EUROPEAN RESPIRATORY journal
OFFICIAL SCIENTIFIC JOURNAL OF THE ERS

Derivation of a high-resolution CT-based, semi-automated radiographic score in tuberculosis and its relationship to bacillary load and antitubercular therapy

Journal:	*European Respiratory Journal*
Manuscript ID	ERJ-00600-2023.R1
Manuscript Type:	Research Letter
Date Submitted by the Author:	13-Jun-2023
Complete List of Authors:	Riou, Catherine; University of Cape Town Faculty of Health Sciences, Wellcome Centre for Infectious Diseases Research in Africa, Institute of Infectious Diseases and Molecular Medicine Du Bruyn, Elsa; University of Cape Town Faculty of Health Sciences, Wellcome Centre for Infectious Diseases Research in Africa, Institute of Infectious Diseases and Molecular Medicine Kim, Hyun Jung; David Geffen School of Medicine at UCLA, Division of Thoracic Radiology da Costa, Irene; David Geffen School of Medicine at UCLA, Division of Thoracic Radiology Lee, Jihey; David Geffen School of Medicine at UCLA, Division of Thoracic Radiology Sher, Alan; National Institutes of Health, Wilkinson, Robert; Division of Medicine, Imperial College London, W2 1PG, UK, Allwood, Brian; Stellenbosch University Department of Medicine, Division of Pulmonology; University of Cape Town Lung Institute, Goldin, Jonathan; David Geffen School of Medicine at UCLA, Division of Thoracic Radiology
Key Words:	Tuberculosis, high resolution computed tomography (HRCT), Bacterial burden, antibacterial therapy, HIV
Abstract:	

Faculty of Health Sciences
Department of Pathology, Division of Medical Virology
Falmouth Building, Anzio Road
OBSERVATORY 7925
Cape Town South Africa
Tel: +27 21 406 6090

13 June 2023

Dear Professor Chalmers,

Thank you for the opportunity to revise our manuscript: “Derivation of a high-resolution CT based, semi-automated radiographic score in tuberculosis and its relationship to bacillary load and antitubercular therapy”.

We have comprehensively addressed all points raised by the three reviewers. This includes the addition of a clinical table to the figure, additional text and the acknowledgment of some limitations.

As requested two versions of the manuscript have been up-loaded. One with changes highlighted in yellow and one “clean” version.

Overall, the conclusions of our manuscript have not changed. We believe the manuscript is substantially improved, and we thank the reviewers for their comments, and your editorial team for your consideration.

Yours sincerely,

Dr Catherine Riou
University of Cape Town
Tel: +27 82 742 9189
cr.riou@uct.ac.za
http://www.idm.uct.ac.za/Catherine_Riou

Point by point response to reviewers.
Manuscript #ERJ-00600-2023
Riou *et. al.* "Derivation of a high-resolution CT based, semi-automated radiographic score in tuberculosis and its relationship to bacillary load and antitubercular therapy."

Reviewer: 1

Comments to the Author
The study aims to address an important area, the quantitation of TB burden in the lung and the relationship between HRCT findings, bacterial load and the host immune response. The strengths are the size of the study, the different clinical groups (HIV negative and positive, TB negative and positive) and the interface with an automated computer analysis of the images. The data are clear and easy to follow. The initial findings of association between score and markers of sputum bacillary load are perhaps relatively predictable, and the paper essentially is a proof of concept of the validity of this approach as a new readout. Overall this is a valuable addition to the literature and provides an new readout for measuring lung infection burden, in particular in the context of HIV. I only have minor comments / clarifications:
>> We thank the reviewer for the positive feedback on our study.

1.1 Can the authors justify the choice of time point of imaging within 7 days of starting ATT?
>>In designing the study, this imaging time point was chosen to reflect the extent of disease at the time of TB diagnosis as closely as possible. The 7-day window was incorporated to logistically facilitate arrangement of CT imaging which was not available at the recruitment site. From an ethical standpoint, we did not wish to delay TB treatment in these patients.

1.2 In fig 1A, there is a subgroup that seems to have normal CAD scores. Can this be commented on and did they have common features, such as being culture negative?
>> Amongst TB patients, those with a "normal" TB-CAD score (i.e., <2) had a higher likelihood of being HIV positive (80%) compared to TB patients with a score >2 (51.3% of HIV-infected patients, p<0.0001, Chi-square test). Moreover, there was also a trend towards a higher proportion of negative sputum Mtb cultures in the "normal" TB-CAD score (<2) group, compared to those with a TB-CAD score >2 (27.2% vs. 12.5% respectively, p=0.067, Chi-square test). These findings support the observation that HIV/TB co-infected patients, especially those at lower CD4 counts, may present with atypical radiographic features of pulmonary TB, including the absence of upper lobe cavities and low sputum bacillary yield (Keiper, Chest, 1995; Palmieri, Infection, 2002; Swaminathan, CID, 2010; Nakiyingi, BMC-Infectious Disease, 2021; Greenberg, Radiology, 1994; Gupta, In J Tuberc Lung Dis, 2013).

A sentence has been added to the text. A sentence has been added to the text.
Page 3, Line 87: "*Of note, about a quarter of TB participants exhibited a TB-CAD score comparable to controls (<2), this sub-group was mostly constituted of HIV-infected participants (80%) and showed a trend towards higher proportion of negative sputum Mtb culture. This supports the observation that HIV/TB co-infection often presents with atypical radiographic features and low sputum bacillary load [11,12].*"

1.3 There seem to be a subgroup where the score paradoxically increases at 6 months (Fig 1H)? Are there any insights from these, such as ARV naïve then starting treatment.
>> Five participants showed an elevation of TB-CAD between BL and W24. No common clinical characteristics were observed in these participants (please refer to the table below). Two were HIV negative and three were viremic HIV positive with CD4 count below 200.

Of the three that were HIV infected two were ART naïve at study start, and one had defaulted treatment. South African guidelines recommend initiation of ART between 2-8 weeks after starting TB treatment, however the exact time point of ART initiation was up to the treating healthcare provider. We unfortunately do not have access to the exact timing of ART initiation in these participants.

TB clinic appointment cards were checked to ensure ongoing follow up with their TB treatment provider. However, pill count or urine INH testing were not conducted and thus non-adherence to TB treatment cannot conclusively be excluded in these participants.

None of these participants experienced treatment failure, TB relapse or TB-IRIS during the 52 week follow up period but were not followed up beyond this.
It is thus difficult to speculate on the possible reasons for the paradoxical increases in TB-CAD scores in these participants.

PID	BMI at BL	BMI at W24	TB-CAD at BL	TB-CAD at W24	Fold change TB-CAD	HIV status	CD4 count at BL	Log_{10} VL at BL	Time sputum conversion
EN-1053	20.4	20.2	0.3	0.6	2.0	Neg	na	na	W8
EN-1112	17.8	17.5	1.2	5.1	4.3	Neg	na	na	W8
EN-1038	20	24.1	2.2	2.6	1.2	Pos	54	4.56	W4
EN-1098	17.4	18.8	3	10.5	3.5	Pos	156	4.78	>W8
EN-1122	21.1	20.7	0.7	0.9	1.3	Pos	112	6.55	W4

A comment has been added:
Page 4, Line 122: "*In five participants, the TB-CAD score increased but none of them experienced treatment failure or relapse during the 52-week follow-up period. While poor treatment adherence was not suspected, this cannot fully be excluded as pill counts or isoniazid urine testing were not conducted*".

Minor comment: Fig 1F the Y axis labels are not consistent, should they all be the same?
>> Thank you for picking-up this mistake, it has been corrected.

Reviewer: 2

Comments to the Author
This is an interesting study evaluating HRCT and mycobacterial burden and responses to ATT treatment. While some of the findings are expected (detailed below), this is one study that can contribute to the growing field of lung destruction caused by TB especially the observation that the TB-CAD score is not normal at the end (week 24) of TB treatment.

