 (58.4 %)	46 (39.7 %)
Program managers	115	0 (0.0 %)	15 (19.5 %)	15 (13.0 %)
Others		0 (0.0 %)	6 (7.8 %)	6 (5.2 %)
If a doctor, highest qualification, n (%)				
MBCHB	49	31 (81.6 %)	7 (63.6 %)	38 (77.6 %)
Masters		4 (10.5 %)	2 (18.2 %)	6 (12.2 %)
Post-graduate Diploma		3 (7.9 %)	2 (18.2 %)	5 (10.2 %)
Health services currently working in				
District, secondary or tertiary hospital		34 (89.5 %)	43 (55.8 %)	77 (67.0 %)
Clinic or primary healthcare	115	4 (10.5 %)	25 (32.5 %)	29 (25.2 %)
Others (NGOs, corporate, research, college, correctional services, retired)		0 (0.0 %)	9 (11.7 %)	9 (7.8 %)
Number of patients put on TB treatment last month, n (%)				
0		5 (13.2 %)	26 (33.8 %)	31 (27.0 %)
1 – 4		20 (52.6 %)	22 (28.6 %)	42 (36.5 %)
5 – 9	115	11 (28.9 %)	13 (16.9 %)	24 (20.9 %)
10 – 19		1 (2.6 %)	9 (11.7 %)	10 (8.7 %)
20 – 30		0 (0.0 %)	3 (3.9 %)	3 (2.6 %)
> 30		1 (2.6 %)	4 (5.2 %)	5 (4.3 %)

SD-Standard deviation, Min.-Minimum, Max.-Maximum.

Table 2
Estimates of therapeutic threshold and the corresponding 95% credible intervals (95% CrI) based on aNGT.

	Clinical setting		Community setting	
	Unadjusted	Adjusted[†]	Unadjusted	Adjusted[†]
Characteristic	Median (95 % CrI) N = 117	Median (95 % CrI) N = 117	Median (95 % CrI) N = 115	Median (95 % CrI) N = 115
Overall	38.0 (36.0, 40.0)	37.7 (35.8, 39.8)	37.9 (35.6, 40.2)	38.2 (35.9, 40.6)
Site				
Lesotho	30.0 (23.2, 37.7)	31.5 (24.1, 39.9)	34.0 (27.2, 41.5)	35.4 (27.7, 43.9)
South Africa	40.2 (33.3, 47.2)	39.8 (33.4, 46.6)	38.8 (32.8, 45.0)	39.2 (33.9, 44.8)
Type of session				
In-person	37.0 (28.1, 46.0)	35.5 (28.5, 43.4)	40.1 (32.7, 47.5)	38.6 (32.3, 45.7)
Online	35.4 (28.4, 42.6)	38.5 (31.9, 45.6)	34.9 (29.4, 40.8)	37.3 (31.8, 42.9)
FP harm first				
No	32.7 (25.6, 40.8)	33.0 (26.4, 40.5)	31.8 (26.6, 37.4)	32.3 (26.3, 39.1)
Yes	39.3 (32.4, 46.2)	40.1 (34.0, 46.5)	42.1 (37.7, 46.5)	42.1 (37.2, 47.5)
Clinical vignettes first				
No	35.0 (28.7, 41.3)	37.9 (32.6, 43.9)	36.5 (30.9, 41.8)	37.0 (32.5, 41.7)
Yes	38.3 (29.8, 47.9)	34.7 (27.5, 43.1)	37.6 (29.7, 47.3)	40.6 (32.9, 48.9)
Profession				
Nurse	39.9 (33.1, 47.2)	39.6 (33.8, 45.9)	37.7 (31.1, 44.8)	38.1 (33.0, 43.6)
Doctor	32.2 (26.8, 37.9)	34.2 (28.2, 41.0)	36.0 (30.1, 42.5)	38.6 (33.2, 44.4)
Others [‡]	38.1 (31.4, 45.1)	38.0 (32.1, 44.6)	36.0 (29.4, 42.9)	36.0 (30.4, 42.1)
Years of experience				
< 5	35.0 (29.6, 41.3)	37.5 (32.4, 42.5)	33.5 (28.9, 38.5)	35.5 (30.9, 40.1)
6–10	34.1 (28.0, 41.0)	35.6 (30.2, 42.0)	37.5 (31.6, 44.0)	38.7 (32.8, 45.2)
> 10	38.5 (31.4, 45.7)	37.8 (32.3, 43.9)	41.8 (36.1, 47.7)	40.6 (35.3, 45.9)
Age (years)				
< 30	33.5 (26.9, 40.9)	–	32.4 (24.5, 41.4)	–
≥ 30	36.2 (30.9, 41.8)	–	37.4 (32.6, 42.3)	–
Sex				
Female	36.3 (31.0, 42.0)	–	36.1 (31.2, 41.2)	–
Male	35.8 (30.5, 41.3)	–	38.3 (33.4, 43.7)	–

