

Tweaking algorithms. Technopolitical issues associated with artificial intelligence based tuberculosis detection in global health

DIGITAL HEALTH
Volume 10: 1–11
© The Author(s) 2024
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/20552076241239778
journals.sagepub.com/home/dhj



Pierre-Marie David^{1,2} , Julien Onno^{1,2,*}, Jessica Pourraz³ and Faiz Ahmad Khan^{2,4,5}

Abstract

Computer-aided detection algorithms based on artificial intelligence are increasingly being tested and used as a means for detecting tuberculosis in countries where the epidemic is still present. Computer-aided detection tools are often presented as a global solution that can be deployed in all the geographical areas concerned by tuberculosis, but at the same time, they need to be adjusted and calibrated according to local populations' characteristics. The aim of this article is to analyze the tensions between the standardization of computer-aided detection algorithms and their local adaptation and the political issues associated with these tensions. We undertook a qualitative analysis of practices associated with tuberculosis detection algorithms in different contexts, contrasting the perspectives of various stakeholders. Algorithms embed the promise of standardization through automation and the bypassing of variable human expertise such as that of radiologists, they are nonetheless objects of local practices that we have characterized as “tweaking.” This work of tweaking reveals how the technology is situated but also the many concerns of the users and workers (insertion in care, control over infrastructure, and political ownership). This should be better considered to truly make computer-aided detection innovative tools for tuberculosis management in global health.

Keywords

Algorithm, artificial intelligence, detection, tuberculosis, tweak, diagnostic

Submission date: 26 September 2023; Acceptance date: 29 February 2024

Introduction

Computer-aided detection (CAD) algorithms based on artificial intelligence (AI) are increasingly being tested and used as a means for detecting tuberculosis (TB), which is one of the top infectious killers in the world with 1.3 million deaths a year and more than 10 million people who fall ill every year.¹ While diagnosing TB has always represented a major medical challenge, diagnostic tools have expanded considerably in recent years. These tools include molecular tests for the detection of TB and drug resistance, various biomarker-based tests for the detection of TB, as well as CAD for TB screening using digital chest radiography, which is the subject of this article. The promises associated with CAD are many and varied: Improved detection, automation, ease of use, enhancing or even replacing radiologists, reduced costs, savings on

staff and materials, etc. CAD is often presented as a global solution that can be deployed in all the geographical areas concerned by TB. At the same time, they need to be

¹Faculty of Pharmacy, University of Montréal, Montréal, Canada

²OBVIA, Observatoire sur les impacts sociétaux de l'IA et du numérique, Québec, Canada

³Centre Émile Durkheim (UMR 5116), Sciences Po Bordeaux, Bordeaux, France

⁴McGill International TB Centre, Research Institute of the McGill University Health Centre, Montréal, Canada

⁵Departments of Medicine & Epidemiology, Biostatistics & Occupational Health, McGill University, Montréal, Canada

*CEThics, ISTC – Lille, France

Corresponding author:

Pierre-Marie David, Faculty of Pharmacy, University of Montréal, Montréal, Canada.

Email: pierre-marie.david@umontreal.ca



calibrated according to local populations' characteristics, in a certain form of tension between standardization and local adaptation. The aim of this article is to analyze the tensions between the standardization of CAD use and their local adaptation, and the technical and political issues associated with these tensions. Those issues, we argue, can be analyzed and understood through the tweaking of algorithms, a usually invisible work, which we define not only as the fine-tuning or optimization of an algorithm's parameters to improve its performance, but as the process which involves a combination of theoretical and practical knowledge of the algorithm itself, an understanding of the social application context and practical experience in engaging with TB care.