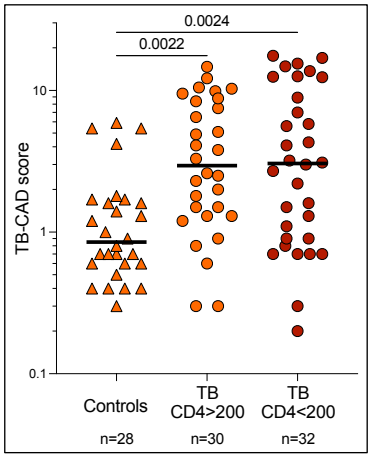
Major comments:
1) While this study focuses on HRCT, it would be wholistic to also have a short mention on CXR scores in the quantification of TB burden. One such score is the Ralph score (Ralph et al BMJ 2010)
>> We agree with the reviewer and have added the following:
Page 2, line 38: *"Ralph et al developed a promising radiographic scoring system [1], with baseline scores being predictive of sputum smear conversion at two months, but it is reliant on skilled readers and has not been systematically validated in predominantly HIV infected study populations of variant CD4 counts"*

2) The authors state that HIV-TB patients do not produce sufficient quality sputum - please cite the relevant literature. Anecdotally that may not be the case especially if HIV-TB patients are

subjected to sputum induction. However, there are certain groups eg. neonates and older adults who definitely do have problems producing quality sputum but this is not the focus of the study.
>> We agree with the reviewer that our wording was unclear, and we have clarified as follows:

Page 2, Line 49: "*However, there is renewed interest in non-sputum-based approaches for treatment monitoring as sputum TB culture is time-consuming (negative results may take up to 42 days in liquid culture) and does not yield results in cases of culture contamination.*"
And
Page 5, Line 150: *"Furthermore, these individuals are often sputum smear negative owing to low bacillary load, leading to treatment monitoring challenges [11]"*

3) line 63 - a Table on the study group demographics would be very helpful for the reader. I note majority of the patients are PLHIV. Key would be to identify how many are immune-reconstituted. What are the number of patients whose CD4 < 200, , < 50 and > 200?
- How do the TB-CAD score perform in those whose CD4 < 50 and those < 200? Data ideally should be reanalysed to factor the CD4 counts.
>> To clarify the clinical characteristics of TB patients, a table was added including HIV viral load, CD4 count and the proportion of participants with CD4 count >200 and CD4 count> 200 (see Fig 1A). Of note, only 6 participants had a CD4 count < 50 cells/mm^3, thus we did not create a category for this sub-group, as the sample size is too small for meaningful statistical comparisons.



TB-CAD was significantly higher in both the CD4 > 200 and CD4 < 200 groups compared to the HIV-infected control group (p = 0.0022 and p = 0.0024, respectively, see figure on the right). Moreover, when HIV-infected TB patients were grouped according to their CD4 count (> 200 cells/mm^3 or < 200 cells/mm^3), we still found an association between TB-CAD and GeneXpert Ct values (CD4 > 200 group: p = 0.017, r = -0.49 and CD4 < 200 group: p = 0.033, r = -0.46). lastly, in those with HIV-TB co-infection, the TB-CAD score did not correlate with absolute CD4 count (p = 0.5, r = 0.08).

See edits on:
Page 3, Line 71: "*Clinical characteristics of TB patients are presented in **Fig 1A**.*"
And
Page 3, Line 84: *"Moreover, when HIV- infected TB participants were grouped according to their absolute CD4 count (>200 or <200 cells/mm^3), the TB-CAD score was significantly elevated in both groups compared to the control group (p=0.0022 and p=0.0024, respectively)".*

4) The finding of TB-CAD score being worse (higher) in those with higher mycobacterial burden (lower Ct values, faster TPP, and of course those who have TB vs healthy) are expected. That TB-CAD also improves with TB treatment longitudinally is also expected.
>> We agree with the reviewer that the associations between the TB-CAD score and measures of bacterial burden was expected. It is thus reassuring that a semi-automated radiographic score could represent an alternative and/or complementary non-sputum-based tool to quantify TB related lung involvement, disease severity and monitor treatment response.

5) CRP is a reflection of host inflammation/infection, so the absence of TB-CAD score correlation in those who are HIV viremic (presumably low CD4?) is also expected.
>> We agree with the reviewer. The group with HIV viraemia had lower CD4 counts compared to those with suppressed viral load ($p<0.0001$). This is in keeping with Lawn *et al*. (Int J Tuberc Lung Dis., 2014) who has shown that higher CRP values in HIV-TB coinfection are more likely to be found in those with lower CD4 counts (below 200 cells/mm^3) and viral loads in excess of 4.5 Log_{10} copies/ml, but did not show association with radiological extent of disease measured on chest X-ray.
Page 3, Line 100: *"However, this correlation was absent in the viremic HIV-1-infected group (p=0.63), possibly due to HIV-associated systemic inflammation."*

6) Some of the correlations are not strong eg Fig F panels (i.e r values of <0.5) so the authors should also moderate their conclusions
>> As suggested, we moderated our conclusion regarding the relationship between the TB-CAD score and phenotypic attributes of Mtb-specific CD4+ T cells.
Page 5, line 140: *"While these correlations are moderate, it is reassuring that Mtb-specific immune responses known to associate with TB disease activity [x,x] also relate to radiographic disease extent."*

7) Line 120: The authors state that the TB-CAD is an objective readout of TB disease in the lung regardless of HIV-1 status. However, the findings are an observation and not validated in independent cohorts of patients, so do clarify. In addition, the authors have not shown if the score varies with CD4 count, so until this analysis is performed, this conclusion has its limitations.
>> As described in point #3, we now showed that the TB-CAD score could be of value even for HIV-infected individuals with CD4 count < 200 cell/mm^3. Nevertheless, we added as a limitation, stating that these findings need to be validated in an independent cohort including participants with low CD4 count.
Page 5, Line 150: *"Further studies are needed to validate our findings in an independent cohort, including participants with CD4 counts <50 cells/mm^3".*

8) Line 124 - 126: The authors again state that TB-HIV patients are unable to produce sufficient sputum without citing evidence. Refer to point 2.
>> Please refer to the response provided for point 2.

Minor comments:
1) The authors should re-term active TB to TB disease with the change in WHO terminology.
>> This has been corrected in the text and the figure.

Reviewer #3

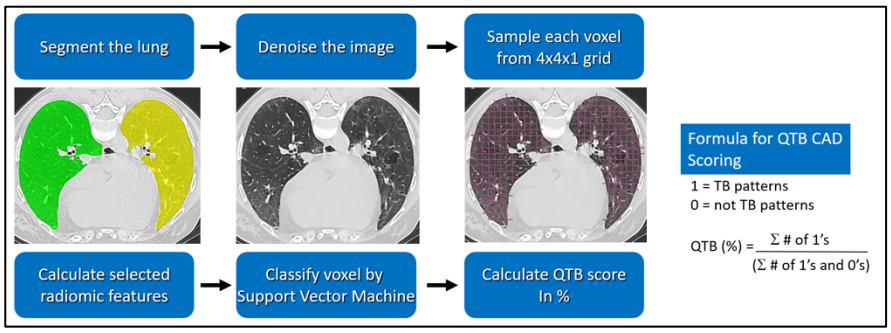
The authors utilise a novel CAD method for CT scans of HIV infected and uninfected people with TB disease to show correlation of this score with various biomarkers of TB disease activity including mycobacterial burden at baseline and "time to sputum positivity". A control population without TB disease or infection (but over half were HIV positive), were used as a comparator group. The authors point out that their findings show benefit over CXR for HIV infected individuals and also over CRP. The CAD score (lower) was found to correlate with those patients

that had an earlier microbiological response to TB treatment and CAD scores were said to be lower for most individuals that underwent a repeat CT at the completion of TB treatment.

Questions for authors;
1. How might other clinicians/researchers reproduce these findings when the methods for CAD are not provided? You mention that your technique would be useful in trial design but without the ability to follow your methods this is not possible. Please provide some general detail in a supplemental file.
>> ERJ Research Letters does not allow any supplementary material. To alleviate this issue and respect the ERJ research letter word limitation, we added a comment in the appendix stating that a detailed method for the generation for the TB-CAD score can be obtained by contacting directly Prof Jonathan Goldin (jgoldin@mednet.ucla.edu) or Dr Grace Hyun J Kim (gracekim@mednet.ucla.edu) (See **page 6, Line 169**)

Please find below a short description of the method.

Generation of the TB-CAD score
We generated new computer-aided diagnostic (CAD) scores for a tuberculosis (TB) by adapting the previously built machine learning for interstitial lung diseases (ILD) and COVID-19 to TB characteristics of high-resolution CT [1, 2]. Quantitative TB CAD score is expressed as a sum of quantitative fibrotic reticulation and consolidation. Quantitative fibrotic reticulation score is one of scores from quantitative ILD, called as quantitative lung fibrosis (QLF) score, that matched with visual ground truth (Area under Curve (AUC) of 0.96) in evaluating 25% or more fibrotic reticulation [1]. Quantitative consolidation was derived from the non-ILD patterns with the high attenuated areas [2]. After the lung segmentation of inspirational HRCT series, five automated steps were used to calculate TB CAD scores: i) denoise based on the noise characteristics in the uniform areas, ii) 4-by-4 grid-sample in each axial image, iii) calculation of selected radiomic features, iv) classification of each voxel, and v) counting the voxels classified as TB and expressing a ratio of TB classified voxels to the total voxels into a percent (See Figure 1 for the detail).



