Development of treatment-decision algorithms for children evaluated for pulmonary tuberculosis: an individual participant data meta-analysis

Kenneth S Gunasekera, Olivier Marcy, Johanna Muñoz, Elisa Lopez-Varela, Moorine P Sekadde, Molly F Franke, Maryline Bonnet, Shakil Ahmed, Farhana Amanullah, Aliya Anwar, Orvalho Augusto, Rafaela Baroni Aurilio, Sayera Banu, Iraj Batool, Annemieke Brands, Kevin P Cain, Lucía Carratalá-Castro, Maxine Caws, Eleanor S Click, Lisa M Cranmer, Alberto L García-Basteiro, Anneke C Hesseling, Julie Huynh, Senjuti Kabir, Leonid Lecca, Anna Mandalakas, Farai Mavhunga, Aye Aye Myint, Kyaw Myo, Dorah Nampijja, Mark P Nicol, Patrick Orikiriza, Megan Palmer, Clemax Couto Sant'Anna, Sara Ahmed Siddiqui, Jonathan P Smith, Rinn Song, Nguyen Thuy Thuong Thuong, Vibol Ung, Marieke M van der Zalm, Sabine Verkuijl, Kerri Viney, Elisabetta G Walters, Joshua L Warren, Heather J Zar, Ben J Marais, Stephen M Graham, Thomas P A Debray, Ted Cohen, James A Seddon

Lancet Child Adolesc Health 2023; 7: 336–46

Published Online March 13, 2023 https://doi.org/10.1016/S2352-4642(23)00004-4

Department of Epidemiology of Microbial Diseases (K S Gunasekera PhD, T Cohen DPH), **Department of Health Policy and Management** (J P Smith PhD), **and Department of Biostatistics** (J L Warren PhD), **Yale School of Public Health, New Haven, CT, USA; Inserm UMR1219, Institut de Recherche pour le Développement EMR 271, GHiGS, University of Bordeaux, Bordeaux, France** (O Marcy MD); **Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, Netherlands** (J Muñoz MSc, T P A Debray PhD); **ISGlobal, Hospital Clínic - Universitat de Barcelona, Barcelona, Spain** (E Lopez-Varela PhD, L Carratalá-Castro MD, A L García-Basteiro PhD); **Centro de Investigação em Saúde de Manhiça, Maputo, Mozambique** (E Lopez-Varela, O Augusto MD, L Carratalá-Castro, A L García-Basteiro); **National Tuberculosis and Leprosy Program, Kampala, Uganda** (M P Sekadde MD); **Department of Global Health and Social Medicine** (M F Franke ScD, L Lecca MD, S A Siddiqui MBA) **and Department of Pediatrics** (R Song MD), **Harvard Medical School, Boston, MA, USA; University of Montpellier, TransVIHMI, Institut de Recherche pour le Développement, Inserm, Montpellier, France**

Summary

Background Many children with pulmonary tuberculosis remain undiagnosed and untreated with related high morbidity and mortality. Recent advances in childhood tuberculosis algorithm development have incorporated prediction modelling, but studies so far have been small and localised, with limited generalisability. We aimed to evaluate the performance of currently used diagnostic algorithms and to use prediction modelling to develop evidence-based algorithms to assist in tuberculosis treatment decision making for children presenting to primary health-care centres.

Methods For this meta-analysis, we identified individual participant data from a WHO public call for data on the management of tuberculosis in children and adolescents and referral from childhood tuberculosis experts. We included studies that prospectively recruited consecutive participants younger than 10 years attending health-care centres in countries with a high tuberculosis incidence for clinical evaluation of pulmonary tuberculosis. We collated individual participant data including clinical, bacteriological, and radiological information and a standardised reference classification of pulmonary tuberculosis. Using this dataset, we first retrospectively evaluated the performance of several existing treatment-decision algorithms. We then used the data to develop two multivariable prediction models that included features used in clinical evaluation of pulmonary tuberculosis—one with chest x-ray features and one without—and we investigated each model's generalisability using internal–external cross-validation. The parameter coefficient estimates of the two models were scaled into two scoring systems to classify tuberculosis with a prespecified sensitivity target. The two scoring systems were used to develop two pragmatic, treatment-decision algorithms for use in primary health-care settings.

Findings Of 4718 children from 13 studies from 12 countries, 1811 (38·4%) were classified as having pulmonary tuberculosis: 541 (29·9%) bacteriologically confirmed and 1270 (70·1%) unconfirmed. Existing treatment-decision algorithms had highly variable diagnostic performance. The scoring system derived from the prediction model that included clinical features and features from chest x-ray had a combined sensitivity of 0·86 [95% CI 0·68–0·94] and specificity of 0·37 [0·15–0·66] against a composite reference standard. The scoring system derived from the model that included only clinical features had a combined sensitivity of 0·84 [95% CI 0·66–0·93] and specificity of 0·30 [0·13-0·56] against a composite reference standard. The scoring system from each model was placed after triage steps, including assessment of illness acuity and risk of poor tuberculosis-related outcomes, to develop treatment-decision algorithms.

Interpretation We adopted an evidence-based approach to develop pragmatic algorithms to guide tuberculosis treatment decisions in children, irrespective of the resources locally available. This approach will empower health workers in primary health-care settings with high tuberculosis incidence and limited resources to initiate tuberculosis treatment in children to improve access to care and reduce tuberculosis-related mortality. These algorithms have been included in the operational handbook accompanying the latest WHO guidelines on the management of tuberculosis in children and adolescents. Future prospective evaluation of algorithms, including those developed in this work, is necessary to investigate clinical performance.

Funding WHO, US National Institutes of Health.

Copyright © 2023 World Health Organization. Published by Elsevier Ltd. All rights reserved. This is an Open Access article published under the CC BY 3.0 IGO license which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. In any use of this article, there should be no suggestion that

WHO endorses any specific organisation, products or services. The use of the WHO logo is not permitted. This notice should be preserved along with the article's original URL.

Introduction

Tuberculosis is a leading cause of mortality among children worldwide,[1] accounting for about 2·5% of the 6 million deaths in children younger than 5 years each year.[2] Modelling suggests that more than 96% of tuberculosis deaths in children younger than 15 years occurred in those not receiving tuberculosis treatment.[3] WHO estimates that fewer than 50% of the 1·1 million children younger than 15 years who develop tuberculosis are diagnosed; the proportion is even lower, about 30%, among children younger than 5 years.[1] Thus, efforts to improve diagnosis, and thereby improve access to tuberculosis treatment, are important to reduce tuberculosis morbidity and deaths in children.

Confirmation of pulmonary tuberculosis in children is challenging because respiratory specimens tend to be paucibacillary.[4] Furthermore, collecting respiratory specimens from young children is invasive and requires resources that are generally concentrated in higher-level health-care centres. Thus, careful symptom review, clinical examination, chest x-ray, and history of *Mycobacterium tuberculosis* exposure can inform treatment decisions in clinical care. However, paediatric clinical expertise to make a diagnosis is often insufficient at primary health-care centres. This limits treatment access and leads to either delays in treatment initiation or no treatment initiation, both of which are associated with poor outcomes, including mortality.[5,6] Facilitation of

Research in context

Evidence before the study

Treatment-decision algorithms relate information gained in the evaluation of children into an assessment of tuberculosis disease risk and empower health-care workers to make appropriate treatment decisions. Studies in primary health-care centres have shown that use of treatment-decision algorithms can improve childhood pulmonary tuberculosis case detection and treatment initiation in settings with a high incidence of tuberculosis. We searched PubMed using the terms ("child*" OR "paediatr*" OR "pediatr*") AND ("tuberculosis" OR "TB") AND ("treatment-decision" OR "algorithm" OR "diagnos*") to identify primary research studies on childhood pulmonary tuberculosis treatment-decision algorithm performance evaluation or development published in any language before June 29, 2022. We additionally consulted several experts in childhood pulmonary tuberculosis diagnosis and management, and we referred to existing, published reviews of treatment-decision algorithms. With respect to performance, several studies have retrospectively estimated the performance of treatment-decision algorithms in a single geographical setting; a subset of these studies have also compared the performance of multiple algorithms using data from a single geographical setting. With respect to development, many existing algorithms have been developed without explicit analysis of data from children with presumptive pulmonary tuberculosis, often developed from expert consensus. Gunasekera and colleagues used model-based approaches to analyse diagnostic evaluations data (eg, clinical history, physical examination, chest radiograph, and results from rapid molecular and culture testing for *Mycobacterium tuberculosis*) collected from children with presumptive pulmonary tuberculosis in a single geographical setting to inform the development of a diagnostic algorithm, whereas Marcy and colleagues and Fourie and colleagues analysed data from multiple geographical settings. However, these studies were relatively small with limited assessment of generalisability.