Abnormality scores, thresholds, and the need for tweaks

In 2021, the World Health Organization (WHO) issued a new recommendation on the use of CAD, approving its use as an alternative to human interpretation of radiographs for the detection of TB.² One of the conditions is specified in the WHO report:

Evaluations showed substantial variation in diagnostic accuracy across different contexts, implying that the use of CAD will require calibration for the purpose and setting in which it will be implemented.²

When CAD is used to analyze a digital chest radiography, for each image, it outputs an abnormality score. Higher scores indicate greater abnormality. To distinguish between presumed active TB and the absence of TB, a detection threshold must be applied. When a CAD score is below the threshold, the possibility of active TB is excluded. When the CAD score is above the threshold, then active TB is detected, and further investigation is required, typically through sputum testing for TB.

There were several reasons why WHO deemed it preferable to use CAD with local calibration. The accuracy of CAD varied between study sites, and within sites between patient subpopulations; for example, it seemed to miss more active TB disease amongst people living with human immunodeficiency virus (HIV), and amongst people who had already had TB, it resulted in more false-positive classifications. Hence, each site or clinic applying CAD should identify the best score to apply as a threshold, and potentially consider identifying separate thresholds for subpopulations. The adjustment practices are important in terms of both individual care and epidemiological impact. In practice, the epidemiological studies needed to calibrate the CAD were left to the research units, often based in the North, which supervised the CAD deployment pilot projects.

Such epidemiological studies are nevertheless very difficult and costly to carry out under real-life conditions. The

high TB burden places where CAD would be most useful are often also those where a paucity of resources makes undertaking rigorous calibration studies potentially quite challenging. To bridge the gap between theoretical requirements and real-life conditions, an abundance of guides and webinars have followed the WHO's recommendation to help users and set the standard for a practice that remains complex. The responsibilities involved in regulating these devices in general, and their calibration in particular, have both technical and political implications. Concrete calibration practices thus lie at the crossroads between the tool's promise of global health and concrete local uses, with their share of economic, social, and material determinants.

The calibration of diagnostic and technological tools is a social and technical process in medicine and global health, involving tensions between standardization and localization.³ Moreover, calibration is often analyzed from epistemological, socio-historical, and analytical perspectives,⁴ but in the case of CAD technologies, through the adjustments to this calibration, another aspect emerges, a socio-political one. In the first part, we will shed some light on the results of our survey, showing how the calibration of CAD tools continues to be uncertain regarding the specificities of TB. In the second part, we will describe practices that we describe as "tweaking," which we define as a variety of ways local actors appropriate technologies, including calibrating them in the field, in parallel with the recommendations of global health institutions. We will then discuss these results, showing how CAD technologies are not automatic solutions, but rather part of work and maintenance procedures that are largely underestimated. This technological workspace reveals the possibilities and limits of political appropriation of algorithms tools in global health.

Conceiving tweaking as labor and care for technologies

The very term calibration can have several meanings. While calibration refers to the calibration and adjustment of equipment to maximize its accuracy, calibration can also be applied to social and political processes, particularly those relating to standardization in global health.⁵ So, beyond the technical aspect, new technological tools bring about calibration phenomena in evaluation and regulation processes. At a public and institutional level, public health interventions can be "calibrated" according to the emergence of new technologies.⁶ Calibrations are also played out between the technical elements that automate detection and human response capabilities, and depend on conditions that play out outside the diagnostic tool alone, local contexts, populations specifications, epidemic particularities, use in general population screening or hospital symptomatic triage, etc.

In this article, we approach detection technologies through a critical anthropology of technologies and related practices. This enables to qualify the idea of automation/standardization of diagnosis by detection algorithms. We explore how technologies and tools for the automatic detection of TB are also at the heart of many human adjustment processes and use. We define “tweaking” as the term to describe and analyze the adjustments and tinkering that take place at every stage in the stabilization of the technical tool in various contexts. We analyze “tweaking” actions in a wider sense, going beyond the technical aspects to the level of regulation and political appropriation to better understand how global algorithms are also localized in the practice. These elements are very much in evidence in discussions about CAD, which depend on infrastructures, local objectives, uses in context, algorithmic models that are constantly being evaluated as new versions are created. Whereas algorithms have been critically analyzed through their construction,⁷ we aim at analyzing algorithms uses and insertion into their context through this tweaking work.