The overall estimates of ThT in the clinical and community settings were similar, difference: −0.5% (95% CrI: −3.5, 2.7) (Table 2 and Fig. 1, first row).

[†] Adjusted for the study site, type of session, order of discussion of harms, order of participation in clinical vignettes, profession and years of experience and standardized to the total population.

[‡] Includes program manager, case manager, clinical associate and pharmacist, "-" Not included in the adjusted model, CrI – Credible Interval.

4.2. Clinical vignettes

The overall estimate of ThT was 27.7 % (95 % CrI: 23.8, 31.3) in the clinical setting and 37.7 (95 % CrI: 33.1, 41.7) in the community setting (Table 3). There was a significant difference between the ThT in the community and the clinical settings, difference: −10.0 % (95 % CrI: −15.5, −4.1) (Fig. 1, second row).

4.3. Adapted nominal group technique versus clinical vignettes

We compared the overall estimates of ThT (adjusted for measured covariates and standardized to the whole population) based on the individuals who completed both the aNGT and the CVs (Table 1).

The estimate of ThT based on the aNGT was higher than the estimate based on the CVs in the clinical setting (Fig. 2, first row). However, there was no difference in the estimate of ThT between the aNGT and CVs in the community setting (Fig. 2, second row).

When comparing results from the doctors only, the estimates of ThT in the clinical and the community settings were similar based on both the aNGT and CVs. Similarly, we did not find a significant difference between the aNGT and CVs in the clinical and community settings.

5. Discussion

We have, for the first time, used the aNGT to estimate the ThT in TB among stable ambulatory patients in a setting where Xpert is often available but chest X-ray and culture are not easily available. We compared the results based on aNGT to those based on CVs.

Based on the aNGT, results demonstrate that when considering stable ambulatory patients for whom a treatment decision is unclear, healthcare workers are willing to accept up to 3 FP cases to avoid 2 FN cases in both the clinical and community settings: i.e., a probability of TB of 40 % is just above the ThT and therefore sufficient to initiate treatment. The findings from CVs were similar. This estimate is higher than the one based on the magnitude and subjective weight of mortality and morbidity due to both the disease and the treatment and the risk and cost of the treatment (11.9 %), and prescriptively based on the literature data (2.7 %) for smear-negative pulmonary TB patients.[3] This illustrates that ThT estimation based on quantifiable factors only is insufficient. However, our findings matched the 36.5 % intuitive estimate of ThT for smear-negative pulmonary TB by the clinicians in the Rwanda study.[3] Our findings also fall within the range of the estimated ThTs for unstable (29.1 %) and stable (74.5 %) HIV-positive patients suspected of having pulmonary TB.[8].

The estimates based on the aNGT in the clinical and community settings did not significantly vary across the study sites and participant characteristics. The findings based on the aNGT largely agreed with those based on CVs. Doctors consistently agreed they are willing to accept 2 FP cases to avoid 1 FN case in both aNGT and CVs in both settings. The estimates of the ThT from other professions were more variable depending on the method and setting. These findings agree with those reported by Basinga et al. 2007.[3] However, Sreeramareddy et al. 2014 established that young physicians considered it less harmful to give treatment to a non-diseased case than to forgo treatment from a diseased case.[4] Also, Moreira et al. 2009 and Sreeramareddy et al. 2014 established that participants from different settings tended to variably intuitively weigh the harms of FP and FN clinical decisions in different settings.[4,11].