Added value of this study

Following a WHO call for data, we identified and collated individual participant data from 13 prospective diagnostic studies from 12 countries including 4718 children with presumptive pulmonary tuberculosis from geographically diverse settings with a high incidence of tuberculosis. We evaluated the performance of existing treatment-decision algorithms and developed multivariable logistic regression models to quantify the contribution of individual features to discriminate tuberculosis from non-tuberculosis. A panel of child tuberculosis experts provided input into performance targets and advised on how to incorporate scores derived from these models into pragmatic treatment-decision algorithms to assist in the evaluation of children presenting with presumptive pulmonary tuberculosis in primary health-care centres.

Implications of all the available evidence

Our findings suggest that evidence-based, pragmatic treatment-decision algorithms can be developed to make sensitive and clinically appropriate decisions to treat a child with pulmonary tuberculosis. Although the specificity does not reach optimal targets for childhood tuberculosis diagnosis, pragmatic treatment-decision algorithms provide clinically relevant guidance that can empower health workers to start children on tuberculosis treatment in the primary health-care setting and could contribute to reducing the case-detection gap in childhood tuberculosis. External, prospective evaluation of these novel algorithms in diverse settings is required, including assessment of their accuracy, feasibility, acceptability, impact, and cost-effectiveness. This work led to a new interim WHO recommendation to support the use of treatment-decision algorithms in the evaluation of children with presumptive tuberculosis in the 2022 updated consolidated guidelines on the management of tuberculosis in children. Two algorithms developed from this work have been included in the WHO operational handbook accompanying these guidelines.

(M Bonnet PhD); **Epicentre, Mbarara, Uganda** (M Bonnet, P Orikiriza PhD); **Department of Paediatrics, Dhaka Medical College Hospital, Dhaka, Bangladesh** (Prof S Ahmed MD); **Indus Hospital & Health Network, Karachi, Pakistan** (F Amanullah MBBS, A Anwar MBBS, I Batool MBBS, S A Siddiqui); **The Aga Khan University Hospital, Karachi, Pakistan** (F Amanullah); **Instituto de Puericultura e Pediatria Martagao Gesteira** (R B Aurilio PhD) **and Faculty of Medicine** (Prof C C Sant'Anna MD), **Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil; Programme on Emerging Infections, Infectious Disease Division, icddr,b, Dhaka, Bangladesh** (S Banu PhD, S Kabir MPH); **Global Tuberculosis Programme, WHO, Geneva, Switzerland** (A Brands MPH, F Mavhunga MPH, S Verkuijl MMed, K Viney PhD); **US Centers for Disease Control and Prevention, Atlanta, GA, USA** (K P Cain MD, E S Click MD, J P Smith); **Department of Clinical Sciences, Liverpool School of Tropical Medicine, Liverpool, UK** (M Caws PhD); **Birat Nepal Medical Trust, Lazmipat, Kathmandu, Nepal** (M Caws); **Department of Pediatrics, Emory School of Medicine, Atlanta, GA, USA** (L M Cranmer MD); **Department of Epidemiology, Emory Rollins School of Public Health, Atlanta, GA, USA** (L M Cranmer); **Children's Healthcare of Atlanta, Atlanta, GA, USA** (L M Cranmer); **Centro de Investigación Biomédica en Red de Enfermedades Infecciosas, Barcelona, Spain** (A L García-Basteiro); **Desmond Tutu TB Centre, Department of Paediatrics and Child Health, Faculty of Medicine and Health Sciences, Stellenbosch University, Tygerberg, South Africa** (E Lopez-Varela, A C Hesseling PhD, Megan Palmer MMed, M M van der Zalm PhD, E G Walters PhD, J A Seddon PhD); **Oxford University Clinical Research Unit, Centre for Tropical Diseases, Ho Chi Minh City, Viet Nam** (J Huynh MBBS, N T Thuong Thuong PhD); **Nuffield Department of Medicine, Centre for Tropical**

Medicine and Global Health (J Huynh, N T Thuong Thuong), and Oxford Vaccine Group, Department of Paediatrics (R Song), University of Oxford, Oxford, UK; Socios En Salud Surcursal Perú, Lima, Perú (L Lecca); Global TB Program, Baylor College of Medicine and Texas Children's Hospital, Houston, TX, USA (Prof A Mandalakas PhD); Clinical Infectious Disease Group, German Center for Infectious Research, Clinical TB Unit, Research Center Borstel, Borstel, Germany (Prof A Mandalakas); Department of Paediatrics, University of Medicine, Mandalay, Myanmar (Prof A A Myint DrMedSc); Department of Paediatrics, University of Medicine, Magway, Myanmar (Prof K Myo DrMedSc); Department of Paediatrics, Mbarara University of Science and Technology, Mbarara, Uganda (D Nampijja MMed); Division of Infection and Immunity, Department of Biomedical Sciences, University of Western Australia, Perth, WA, Australia (Prof M P Nicol PhD); Department of Microbiology, Division of Basic Medical Sciences, School of Medicine, University of Global Health Equity, Kigali, Rwanda (P Orikiriza); University of Health Sciences, Phnom Penh, Cambodia (V Ung MD); National Pediatric Hospital, Phnom Penh, Cambodia (V Ung); School of Public Health (K Viney) and The Children's Hospital at Westmead Clinical School, Faculty of Medicine and Health (Prof B J Marais PhD), University of Sydney, Sydney, NSW, Australia; Directorate of Integrated Laboratory Medicine, Institute of Human Genetics, Newcastle upon Tyne NHS Foundation Trust, Newcastle upon Tyne, UK (E G Walters); Department of Paediatrics and Child Health, Red Cross Children's Hospital, and SA-MRC Unit on Child and Adolescent Health, University of Cape Town, Cape Town, South Africa (Prof H J Zar PhD); Department of Paediatrics and Murdoch Children's Research Institute, Royal Children's Hospital, University of Melbourne, Melbourne, VIC, Australia

appropriate diagnostic assessment with rapid treatment initiation at primary health-care settings where children initially present could contribute to reductions in tuberculosis-related morbidity and mortality.

Treatment-decision algorithms aim to standardise clinical assessment and decision making. Algorithms relate information gained in the evaluation of children into an assessment of tuberculosis disease risk and empower health-care workers to make appropriate treatment decisions. Adoption of an algorithmic approach to treatment decision making has been shown to improve childhood tuberculosis case detection and treatment access at primary health-care settings.[7,8] However, these algorithms were developed using consensus expert opinion rather than analysis of data.