Supplementary Figure 1: Computer-aided algorithm (CAD) and automatic pipeline of Quantitative TB (TB-CAD) score

[1] Kim HG, Tashkin DP, Clements PJ, Li G, Brown MS, Elashoff R, Gjertson DW, Abtin F, Lynch DA, Strollo DC, Goldin JG. A computer-aided diagnosis system for quantitative scoring of extent of lung fibrosis in scleroderma patients. Clin Exp Rheumatol 2010; 28 (5 Suppl 62): S26-35.

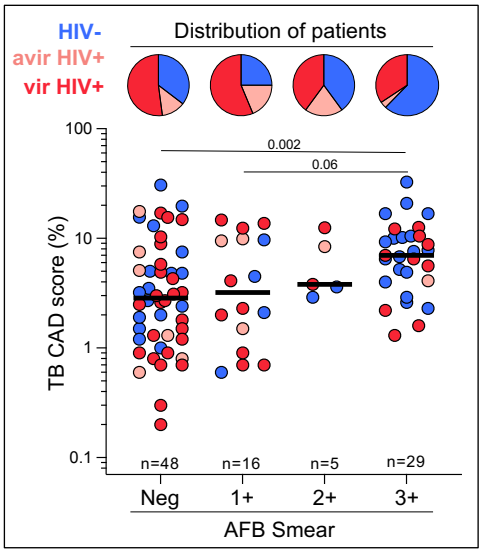
[2] Dolinay T, Jun D, Maller A, Chung A, Grimes B, Hsu L, Nelson D, Villagas B, Kim GHJ, Goldin J. Quantitative image analysis in COVID-19 acute respiratory distress syndrome: a cohort observational study. F1000Res. 2022 May 24;10:1266. doi: 10.12688/f1000research.75311.2. PMID: 37224317; PMCID: PMC10182379.

2. How were cases (of TB) selected? consecutive, at random etc. b. Were children included?

>> TB patients were recruited from the Ubuntu Clinic, Site B, Khayelitsha (Cape Town, South Africa). All participants were adults (age ≥ 18 years, legal age in South Africa) and recruitment was consecutive. Patients attending the clinic who tested sputum Xpert MTB/RIF (Cepheid) positive and were subsequently being commenced on TB treatment were recruited to the study and enrolled once informed consent was provided. Due to word limitation, we included a refence where further description of the cohort is provided (#X, Riou C et al. "Disease extent and anti-tubercular treatment response correlates with Mycobacterium tuberculosis-specific CD4 T-cell phenotype regardless of HIV-1 status". *Clin Transl Immunology* 2020: 9(9): e1176).
Page 2, Line 63: "*All participants were adults and recruited at the Site B Clinic, Khayelitsha, Cape Town, South Africa and provided written informed consent [8]"*

3. HIV status is provided but not other demographic data such as AFB smear status at baseline (I note the cycling time and time to culture positivity variables, but does baseline AFB smear status conversion also correlate?
>> To clarify the clinical characteristics of TB participants, a table has been added (see Fig. 1A) including HIV VL and CD4 count. Due to word limit, we did not include AFB smear data. Smear status was available for 98 out of the 104 TB participants.

The figure presented on the right, depicts the TB-CAD score in TB participants grouped according to their AFB smear status at baseline. As expected, patients with negative smear were enriched in viremic HIV-infected individuals (see pies). The TB-CAD score was significantly elevated in patients with 3+ sputum smear compared to negative smear (p=0.002).



A sentence was added to report these results.
Page 3, Line 96: "*Additionally, the TB-CAD score was significantly higher in TB participants with 3+ AFB smear compared those who were smear negative (median: 7 vs 2.85, p=0.0002, data not shown)".*

Were cases of MDR-TB included in this study?
>> All TB participants were drug sensitive. This information is included in the participant description section (**Page 3, line 68**).

4. 80 patients underwent repeat CT scans at the completion of TB treatment. How were these patients selected.
>> The 80 participants that underwent repeat scan were not specifically selected.

The 24 participants did not undergo treatment completion scan for the following reason:
- Lost to follow up (n=9),
- Declined repeat HRCT (n=3)
- Meeting exclusion criteria (n=9) including out of window for the visit, pregnancy, no available Xpert MTB/RIF cycle threshold (Ct) value at baseline.
- Death (n=3) as far as the clinical study team could ascertain, the cause of death was not related to TB.

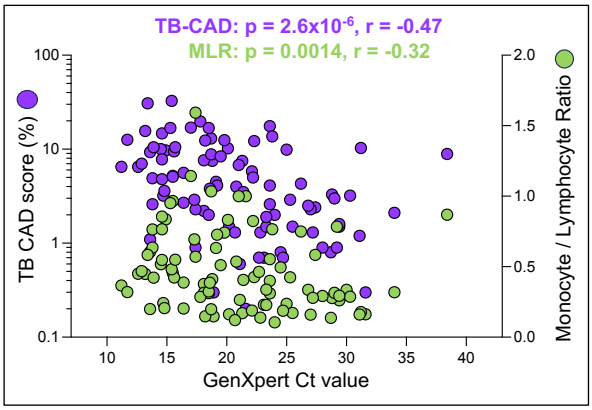
This was reworded.

Page 4, line 119: "*A subset of 80 eligible participants underwent repeat HRCT…..*"

5. Monocytes but not CRP correlated with CAD scores. Could you not just use FBC and differential white cell count instead of CAD? If not what are the benefits (given the stated aims), of CAD over this biomarker?

>> Of all components of the full blood count the monocyte to lymphocyte ratio (MLR) has been the most extensively studied in its relation to TB related disease activity. Several studies, especially involving HIV-TB coinfected individuals and children, have implicated the MLR as potential TB biomarker (Naranbhai, J Infect Di, 2014; Rakotosamimanana, Eur Respir J, 2015; La Manna, PLoS One, 2017; Choudhary, J Acquir Immune Defic Syndr, 2019; Adane, Medicine, 2022; Kissling, Pediatr Infect Dis J, 2023).

In our cohort, we also found an association between MLR and GeneXpert Ct values ($p = 0.0014$, $r = -0.32$, see figure on the right). However, the TB-CAD score showed a stronger association with GeneXpert Ct values ($p = 2.6x10^{-6}$, $r = -0.47$), , thus offering added advantage of more closely corresponding to Mtb load.



Furthermore, La Manna *et al*. (La Manna, PLoS One, 2017) compared the MLR of healthy QFT negative donors to that of cured TB patients and found no statistically significant difference. In our study those who completed TB treatment still had higher TB-CAD scores than those in the healthy control group (Figure I, p=0.0024), thus suggesting that TB-CAD may offer a more sensitive measure of TB related lung involvement, with residual changes still being apparent at the end of treatment. Similarly, Malherbe *et al*. (Malherbe, Nature Med, 2016) using PET-CT has previously shown that a subset of TB patients still showed imaging patterns in keeping with active TB disease after curative TB treatment.

To reflect these results, we have added the following to the text:

Page 4, Line 103: "*As the monocyte to lymphocyte ratio (MLR) holds promise as TB biomarker[X], we also compared the relationship between sputum Mtb load and MLR or TB-CAD. We indeed found a significant correlation between MLR and Xpert Ct values ($p=0.0014$, $r=-0.32$). However, the TB-CAD showed a stronger correlation with Xpert Ct values ($p=2.6x10^{-6}$, $r=-0.47$), thus offering added advantage of more closely corresponding to Mtb load".*

6. Figure 1h. What were the characteristics of those 4(?) patients that had a worsening in TB CAD score? Were they HIV positive with IRIS? Was prednisolone used.

>> Five participants showed an elevation of TB-CAD between BL and W24. No common clinical characteristics were observed in these participants (please refer to the table below). Two were HIV negative and three were viremic HIV positive with CD4 count below 200.
Of the three that were HIV infected two were ART naïve at study start, and one had defaulted treatment. South African guidelines recommend initiation of ART between 2-8 weeks after starting TB treatment, however the exact time point of ART initiation was up to the treating healthcare provider. We unfortunately do not have access to the exact timing of ART initiation in these participants.

TB clinic appointment cards were checked to ensure ongoing follow up with their TB treatment provider. However, pill count or urine INH testing were not conducted and thus non-adherence to TB treatment cannot conclusively be excluded in these participants.

None of these participants experienced treatment failure, TB relapse or TB-IRIS during the 52 week follow up period but were not followed up beyond this.
It is thus difficult to speculate on the possible reasons for the paradoxical increases in TB-CAD scores in these participants.