The elicitation and discussion of the different harms of FP and FN clinical decisions among the participants in the aNGT sessions, though in different study sites, might have mitigated the differences because the participants are expected to include more possible harms as opposed to when an individual is solely considering treatment. Nevertheless, both the number and nature of harms included varied considerably between aNGT sessions.

Though not compelling enough, the estimate of ThT based on the CVs was higher in the community setting than in the clinical setting. This is likely because the community setting involves 'healthy' individuals who are not at risk of severe harm, and withholding treatment at one moment in time does not imply that treatment is never offered. This reflects that the healthcare workers are implicitly giving more weight to the harm of commission (giving treatment to a non-diseased case) in the community setting compared to the clinical setting.

The overall estimates of ThT in the clinical setting based on the aNGT and CVs were significantly different. This difference is in line with similar comparisons of other decision thresholds.[17] The low estimate of ThT in the clinical setting based on CVs tended in the direction of the estimate of ThT for unstable HIV-positive patients suspected of pulmonary TB in South Africa.[8] However, our hypothetical patients were typically stable and ambulatory. Overall, healthcare workers indicated very often that they were willing to initiate treatment in case of doubt. This implies that they give considerably more weight to the harm of FN than FP clinical decisions. However, not all empiric decisions are made purely based on probability. Cognitive biases, heuristics, 'rules of thumb', and personality (e.g., risk aversion) may influence a physician's decision leading to diagnostic inaccuracies and medical errors.[22,23]. Most importantly, physicians are very different in the way they deal with harm and are risk-averse.[24].

In the CVs, the participants overestimated the probability of TB for the hypothetical cases. As a result, the estimates of ThT in the clinical and community settings were higher than those based on the expert-elicited probability of TB. It is unclear whether participants truly gave their best estimate of the probability of TB, or may have subconsciously

Table 3
Estimates of therapeutic threshold and the corresponding 95% credible intervals (95% CrI) based on clinical vignettes.

	Clinical setting		Community setting	
	Unadjusted	Adjusted†	Unadjusted	Adjusted†
Characteristic	Median (95 % CrI) N = 130	Median (95 % CrI) N = 130	Median (95 % CrI) N = 130	Median (95 % CrI) N = 130
Overall	27.6 (23.5 31.5)	27.7 (23.8, 31.3)	37.6 (32.9, 41.8)	37.7 (33.1, 41.7)
Site				
Lesotho	29.9 (23.1, 36.4)	30.2 (24.9, 34.6)	38.0 (32.9, 41.8)	41.2 (36.5, 45.0)
South Africa	26.3 (21.2, 31.3)	26.5 (21.1, 31.2)	37.4 (31.6, 42.6)	35.3 (27.3, 41.9)
Clinical vignettes first				
No	25.9 (21.0, 30.3)	26.5 (22.0, 30.4)	35.2 (29.6, 40.0)	37.2 (31.7, 41.9)
Yes	33.4 (25.3, 41.4)	33.8 (23.2, 44.1)	45.5 (37.0, 53.7)	45.1 (33.0, 55.5)
Profession				
Nurse	24.5 (18.0, 30.6)	24.4 (16.9, 30.7)	37.4 (30.3, 44.0)	34.0 (21.6, 43.6)
Doctor	34.2 (28.4, 39.8)	34.4 (29.6, 38.7)	40.9 (34.4, 47.2)	42.5 (39.0, 45.7)
Others ‡	17.5 (7.5, 26.8)	20.4 (8.7, 29.2)	29.7 (18.4, 39.6)	25.2 (0.1, 39.0)
Years of experience				
< 5	28.5 (22.4, 34.3)	28.4 (21.9, 34.4)	37.1 (30.4, 41.9)	38.5 (32.6, 43.2)
6–10	22.1 (13.7, 30.1)	22.8 (15.1, 28.2)	34.1 (25.2, 42.5)	34.4 (25.7, 40.2)
> 10	30.9 (23.5, 38.1)	31.3 (23.5, 38.6)	41.4 (33.6, 48.8)	39.4 (26.2, 49.8)
Age (years)				
< 30	33.3 (22.0, 44.6)	–	45.3 (33.7, 56.8)	–
≥ 30	26.9 (22.2, 31.2)	–	36.5 (31.4, 41.1)	–
Sex				
Female	25.7 (20.6, 30.6)	–	35.6 (29.7, 40.7)	–
Male	31.3 (24.5, 38.1)	–	41.4 (33.9, 48.4)	–