Recent approaches for algorithm generation have used data from cross-sectional childhood tuberculosis diagnostic studies to quantify the contribution of clinical characteristics to the risk of tuberculosis disease.[9–11] Evidence-based approaches are objective and offer the potential for validation; however, existing studies have been small and not generalisable. We assembled individual participant data from children investigated for presumptive pulmonary tuberculosis. We then aimed to evaluate the performance of currently used diagnostic algorithms and to develop evidence-based algorithms to assist in tuberculosis treatment decision making for children younger than 10 years presenting to primary health-care settings. This work was conducted to inform the 2022 WHO guidelines for the management of tuberculosis in children and adolescents and the accompanying WHO operational handbook.[12,13]

Methods

Establishment of individual participant data

We identified potential sources of individual participant data through responses to a WHO public call for data on the management of tuberculosis in children and adolescents in July, 2020,[14] and through referral from childhood tuberculosis experts. Studies were eligible for inclusion if they prospectively recruited consecutive participants younger than 10 years attending health-care centres in countries with a high tuberculosis incidence for clinical evaluation of pulmonary tuberculosis and if they provided standardised reference classifications of pulmonary tuberculosis. We used an age cutoff of younger than 10 years to be consistent with the WHO definition of a child and to acknowledge that tuberculosis in children aged 10 years or older presents more similarly to adult tuberculosis and that adult diagnostic approaches are commonly used in this age group. We preferred for studies to have used the US National Institutes of Health (NIH) clinical case definitions of intrathoracic tuberculosis in children for diagnostic studies,[15] which classifies tuberculosis as follows: confirmed tuberculosis as culture-confirmed or Xpert MTB/RIF-confirmed *M tuberculosis* from respiratory specimens; unconfirmed tuberculosis as having symptoms, chest x-ray findings, immune tests of *M tuberculosis* sensitisation suggestive of tuberculosis, and follow-up to assess response to tuberculosis treatment (or without resolution of symptoms in the absence of tuberculosis treatment); and unlikely tuberculosis as not meeting criteria for either confirmed or unconfirmed tuberculosis. To ensure greater geographical representation, we also accepted data from high-quality studies that provided reference classifications using a previous NIH clinical case definition (in which the categories of probable and possible tuberculosis were combined into the unconfirmed tuberculosis category) and those that classified children using similar, prespecified definitions of confirmed, unconfirmed, and unlikely tuberculosis. Quality assessment was performed using a modified version of the Newcastle-Ottawa scale for cohort studies.[16]

After identification of eligible studies, we requested individual participant data including details from the clinical history, physical examination, chest x-ray, and results from rapid molecular and culture testing for *M tuberculosis* performed on respiratory specimens collected at study entry (appendix pp 3–6). All data assembly and analysis were carried out using R software (version 4.1.1). To account for the uncertainty associated with incomplete data, we used multilevel multiple imputation by chained equations (MICE) implemented in the mice package to generate 100 imputed datasets (appendix p 7).[17] This study was approved by the Stellenbosch University (Cape Town, South Africa) Health Research Ethics Committee (reference number X21/02/003) and the Yale University (New Haven, CT, USA) Institutional Review Board (reference number 2000028046). All collaborating investigators confirmed institutional ethical approval for their original data collection.

Evaluation of existing treatment-decision algorithms

We identified existing treatment-decision algorithms and scores (henceforth referred to as algorithms) to guide the evaluation of children with presumptive pulmonary tuberculosis through consultation with members of the WHO Guideline Development Group on the management of tuberculosis in children and adolescents. We defined a composite reference standard that includes confirmed and unconfirmed pulmonary tuberculosis to evaluate the performance of these algorithms. We carried out a sensitivity analysis of performance using a reference standard of confirmed pulmonary tuberculosis only (excluding children with unconfirmed tuberculosis). We used the reitsma function from the R package mada to pool study-level sensitivity and specificity estimates with 95% CIs using a bivariate random-effects meta-analysis (appendix p 8).[18,19]

Prediction model development and validation

We developed a multivariable logistic regression model to predict pulmonary tuberculosis using the composite reference standard in accordance with the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis standards.[20] We prespecified our model to include predictors from clinical and chest x-ray features commonly considered in the evaluation of presumptive childhood pulmonary tuberculosis in primary health-care settings with less than 50% missingness in our individual participant dataset. We also built a model without chest x-ray data to inform predictions in health-care centres without access to radiology services.

We adopted an internal–external cross-validation framework to estimate model parameters and assess generalisability.[21] Briefly, this leave-one-study-out approach built the prediction model on $n-1$ studies (n being the total number of studies included in the individual participant dataset) and validated using the remaining study, and is repeated for each hold-out study. Given that model performance was expected to vary across each of the hold-out studies, the regression coefficients of the n prediction models were subsequently meta-analysed to produce a single, summary prediction model. This approach was implemented in the metapred function of package metamisc, which accomplished meta-analysis via linear models.[21,22] To account for missing data, we generated a prediction model as described previously from each of the 100 imputed datasets and then used Rubin's rules to pool the regression coefficients and SEs to generate a final, single prediction model and compute odds ratios (ORs) with 95% CIs.[23] Using the leave-one-study-out approach, we estimated the c statistic (also known as the area under the receiver operating characteristic curve) to assess the model's ability to distinguish between children with tuberculosis and unlikely tuberculosis, the calibration intercept, and the observed-to-expected (O:E) ratio, to assess whether there were studies in which the model over-predicted or under-predicted tuberculosis.

Algorithm development

To generate clinically and programmatically implementable algorithms, we first converted the coefficient estimates of the parameters from each prediction model into a respective scoring system. We then placed the scoring system after several triage steps to guide health-care workers on its appropriate use, leading to two complete treatment-decision algorithms. We describe these steps in additional detail as follows.

We scaled the coefficient estimates for the parameters of the final prediction models (developed from all n studies) to estimate scores for each parameter such that a combined score of more than 10 corresponded to classification of tuberculosis at fixed sensitivities of 90%, 85%, 80%, 75%, and 70% (appendix p 9). To estimate the sensitivity and specificity of the scoring system in classifying tuberculosis using the composite reference standard, study-level sensitivities and specificities were pooled using the bivariate normal model of Reitsma and colleagues (implemented in the mada package) accounting for uncertainty introduced by imputation of missing data.[18,19] As a sensitivity analysis, we evaluated the performance of the score against a reference standard of confirmed pulmonary tuberculosis only.

(Prof S M Graham PhD); **Burnet Institute, Melbourne, VIC, Australia** (Prof S M Graham); **Department of Infectious Diseases, Imperial College London, London, UK** (J A Seddon)

Correspondence to:
Dr Kenneth S Gunasekera, Department of Epidemiology of Microbial Diseases, Yale School of Public Health, New Haven, CT 06510, USA
kenneth.gunasekera@yale.edu

See **Online** for appendix

	Sample size	Country	Age <2 years	Age <5 years	HIV	Severely acutely malnourished	Confirmed pulmonary tuberculosis	Unconfirmed pulmonary tuberculosis	Unlikely pulmonary tuberculosis	Newcastle-Ottawa Scale*
Aurilio et al (2020)	50	Brazil	21 (42·0%)	31 (62·0%)	6 (12·0%)	0	9 (18·0%)	11 (22·0%)	24 (48·0%)	5
Giang et al (2015)	113	Viet Nam	86 (76·1%)	106 (93·8%)	0	8 (7·1%)	20 (17·7%)	77 (68·1%)	16 (14·2%)	5
Hamid et al (2019)	445	Pakistan	41 (9·2%)	175 (39·3%)	0	26 (5·8%)	0	29 (6·5%)	416 (93·5%)	5
Kabir et al (2020)	402	Bangladesh	219 (54·5%)	296 (73·6%)	0	93 (23·1%)	63 (15·7%)	36 (9·0%)	303 (75·4%)	4
López-Varela et al (2015)	789	Mozambique	549 (69·6%)	789 (100·0%)	104 (13·2%)	68 (8·6%)	13 (1·6%)	128 (16·2%)	648 (82·1%)	4
Marcy et al (2019)	338	Burkina Faso, Cambodia, Cameroon, and Viet Nam	78 (23·1%)	142 (42·0%)	338 (100%)	64 (18·9%)	41 (12·1%)	155 (45·9%)	142 (42·0%)	5
Myo et al (2018)	223	Myanmar	72 (32·3%)	150 (67·3%)	27 (12·1%)	46 (20·6%)	27 (12·1%)	84 (37·7%)	112 (50·2%)	5
Orikiriza et al (2018)	338	Uganda	124 (36·7%)	222 (65·7%)	101 (29·9%)	41 (12·1%)	12 (3·6%)	145 (42·9%)	167 (49·4%)	5
Orikiriza et al (2022)	217	Uganda	157 (72·4%)	196 (90·%)	70 (32·%)	108 (50·%)	12 (6·%)	58 (26·7%)	125 (57·6%)	4
Song et al (2021)	300	Kenya	146 (48·7%)	300 (100·0%)	73 (24·3%)	8 (2·7%)	31 (10·3%)	65 (22·3%)	170 (56·7%)	4
Valencia et al (2017)	142	Mozambique	59 (41·5%)	95 (66·9%)	70 (49·3%)	27 (19·0%)	5 (3·5%)	28 (19·7%)	109 (76·8%)	5
Walters et al (2017)	595	South Africa	389 (65·4%)	548 (92·1%)	70 (11·8%)	18 (3·0%)	119 (20·0%)	180 (30·3%)	283 (47·6%)	5
Zar et al (2019)	766	South Africa	362 (47·3%)	603 (78·7%)	137 (17·9%)	32 (4·2%)	189 (24·7%)	274 (35·8%)	303 (39·6%)	5
Total individual participant data	4718	..	2303 (48·8%)	3653 (77·4%)	996 (21·1%)	539 (11·4%)	541 (11·4%)	1270 (26·9%)	2818 (59·7%)	..