PID	BMI at BL	BMI at W24	TB-CAD at BL	TB-CAD at W24	Fold change TB-CAD	HIV status	CD4 count at BL	Log_{10} VL at BL	Time sputum conversion
EN-1053	20.4	20.2	0.3	0.6	2.0	Neg	na	na	W8
EN-1112	17.8	17.5	1.2	5.1	4.3	Neg	na	na	W8
EN-1038	20	24.1	2.2	2.6	1.2	Pos	54	4.56	W4
EN-1098	17.4	18.8	3	10.5	3.5	Pos	156	4.78	>W8
EN-1122	21.1	20.7	0.7	0.9	1.3	Pos	112	6.55	W4

A comment has been added:
Page 4, Line 122: "*In five participants, the TB-CAD score increased but none of them experienced treatment failure or relapse during the 52-week follow-up period. While poor treatment adherence was not suspected, this cannot fully be excluded as pill counts or isoniazid urine testing were not conducted*".

Derivation of a high-resolution CT-based, semi-automated radiographic score in tuberculosis and its relationship to bacillary load and antitubercular therapy

Catherine Riou[1,2*], Elsa du Bruyn[1*], Grace Hyun J. Kim[3,4*], Irene da Costa[3,4], Jihey Lee[3,4], Alan Sher[5], Robert J. Wilkinson[1,6,7,8†], Brian W. Allwood[9†], Jonathan Goldin[3,4†].

[1] Wellcome Centre for Infectious Disease Research in Africa and Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Observatory, Cape Town, South Africa

[2] Division of Medical Virology, Department of Pathology, University of Cape Town, Observatory, Cape Town, South Africa

[3] Department of Radiology, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA

[4] UCLA Center for Computer Vision and Imaging Biomarkers, Los Angeles, CA, USA

[5] Immunobiology Section, Laboratory of Parasitic Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, USA

[6] Department of Medicine, University of Cape Town, Observatory, Cape Town, South Africa

[7] Department of Infectious Diseases, Imperial College London, W12 0NN, UK

[8] The Francis Crick Institute, London, NW1 1AT, UK

[9] Division of Pulmonology, Department of Medicine, Stellenbosch University and Tygerberg Hospital, Cape Town, South Africa

[*] and [†] These authors equally contributed to the work.

Correspondence to Dr Catherine Riou and Dr Elsa du Bruyn, University of Cape Town, Observatory, Cape Town, South Africa. cr.riou@uct.ac.za and eldubruyn@gmail.com

Word count
Title: 20 words
Main text: 1500 words
References: 15

To the Editor:

Efforts to curb the TB pandemic remain hindered by the lack of objective measures to quantify disease severity and track treatment success that are valid in both HIV-1-infected and -uninfected TB patients. Ralph *et al* developed a promising radiographic scoring system [1], with baseline scores being predictive of sputum smear conversion at two months, but it is reliant on skilled readers and has not been systematically validated in predominantly HIV infected study populations of variant CD4 counts. Superior to conventional chest radiography, High-resolution computed tomography (HRCT) is a highly sensitive tool to track endobronchial TB disease extent [2]. Although recent studies have made progress in development of semi-automated methods for TB diagnosis and quantification of disease severity on chest CT [3, 4], none have assessed how these radiographic scoring systems relate to mycobacterial burden, the immune response to TB and whether they can reliably quantify the effect of antitubercular therapy (ATT) on disease affected lung. Sputum culture conversion at two months remains the gold standard measure of efficacy in trials of new TB drugs and regimens [5]. However, there is renewed interest in non-sputum-based approaches for treatment monitoring as sputum TB culture is time-consuming (negative results may take up to 42 days in liquid culture) and does not yield results in cases of culture contamination.

In the present study, we derived a computer-assisted, semi-automated quantitative radiographic scoring system (TB-CAD) applied to chest HRCT. Briefly, HRCTs underwent quantitative analysis by applying texture-based computer-aided diagnosis (CAD) at the Center for Computer Vision and Imaging Biomarkers (University of California). A computer assisted, semi-automated quantitative radiographic score of TB disease extent, called TB-CAD, was developed by modification of a previously developed algorithm [6] to detect and quantify areas of abnormality including, cavitation, consolidation, nodules, scarring and airway disease. The algorithm was run after a segmentation algorithm isolated the lung parenchyma and the TB-CAD score was calculated as the percentage of pixels of TB related abnormality present within the lungs on chest HRCT of patients with pulmonary TB and healthy controls. A detailed method is provided as supplementary material. All participants were adults and recruited at the Site B Clinic, Khayelitsha, Cape Town, South Africa and provided written informed consent [7]. The study was approved by the UCT Human Research Ethics Committee (HREC:050/2015).

This study included 104 participants with newly diagnosed, drug-sensitive TB who tested sputum Xpert MTB/RIF (Cepheid) positive and underwent HRCT within 7 days of initiating ATT. 59.6% were HIV positive (n=62): 12 participants being aviremic and 50 having a detectable HIV-1 viral load. Clinical characteristics of TB patients are presented in **Fig 1A**. Eighty participants underwent repeat HRCT after completing antitubercular therapy (ATT). A group of 48 healthy controls (i.e., asymptomatic, IFN-γ release assay and sputum Xpert MTB/RIF negative and no previous TB) also underwent HRCT. 59.6% of the control participants were HIV positive (n=28), with 13 being aviremic (median CD4 count: 460 cell/mm^3) and 15 being viremic (median: 3.83 Log_{10} mRNA copies/ml, IQR: 1.91-4.29) and a median CD4 count of 364 cells/mm^3. The median age was comparable between the TB and control groups (36 vs 37 years old, respectively).

We examined the relationship between the TB-CAD score and TB disease activity, soluble inflammatory markers and the *Mycobacterium tuberculosis* (Mtb)-specific CD4 T cell profile in blood using flow cytometry [7], in a subset of the TB participants (n=60). The TB-CAD score was significantly higher in the pulmonary TB group at baseline compared to the healthy controls (median: 3.9 vs 0.8, p<0.0001, **Fig. 1B**). Moreover, when HIV- infected TB participants were grouped according to their absolute CD4 count (>200 or <200 cells/mm^3), the TB-CAD score was significantly elevated in both groups compared to the control group (p=0.0022 and p=0.0024, respectively). Of note, about a quarter of TB patients exhibited a TB-CAD score comparable to controls (<2), this sub-group was mostly constituted of HIV-infected participants (80%) and showed a trend towards higher proportion of negative Mtb culture compared to those with a TB-CAD score >2. This supports the observation that HIV/TB co-infection often presents with limited lung involvement, indicating a paucibacillary nature [8, 9].

There was a significant inverse correlation between TB-CAD score and sputum Xpert MTB/RIF cycle threshold (Ct) value at baseline (p=2.6x10^{-6}, r=-0.47), as well as sputum culture time to positivity (p=1.5x10^{-6}, r=-0.44), irrespective of HIV-1 status (**Fig. 1C and D**). Additionally, the TB-CAD score was significantly higher in TB participants with 3+ AFB smear compared those who were smear negative (median: 7 vs 2.85, p=0.0002, data not shown). The TB-CAD score correlated with plasma C-Reactive Protein (CRP) levels at baseline in the HIV-1-uninfected and aviremic HIV-1 infected TB groups (p=0.0001, r=0.56 and p=0.004, r=0.78) (**Fig. 1E**). However, this correlation was absent in the viremic HIV-1-infected group (p=0.63), possibly due to HIV-associated systemic inflammation. TB-CAD

scores correlated with blood monocyte counts of HIV-1 uninfected and HIV-1 infected participants (p=0.0003, r=0.53 and p=0.006, r=0.39) (**Fig. 1F**). As the monocyte to lymphocyte ratio (MLR) holds promise as TB biomarker [10], we also compared the relationship between sputum Mtb load and MLR or TB-CAD. We indeed found a significant correlation between MLR and Xpert Ct values (p=0.0014, r=-0.32). However, the TB-CAD showed a stronger correlation with Xpert Ct values ($p=2.6x10^{-6}$, r=-0.47), thus offering added advantage of more closely corresponding to Mtb load.

Our group and others have shown that the activation, memory differentiation and functional profile of Mtb-specific CD4 T cells relates TB disease activity [7, 11, 12]. We found a positive association between the TB-CAD score and the expression of the activation marker HLA-DR on Mtb-specific CD4 T cells (p=0.0008, r=0.41), while moderate negative associations were observed with the expression of the memory marker CD27 and the TNF super family member CD153 on Mtb-specific CD4 T cells (p=0.008, r=-0.34 and p=0.003, r=-0.37) (**Fig. 1G**).