† Adjusted for study site, order of participation in the aNGT and clinical vignettes, profession, and years of experience and standardizing for the whole population

‡ Includes program manager, case manager, clinical associate and pharmacist. The 95% CrI for the adjusted estimates in the community setting is very wide because of the small sample size for this group, "-" Not included in the adjusted model, CrI – Credible Interval.

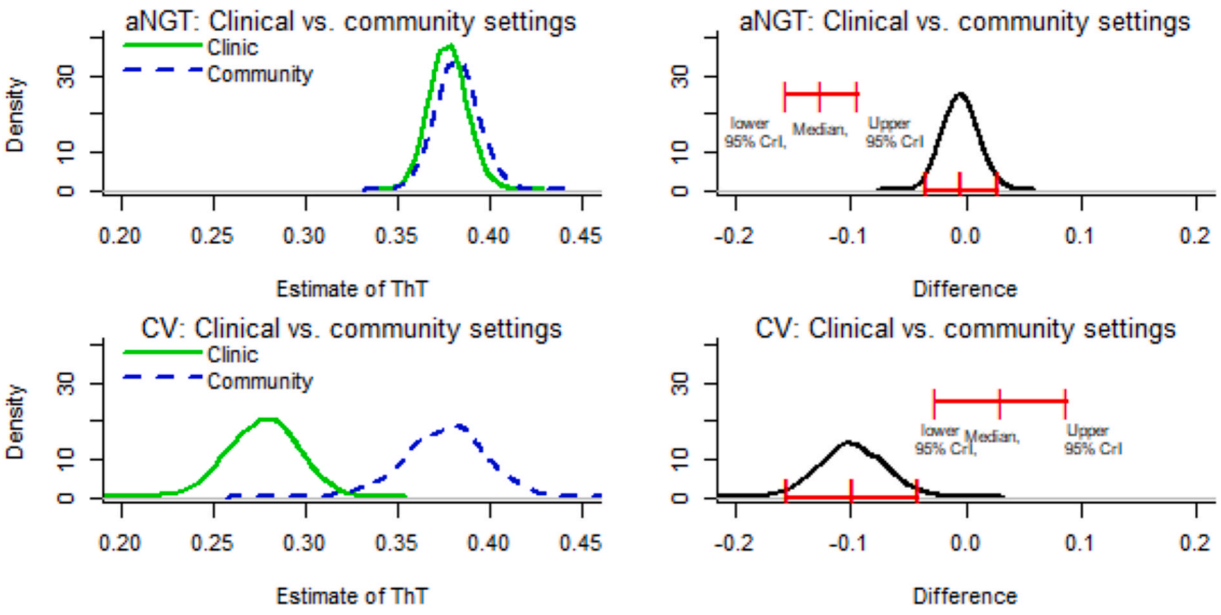


Fig. 1. Comparison of the overall standardized estimates of ThT in the clinical setting and community setting for the participants in the adapted nominal group technique (first row) and the participants in the clinical vignettes (second row).