Indeed, the automation generated by the widespread use of AI is leading to a redistribution of tasks in the workplace. Two types of carers seem to be emerging with the arrival of algorithms in healthcare infrastructures. Firstly, the “northern” carers, from the academic world, who are able to promote the technology through efficiency studies. Then there are the “workers in the shadows” such as field technicians—whose work may be devalued, like click workers⁸—who use CAD in conditions other than the laboratory. One of the aims of this article is to shed light on these two types of work and their differential valuation. One of the hypotheses is that “middle workers” specifically construct the technological universe specific to CAD through tweakings; a work that is often little valued or even invisible, and which is just as important as the preliminary design work carried out by developers.⁹

Methods

We therefore undertook a qualitative inquiry of these tweaking practices in different contexts from 2020 to August 2023. The first step was to analyze the documents providing a framework for calibration, in particular those provided by the WHO, Foundation for Innovative New Diagnostics (FIND which has become Diagnosis for all), the StopTB partnership, an international partnership specializing in TB issues, TB funding and in the evaluation of new diagnostic tools. These documents contain several elements, including scientific findings and practical recommendations. We also watched and analyzed online webinars dealing with the calibration of detection tools and the infrastructural conditions that enable CAD to operate in the different contexts studied.

Our method also consisted of contrasting the perspectives of various stakeholders ($n = 43$) on calibration procedures and times (Table 1). We conducted a series of interviews with

Table 1. Research participants.

Perspective	Institution	Location	Nb of interviews
Computer-aided detection (CAD) developers ($n = 12$)	Company A	Anonymised	3
	Company B	Anonymised	6
	Company C	Anonymised	2
	Company D	Anonymised	1
Regulators ($n = 3$)	International Health Organization A	Anonymised	1
	International Health Organization B	Anonymised	2
Facilitators ($n = 13$)	International Health Organization A	Geneva	1
	International Health Organization B	Geneva	4
	International Development Organization A	Anonymised	2
	International Development Organization B	Anonymised	4
	International NGO headquarters	Europe	2
Users ($n = 15$)	NGO	Eastern Africa	1
	NGO	Southern Africa	2
	NGO	South East Asia	5
	NGO	Western Africa	4
	Academic	Africa	2
	Academic	North America	1

regulators (International health organizations) and facilitators (Global Fund, StopTB Partnership, FIND), giving an overview of how (a) the score, (b) the threshold, and (c) calibration as an iterative process were understood. Another series of interviews was conducted with leading users of the technologies located in Africa, Europe, South, and Southeast Asia, which provided information on the complexity of calibration and the methods used, which are much more iterative and dependent on local contexts.

The main limitation of this research is that we were not able to observe first-hand the tweaking work on site and the practices of the people working with CAD. Even though we tried, through interviews, to situate people in their contexts, there is a lack of in situ observations that could be the subject of future research. We feel, however, that we have addressed the main issues revealed by tweaking practices through the concerns expressed by users in our interviews, or through online workshops and trainings on the use of CAD which we attended to.

Results

Uncertainties and calibration

CAD algorithms classify digital chest images, resulting in a numerical score. The software enables a score from 0 to 100 to be established for each image. The score generated by CAD software is often referred to as an abnormality score. However, in the discourse and recommendations of global health actors, it is repeatedly stated that the abnormality score should not in itself be considered as a probability, as these abnormality scores are not necessarily linearly linked to an objective probability that the images contain signs of TB. Based on this abnormality score, the software's associated decision is based on a cut-off point. Otherwise known as the threshold score or operating point, it determines the boundary between images with "absent abnormalities" and those with "suggestive abnormalities" of TB. The key to calibration for local practices, whether screening in the general population or hospital triage, lies in setting this threshold. The score threshold is representative of the test's expectations of sensitivity (% of truly positive patients) and specificity (% of truly non-diseased negatives). Since 2014, the TB community has considered acceptable tests with at least 90% sensitivity and 70% specificity.