Additional study-level details and relevant references are available in the appendix (pp 12–24). *Modified version for cohort studies; the highest score is 5, which indicates lower risk of bias.

Table 1: **Study-level descriptions of data included in the individual participant dataset**

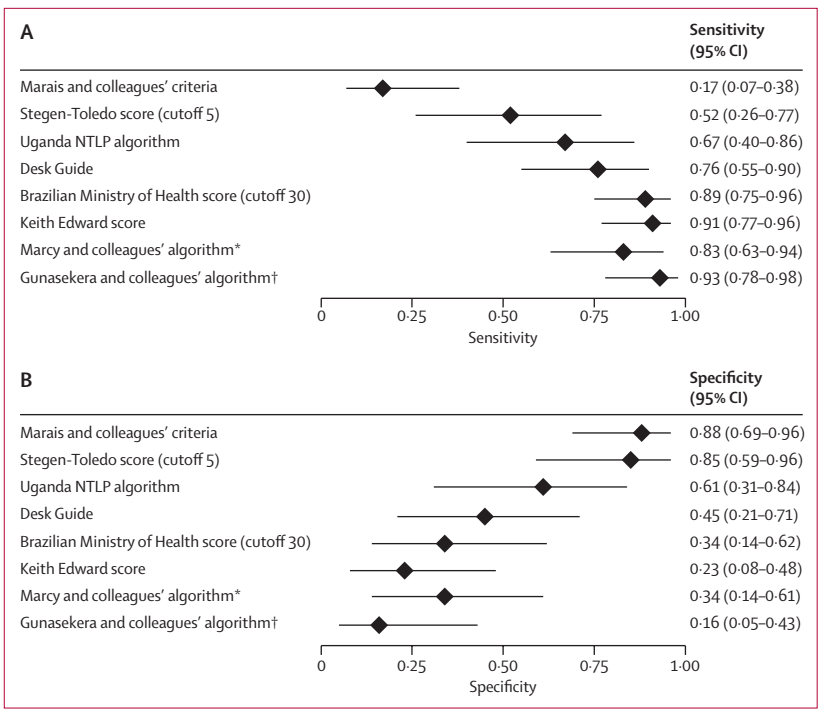

Figure 1: **Performance of existing treatment-decision algorithms at classifying tuberculosis**
Retrospective estimates of the pooled sensitivity (A) and specificity (B) of eight algorithms to guide decisions to treat children with presumptive pulmonary tuberculosis, had they been used to evaluate the children for whom we have individual participant data records. The reference classification of pulmonary tuberculosis included bacteriologically confirmed pulmonary tuberculosis and unconfirmed pulmonary tuberculosis. Modifications were made to the algorithms to maximise the use of the available individual participant dataset. NTLP=National TB and Leprosy Program. *Performance estimates from Marcy and colleagues. The algorithm was derived from only HIV-positive children in the individual participant dataset that excludes data from the cohort comprising HIV-positive children from Burkina Faso, Cambodia, Cameroon, and Viet Nam (from which the algorithm was developed). †Performance estimates by Gunasekera and colleagues. The algorithm was derived from only HIV-negative children in the individual participant dataset that excludes data from the South Africa population (from which the algorithm was developed).

We worked with staff from the WHO Global TB Programme to identify a group of experts in childhood tuberculosis (henceforth referred to as the expert group; appendix p 10) to advise on selection of a sensitivity performance target to develop the scoring system and development of triage steps before the scoring system to guide its appropriate use at primary health-care centres.

Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation, writing of the report, or the decision to submit.

Results

18 studies were identified as having potentially appropriate data, largely sourced from diagnostic evaluation studies (appendix p 11). The study investigators for two studies were unable to provide data in the necessary timeline, and an additional three studies did not meet the inclusion criteria. From the 13 included studies carried out in 12 countries, 4718 individual participant data records from children younger than 10 years with presumptive pulmonary tuberculosis were available (table 1). The data were predominantly collected at secondary, tertiary, or referral health-care centres; additional study-level details and relevant references are available in the appendix (pp 12–24). Although each study was required to include children with presumptive pulmonary tuberculosis, studies differed slightly with respect to inclusion criteria, variable definitions, and reference classification of tuberculosis (appendix pp 25–39). Of all 4718 children, 541 (11·5%) were classified as having confirmed tuberculosis, 1270 (26·9%) as having unconfirmed tuberculosis, and 2818 (59·7%) unlikely tuberculosis (appendix pp 40–65). Many demographic and clinical characteristics were similar between children belonging to these groups (appendix pp 66–67). All contributing studies had quality assessment scores of 4 out of 5 or 5 out of 5 (table 1; appendix p 68).

We evaluated the performance of eight existing treatment-decision algorithms. One of these algorithms was evaluated only on data from children living with HIV, and another evaluated only on data from children without HIV. Given that some algorithms considered variables that were not available in our individual participant dataset, we elected to estimate their performance using only the data that were available, without considering the contribution of those missing variables. References to each algorithm and details on which variables were not considered in this analysis are available in the appendix (pp 69–76). The sensitivities varied from 0·17 (95% CI 0·07–0·38) to 0·93 (0·78–0·98), with specificities varying from 0·88 (0·69–0·96) to 0·16 (0·05–0·43) when evaluated against the composite reference standard (figure 1; appendix pp 77–84). A sensitivity analysis evaluating performance to discriminate confirmed tuberculosis from unlikely tuberculosis showed marginally higher sensitivities and similar specificities to the performance in the entire dataset (appendix p 85).

ORs and 95% CIs of the predictors included in the model are shown in table 2. For the prediction model including chest x-ray features, the pooled *c* statistic was 0·71 (95% CI 0·66 to 0·76), the calibration intercept was –0·18 (–0·76 to 0·41), and the O:E ratio was 0·90 (0·64 to 1·26). Additional internal–external cross-validation *c* statistic, calibration intercept, and O:E ratio estimates are included in the appendix (pp 86–87). For the prediction model without chest x-ray features, the pooled *c* statistic was 0·65 (95% CI 0·59 to 0·71), the calibration intercept was –0·14 (–0·74 to 0·46), and the O:E ratio was 0·92 (0·60 to 1·43), with additional estimates included in the appendix (pp 88–90).

The appendix shows the scores derived from the model prediction coefficients that correspond to classification of all tuberculosis with respective sensitivities of 90%, 85%, 80%, 75%, and 70% (pp 91–92), and the study-level and summary performance of these scores in classifying tuberculosis (pp 93–97).

To balance the consequences of untreated tuberculosis versus the consequences of overtreatment, the expert group recommended a sensitivity threshold of 85% in classifying tuberculosis using the composite reference standard, resulting in the development of a score with a sensitivity of 0·86 (95% CI 0·68–0·94) and a specificity of 0·37 (0·15–0·66; figure 2). An analysis of the performance in classifying confirmed tuberculosis versus unlikely tuberculosis showed a sensitivity of 0·88 (95% CI 0·71–0·95) and specificity of 0·37 (0·15–0·67; appendix p 98). Under a sensitivity threshold of 85%, the model that included only features from the baseline clinical evaluation (without chest x-ray findings) had a sensitivity of 0·84 (95% CI 0·66–0·93) and specificity of 0·30 (0·13–0·56) in classifying tuberculosis (appendix pp 99–100).