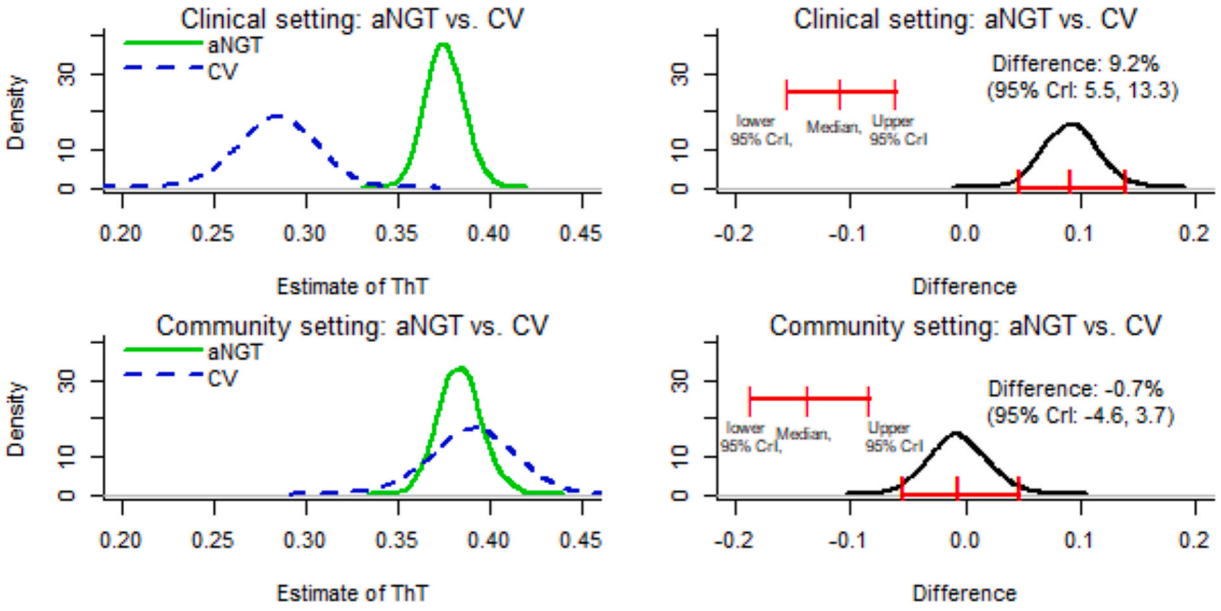


Fig. 2. Comparison of the overall standardized estimates of ThT based on the adapted nominal group technique (aNGT) and clinical vignettes (CVs) in the clinical (first row) and community settings (second row) among the individuals who completed both the aNGT and the CVs.

given higher probabilities to "justify" wanting to treat. Unlike the estimates of ThT based on the expert-elicited probability of TB that were different between the clinical and community settings, the participant-elicited probability of TB produced similar estimates of ThT in the clinical and community settings. The estimates changed slightly when restricted to professional doctors only (Table A.5 in the Appendix). These estimates can be interpreted as providing an upper bound of ThT in the clinical and community setting based on CVs.

6. Limitations

The aNGT is time-consuming and participants who are involved in the routine management of TB patients may not easily get up to 3 h to participate in the entire aNGT session. This might have also negatively affected the estimate of ThT for the community setting, which was always the last activity in every aNGT session.

Conducting online aNGT exercises is challenging for the participants. Firstly, they may often get distracted, resulting in them missing important information. All online exercises are affected by bad internet connectivity or power outages in places where the electricity supply is not stable. The latter was often experienced in South Africa due to load shedding that was affecting the whole country. On the other hand, Lesotho had poor internet connectivity during the time we implemented the online aNGT sessions. All these challenges affect the quality of the exercise and the data obtained.

In the aNGT, the participants were allowed to elicit, discuss and rate the harms of FP and FN treatment decisions before weighting them to obtain the relative weights of the harms in a clinical setting. Weighing

the harms in the context of the community setting immediately after weighing the harms in the context of the clinical setting might be the reason for the lack of difference between the estimates of ThT in the two settings. This might have inadvertently introduced bias in the estimate of ThT in the community setting. This can be avoided by separately allowing elicitation, discussion, rating and weighing of the harms for the two settings. Different groups of participants with similar characteristics may be used to separately estimate the ThT in the clinical and community settings.

The expert and participant elicited probability of TB may be subject to bias, and using cohort studies to estimate the probability of TB in individuals with similar profiles may also be biased. Cohort studies use imperfect reference standards such as culture or a combination of culture and/or Xpert® MTB/RIF to infer the presence of TB. These diagnostic tests are known to have imperfect sensitivity for TB.[18] Additionally, it cannot easily identify extrapulmonary TB if the correct specimen is not used. Therefore, using alternative strategies that mitigate the bias attributable to imperfect reference tests may need to be considered to resolve the problem.

