PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Use of targeted mobile X-ray screening and computer-aided detection software to identify tuberculosis among high-risk groups in Romania: descriptive results of the E-DETECT TB active case-finding project
AUTHORS	Mahler, Beatrice; de Vries, Gerard; van Hest, Rob; Gainaru, Dan; Menezes, Dee; Popescu, Gilda; Story, Alistair; Abubakar, Ibrahim

VERSION 1 – REVIEW

REVIEWER	Fumihiro Yamaguchi Showa University Fujigaoka Hospital, Respiratory Medicine
REVIEW RETURNED	06-Feb-2021

GENERAL COMMENTS	The present study investigated the association between active tuberculosis and the detection system in Romania. Results, Page 12/26, Line 241-248. The paragraph makes me confused. Where did the number "194" come from? I would suggest to draw a flowchart to describe how patients were selected from the database.

REVIEWER	Sarah Haile University of Zurich, Epidemiology, Biostatistics and Prevention Institute
REVIEW RETURNED	19-Jan-2021

GENERAL COMMENTS	Thank you for the opportunity to review this manuscript. As a statistician, I have primarily comments related to the statistical methods. I could not find a section on statistical methods. Please add this. In the 1st paragraph of the section "CAD4TB scores of CXRs and classification of the human reader", as well as Table 2, I would have liked to see boxplots or histograms corresponding to these reported results. A scatter plot corresponding to Table 2 would also be helpful to readers (perhaps with some jitter in the values, or using beeswarm plots or similar). Related to that paragraph, were the measurements for CAD4TB versions 5 and 6 comparable? A method such as that described by Bland and Altman 1986 (https://www-users.york.ac.uk/~mb55/meas/ba.pdf). In Table 2, it is noted that some scores of 100 were "in fact due to incorrect positioning". That statement deserves a longer statement in the paper. Is it known which subjects had incorrect positioning? How does removing them affect the results?

	In the first paragraph on pg 11, cutoffs for abnormalities are discussed. Would ROC curves be useful here?
	What does a negative score mean? A faulty measurement of the device?
	In Table 4, it should be nevertheless possible to provide a confidence interval for staff. In R, for example > binom.test(0, 733)$conf.int * 100000 [1] 0.0000 501.9935 Please verify that the statistical methods also describe how these confidence intervals were calculated.
	A less statistical question: Is there a general testing strategy in these cases that does not involve performing all described tests? Or is there an approach for determining in a systematic way who is determined to have TB?

REVIEWER	Fumihiro Yamaguchi Showa University Fujigaoka Hospital, Respiratory Medicine
REVIEW RETURNED	06-Feb-2021

GENERAL COMMENTS	The present study investigated the association between active tuberculosis and the detection system in Romania. Results, Page 12/26, Line 241-248. The paragraph makes me confused. Where did the number "194" come from? I would suggest to draw a flowchart to describe how patients were selected from the database.

VERSION 1 – AUTHOR RESPONSE

Authors' response to comments of the reviewers:

Reviewer: 1 (Dr. Simon Walusimbi, Makerere University, World Health Organization Country Office for Uganda)
Comments to the Author:
1. Screening for TB in high-risk populations is an efficient strategy for TB control. The authors, therefore, set out to implement a TB screening programme using innovative TB diagnostic tools based on evidence from the E-DETECT consortium. However, the research question and hypothesis for the presented work are not clear. This limits the scholarly value of the project results.
   Authors' response: Many thanks for this comment. We have now revised the research question in the abstract and the introduction ("We describe the implementation and assessed the yield, the bacteriological results and treatment outcome findings of the E-DETECT TB active case-finding project in Romania and compared the results of CAD to human reading in the algorithm used").
   In response to the reviewer's remarks, we have further clarified the "study design" paragraph in the Methods section, now reading "*Descriptive study based upon operational research during a TB control intervention*".
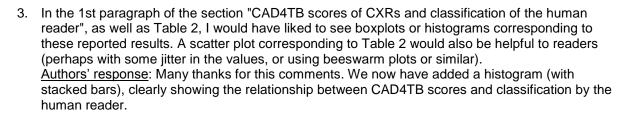
Reviewer: 2 (Dr. Sarah Haile, University of Zurich)
Comments to the Author:
Thank you for the opportunity to review this manuscript. As a statistician, I have primarily comments related to the statistical methods.
2. I could not find a section on statistical methods. Please add this.
   Authors' response: We have added a paragraph on the statistical analysis in the Methods section, i.e. "*All results are expressed as proportions. The prevalence rates were calculated for each risk group by dividing the yield with the number of people screened per 100.000 people. Confidence intervals were calculated using the Wilson score given some of the smaller cell sizes.*"

3. In the 1st paragraph of the section "CAD4TB scores of CXRs and classification of the human reader", as well as Table 2, I would have liked to see boxplots or histograms corresponding to these reported results. A scatter plot corresponding to Table 2 would also be helpful to readers (perhaps with some jitter in the values, or using beeswarm plots or similar).
   Authors' response: Many thanks for this comments. We now have added a histogram (with stacked bars), clearly showing the relationship between CAD4TB scores and classification by the human reader.