To adapt the scores into treatment-decision algorithms to be used at primary health-care centres, the expert group recommended the following triage steps before tuberculosis classification using the score: identifying children with clinical symptoms and signs requiring urgent referral to higher levels of health care, and stratifying children by risk of mortality and progression of tuberculosis. Children at higher risk were defined by the expert group as those who belonged to any of the following categories: younger than 2 years, severely malnourished, or living with HIV. These children would be evaluated using the score at the time of the initial evaluation. Children not meeting this definition would be treated for the most likely non-tuberculosis condition and complete re-evaluation in 1–2 weeks; those with persistent or worsening symptoms at follow-up would be evaluated using the score. The expert group additionally recommended to pursue bacteriological testing, whenever available, on respiratory or stool specimens with rapid molecular diagnostics for all children and urine lateral flow assays for HIV-positive children, to align with existing WHO recommendations.[24]

The expert group recommendations resulted in the development of a treatment-decision algorithm (figure 3), in which children younger than 10 years with presumptive pulmonary tuberculosis are triaged by risk of tuberculosis-related morbidity and mortality before being evaluated for the presence of clinical and chest x-ray features to assign a score corresponding to tuberculosis risk. A total score of more than 10 results in classification of tuberculosis with a sensitivity of 85%. A known exposure to tuberculosis alone has a score of more than 10, which provides sufficient evidence to justify sensitive treatment decision making on its own; thus, this feature was placed above the other scored elements in the algorithm. The same parameters were used to construct the treatment-decision algorithm from the model without chest x-ray features (appendix p 101).

	Coefficient (95% CI)	Odds ratio (95% CI)	p value*
Intercept	−1·92 (−2·58 to −1·25)	..	..
Cough duration ≥2 weeks†	0·17 (−0·09 to 0·43)	1·19 (0·91 to 1·54)	0·86
Fever duration ≥2 weeks‡	0·45 (0·16 to 0·74)	1·57 (1·18 to 2·09)	0·25
Lethargy	0·25 (0·02 to 0·48)	1·28 (1·02 to 1·62)	0·66
Weight loss	0·22 (−0·03 to 0·48)	1·25 (0·97 to 1·62)	0·75
History of documented tuberculosis exposure	1·43 (0·87 to 2·00)	4·20 (2·39 to 7·38)	<0·0001
Haemoptysis	0·34 (−0·37 to 1·05)	1·40 (0·69 to 2·86)	0·78
Night sweats	0·20 (0·02 to 0·38)	1·22 (1·02 to 1·47)	0·71
Peripheral lymphadenopathy	0·35 (0·13 to 0·57)	1·42 (1·14 to 1·77)	0·35
Temperature ≥38°C	0·00 (−0·25 to 0·26)	1·00 (0·78 to 1·30)	>0·999
Tachycardia	0·15 (−0·13 to 0·42)	1·16 (0·88 to 1·53)	0·90
Tachypnoea	−0·05 (−0·27 to 0·16)	0·95 (0·77 to 1·18)	0·98
Cavities on chest x-ray	0·47 (−0·11 to 1·05)	1·60 (0·90 to 2·85)	0·53
Intrathoracic lymphadenopathy on chest x-ray	1·46 (1·00 to 1·92)	4·32 (2·73 to 6·85)	<0·0001
Opacities on chest x-ray	0·43 (0·02 to 0·84)	1·54 (1·02 to 2·32)	0·45
Miliary infiltrate on chest x-ray	1·27 (0·57 to 1·97)	3·56 (1·76 to 7·19)	0·0002
Pleural effusion on chest x-ray	0·64 (0·20 to 1·09)	1·90 (1·22 to 2·96)	0·13

The estimate provided for each predictor is computed against a reference that reflects the absence of that feature. *Calculated using Rubin's rules for multiple imputed data. †Absence is no cough or cough lasting less than 2 weeks. ‡Absence is no fever or fever lasting less than 2 weeks.

Table 2: **Estimates from logistic regression prediction model to classify pulmonary tuberculosis using variables from initial evaluation**

Discussion

We assembled a large individual participant dataset from nearly 5000 children from geographically diverse, high tuberculosis incidence settings to evaluate existing treatment-decision algorithms and develop new evidence-based treatment-decision algorithms to guide evaluation of children with presumptive pulmonary tuberculosis. As far as we are aware, this work describes the largest analysis to date of the best currently available individual participant data to provide practical guidance to health-care workers in primary health-care settings to identify which clinical features, with or without chest x-ray assessment, indicate whether initiation of tuberculosis treatment is warranted.

Previous work in childhood tuberculosis algorithm development has been from smaller studies with limited investigation into generalisability. This analysis leverages the clustered nature of the individual participant dataset using an internal–external cross-validation framework to allow for more generalisable model parameter estimation and investigation into model validity. Furthermore, the algorithms were developed closely with members of the WHO and experts in childhood tuberculosis to prioritise sensitive treatment decision making to address the global burden of child mortality associated with untreated childhood tuberculosis. The newly developed algorithms were incorporated into the WHO operational handbook to support implementation of the new consolidated guidelines.[12,13]

Of the clinical features, only reported exposure to tuberculosis was independently sufficient to meet the threshold for treatment initiation. This was true even in the model-based score without chest x-ray features, suggesting that none of the common clinical features

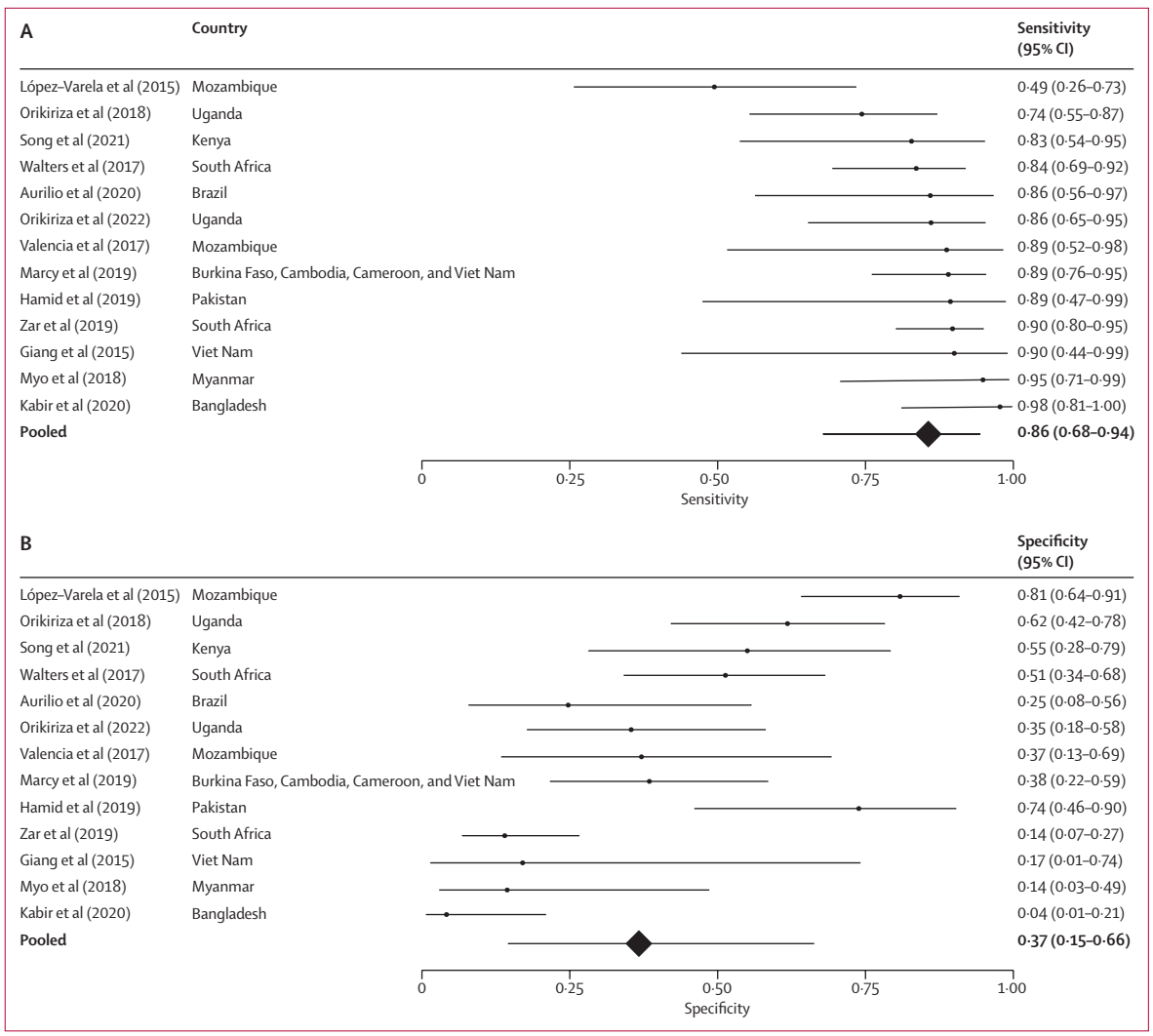

Figure 2: **Forest plot depicting performance of scaled scores from prediction model to classify tuberculosis with 85% sensitivity**
Study-level and pooled estimates of the (A) sensitivity and (B) specificity of classifying tuberculosis (composite reference standard: bacteriologically confirmed pulmonary tuberculosis and unconfirmed pulmonary tuberculosis) of the scores derived from the prediction model developed from the individual participant dataset to classify tuberculosis with 85% sensitivity.