However, these scores—and therefore, the thresholds—are not standardized in the same way between different developers' CAD softwares and versions. In fact, CAD software is developed by several companies whose algorithms have been constructed differently. The abnormality score distributions will therefore differ. There is also substantial variability within the same software—successive version updates have an impact on the abnormality score distribution, as the algorithms are refined several times,

meaning that repeat calibration would be needed after every software update. Questions about scores and thresholds have led to a series of discussions and even misunderstandings about how the algorithms work. The global health community dealing with TB then discussed the meaning of this score, generally out of 100, which became equivocal with a rate or a percentage.

The abnormality scores is not a percentage. It does correlate. As you go higher, generally speaking, there is almost always a higher positive predictive value or there's going to be a higher positivity rate at that score, if you're doing a large population. But it is not $90 = 90\%$. And there was a lot of education that needed to go on early when you gave a score that was numeric close to 100, yet that person did not have TB. And not only did that person not have TB, more than 50% of those people did not have TB. [...] It is not intuitive. (Interview 1, International TB organization)

Finally, there was another misunderstanding in the global health community regarding the thresholds to be applied according to the scores obtained, leading to a presumption of TB or not. The continuous variable of the abnormality score then becomes a "yes" or "no" variable, leading to confirmatory biological tests (molecular biology confirmation tests such as GeneXpert), or not. The reduction of a complex reality such as TB to a single figure is part of a complex social process that generates a certain number of doubts on the part of those involved in the technology and the care associated with it. As analyzed in the literature, the level of trust placed in AI systems conditions the use that can be made of them.¹⁰ What is known as calibration is also at play in the expectations raised by usage compared with the actual possibilities of the tools using AI. Two biases are often considered in practice: A so-called over-reliance bias and an algorithmic aversion bias, i.e., a complete refusal to use algorithms.^{11,12} The process of building trust in the tool is a gradual one, involving many players. The first stage in this process of building trust has been ensured by numerous scientific publications evaluating CAD. But this issue of calibration and finding a "sweet spot" in the effectiveness of tools remains a challenge at the intersection of knowledge, confidence, the actual practices of users, and the health systems in which CAD are applied.

A global recommendation and various practices. The crucial stage of calibration raises questions for tool designers, evaluators and technology users alike. Within the calibration recommendation module proposed by the WHO,¹³ a certain number of conditions must be met for this adjustment of the technologies. The WHO recommendation of 2021 clearly states that one of the conditions for the use of CAD is the reference to epidemiological studies. CAD

score is then adjusted to the context and characteristics of the populations concerned (age, HIV co-infections in particular). These studies require usable clinical databases and demographics, and access to standardized bacteriological tests such as GeneXpert.

However, this method of calibration based on large-scale epidemiological studies appears to be a particularly time-consuming and cost-demanding stage, and the players interviewed do not necessarily have the same point of view on this stage. Some consider it too restrictive and may discourage future users. Faced with this situation, iterative calibration based on the realities on the ground is sometimes preferred. The interviewees, faced with the problem of determining thresholds, also gave concrete examples of how they are used:

An easy way to understand this: a 70-year-old person with a CAD score of 80, is not the same as a 20-year-old person with a CAD score of 50 or 60. But right now if I take a CAD score cut-off, I will be testing the 70-year-old person and not the 20-year-old person with a 60 cut-off. So it doesn't make any sense and it isn't efficient. It is not going to be scalable. (Interview 2, TB Organization in South Asia)

In the end, far from the WHO's initial recommendations, calibration is more of an "art," a human trade-off based on economic and diagnostic resources. Calibration is part of a chain of relationships involving radiologists in the field, screening program policies, national TB program choices, etc. The adjustment of CAD tools appears "messy" and "still evolving" for some of the participants in a context where resources and time are limited, which more broadly reflects the emergency context surrounding TB epidemics.