7. Future work

Focus group discussions (FGDs) and in-depth interviews (IDIs) are aimed at assessing the acceptability of the estimate of ThT and the aNGT method among the healthcare workers. Through FGDs and IDIs, we will gain a better understanding of the factors that influence the individual perceptions of the harms that drive the ThT and how they can be incorporated into consensus-building methods such as aNGT. Participant observation by the social scientist during the aNGT sessions will help highlight some of the nonverbal communications that may explain the variability in the estimates of ThT between the different aNGT sessions and settings. The findings of the qualitative study will be presented and published separately.

8. Conclusion

The estimate of ThT based on aNGT is consistent with the one estimated using CVs. While several studies have alluded that ThT is setting-dependent and is driven by participant characteristics, we found that the estimates of ThT based on aNGT were relatively similar between sites and stable across participant characteristics. Additional studies in other settings need to be conducted to assess the reproducibility and consistency of the method.

CRediT authorship contribution statement

Alfred Kipyegon Keter: Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Alastair Van Heerden:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization. **Tom Decroo:** Writing – review & editing, Methodology, Conceptualization. **Tom Boyles:** Writing – review & editing, Methodology, Conceptualization. **Shannon Bosman:** Writing – review & editing, Project administration. **Thandanani Madonsela:** Writing – review & editing, Investigation. **Lindani Innocent Msimango:** Writing – review & editing, Investigation. **Lenika Naiken:** Writing – review & editing, Project administration. **Carlos Kiyan:** Writing – review & editing, Data curation. **Mashaete Kamele:** Writing – review & editing, Project administration. **Irene Ayakaka:** Writing – review & editing, Project administration. **Klaus Reither:** Writing – review & editing, Methodology, Funding acquisition, Conceptualization. **Bart Karl Mario Jacobs:** Writing – review & editing, Visualization, Supervision, Methodology, Investigation, Conceptualization. **Lutgarde Lynen:** Writing – review & editing, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization.

Funding

This project is part of the European Union and Developing Countries Clinical Trials Partnership (EDCTP2) programme supported by the European Union (grant number: RIA2018D-2498).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A

.

Table A1
A six-point Likert scale ranging from 0 to 5 with possible interpretation.

Scale	Interpretation
0	not a factor that needs to be considered
1	a factor that should be considered but not given much weight in the treatment decision
2	a reasonably important factor that may influence the treatment decision
3	an important factor that will typically influence the treatment decision
4	a very important factor with a major impact on the treatment decision
5	a most important factor which will be a driving force in the treatment decision

Table A2
Probability of TB for each clinical vignette.

Clinical vignette number	Experts	All participants	Doctor participants
A1	0.25	0.59	0.45
A2	0.30	0.80	0.74
A3	0.40	0.63	0.52
A4	0.40	0.74	0.64
A5	0.50	0.70	0.61
A6	0.30	0.63	0.45
A7	0.70	0.80	0.74
A8	0.20	0.66	0.59
A9	0.02	0.47	0.25
A10	0.50	0.75	0.62
B1	0.95	0.88	0.85
B2	0.35	0.72	0.57
B3	0.26	0.52	0.34
B4	0.35	0.67	0.56
B5	0.53	0.70	0.65
B6	0.60	0.68	0.57
B7	0.64	0.80	0.74
B8	0.37	0.74	0.59
B9	0.63	0.77	0.71
B10	0.55	0.66	0.55

A = clinical, B = population screening.

Table A3
Summary of the number of participants and sessions.

Characteristic	n (%)
Number of participants in the aNGT sessions (n = 123)	
Lesotho	40 (33.0 %)
South Africa	83 (67.0 %)
Number of aNGT sessions (n = 17)	
Lesotho	7 (41.0 %)
South Africa	10 (59.0 %)
Number of participants for clinical vignettes (n = 130)	
Lesotho	45 (34.6 %)
South Africa	85 (65.4 %)
Number of participants in both aNGT & clinical vignettes (n = 115)	
Lesotho	38 (33.0 %)
South Africa	77 (67.0 %)

Table A4
Distribution of the participants who participated in clinical vignettes only and those who participated in both the clinical vignettes and aNGT.