We compared TB-CAD scores to the earliest timepoint where a negative sputum culture result was registered (n=78). Those who were culture negative at baseline or at week 2 or 4 of ATT had significantly lower TB-CAD scores at baseline compared to those who only culture converted at or after week 8 ($p=8.8x10^{-6}$, **Fig. 1H**). A subset of 80 eligible participants underwent repeat HRCT with TB-CAD scoring after ATT completion (**Fig. 1I**). TB-CAD scores significantly decreased after ATT in most participants (median: 4.65 vs 1.55, $p=2x10^{-15}$). In five participants, the TB-CAD score increased but none of them experienced treatment failure or relapse during the 52-week follow-up period. While poor treatment adherence was not suspected, this cannot fully be excluded as pill counts or isoniazid urine testing were not conducted. However, TB-CAD scores only partially normalized, as TB-CAD scores post-ATT remained significantly higher than those of the control group (p=0.0024, **Fig. 1I**). The median fold change of TB-CAD scores between baseline and post-ATT was comparable, regardless of HIV status (**Fig. 1J**).

Overall, we show that the TB-CAD score correlated with TB bacillary load, as evidenced by its inverse correlation to both sputum Xpert Ct value and culture time to positivity, whilst significantly declining post-ATT, regardless of HIV status. Furthermore, TB-CAD scores correlated with blood monocyte count and CRP, both of which are markers of systemic inflammation, usually elevated in TB, and the latter poorly prognostic [13, 14].

Focusing on more specific readouts of TB disease activity, we evaluated the Mtb-specific CD4+ T-cell profile in relation to the TB-CAD score. CD153 has been implicated as marker of protection, with the proportion of Mtb-specific CD4 T cells expressing CD153 significantly lower in active compared to latent TB [11]. It is thus noteworthy that higher TB-CAD scores were associated with a more differentiated ($CD27^{low}$), highly activated ($HLA\text{-}DR^{high}$) Mtb-specific CD4+ T-cell profile with low CD153 expression. While these correlations are moderate, it is reassuring that Mtb-specific immune responses known to associate with TB disease activity [7, 11] also relate to radiographic disease extent.

The main limitations of our study include the absence of an IGRA positive control group, and the significant differences in ART uptake between the HIV-1 infected study groups. However, this does not detract from our main finding that the TB-CAD score offers an objective, quantitative readout of TB disease extent in the lung, regardless of HIV-1 status. This is significant as chest radiographs, commonly used for diagnostic and treatment monitoring purposes in TB, can frequently be normal or display non-specific features in TB-HIV coinfected individuals, specifically in those with low CD4 counts, who in turn represent those at highest risk of TB related mortality [8, 9]. Furthermore, these individuals are often sputum smear negative owing to low bacillary load, leading to treatment monitoring challenges [15]. This highlights the importance of investigating non-sputum-based technologies. Further studies are needed to validate our findings in an independent cohort, including participants with CD4 counts <50 cells/mm^3 and determine whether TB-CAD scores at earlier timepoints during ATT can predict outcome. We envisage that the TB-CAD score has potential utility in the clinical trial setting where real time, quantitative data reflecting efficacy of experimental drugs in reducing mycobacterial burden and inflammation, in both HIV-1-infected and -uninfected participants, is critical.

Acknowledgements: We would like to thank all the study participants as well as Lorraine Brydon and her dedicated team of radiographers at Morton and Partners at Vincent Pallotti Hospital, Cape Town, South Africa.

Contributors: E.d.B., C.R., R.J.W. and B.A. designed the study. E.d.B. recruited the study participants. E.d.B. and C.R. performed the whole blood assay. E.d.B. and C.R. performed the flow experiments, data analysis and interpretation. G.H.K., I.D., J.L., and J.G. derived the TB-CAD score, performed the quantitative HRCT analysis and curated the radiographic data. R.J.W., A.S. and C.R. obtained funding to support the project. C.R. and E.d.B. wrote the manuscript with all authors contributing to providing critical feedback.

A detailed method for the generation for the TB-CAD score can be obtained by contacting directly Prof Jonathan Goldin (jgoldin@mednet.ucla.edu) or Dr Grace Hyun J Kim (gracekim@mednet.ucla.edu).

Funding: This work was supported by grants from the National Institutes of Health (NIH) (U01AI115940 to R.J.W.) and (R21AI148027 to C.R). R.J.W. is supported by the Francis Crick Institute, which receives funds from Cancer Research UK (FC00110218), the Wellcome trust (FC00110218) and the UK Medical Research Council (FC00110218). R.J.W. is also supported by Wellcome (203135), and EDCTP (RIA2017T-2004). C.R. is supported by the European and Developing Countries Clinical Trials Partnership (EDCTP) programme supported under the European Union's Horizon 2020 programme (TMA2017SF-1951/TB-Spec, to C.R.). For the purposes of open access, the authors have applied a CC BY public copyright license to any author-accepted version arising from this submission.

Competing interests: None declared.

Figure legends:

Figure 1: TB-CAD score relationship with TB disease activity and treatment response in HIV-uninfected and HIV-infected participants.

A) Clinical characteristics of TB participants grouped according to their HIV status (HIV-, aviremic HIV+ and viremic HIV+). **B)** TB-CAD score in heathy controls (n=48) and pulmonary TB patients (n=104) at baseline. Bars represent medians. Statistical comparison was performed using the Mann-Whitney test. **C-G)** Relationship between TB-CAD score and Xpert Ct value (C), time to Mtb culture positivity, TTP (D), plasma CRP (E), blood monocyte absolute count (F), and the expression of HLA-DR, CD153 and CD27 on IFN-g producing Mtb-specific CD4+ T cells (G). Correlations were tested by a two-tailed non-parametric Spearman rank test. **H)** Relationship between TB-CAD score and time to Mtb culture conversion. Bars represent median. Statistical comparisons were defined using a Kruskal-Wallis test, adjusted for multiple comparisons (Dunn's test). **I)** Evolution of the TB-CAD score between baseline and 24-week post-ATT initiation in pulmonary TB patients (n=80). Bars represent medians. Statistical comparison was performed using the paired Wilcoxon ranked test. **J)** Fold change in the TB-CAD score between baseline and 24-week post-ATT initiation in patients grouped based on their HIV status. Bars represent medians. Statistical comparisons were defined using a Kruskal-Wallis test, adjusted for multiple comparisons (Dunn's test).

References

1. Ralph AP, Ardian M, Wiguna A, Maguire GP, Becker NG, Drogumuller G, Wilks MJ, Waramori G, Tjitra E, Sandjaja, Kenagalem E, Pontororing GJ, Anstey NM, Kelly PM. A simple, valid, numerical score for grading chest x-ray severity in adult smear-positive pulmonary tuberculosis. *Thorax* 2010: 65(10): 863-869.
2. Skoura E, Zumla A, Bomanji J. Imaging in tuberculosis. *Int J Infect Dis* 2015: 32: 87-93.
3. Yan C, Wang L, Lin J, Xu J, Zhang T, Qi J, Li X, Ni W, Wu G, Huang J, Xu Y, Woodruff HC, Lambin P. A fully automatic artificial intelligence-based CT image analysis system for accurate detection, diagnosis, and quantitative severity evaluation of pulmonary tuberculosis. *Eur Radiol* 2022: 32(4): 2188-2199.
4. Li K, Jiang Z, Zhu Y, Fan C, Li T, Ma W, He Y. A valuable computed tomography-based new diagnostic tool for severe chest lesions in active pulmonary tuberculosis: combined application of influencing factors. *Sci Rep* 2020: 10(1): 2023.
5. Rockwood N, du Bruyn E, Morris T, Wilkinson RJ. Assessment of treatment response in tuberculosis. *Expert Rev Respir Med* 2016: 10(6): 643-654.
6. Allwood BW, Maasdorp E, Kim GJ, Cooper CB, Goldin J, van Zyl-Smit RN, Bateman ED, Dawson R. Transition from Restrictive to Obstructive Lung Function Impairment During Treatment and Follow-Up of Active Tuberculosis. *Int J Chron Obstruct Pulmon Dis* 2020: 15: 1039-1047.
7. Riou C, Du Bruyn E, Ruzive S, Goliath RT, Lindestam Arlehamn CS, Sette A, Sher A, Barber DL, Wilkinson RJ. Disease extent and anti-tubercular treatment response correlates with Mycobacterium tuberculosis-specific CD4 T-cell phenotype regardless of HIV-1 status. *Clin Transl Immunology* 2020: 9(9): e1176.
8. Keiper MD, Beumont M, Elshami A, Langlotz CP, Miller WT, Jr. CD4 T lymphocyte count and the radiographic presentation of pulmonary tuberculosis. A study of the relationship between these factors in patients with human immunodeficiency virus infection. *Chest* 1995: 107(1): 74-80.
9. Nakiyingi L, Bwanika JM, Ssengooba W, Mubiru F, Nakanjako D, Joloba ML, Mayanja-Kizza H, Manabe YC. Chest X-ray interpretation does not complement Xpert MTB/RIF in diagnosis of smear-negative pulmonary tuberculosis among TB-HIV co-infected adults in a resource-limited setting. *BMC Infect Dis* 2021: 21(1): 63.
10. Adane T, Melku M, Ayalew G, Bewket G, Aynalem M, Getawa S. Accuracy of monocyte to lymphocyte ratio for tuberculosis diagnosis and its role in monitoring anti-tuberculosis treatment: Systematic review and meta-analysis. *Medicine (Baltimore)* 2022: 101(44): e31539.
11. Du Bruyn E, Ruzive S, Lindestam Arlehamn CS, Sette A, Sher A, Barber DL, Wilkinson RJ, Riou C. Mycobacterium tuberculosis-specific CD4 T cells expressing CD153 inversely associate with bacterial load and disease severity in human tuberculosis. *Mucosal Immunol* 2021: 14(2): 491-499.
12. Musvosvi M, Duffy D, Filander E, Africa H, Mabwe S, Jaxa L, Bilek N, Llibre A, Rouilly V, Hatherill M, Albert M, Scriba TJ, Nemes E. T-cell biomarkers for diagnosis of tuberculosis: candidate evaluation by a simple whole blood assay for clinical translation. *Eur Respir J* 2018: 51(3).
13. Lawn SD, Kerkhoff AD, Vogt M, Wood R. Diagnostic and prognostic value of serum C-reactive protein for screening for HIV-associated tuberculosis. *Int J Tuberc Lung Dis* 2013: 17(5): 636-643.
14. La Manna MP, Orlando V, Dieli F, Di Carlo P, Cascio A, Cuzzi G, Palmieri F, Goletti D, Caccamo N. Quantitative and qualitative profiles of circulating monocytes may help identifying tuberculosis infection and disease stages. *PLoS One* 2017: 12(2): e0171358.