Calibration responsibility. In most cases, developers state that calibration is the responsibility of the "customer" (national TB programs, pilot projects, etc.). Some companies state that they do not carry out the calibration stage, while others provide more direct support to their customers. The question of who is responsible for calibrating the tools came up regularly in the interviews. There are major issues at stake, insofar as this technical procedure prefigures, in part, an individual's entry into a care process:

And then there were questions on CAD threshold. So whose responsibility is it to define the CAD threshold? Should implementers be budgeting and doing that operational research to find the CAD threshold, or should the companies already be doing that, and then telling you in your context that this is the context? I remember questions like that from that session, yeah. (Interview 3, International health organization)

The clear answer to this question from global health implementers is that it is up to the group implementing the technology, i.e., the customers, to decide:

If you talk to the groups, and you probably will, who have done the biggest scale of this stuff, they never did these studies. They can't even tell you what their optimal threshold was, but they can tell you very well how many people they can screen per day and where they were looking at to get. And they would move it up when they didn't have as much money and move it down when they did have money. They knew that they were missing people. They could do some studies and say "if we retest people from range 40 to 50, we can quantify that we are missing". They could tell you that, but it didn't matter, because they didn't have enough tests or enough space. So it is really up to the group that is implementing. It is not regulated in that way. (Interview 4, International TB Organization)

Responsibility for calibration is therefore not at all clear and ultimately lies with users, who are rarely equipped to calibrate CAD in the way recommended by the WHO. Ultimately, the impossibility of standardizing calibration has led to a proliferation of practice guides and webinars aimed at sharing local practices and providing "bottom-up" guidance. Many users are also highly critical of the standardization recommendations. There are several reasons for this. Firstly, it slows down the implementation of this tool, and in an "accelerating global health" context,¹⁴ precautions are often seen as obstacles. Secondly, some stakeholders point out that these recommendations were based on certain systems that did not provide a default threshold. However, many designers now offer default thresholds. It is then up to the user to optimize their use, if they can. This situation accentuates inequalities in health, since the best-equipped users will be able to calibrate, while those with less time and resources will use default thresholds and manage to do as they can. Dangers of changes in thresholds with version updates have also been pointed in the recent literature.¹⁵

A new evaluation infrastructure. Noting the difficulties of calibration, some organizations promoting the use of CAD have begun to offer evaluation platforms, facilitating the validation of CAD tools for different populations. In this respect, the platforms are also infrastructures that aggregate datasets from different sources, particularly in the context of migration or active case finding. This is the case of FIND, the institution at the center of the CAD promotion and evaluation network. FIND has set up a validation platform that allows them to independently evaluate AI-based diagnostics including CAD software for x-ray or other images and predominantly for TB but also for other diseases. Such evaluation platforms are therefore concrete responses

to calibration issues. This infrastructure is also positioned as the heart of the WHO's prequalification process of CAD. The agreement between the WHO and FIND enables this infrastructure to be positioned in this way.

Relying on such a platform, the WHO is in fact in the process of setting in motion a long-awaited prequalification process aligned on the one used for medicines¹⁶ to validate CAD products and their uses. It should be remembered that this process is unprecedented on the scale of the WHO and global health, while certain jurisdictions such as Canada, the USA, and Europe are in the process of adopting rules akin to this prequalification for AI-based detection tools. WHO working groups on the issue of prequalification have embarked on discussions as well as multiple assessments that are forms of regulatory tweaking at several scales given the difficulties and obstacles that remain significant, particularly in the case of AI technologies that are constantly evolving.

How does a PQ [prequalification] work when a software changes every year? And between the versions, the underlying algorithm can change dramatically. You never know. We will never know because it is a business secret and engineers themselves don't even know. So the only way to assess which projects are used or whether they're up to different versions, which versions they are up to, it is to see how they actually perform. I don't want to say forget about how it works, but if it works, then it works. (interview 5, International TB Organization)

The people in charge of regulation are aware of the task, its technical difficulty and its political importance for the WHO. Regulation is essential to the legitimization of global health institutions, which are forced to adapt to new and changing tools.¹⁷ However, while the international infrastructure for this regulation is proving difficult to put in place, local and national players are more directly affected by the need to cope with their CAD calibrations.