	Participation in aNGT, CV and both	
	CV	aNGT and CV
Characteristics	n = 130	n = 115
Age (years) ǂ		
Mean (SD)	40.5 (9.6)	40.5 (9.5)
Range (Min. − Max.)	26–66	26–66
Male, n (%)	44 (33.8 %)	37 (32.2 %)
Years of experience in TB treatment, n (%)		
< 1 Year	9 (6.9 %)	5 (4.3 %)
1–5 Years	49 (37.7 %)	44 (38.3 %)
6–10 Years	33 (25.4 %)	30 (26.1 %)
> 10 Years	39 (30.0 %)	36 (31.3 %)
Profile, n (%)		
Doctors	58 (44.6 %)	49 (42.6 %)
Nurses	49 (37.7 %)	46 (39.7 %)
Program managers	16 (12.3 %)	15 (13.0 %)
Others	7 (5.4 %)	6 (5.2 %)
If a doctor, highest qualification, n (%)		
MBCHB	46 (79.3 %)	38 (77.6 %)
Masters	6 (10.3 %)	6 (12.2 %)
Post-graduate Diploma	6 (10.3 %)	5 (10.2 %)
Health services currently working in		
District, secondary or tertiary hospital	89 (68.5 %)	77 (67.0 %)
Clinic or primary healthcare	31 (23.8 %)	29 (25.2 %)
Others (NGOs, corporate, research, college, correctional services, retired)	10 (7.7 %)	9 (7.8 %)
Number of patients put on TB treatment last month, n (%)		

(*continued on next page*)

Table A4 (*continued*)

	Participation in aNGT, CV and both	
	CV	aNGT and CV
0	38 (29.2 %)	31 (27.0 %)
1 – 4	43 (33.1 %)	42 (36.5 %)
5 – 9	27 (20.8 %)	24 (20.9 %)
10 – 19	14 (10.8 %)	10 (8.7 %)
20 – 30	3 (2.3 %)	3 (2.6 %)
> 30	5 (3.8 %)	5 (4.3 %)

‡ One participant had an outlier age of 13 years. This value was excluded from the summary statistics.

The demographic characteristics were only captured in the Formsite® Therefore, the participants who participated in the aNGT alone did have their demographic characteristics captured. This explains why we do not have separate summaries for the participants in aNGT only.

Table A5

Estimates of therapeutic threshold and the corresponding 95 % credible intervals (95 % CrI) based on the participant-elicited probability of TB in the clinical vignettes.

	Clinical setting	Community setting
	Adjusted†	Adjusted†
Characteristic	Median (95 % CrI)	Median (95 % CrI)
Overall	62.1 (59.1, 64.9)	60.5 (57.9, 62.9)
Doctors only	56.3 (52.7, 59.6)	53.5 (50.3, 56.4)

† Adjusted for study site, order of participation in the aNGT and clinical vignettes, profession, and years of experience and standardizing for the whole population

Appendix B. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jctube.2025.100529.

References

[1] Abbafati C, Abbas KM, Abbasi-Kangevari M, Abd-Allah F, Abdelalim A, Abdollahi M, et al. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet 2020;396:1204–22. https://doi.org/10.1016/S0140-6736(20)30925-9.

[2] World Health Organization. Global Tuberculosis Report 2021. Geneva: World Health Organization; 2021.

[3] Basinga P, Moreira J, Bisoffi Z, Bisig B, Van Den Ende J. Why are clinicians reluctant to treat smear-negative tuberculosis? An inquiry about treatment thresholds in Rwanda. Med Decis Making 2007;27:53–60. https://doi.org/10.1177/0272989X06297104.

[4] Sreeramareddy CT, Rahman M, Harsha Kumar HN, Shah M, Hossain AM, Sayem MA, et al. Intuitive weights of harm for therapeutic decision making in smear-negative pulmonary Tuberculosis: An interview study of physicians in India, Pakistan and Bangladesh. BMC Med Inform Decis Mak 2014;14:1–9. https://doi.org/10.1186/1472-6947-14-67.

[5] Theron G, Peter J, Dowdy D, Langley I, Squire SB, Dheda K. Do high rates of empirical treatment undermine the potential effect of new diagnostic tests for tuberculosis in high-burden settings? Lancet Infect Dis 2014;14:527–32. https://doi.org/10.1016/S1473-3099(13)70360-8.