could independently inform highly sensitive and specific treatment decisions. While results from tuberculin skin testing were used by studies to inform classification of tuberculosis and to improve imputation of missing data, we did not include these in the algorithms given operational limitations in using tuberculin skin testing at scale in high-burden settings. Of the chest x-ray features included in the algorithm, the presence of intrathoracic lymphadenopathy and a miliary pattern, respectively, were independently sufficient to start treatment. It is worth noting that inclusion of chest x-ray features only increased the specificity of the score slightly as compared with the score developed from the model with clinical features only. Chest x-ray has additional utility in guiding childhood pulmonary tuberculosis treatment duration for severe versus non-severe disease,[25] in monitoring tuberculosis treatment response (including associated complications and sequelae), and in the diagnosis of other non-tuberculosis intrathoracic pathology.

The decision to prioritise sensitivity in our algorithm development is crucial to initiate appropriate treatment in more children with tuberculosis; however, many children might be incorrectly treated for tuberculosis given the resulting specificity. No test or algorithm meets the WHO target sensitivity and specificity for a confirmatory diagnostic test for childhood pulmonary tuberculosis.[26] Thus, the expert group advised to develop an algorithm with a minimum sensitivity target of 85% as an acceptable balance between sensitivity and the resulting specificity. Given the estimated specificity of 37% for the scored part of the algorithm, it is likely that children without tuberculosis will be started on

4–6 months of treatment for tuberculosis and exposed to risk of adverse drug events. However, given the severe consequences of a missed tuberculosis diagnosis and the low rates of severe adverse drug events in this age group,[27] this trade-off of improved sensitivity for poorer specificity is reasonable. Additionally, it is noteworthy that current WHO recommendations for preventive treatment include at a minimum 3–4 months of treatment. It is also true that overtreatment for tuberculosis might result in delayed diagnosis of non-tuberculosis disease. Future study of the relative cost of false positive versus false negative classification at varying prevalence of tuberculosis could inform sensitivity threshold selection in subsequent algorithm development.

We note that the model-based scoring component of the algorithm demonstrates considerable study-level heterogeneity in sensitivity and specificity. Although this individual participant dataset is the largest of its size compiled to date, there were not enough studies to quantitatively describe the features that drive the observed heterogeneity. Given that we used data made available to WHO following a public call rather than conducting a systematic review, it is possible that some diagnostic studies might have been excluded. The inclusion of more data from existing, ongoing, and future studies could allow meta-regression to describe study-level sources of heterogeneity. Heterogeneity might have been driven in part by varied tuberculosis prevalence in the cohorts included as well as heterogeneities in disease presentation.

Several existing algorithms evaluated have similar performance to ours. Although we are unable to formally compare our newly developed algorithm with existing algorithms, the similar performance suggests that there might be several algorithms that a public health programme could consider to suit specific settings, available resources, and other implementation considerations. Future analysis of available data, including those obtained through a systematic review, might provide the opportunity to revise and calibrate the model and further interrogate possible sources of heterogeneity. This could lead to investigation and development of algorithms that might perform better among different subgroups of children, including those at higher risk for tuberculosis-associated mortality.

We considered it important to evaluate existing treatment-decision algorithms and develop new algorithms using a composite standard rather than solely a microbiological standard, given the high proportion of children treated for tuberculosis without bacteriological confirmation, even in the best resourced settings, which reflects the paucibacillary nature of disease in most young children. However, this reference standard remains imperfect, and misclassification might occur.[28] The underlying composition of the unconfirmed tuberculosis group might represent a heterogeneous group in which some children have tuberculosis, and some have other causes for their observed symptoms and signs. Additionally, it is possible that inclusion of unconfirmed pulmonary tuberculosis biased the estimation of the

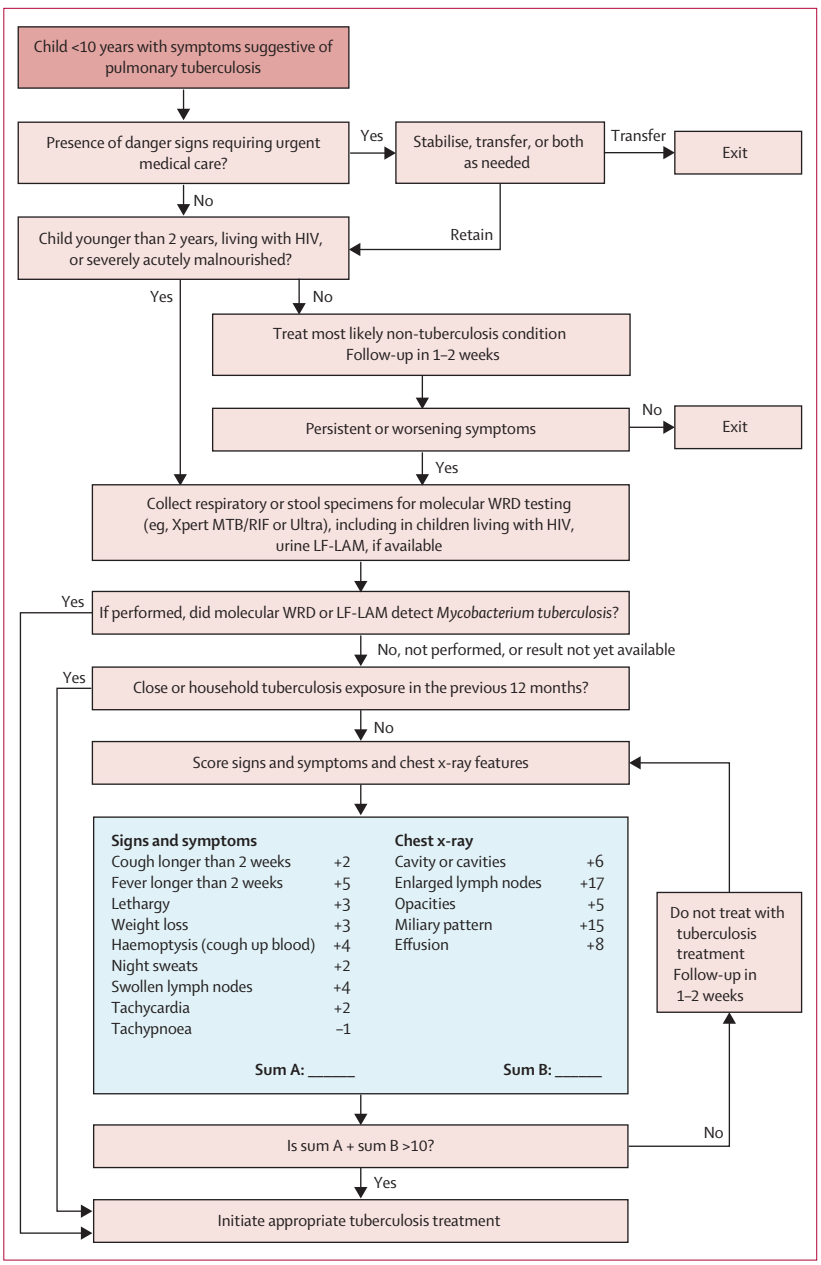

Figure 3: **Treatment-decision algorithm including chest x-ray features derived from the prediction model**
Tuberculosis treatment-decision algorithm for use among children younger than 10 years with symptoms suggestive of pulmonary tuberculosis, reproduced from the operational handbook accompanying the WHO consolidated guidelines on the management of tuberculosis in children and adolescents.[12,13] Selection steps before entering the scoring system reflect recommendations from the WHO expert panel to enrich the probability of tuberculosis among the population of children proceeding through the algorithm to the model such that the probability would more closely reflect the preselected population producing the data from which the prediction model was built, while balancing the consequences of untreated tuberculosis in children at high risk. Scores associated with features from clinical history and physical exam and chest x-ray translate to risk of tuberculosis and are scaled from the prediction model developed from the individual participant dataset. Guidance on the practical use of this algorithm is outlined in the WHO operational handbook.[13] LF-LAM=lateral flow urine lipoarabinomannan assay. WRD=WHO-recommended rapid diagnostic test.

prediction model parameters, especially those used to classify the unconfirmed group. Although this is a limitation of our study, the similar performance estimates of the score developed in the primary analysis using both the composite and confirmed tuberculosis reference standards suggest that this might not be a major issue.