15. Peters JS, Andrews JR, Hatherill M, Hermans S, Martinez L, Schurr E, van der Heijden Y, Wood R, Rustomjee R, Kana BD. Advances in the understanding of Mycobacterium tuberculosis transmission in HIV-endemic settings. *Lancet Infect Dis* 2019: 19(3): e65-e76.

Derivation of a high-resolution CT-based, semi-automated radiographic score in tuberculosis and its relationship to bacillary load and antitubercular therapy

Catherine Riou[1,2*], Elsa du Bruyn[1*], Grace Hyun J. Kim[3,4*], Irene da Costa[3,4], Jihey Lee[3,4], Alan Sher[5], Robert J. Wilkinson[1,6,7,8†], Brian W. Allwood[9†], Jonathan Goldin[3,4†].

[1] Wellcome Centre for Infectious Disease Research in Africa and Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Observatory, Cape Town, South Africa

[2] Division of Medical Virology, Department of Pathology, University of Cape Town, Observatory, Cape Town, South Africa

[3] Department of Radiology, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA

[4] UCLA Center for Computer Vision and Imaging Biomarkers, Los Angeles, CA, USA

[5] Immunobiology Section, Laboratory of Parasitic Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, USA

[6] Department of Medicine, University of Cape Town, Observatory, Cape Town, South Africa

[7] Department of Infectious Diseases, Imperial College London, W12 0NN, UK

[8] The Francis Crick Institute, London, NW1 1AT, UK

[9] Division of Pulmonology, Department of Medicine, Stellenbosch University and Tygerberg Hospital, Cape Town, South Africa

[*] and [†] These authors equally contributed to the work.

Correspondence to Dr Catherine Riou and Dr Elsa du Bruyn, University of Cape Town, Observatory, Cape Town, South Africa. cr.riou@uct.ac.za and eldubruyn@gmail.com

Word count
Title: 20 words
Main text: 1500 words
References: 15

To the Editor:

Efforts to curb the TB pandemic remain hindered by the lack of objective measures to quantify disease severity and track treatment success that are valid in both HIV-1-infected and -uninfected TB patients. Ralph *et al* developed a promising radiographic scoring system [1], with baseline scores being predictive of sputum smear conversion at two months, but it is reliant on skilled readers and has not been systematically validated in predominantly HIV infected study populations of variant CD4 counts. Superior to conventional chest radiography, High-resolution computed tomography (HRCT) is a highly sensitive tool to track endobronchial TB disease extent [2]. Although recent studies have made progress in development of semi-automated methods for TB diagnosis and quantification of disease severity on chest CT [3, 4], none have assessed how these radiographic scoring systems relate to mycobacterial burden, the immune response to TB and whether they can reliably quantify the effect of antitubercular therapy (ATT) on disease affected lung. Sputum culture conversion at two months remains the gold standard measure of efficacy in trials of new TB drugs and regimens [5]. However, there is renewed interest in non-sputum-based approaches for treatment monitoring as sputum TB culture is time-consuming (negative results may take up to 42 days in liquid culture) and does not yield results in cases of culture contamination.

In the present study, we derived a computer-assisted, semi-automated quantitative radiographic scoring system (TB-CAD) applied to chest HRCT. Briefly, HRCTs underwent quantitative analysis by applying texture-based computer-aided diagnosis (CAD) at the Center for Computer Vision and Imaging Biomarkers (University of California). A computer assisted, semi-automated quantitative radiographic score of TB disease extent, called TB-CAD, was developed by modification of a previously developed algorithm [6] to detect and quantify areas of abnormality including, cavitation, consolidation, nodules, scarring and airway disease. The algorithm was run after a segmentation algorithm isolated the lung parenchyma and the TB-CAD score was calculated as the percentage of pixels of TB related abnormality present within the lungs on chest HRCT of patients with pulmonary TB and healthy controls. A detailed method is provided as supplementary material. All participants were adults and recruited at the Site B Clinic, Khayelitsha, Cape Town, South Africa and provided written informed consent [7]. The study was approved by the UCT Human Research Ethics Committee (HREC:050/2015).

This study included 104 participants with newly diagnosed, drug-sensitive TB who tested sputum Xpert MTB/RIF (Cepheid) positive and underwent HRCT within 7 days of initiating ATT. 59.6% were HIV positive (n=62): 12 participants being aviremic and 50 having a detectable HIV-1 viral load. Clinical characteristics of TB patients are presented in **Fig 1A**. Eighty participants underwent repeat HRCT after completing antitubercular therapy (ATT). A group of 48 healthy controls (i.e., asymptomatic, IFN-γ release assay and sputum Xpert MTB/RIF negative and no previous TB) also underwent HRCT. 59.6% of the control participants were HIV positive (n=28), with 13 being aviremic (median CD4 count: 460 cell/mm^3) and 15 being viremic (median: 3.83 Log_{10} mRNA copies/ml, IQR: 1.91-4.29) and a median CD4 count of 364 cells/mm^3. The median age was comparable between the TB and control groups (36 vs 37 years old, respectively).

We examined the relationship between the TB-CAD score and TB disease activity, soluble inflammatory markers and the *Mycobacterium tuberculosis* (Mtb)-specific CD4 T cell profile in blood using flow cytometry [7], in a subset of the TB participants (n=60). The TB-CAD score was significantly higher in the pulmonary TB group at baseline compared to the healthy controls (median: 3.9 vs 0.8, p<0.0001, **Fig. 1B**). Moreover, when HIV- infected TB participants were grouped according to their absolute CD4 count (>200 or <200 cells/mm^3), the TB-CAD score was significantly elevated in both groups compared to the control group (p=0.0022 and p=0.0024, respectively). Of note, about a quarter of TB patients exhibited a TB-CAD score comparable to controls (<2), this sub-group was mostly constituted of HIV-infected participants (80%) and showed a trend towards higher proportion of negative Mtb culture compared to those with a TB-CAD score >2. This supports the observation that HIV/TB co-infection often presents with limited lung involvement, indicating a paucibacillary nature [8, 9].

There was a significant inverse correlation between TB-CAD score and sputum Xpert MTB/RIF cycle threshold (Ct) value at baseline (p=2.6x10^{-6}, r=-0.47), as well as sputum culture time to positivity (p=1.5x10^{-6}, r=-0.44), irrespective of HIV-1 status (**Fig. 1C and D**). Additionally, the TB-CAD score was significantly higher in TB participants with 3+ AFB smear compared those who were smear negative (median: 7 vs 2.85, p=0.0002, data not shown). The TB-CAD score correlated with plasma C-Reactive Protein (CRP) levels at baseline in the HIV-1-uninfected and aviremic HIV-1 infected TB groups (p=0.0001, r=0.56 and p=0.004, r=0.78) (**Fig. 1E**). However, this correlation was absent in the viremic HIV-1-infected group (p=0.63), possibly due to HIV-associated systemic inflammation. TB-CAD

scores correlated with blood monocyte counts of HIV-1 uninfected and HIV-1 infected participants (p=0.0003, r=0.53 and p=0.006, r=0.39) (**Fig. 1F**). As the monocyte to lymphocyte ratio (MLR) holds promise as TB biomarker [10], we also compared the relationship between sputum Mtb load and MLR or TB-CAD. We indeed found a significant correlation between MLR and Xpert Ct values (p=0.0014, r=-0.32). However, the TB-CAD showed a stronger correlation with Xpert Ct values ($p=2.6x10^{-6}$, r=-0.47), thus offering added advantage of more closely corresponding to Mtb load.