Algorithmic tweakings and political stakes

Economical conditions associated with CAD and other technologies. CAD detection technology is intimately linked with another technology, in this case, biomolecular diagnostics, GeneXpert testing. As the entanglement of CAD in a more global network of players and technologies becomes more apparent, the promise of automation, so central to the rhetoric about the potential of AI, becomes less obvious:

The linkage to confirmation tests is to be honest, a weak point. (Interview 6, Local TB Organization).

Adjusting calibration means finding what is convenient, or what is pragmatic. Participants tweak and adjust to their

own experiences, demonstrating their ability to appropriate CAD tools both practically and politically. One of the most important conditions for determining detection thresholds is the availability of GeneXpert cartridges. For example, an interviewee based in West Africa highlights how calibration is carried out according to the economic resources available:

When we started at 35, we had a huge pool eligible to go for GeneXpert Machine. So our sample size became very huge. So we had to increase the bar to 50. (...) We had two machines, one was 50 and one was 60. We tried those separate thresholds to see the sensitivity of the machine and how many presumptive cases we were getting. And we had them different for different periods. [...] Based on those trials for thresholds, we eventually agreed and said "ok, 50 was an acceptable threshold that would give us good presumptive cases that would eventually get us the people to be tested using the GeneXpert machine". So again, this is from our country experience. (Interview 6, Local TB Organization)

In these cases, the calibration experiments show precisely how CAD can be tools for adapting to existing conditions, rather than revolutionary tools for creating a radically new reality. Indeed, CAD can act here as a palliative to a structural shortage of GeneXpert tests, which cost from \$8 to \$10 per test. Beyond this link with GeneXpert testing, these practices also raise questions about the ways in which CAD could adapt to other systemic difficulties, such as access to anti-TB treatment, or more structurally, the lack of data infrastructure in a given territory.

« *Tweaking* » as *political ownership*. Other tweaking is carried out and is more directly political. An international NGO in the North found itself in a situation, which led it to view the calibration methodology proposed by the WHO as an unacceptable dependency. For this NGO, allowing themselves to make calibration adjustments also became a way of appropriating the technology:

This is WHO's recommendation for calibration. They asked for a calibration study. It's impossible to do that, and it's a bad WHO recommendation. It's going to prevent people from using CAD. We don't do that. Even in the country where we operate, they did a trial and they saw what it produced. The protocol they're suggesting is very difficult. Our study is already focused on feasibility and accessibility. The first thought is that it's not feasible to do a calibration study.

(...) In fact, for us, when we do mass screening, you have to decide where you can put the test, so that you have enough Xpert machines to test all the samples you're going to have. If you don't have enough Xpert machines, there's no point (Interviewee 7, international Health NGO).

Beyond the technical act and the economic quantities involved, calibration constitutes a political practice of technology appropriation; a way of giving oneself the possibility of breaking a dependence that comes with this technology. Other local NGOs have come to the same conclusion. As part of a large-scale screening program in South Asia, threshold adjustment has become a public health necessity as compared to standard practices accompanying CAD tools:

(...) it was also very different in men and women. It was very different in different age groups. So those thresholds, having 1 CAD and 1 threshold where this is positive or this is negative which is based on an internal cut-off point, is completely inappropriate. Which is why an image-based CAD on its own is actually far superior than verbally asking symptoms, but it is still a very crude tool for the 21st century. So you need layers of other data feeding into that decision about whether or not a person should be asked to get a diagnostic test, and people don't have the data first of all. And when they do have the data, for product manufacturers, somebody who is just selling CAD, why should I create this hassle for myself saying "oh I'm going to tweak it differently for all your different populations". They don't need to; nobody is asking them why they are spending 3 million dollars on gene expert cartridges. But they are asking us. We had to tweak the system. (Interview 8, South Asia TB Organization)

Tweaking thus appeared to be a condition of technological development in the sense of its ability to adapt to screening contexts. From this perspective, the calibration of CAD is also part of a concern for the care of beneficiary populations. The use of CAD by this NGO in South Asia is particularly revealing of a form of political ownership of a tool so that it produces the expected care effects. This question of responsibility that emerges from this NGO's use of CAD goes well beyond global recommendations and announcement effects. Tweaking strategies then appear as part of a political relationship to the technology that is framed by some CAD users.