4. Related to that paragraph, were the measurements for CAD4TB versions 5 and 6 comparable? A method such as that described by Bland and Altman 1986 (https://www-users.york.ac.uk/~mb55/meas/ba.pdf).
   Authors' response: This is indeed an interesting question, but assessment and comparing of the two CAD4TB versions was not part of this study, and is the responsibility of the supplier.
   For the interest of the reviewer (and our interest) we compared the results of version 5 and the re-reads by version 6 in our database (Total 3345, excluding 4 records with a '-1' score and 20 records with a '100 score' (see below) resulted in 3321 records).
   • The difference between the two scores (CAD4TBv5 – CAD4TBv6) was median 0, with IQR between -8 to +7, i.e. 50% of the CXRs had scores 8 or less points different from each other.
   • 90% of CXRs were between -19 (extreme difference -42) and +16 (extreme difference +48).

5. In Table 2, it is noted that some scores of 100 were "in fact due to incorrect positioning". That statement deserves a longer statement in the paper. Is it known which subjects had incorrect positioning? How does removing them affect the results?
   Authors' response: This was erroneously reported. We have deleted the statement.
   Twenty of the CXRs had a score of 100 in the version 5, but none had a score of 100 after re-read by version 6 (median score of these 20 in CAD4TB version 6, 70 (range 51-99, IQR 66-75).
   None of the CXRs only read by version 6 had a score of 100.

6. In the first paragraph on pg 11, cutoffs for abnormalities are discussed. Would ROC curves be useful here?
   Authors' response: Threshold score of CAD4TB are set by the supplier representing the best trade-off between sensitivity and specificity selected after training CAD4TB on a validation set of CXRs, i.e. based on ROC curves. Since we were not able to base further TB examination (GeneXpert testing) on the CAD4TB curves, but had to rely on the human reader classification, ROC curves in this study were not useful.

7. What does a negative score mean? A faulty measurement of the device?
   Authors' response: Three CXRs had a negative score in the CAD4TB version 6 reading, because they were rejected by the quality check of CAD4TB. We have clarified this in the flowchart (Figure 1) and in the subscript of Figure 2: "*Excluding 3 CXRs that produced a negative CAD4TB score due to anatomical anomalies or wide framing and collimation of structures outside the areas of interest*."

8. In Table 4, it should be nevertheless possible to provide a confidence interval for staff. In R, for example > binom.test(0, 733)$conf.int * 100000 [1]   0.0000 501.9935. Please verify that the statistical methods also describe how these confidence intervals were calculated.
   Authors' response: Upon further consultation with our statisticians, we have revised our CIs to use Wilson score rather than Binomial Exact, as we feel this is more appropriate given some of the smaller cell sizes.
   We have added a subparagraph to the Methods describing the statistical analysis, i.e. "*All results are expressed as proportions. The prevalence rates were calculated for each risk group by dividing the yield with the number of people screened per 100.000 people. Confidence intervals were calculated using the Wilson score given some of the smaller cell sizes.*"

9. A less statistical question: Is there a general testing strategy in these cases that does not involve performing all described tests? Or is there an approach for determining in a systematic way who is determined to have TB?
   Authors' response: Screening strategies for active TB are described on the WHO guidance "Systematic screening for active tuberculosis" (https://apps.who.int/iris/bitstream/handle/10665/84971/9789241548601_eng.pdf;jsessionid=6A9

C655D7810C8114D4485DB85E59981?sequence=1). For screening various algorithms are possible and the document states "The choice of algorithm for screening and diagnosis should be based on an assessment of the accuracy of the algorithm for each risk group considered, as well as the availability, feasibility and cost of the tests"). We have chosen for the most direct, simple, fast and (relatively) cheap algorithm (namely initial radiology) for detecting (suspicion of) active intrathoracic TB in the context of the "hard-to-find" and "hard-to-reach" risk groups examined and ignored symptom screening (as less specific for these risk groups) or performing routine microbiology (smear, PCR and culture) unless upon indication.