Our group and others have shown that the activation, memory differentiation and functional profile of Mtb-specific CD4 T cells relates TB disease activity [7, 11, 12]. We found a positive association between the TB-CAD score and the expression of the activation marker HLA-DR on Mtb-specific CD4 T cells (p=0.0008, r=0.41), while moderate negative associations were observed with the expression of the memory marker CD27 and the TNF super family member CD153 on Mtb-specific CD4 T cells (p=0.008, r=-0.34 and p=0.003, r=-0.37) (**Fig. 1G**).

We compared TB-CAD scores to the earliest timepoint where a negative sputum culture result was registered (n=78). Those who were culture negative at baseline or at week 2 or 4 of ATT had significantly lower TB-CAD scores at baseline compared to those who only culture converted at or after week 8 ($p=8.8x10^{-6}$, **Fig. 1H**). A subset of 80 eligible participants underwent repeat HRCT with TB-CAD scoring after ATT completion (**Fig. 1I**). TB-CAD scores significantly decreased after ATT in most participants (median: 4.65 vs 1.55, $p=2x10^{-15}$). In five participants, the TB-CAD score increased but none of them experienced treatment failure or relapse during the 52-week follow-up period. While poor treatment adherence was not suspected, this cannot fully be excluded as pill counts or isoniazid urine testing were not conducted. However, TB-CAD scores only partially normalized, as TB-CAD scores post-ATT remained significantly higher than those of the control group (p=0.0024, **Fig. 1I**). The median fold change of TB-CAD scores between baseline and post-ATT was comparable, regardless of HIV status (**Fig. 1J**).

Overall, we show that the TB-CAD score correlated with TB bacillary load, as evidenced by its inverse correlation to both sputum Xpert Ct value and culture time to positivity, whilst significantly declining post-ATT, regardless of HIV status. Furthermore, TB-CAD scores correlated with blood monocyte count and CRP, both of which are markers of systemic inflammation, usually elevated in TB, and the latter poorly prognostic [13, 14].

Focusing on more specific readouts of TB disease activity, we evaluated the Mtb-specific CD4+ T-cell profile in relation to the TB-CAD score. CD153 has been implicated as marker of protection, with the proportion of Mtb-specific CD4 T cells expressing CD153 significantly lower in active compared to latent TB [11]. It is thus noteworthy that higher TB-CAD scores were associated with a more differentiated ($CD27^{low}$), highly activated ($HLA\text{-}DR^{high}$) Mtb-specific CD4+ T-cell profile with low CD153 expression. While these correlations are moderate, it is reassuring that Mtb-specific immune responses known to associate with TB disease activity [7, 11] also relate to radiographic disease extent.

The main limitations of our study include the absence of an IGRA positive control group, and the significant differences in ART uptake between the HIV-1 infected study groups. However, this does not detract from our main finding that the TB-CAD score offers an objective, quantitative readout of TB disease extent in the lung, regardless of HIV-1 status. This is significant as chest radiographs, commonly used for diagnostic and treatment monitoring purposes in TB, can frequently be normal or display non-specific features in TB-HIV coinfected individuals, specifically in those with low CD4 counts, who in turn represent those at highest risk of TB related mortality [8, 9]. Furthermore, these individuals are often sputum smear negative owing to low bacillary load, leading to treatment monitoring challenges [15]. This highlights the importance of investigating non-sputum-based technologies. Further studies are needed to validate our findings in an independent cohort, including participants with CD4 counts <50 cells/mm^3 and determine whether TB-CAD scores at earlier timepoints during ATT can predict outcome. We envisage that the TB-CAD score has potential utility in the clinical trial setting where real time, quantitative data reflecting efficacy of experimental drugs in reducing mycobacterial burden and inflammation, in both HIV-1-infected and -uninfected participants, is critical.

Acknowledgements: We would like to thank all the study participants as well as Lorraine Brydon and her dedicated team of radiographers at Morton and Partners at Vincent Pallotti Hospital, Cape Town, South Africa.

Contributors: E.d.B., C.R., R.J.W. and B.A. designed the study. E.d.B. recruited the study participants. E.d.B. and C.R. performed the whole blood assay. E.d.B. and C.R. performed the flow experiments, data analysis and interpretation. G.H.K., I.D., J.L., and J.G. derived the TB-CAD score, performed the quantitative HRCT analysis and curated the radiographic data. R.J.W., A.S. and C.R. obtained funding to support the project. C.R. and E.d.B. wrote the manuscript with all authors contributing to providing critical feedback.
A detailed method for the generation for the TB-CAD score can be obtained by contacting directly Prof Jonathan Goldin (jgoldin@mednet.ucla.edu) or Dr Grace Hyun J Kim (gracekim@mednet.ucla.edu).

Funding: This work was supported by grants from the National Institutes of Health (NIH) (U01AI115940 to R.J.W.) and (R21AI148027 to C.R). R.J.W. is supported by the Francis Crick Institute, which receives funds from Cancer Research UK (FC00110218), the Wellcome trust (FC00110218) and the UK Medical Research Council (FC00110218). R.J.W. is also supported by Wellcome (203135), and EDCTP (RIA2017T-2004). C.R. is supported by the European and Developing Countries Clinical Trials Partnership (EDCTP) programme supported under the European Union's Horizon 2020 programme (TMA2017SF-1951/TB-Spec, to C.R.).

Competing interests: None declared.

Figure legends:

Figure 1: TB-CAD score relationship with TB disease activity and treatment response in HIV-uninfected and HIV-infected participants.

A) Clinical characteristics of TB participants grouped according to their HIV status (HIV-, aviremic HIV+ and viremic HIV+). **B)** TB-CAD score in heathy controls (n=48) and pulmonary TB patients (n=104) at baseline. Bars represent medians. Statistical comparison was performed using the Mann-Whitney test. **C-G)** Relationship between TB-CAD score and Xpert Ct value (C), time to Mtb culture positivity, TTP (D), plasma CRP (E), blood monocyte absolute count (F), and the expression of HLA-DR, CD153 and CD27 on IFN-g producing Mtb-specific CD4+ T cells (G). Correlations were tested by a two-tailed non-parametric Spearman rank test. **H)** Relationship between TB-CAD score and time to Mtb culture conversion. Bars represent median. Statistical comparisons were defined using a Kruskal-Wallis test, adjusted for multiple comparisons (Dunn's test). **I)** Evolution of the TB-CAD score between baseline and 24-week post-ATT initiation in pulmonary TB patients (n=80). Bars represent medians. Statistical comparison was performed using the paired Wilcoxon ranked test. **J)** Fold change in the TB-CAD score between baseline and 24-week post-ATT initiation in patients grouped based on their HIV status. Bars represent medians. Statistical comparisons were defined using a Kruskal-Wallis test, adjusted for multiple comparisons (Dunn's test).