Given that our algorithms are intended to guide decisions to treat children in primary health-care centres, it is a limitation that our individual participant dataset was derived from primarily tertiary and referral health centres. We are not aware of studies that provide this quality of diagnostic evaluation data from presumptive childhood tuberculosis in primary health-care centres. However, in several studies, children presenting at primary health-care settings were directly referred for study evaluation, providing some degree of reassurance as to the generalisability of results. The pre-test probability of tuberculosis is likely to be substantially lower among children attending primary health-care centres and the clinical presentation might be different as compared with tertiary and referral centres from which the data were obtained due to differences in tuberculosis prevalence. These are important given that many children with tuberculosis first present to primary health-care centres.[29] We believe that the risk stratification and delayed entry of lower risk children with presumptive tuberculosis (who should be able to tolerate the delay) is a practical attempt to safely raise the pre-test probability when implementing the algorithm in primary health-care centres.

It should be noted that although these performance estimates relate to the scored component of the algorithm, the overall sensitivity and specificity of the whole algorithm, which includes the triage steps, remain unknown and should be evaluated prospectively. As low-risk children are made to wait before being evaluated with the scored part of the algorithm, symptoms in some with diagnoses other than tuberculosis will resolve, probably improving specificity. Prospective evaluation of the entire algorithms in primary health-care settings will be crucial to determining their utility in improving case finding and reducing the mortality associated with untreated tuberculosis. Prospective studies of algorithm acceptability and feasibility are also indicated.

There are inherent limitations to developing a prediction model using data from multiple cohorts for a disease with an imperfect diagnostic gold standard. Study-level inclusion criteria varied, which affects the baseline tuberculosis prevalence and applicability of the score prediction estimates. Additionally, prediction variable definitions varied among the included studies—eg, history of weight loss was variably defined as caregiver-reported history of weight loss, objective weight loss, or deviation from previous growth trajectory. This heterogeneity is also true for the study-level reference classifications, especially for unconfirmed tuberculosis. Some studies used a previous version of the NIH reference classification, which included probable and possible tuberculosis categories that we reclassified as unconfirmed tuberculosis, despite the limitations of using this approach.[30] Furthermore, studies contributing chest x-ray data included interpretations of managing health-care providers or expert radiologists, depending on the study setting. These might contribute to heterogeneities in estimating the association between the predictors and the outcome of tuberculosis. Notably, a high degree of missingness in the individual participant dataset limited the variables available to evaluate existing algorithms and include in algorithm development. Protocol standardisation for childhood tuberculosis diagnostic evaluation will reduce heterogeneity in variable definition and assist future attempts to consolidate data for algorithm development and evaluation. Finally, we note that using a prespecified prediction model, as we did, might lead to overfitting.[31] Despite a reasonable summary O:E ratio for our model, the heterogeneity in study-level O:E ratio demonstrated in our internal–external cross-validation suggests that overfitting might be an issue. As more data become available, future investigation into the causes driving heterogeneity as well as other methods of prediction model feature selection might inform more nuanced use of this algorithm within specific contexts and populations.

Pragmatic treatment-decision algorithms can lead to better detection of tuberculosis in children, with improved access to early treatment and reduced tuberculosis morbidity and mortality. Although we developed these algorithms using a thorough modelling analysis of a large, high-quality individual participant dataset, the disappointing specificity of the scoring component suggests that improved diagnostic tools, such as computer-assisted interpretation of chest x-ray and biomarkers specific to tuberculosis, will be necessary to meet sensitivity and specificity targets. As these diagnostic tools become available, their data might be incorporated into treatment-decision algorithms to improve the specificity of the algorithms while maintaining high sensitivity.

Treatment-decision algorithms are now conditionally recommended by WHO in the evaluation of children with presumptive tuberculosis, which could lead to improved diagnostic capacity and treatment initiation at primary health-care centres where childhood tuberculosis expertise might be lacking. This work represents a paradigm shift in pragmatic and evidence-based approaches using advanced analytical methods to develop algorithms that draw on the best globally available data. This approach can be further improved and interrogated as additional data and diagnostic tools become available.

Contributors

KSG, OM, and JAS conceptualised the study. KSG, OM, EL-V, MFF, MB, SA, FA, AA, OA, RBA, SB, IB, KPC, LC-C, MC, ESC, ALG-B, ACH, JH, SK, LL, AAM, KM, DN, MPN, PO, MP, CCS'A, SAS, JPS, RS, NTTT, VU, MMvdZ, EGW, and HJZ contributed to data curation. KSG, OM, JM, JLW, TPAD, TC, and JAS did the formal analysis. KSG, AB, FM, SV, KV, TC, and JAS were involved in funding acquisition. KSG, OM, JM, EL-V,

MPS, MFF, MB, SA, FA, AA, OA, RBA, SB, IB, AB, KPC, LC-C, MC, ESC, ALG-B, ACH, JH, SK, AM, FM, AAM, KM, DN, MPN, PO, MP, CCS'A, SAS, JPS, RS, NTTT, VU, MMvdZ, SV, KV, EGW, JLW, HJZ, BJM, SMG, TPAD, TC, and JAS were involved in investigation. KSG, OM, JM, AB, FM, SV, KV, TPAD, TC, and JAS contributed to the methodology. JM and TPAD developed the software. KSG and JAS verified the data. KSG was responsible for visualisation of data. KSG, TC, and JAS wrote the original draft of the report. All authors contributed to review and editing of the report. All authors had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Declaration of interests
We declare no competing interests.

Data sharing
Data are available upon written request to the corresponding author.