[6] Pauker SG, Kassirer JP. The Threshold Approach to Clinical Decision Making. NEJM 1980;302:1109–17.

[7] Pauker SG, Kassirer JP. Therapeutic Decision Making: A Cost-Benefit Analysis. NEJM 1975;293:229–34.

[8] Boyles T, Locatelli I, Senn N, Ebell M. Determining clinical decision thresholds for HIV-positive patients suspected of having tuberculosis. Evid Based Med 2017;22:132–8. https://doi.org/10.1136/ebmed-2017-110718.

[9] de Rooij ML, Lynen L, Decroo T, Henriquez-Trujillo AR, Boyles T, Jacobs BKM. The therapeutic threshold in clinical decision-making for TB. Int Health 2023. https://doi.org/10.1093/inthealth/ihad002.

[10] von Neumann J, Morgenstern O. Theory of Games and Economic Behavior. Princeton University Press; 1944.

[11] Moreira J, Bisig B, Muwawenimana P, Basinga P, Bisoffi Z, Haegeman F, et al. Weighing Harm in Therapeutic Decisions of Smear-Negative Pulmonary Tuberculosis. Med Decis Making 2009;29:380–90. https://doi.org/10.1177/0272989X08327330.

[12] Ebell MH, Locatelli I, Senn N. A novel approach to the determination of clinical decision thresholds. BMJ 2015;20:41–7. https://doi.org/10.1136/ebmed-2014-110140.

[13] Ebell MH, Locatelli I, Senn N. Correction: A novel approach to the determination of clinical decision thresholds. Evidence Based Medicine 2015;20:104–104. doi: 10.1136/ebmed-2014-110140corr1.

[14] Djulbegovic B, Hamm RM, Mayrhofer T, Hozo I, Van Den Ende J. Rationality, practice variation and person-centred health policy: A threshold hypothesis. J Eval Clin Pract 2015;21:1121–4. https://doi.org/10.1111/jep.12486.

[15] Harvey N, Holmes CA. Nominal group technique: An effective method for obtaining group consensus. Int J Nurs Pract 2012;18:188–94. https://doi.org/10.1111/j.1440-172X.2012.02017.x.

[16] McMillan SS, King M, Tully MP. How to use the nominal group and Delphi techniques. Int J Clin Pharm 2016. https://doi.org/10.1007/s11096-016-0257-x.

[17] Jacobs BKM, Keter AK, Henriquez-Trujillo AR, Trinchan P, de Rooij ML, Decroo T, et al. Piloting a new method to estimate action thresholds in medicine through intuitive weighing. BMJ. Evid Based Med 2023. https://doi.org/10.1136/bmjebm-2023-112350. bmjebm -2023-112350.

[18] Keter AK, Lynen L, Van Heerden A, Wong E, Reither K, Goetghebeur E, et al. Evaluation of tuberculosis diagnostic test accuracy using Bayesian latent class analysis in the presence of conditional dependence between the diagnostic tests used in a community-based tuberculosis screening study. PLoS One 2023;18:e0282417. https://doi.org/10.1371/journal.pone.0282417.

[19] Plasencia CM, Alderman BW, Barón AE, Rolfs RT, Boyko EJ. A Method to Describe Physician Decision Thresholds and Its Application in Examining the Diagnosis of Coronary Artery Disease Based on Exercise Treadmill Testing. Med Decis Making 1992;12:204–12. https://doi.org/10.1177/0272989X9201200306.

[20] Plummer M. "JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling 2003.

[21] R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria: 2020.

[22] Whelehan DF, Conlon KC, Ridgway PF. Medicine and heuristics: cognitive biases and medical decision-making. Ir J Med Sci 2020;189:1477–84. https://doi.org/10.1007/s11845-020-02235-1.

[23] Djulbegovic B, van den Ende J, Hamm RM, Mayrhofer T, Hozo I, Pauker SG. When is rational to order a diagnostic test, or prescribe treatment: The threshold model as an explanation of practice variation. Eur J Clin Invest 2015;45:485–93. https://doi.org/10.1111/eci.12421.

[24] Felder S, Mayrhofer T. Risk Preferences. Med Decis Making 2014;34:33–41. https://doi.org/10.1177/0272989X13493969.