References

1. Ralph AP, Ardian M, Wiguna A, Maguire GP, Becker NG, Drogumuller G, Wilks MJ, Waramori G, Tjitra E, Sandjaja, Kenagalem E, Pontororing GJ, Anstey NM, Kelly PM. A simple, valid, numerical score for grading chest x-ray severity in adult smear-positive pulmonary tuberculosis. *Thorax* 2010: 65(10): 863-869.
2. Skoura E, Zumla A, Bomanji J. Imaging in tuberculosis. *Int J Infect Dis* 2015: 32: 87-93.
3. Yan C, Wang L, Lin J, Xu J, Zhang T, Qi J, Li X, Ni W, Wu G, Huang J, Xu Y, Woodruff HC, Lambin P. A fully automatic artificial intelligence-based CT image analysis system for accurate detection, diagnosis, and quantitative severity evaluation of pulmonary tuberculosis. *Eur Radiol* 2022: 32(4): 2188-2199.
4. Li K, Jiang Z, Zhu Y, Fan C, Li T, Ma W, He Y. A valuable computed tomography-based new diagnostic tool for severe chest lesions in active pulmonary tuberculosis: combined application of influencing factors. *Sci Rep* 2020: 10(1): 2023.
5. Rockwood N, du Bruyn E, Morris T, Wilkinson RJ. Assessment of treatment response in tuberculosis. *Expert Rev Respir Med* 2016: 10(6): 643-654.
6. Allwood BW, Maasdorp E, Kim GJ, Cooper CB, Goldin J, van Zyl-Smit RN, Bateman ED, Dawson R. Transition from Restrictive to Obstructive Lung Function Impairment During Treatment and Follow-Up of Active Tuberculosis. *Int J Chron Obstruct Pulmon Dis* 2020: 15: 1039-1047.
7. Riou C, Du Bruyn E, Ruzive S, Goliath RT, Lindestam Arlehamn CS, Sette A, Sher A, Barber DL, Wilkinson RJ. Disease extent and anti-tubercular treatment response correlates with Mycobacterium tuberculosis-specific CD4 T-cell phenotype regardless of HIV-1 status. *Clin Transl Immunology* 2020: 9(9): e1176.
8. Keiper MD, Beumont M, Elshami A, Langlotz CP, Miller WT, Jr. CD4 T lymphocyte count and the radiographic presentation of pulmonary tuberculosis. A study of the relationship between these factors in patients with human immunodeficiency virus infection. *Chest* 1995: 107(1): 74-80.
9. Nakiyingi L, Bwanika JM, Ssengooba W, Mubiru F, Nakanjako D, Joloba ML, Mayanja-Kizza H, Manabe YC. Chest X-ray interpretation does not complement Xpert MTB/RIF in diagnosis of smear-negative pulmonary tuberculosis among TB-HIV co-infected adults in a resource-limited setting. *BMC Infect Dis* 2021: 21(1): 63.
10. Adane T, Melku M, Ayalew G, Bewket G, Aynalem M, Getawa S. Accuracy of monocyte to lymphocyte ratio for tuberculosis diagnosis and its role in monitoring anti-tuberculosis treatment: Systematic review and meta-analysis. *Medicine (Baltimore)* 2022: 101(44): e31539.
11. Du Bruyn E, Ruzive S, Lindestam Arlehamn CS, Sette A, Sher A, Barber DL, Wilkinson RJ, Riou C. Mycobacterium tuberculosis-specific CD4 T cells expressing CD153 inversely associate with bacterial load and disease severity in human tuberculosis. *Mucosal Immunol* 2021: 14(2): 491-499.
12. Musvosvi M, Duffy D, Filander E, Africa H, Mabwe S, Jaxa L, Bilek N, Llibre A, Rouilly V, Hatherill M, Albert M, Scriba TJ, Nemes E. T-cell biomarkers for diagnosis of tuberculosis: candidate evaluation by a simple whole blood assay for clinical translation. *Eur Respir J* 2018: 51(3).
13. Lawn SD, Kerkhoff AD, Vogt M, Wood R. Diagnostic and prognostic value of serum C-reactive protein for screening for HIV-associated tuberculosis. *Int J Tuberc Lung Dis* 2013: 17(5): 636-643.
14. La Manna MP, Orlando V, Dieli F, Di Carlo P, Cascio A, Cuzzi G, Palmieri F, Goletti D, Caccamo N. Quantitative and qualitative profiles of circulating monocytes may help identifying tuberculosis infection and disease stages. *PLoS One* 2017: 12(2): e0171358.

15. Peters JS, Andrews JR, Hatherill M, Hermans S, Martinez L, Schurr E, van der Heijden Y, Wood R, Rustomjee R, Kana BD. Advances in the understanding of Mycobacterium tuberculosis transmission in HIV-endemic settings. *Lancet Infect Dis* 2019: 19(3): e65-e76.

Figure 1

A

	TB patients		
	HIV-	Avir HIV+	Vir HIV+
Number of participants	42	12	50
Log_{10} HIV VL (copies mRNA/ml), Median [IQR]	na	<20	4.91 [4.56-5.36]
CD4 count (cells/mm³), Median [IQR]	nd	519 [358-775]	159 [71-274]
CD4 count >200 cells/mm³ (%, n)	-	100% (n=12)	40% (n=20)
CD4 count <200 cells/mm³ (%, n)	-	0%	60% (n=30)

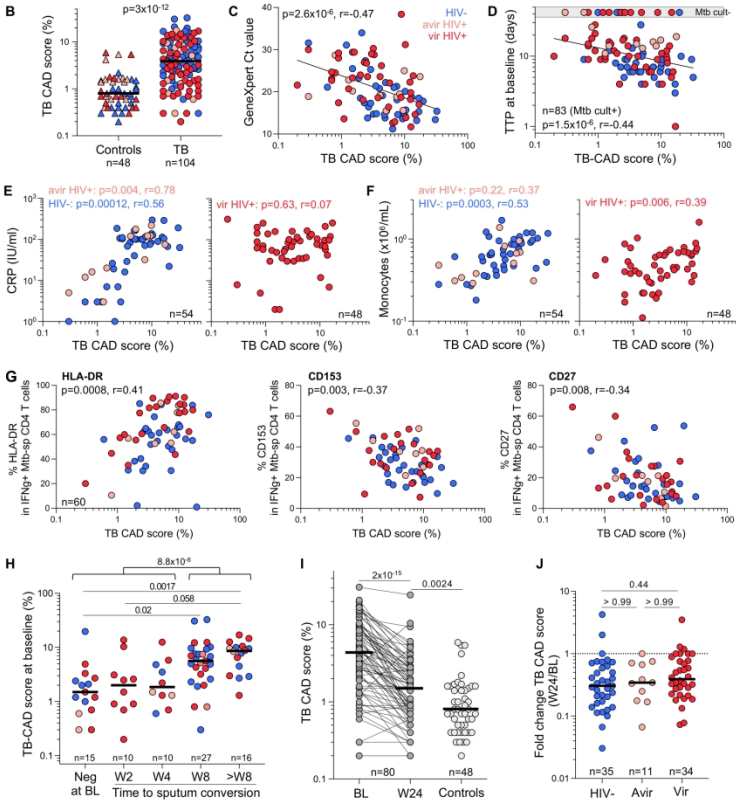


Figure 1

190x275mm (300 x 300 DPI)

Reporting checklist for case-control study.

Based on the STROBE case-control guidelines.

Instructions to authors

Complete this checklist by entering the page numbers from your manuscript where readers will find each of the items listed below.

Your article may not currently address all the items on the checklist. Please modify your text to include the missing information. If you are certain that an item does not apply, please write "n/a" and provide a short explanation.

Upload your completed checklist as an extra file when you submit to a journal.

In your methods section, say that you used the STROBE case-controlreporting guidelines, and cite them as:

von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies.

		Reporting Item	Page Number
Title and abstract			
Title	#1a	Indicate the study's design with a commonly used term in the title or the abstract	1
Abstract	#1b	Provide in the abstract an informative and balanced summary of what was done and what was found	n/a
Introduction			
Background / rationale	#2	Explain the scientific background and rationale for the investigation being reported	2
Objectives	#3	State specific objectives, including any prespecified hypotheses	2
Methods			
Study design	#4	Present key elements of study design early in the paper	2-3
Setting	#5	Describe the setting, locations, and relevant dates, including periods of	2

		recruitment, exposure, follow-up, and data collection	
Eligibility criteria	#6a	Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls. For matched studies, give matching criteria and the number of controls per case	2
Eligibility criteria	#6b	For matched studies, give matching criteria and the number of controls per case	2
	#7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	2 and 4
Data sources / measurement	#8	For each variable of interest give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group. Give information separately for cases and controls.	2
Bias	#9	Describe any efforts to address potential sources of bias	4
Study size	#10	Explain how the study size was arrived at	n/a
Quantitative variables	#11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why	3
Statistical methods	#12a	Describe all statistical methods, including those used to control for confounding	6
Statistical methods	#12b	Describe any methods used to examine subgroups and interactions	n/a
Statistical methods	#12c	Explain how missing data were addressed	n/a
Statistical methods	#12d	If applicable, explain how matching of cases and controls was addressed	n/a
Statistical methods	#12e	Describe any sensitivity analyses	n/a
Results			
Participants	#13a	Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed. Give information separately for cases and controls.	2
Participants	#13b	Give reasons for non-participation at each stage	n/a
Participants	#13c	Consider use of a flow diagram	n/a

Descriptive data	#14a	Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders. Give information separately for cases and controls	2
Descriptive data	#14b	Indicate number of participants with missing data for each variable of interest	2
Outcome data	#15	Report numbers in each exposure category, or summary measures of exposure. Give information separately for cases and controls	2
Main results	#16a	Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	n/a
Main results	#16b	Report category boundaries when continuous variables were categorized	3-4
Main results	#16c	If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	n/a
Other analyses	#17	Report other analyses done—e.g., analyses of subgroups and interactions, and sensitivity analyses	3-4
Discussion			
Key results	#18	Summarise key results with reference to study objectives	3
Limitations	#19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias.	4
Interpretation	#20	Give a cautious overall interpretation considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence.	4
Generalisability	#21	Discuss the generalisability (external validity) of the study results	4
Other Information			
Funding	#22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	5