[...] we are determined that we are not going to find ourselves in a situation dependent on the companies again. And we are going to be able to tweak it. So just like I said, one part of this is tweaking by population. You need to tweak the algorithms by the different populations. (Interviewee 8, South Asia TB Organization)

Calibration through tweaking is thus a way of emancipating oneself from a technological dependency in which users find themselves indirectly placed by the WHO's recommendation. In fact, this NGO concluded that it would have to develop its own algorithms for the sake of better care, but also for the sake of appropriating the

technology. On the other hand, if a country or its national instances in the ministries do not have the technical and human capacity to adjust these algorithms to their context, the tools could end up being used in a sub-optimal way, or not being used at all. Local tweaking and political ownership are therefore essential to the deployment of CAD for global health. Control over the infrastructure necessary for CAD use also raises political ownership issues.

Managing clouds and infrastructure. Alongside tweaking practices, we can also categorize more infrastructural adjustments. Indeed, there is also a need for cloud management to handle the mass of data generated by screening projects. Here too, several issues arise, such as the difficulty for national programs to store and secure data through clouds, which require infrastructure, national policies, and, more generally, political sensitivity to these issues. This infrastructure issue is crucial insofar as it can bring care inequities and political dependencies toward technologies and the donors associated with them.¹⁸ This infrastructural issue is directly observed by the developers of detection algorithms:

For example, in this African Country, we have many systems that are connected to our cloud sending images everyday. Then those images are viewable anywhere in the world. There are no radiologists there. (...) And so, web-based security: in my opinion, our cloud service is hyper-compliant and does all of those things. So, you do your best to maintain the best possible security you can. But in a location where there is no access to specialty care, the people in this country are not concerned that somebody is going to accidentally look at their chest x-ray. Nobody cares. The programs care, the funders in Europe care, but the person who has TB and is potentially dying absolutely doesn't care. No patient or operator of the x-ray machine has ever asked me about cyber security. (...) At the end of the day, the people that are running the programs who are out on the ground everyday, they don't care. (Interview 9, International technological developer)

The infrastructural issue is mentioned here by way of cyber security and data anonymity. From this point of view, the lack of concern on the part of local ministries would be an incentive to delegate this infrastructural side to developers or other private companies. This point is also made in the context of large-scale pilot projects, each of which is developing its own infrastructure:

I think that is a major challenge for the government to upgrade. The owner of the data would be the government. So, the epidemiology bureau (EB) should have this capacity to support these things. Some of the

infrastructure is provided by the government, but sometimes they cannot keep up with the infrastructure requirements so they request from donors to help. (interview 10, International TB NGO)

On a national and international level, more and more data brokers are investing in those parts of the world where the digitization of human and social activities is taking place. These private companies are penetrating the interstices and limits of national infrastructures in the countries most affected by TB. In this African Country, for example, CAD developer and a private national IT company (ITC) have joined forces to offer integrated solutions. The aim is to offer complete solutions integrated with the cloud-based data infrastructure model, enabling both the deployment of algorithmic tools and a form of appropriation/enclosure via a local data manager:

ITC is part of the package that we have bought into through our funding. So, it is independent of the Department of Health. When we got the mobile units and set everything up, we did everything through this ITC company. So, the software we are using is the CAD software. That is the software on our van. We've had quite a few meetings with them to discuss various things but they are not ITC. ITC actually constructed the whole van. They put the x-ray machine in and synced all the systems up. They came as a whole package. We got it as a package, and as part of the package, ITC's software and all of that is on the package. The funding and the costs/expenses of it are running independently of the Department of Health. (...) If we pick up a patient with a presumptive x-ray, we work together with the doctors and the clinicians in the Department of Health structure to make the management decisions going forward. So, there is like a partnership but we are sort of independent to the Department of Health. (Interview 11, International TB NGO in Africa)