Reviewer: 3 (Dr. Fumihiro Yamaguchi, Showa University Fujigaoka Hospital)
Comments to the Author:

10. The present study investigated the association between active tuberculosis and the detection system in Romania. Results, Page 12/26, Line 241-248. The paragraph makes me confused. Where did the number "194" come from? I would suggest to draw a flowchart to describe how patients were selected from the database.

Authors' response: The planning was that all clients with a CAD4TB scores above a certain threshold would be requested to produce sputum for on-the-spot examination with the GeneXpert in the MXU. Biosafety requirements in Romania however prevented this fast-track algorithm. It was then decided that GeneXpert examination would be performed on request of the chest physician reading the CXRs, usually when individuals were classified as "high suspicion" or "low suspicion" of active intrapulmonary TB, to confirm or rule out disease. For logistical reasons (CXRs were read later, and some clients could not be traced, or due to absence of productive cough not all requests could be met).

We have now changed the first sentence and paragraph of the GeneXpert paragraph of the Results section to clarify.

Review comments on manuscript titled active case finding among risk groups in Romania. Results from the E-DETECT TB

General comments:
Screening for TB in high-risk populations is an efficient strategy for TB control. The authors, therefore, set out to implement a TB screening programme using innovative TB diagnostic tools based on evidence from the E-DETECT consortium. However, the research question and hypothesis for the presented work are not clear. This limits the scholarly value of the project results.

Abstract
The abstract needs improvement to ensure results are not part of the interventions/methods. Also, the research question (s) need to be included in the abstract.

Authors' response: We have now revised the research question in the abstract and the introduction ("We describe the implementation and assessed the yield, the bacteriological results and treatment outcome findings of the E-DETECT TB active case-finding project in Romania by comparing the results of CAD to human reading in the algorithm used").

In response to the reviewer's remarks, we have further clarified the "study design" paragraph in the Methods section, now reading "Descriptive study based upon operational research during a TB control intervention".

Rationale for the study.
It appears this was a programme activity designed to test the feasibility of the EDETECT project. The study design employed for the presented work is not specified.

Authors' response: We have now revised the research question in the abstract and the introduction and clarified the "study design" paragraph in the Methods section, now reading "Descriptive study based upon operational research during a TB control intervention".

Literature:
There is limited literature referenced on CAD/Automated CXR imaging. A review of the literature could have helped to shape the research questions and study design.

Authors' response: We have added two publications on CAD4TB to the paragraph dealing with this topic in he discussion.

Methods
The cut-off threshold for the CAD4TB software, which qualified participants for GeneXpert testing is not specified. This would be useful for comparison with the human readers (chest physicians).
Authors' response: We have added to the body of the text (line 175): "*for this evaluation we did not use a threshold score set by the supplier representing the best trade-off between sensitivity and specificity selected after training CAD4TB on a validation set of CXRs or retrospectively selected a score that matched the performance of a human reader but empirically chose CAD4TB scores 50 respectively 60*".

Results
If the research questions are clarified, then the results should be presented in line with the research questions. The table 2 needs to be re-organized to enable better understanding of the results. It appears a 2x2 table was intended for the analysis, but the reference/gold standard is not indicated.
Authors' response: We have rearranged Table 2 as requested by the editor and one of the reviewers into a histogram (with stacked bars) clearly showing the proportion of normal and abnormal CXRs increasing with increasing CAD4TB scores.

VERSION 2 – REVIEW

REVIEWER	Sarah Haile University of Zurich, Epidemiology, Biostatistics and Prevention Institute
REVIEW RETURNED	12-Apr-2021

GENERAL COMMENTS	Thank you for the revised manuscript. Where did the categories shown in Figure 2 come from? Would a boxplot with individual points not be possible here? (as seen e.g. here https://www.r-graph-gallery.com/89-box-and-scatter-plot-with-ggplot2.html)

VERSION 2 – AUTHOR RESPONSE

Reviewer: 2. Dr. Sarah Haile, University of Zurich
Comments to the Author:
1. Where did the categories shown in Figure 2 come from?
    Authors' response: We have added in the text the rationale for the categories: *The CAD4TB scores were grouped into three categories, based on the two threshold scores (50 and 60) mostly used by the supplier to compare the results with the human reading.*

2. Would a boxplot with individual points not be possible here? (as seen e.g. here https://www.r-graph-gallery.com/89-box-and-scatter-plot-with-ggplot2.html)
    Authors' response: We agree that a boxplot contributes to the explanation of the results. The actual numbers are needed to support the discussion. We suggest to re-install the table included in the first submission (with few minor changes explained in the previous authors' response) together with the boxplot.