Acknowledgments
This work was supported by the WHO Global Tuberculosis Programme (GTB) and the Eunice Kennedy Shriver National Institute of Child Health and Human Development of the US National Institutes of Health (NIH). It was used to inform WHO consolidated guidelines on tuberculosis management in children and adolescents and the accompanying operational handbook, both published by WHO in March, 2022. This research was also supported in part by the President's Emergency Plan for AIDS Relief (PEPFAR) through the US Centers for Disease Control and Prevention (CDC). Staff from the WHO GTB and from the CDC are co-authors of this paper, offered feedback on the methodology and investigation, reviewed the manuscript, and made the decision to submit. The authors alone are responsible for the views expressed in this Article, and they do not necessarily represent the views, decisions or policies of the NIH, WHO, CDC, and PEPFAR. KSG was supported by the NIH through the Eunice Kennedy Shriver National Institute of Child Health and Human Development (F30HD105440) as well as the Yale Medical Scientist Training Program (T32GM007205). EL-V has received funding from the European Respiratory Society and the European Union's H2020 research and innovation program under Marie Sklodowska-Curie grant agreement number 847462. ISGlobal receives support from the Spanish Ministry of Science and Innovation through the Centro de Excelencia Severo Ochoa 2019–2023 Program (CEX2018-000806-S). SA, SB, and SK received funding support from the US Agency for International Development through Research for Decision Makers' Activity, and this Article was prepared through the Structured Operational Research and Training Initiative, a global partnership led by the Special Program for Research and Training in Tropical Diseases at WHO. LMC was supported by the NIH through the National Institute of Allergy and Infectious Disease (K23 AI143479). RS was supported by the NIH through the Eunice Kennedy Shriver National Institute of Child Health and Human Development (K23HD072802). MMvdZ is supported by a career development grant from the EDCTP2 programme supported by the EU (TMA2019SFP-2836 TB lung-FACT2), by the NIH Fogarty International Center (award number K43TW011028), and by funding from the South African Medical Research Council. EGW was supported by a scholarship for doctoral studies from the Medical Research Council (MRC) of South Africa under MRC Clinician Researcher Programme, the Faculty of Medicine and Health Sciences at Stellenbosch University (Early Career Grant and Temporary Research Assistantship grant), the Harry Crossley Foundation, and the South African National Research Foundation (Thuthuka programme funding for doctoral students). HJZ and MPN received funding from the NIH (R01HD058971) and the Regional Prospective Observational Research in Tuberculosis Consortium, co-funded by MRC of South Africa and the NIH. TC and JLW were supported by the NIH through the National Institute of Allergy and Infectious Disease (R01AI147854 to TC and R01AI137093 to JLW). JAS was supported by a Clinician Scientist Fellowship jointly funded by the UK MRC and the UK Department for International Development (DFID) under the Concordat agreement (MR/R007942/1). We acknowledge the patients and their caregivers for their willingness to participate in each individual study. We hope that their contribution will benefit other families affected by tuberculosis. We thank the following individuals who contributed to execution of this work: Albert Okumu, Andrea T Cruz, Carlos M Perez-Velez, Chad Heilig, Colleen Wright, Elisha Okeyo, Ha Thi Minh Đang, Hendrik Simon Schaaf, James Orwa, Lazarus Odeny, Lesley Workman, Maria R. Jaswal, Mariaem Andres, Mark Fajans, Moe Zaw, Parisa Hariri, Patrice Ahenda, Prisca Rabuogi, Rose Abwunza, Rumana Nasrin, Sabrina Choudhury, Scott Lee, Shaikh Shumail, Shoaib Ahmed, Susan Musau, Syed Mohammad Mazidur Rahman, Walter Mchembere, Yi Kyaw, Leonardo Martinez, Vivian Cox, Bryan Vonasek, and Alexander Kay. Finally, we acknowledge the members of the WHO Guideline Development Group who considered this evidence to inform the 2022 WHO consolidated guidelines on the management of tuberculosis in children and adolescents.

References

1 WHO. Global tuberculosis report 2022. Geneva: World Health Organization, 2022.
2 Perin J, Mulick A, Yeung D, et al. Global, regional, and national causes of under-5 mortality in 2000–19: an updated systematic analysis with implications for the Sustainable Development Goals. *Lancet Child Adolesc Health* 2022; **6:** 106–15.
3 Dodd PJ, Yuen CM, Sismanidis C, Seddon JA, Jenkins HE. The global burden of tuberculosis mortality in children: a mathematical modelling study. *Lancet Glob Health* 2017; **5:** e898–906.
4 Detjen AK, DiNardo AR, Leyden J, et al. Xpert MTB/RIF assay for the diagnosis of pulmonary tuberculosis in children: a systematic review and meta-analysis. *Lancet Respir Med* 2015; **3:** 451–61.
5 Naidoo P, Theron G, Rangaka MX, et al. The South African tuberculosis care cascade: estimated losses and methodological challenges. *J Infect Dis* 2017; **216** (suppl 7)**:** S702–13.
6 Wobudeya E, Jaganath D, Sekadde MP, Nsangi B, Haq H, Cattamanchi A. Outcomes of empiric treatment for pediatric tuberculosis, Kampala, Uganda, 2010–2015. *BMC Public Health* 2019; **19:** 446.
7 Zawedde-Muyanja S, Nakanwagi A, Dongo JP, et al. Decentralisation of child tuberculosis services increases case finding and uptake of preventive therapy in Uganda. *Int J Tuberc Lung Dis* 2018; **22:** 1314–21.
8 Van Brusselen D, Simons E, Luendo T, et al. Improving pediatric TB diagnosis in North Kivu (DR Congo), focusing on a clinical algorithm including targeted Xpert MTB/RIF on gastric aspirates. *Confl Health* 2020; **14:** 26.
9 Fourie PB, Becker PJ, Festenstein F, et al. Procedures for developing a simple scoring method based on unsophisticated criteria for screening children for tuberculosis. *Int J Tuberc Lung Dis* 1998; **2:** 116–23.
10 Gunasekera KS, Walters E, van der Zalm MM, et al. Development of a treatment-decision algorithm for human immunodeficiency virus-uninfected children evaluated for pulmonary tuberculosis. *Clin Infect Dis* 2021; **73:** e904–12.
11 Marcy O, Borand L, Ung V, et al. A treatment-decision score for HIV-infected children with suspected tuberculosis. *Pediatrics* 2019; **144:** e20182065.
12 WHO. WHO consolidated guidelines on tuberculosis. Module 5: management of tuberculosis in children and adolescents. Geneva: World Health Organization, 2022.
13 WHO. WHO operational handbook on tuberculosis. Module 5: management of tuberculosis in children and adolescents. Geneva: World Health Organization, 2022.
14 WHO. WHO public call for data on the management of TB in children and adolescents. July 24, 2020. https://www.who.int/news-room/articles-detail/who-public-call-for-data-on-the-management-of-tb-in-children-and-adolescents (accessed Sept 13, 2022).
15 Graham SM, Cuevas LE, Jean-Philippe P, et al. Clinical case definitions for classification of intrathoracic tuberculosis in children: an update. *Clin Infect Dis* 2015; **61** (suppl 3)**:** S179–87.
16 Wells GA, Shea B, O'Connell D, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. 2014. https://www.ohri.ca/programs/clinical_epidemiology/oxford.asp (accessed Sept 13, 2022).
17 van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate imputation by chained equations in R. *J Stat Softw* 2011; **45:** 1–67.
18 Reitsma JB, Glas AS, Rutjes AWS, Scholten RJPM, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol* 2005; **58:** 982–90.

19 Doebler P. mada: meta-analysis of diagnostic accuracy. 2020. https://cran.r-project.org/web/packages/mada/mada.pdf (accessed Sept 13, 2022).

20 Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ* 2015; **350:** g7594.

21 Debray TP, Damen JA, Riley RD, et al. A framework for meta-analysis of prediction model studies with binary and time-to-event outcomes. *Stat Methods Med Res* 2019; **28:** 2768–86.

22 Debray T, de Jong V. metamisc: meta-analysis of diagnosis and prognosis research studies. R package version 0.2.4/r584. 2021. https://R-Forge.R-project.org/projects/metamisc/ (accessed Sept 13, 2022).

23 Heymans MW, Eekhout I. Applied missing data analysis with SPSS and (R). Studio, 2019. https://bookdown.org/mwheymans/bookmi/ (accessed Sept 13, 2022).

24 WHO. WHO consolidated guidelines on tuberculosis. Module 3: diagnosis. Rapid diagnostics for tuberculosis detection, 2021 update. Geneva: World Health Organization, 2021.

25 Turkova A, Wills GH, Wobudeya E, et al. Shorter treatment for nonsevere tuberculosis in African and Indian children. *N Engl J Med* 2022; **386:** 911–22.

26 WHO. High-priority target product profiles for new tuberculosis diagnostics: report of a consensus meeting. Geneva: World Health Organization, 2014.

27 Frydenberg AR, Graham SM. Toxicity of first-line drugs for treatment of tuberculosis in children: review. *Trop Med Int Health* 2009; **14:** 1329–37.

28 Wong M, Coit JM, Mendoza M, et al. Incident tuberculosis diagnoses in children at high risk for disease. *Open Forum Infect Dis* 2021; **8:** ofab075.

29 Zawedde-Muyanja S, Reuter A, Tovar MA, et al. Provision of decentralized TB care services: a detect-teat-prevent strategy for children and adolescents affected by TB. *Pathogens* 2021; **10:** 1568.

30 Marcy O, Goyet S, Borand L, et al. Tuberculosis diagnosis in HIV-infected children: comparison of the 2012 and 2015 clinical case definitions for classification of intrathoracic tuberculosis disease. *J Pediatr Infect Dis* 2022; **11:** 108–14 .

31 Steyerberg E. Clinical prediction models: a practical approach to development, validation, and updating. New York, NY: Springer-Verlag New York, 2009.