In this case, it's a private national company that manages the operability of the algorithms and, more generally, the infrastructure they require. This extract clearly expresses the advantages of these solutions integrated with data management and even practical management aspects, including the organization of vans for screening campaigns. On the other hand, the use of these algorithms needs to be integrated into the national system, particularly for the confirmation of presumptive cases. This brings us back to some of the limiting factors mentioned above. Beyond this technical aspect, the question of local ownership of these tools is raised by the involvement of new private companies in healthcare systems, which are becoming necessary for the infrastructural management of these tools and their data.

Discussion

The promises of standardization and the realities of human needs for technology

Standardization in medicine in general, and in the field of global health in particular, is always subject to interactions with local practices.^{5,19} So, while the intrinsic power of algorithms holds out the promise of standardization through automation and the bypassing of variable human expertise such as that of radiologists, they are nonetheless objects of local practices that we have characterized as "tweaking." This "localization," in tension with standardization, is often the way in which technologies make sense in global health. More specifically, the field of TB has been particularly marked by these issues for over a decade of innovations, such as new technologies and standards that aim to reduce epidemic and improve treatment.²⁰ From this point of view, CAD represents an additional layer of technology to the supposedly universal promise of improving TB treatment. What is also striking about this technology is its claim to improve detection as such, while at the same time making the other technologies introduced in recent years (GeneXpert, treatments for resistant TB, etc.) work better. From this point of view, CAD is sometimes described as the missing link in the optimal use of technologies to end the TB epidemic. Yet our results show that these technologies are also localized through these tweaking practices.

More specifically, CAD is also intrinsically conducive to these localized practices through the uncertainties that accompany them. Indeed, the issues surrounding calibration include a series of trials and errors on the part of global health institutions, echoing the "algorithmic doubt" but also the algorithmic calculation of doubt put forward by Amoore.²¹

Though this arrangement of probabilities contains within it a multiplicity of doubts in the model, the algorithm nonetheless condenses this multiplicity to a single output with a numeric value between 0 and 1. In short, the single output of the machine learning algorithm is rendered as a decision placed beyond doubt; a risk score or target that is to be actioned.

The challenge of calibration is to ensure that the machine's result, the score, leads to a decision in the care pathway. For CAD, and depending on the field of use, there are still radiologists, which raised the question of doubt and responsibility. Tweaking the algorithm, which brings the human element back into the decision-making process, is sometimes seen as a way of regaining control over the algorithm's characteristics. Furthermore, the tweaking or iterative exploration of machine/deep learning algorithms

can be found in other fields where operators, test and experiment.²¹

In the context of TB treatment, AI devices are being used and deployed for multiple purposes. Embedded in vehicles or backpacks, they can screen communities far from hospitals, and mass screening interventions. These projects, whether local or global, constitute instruments for global health, as well as a tool at the center of a network of actors, technologies, and institutions which constitute this field.¹⁷ More broadly, this new type of technical tool raises anew the question of the link between, on the one hand, screening and its effectiveness, and on the other, its real effects on the care and management of the individuals detected. Even more so as CAD are presented (or questioned, for that matter) as game-changing tools in the fight against TB.

Calibration is a way of making measurements universalizable, but in the case of CAD, it's more a way of making the instrument universalizable. The mirage of this universalizability in global health is maintained by pilot projects in archipelagos,^{5,19,22} which are carried out in underprivileged conditions to demonstrate that the tool is "globalizable." If medicine was built on the representation of a standardized and normalized body,⁵ AI today seems to reproduce forms of standardization at another level, that of the global management of populations, while at the same time reproducing the idea of a body without context or history, which can be analyzed by a standardized machine, even though there are differential immunities or specific conditions which mean that the tool cannot be used in the same way on all bodies.

Competition between technology care and patient care?

From this point of view, CAD can be analyzed between standardization and localization³ through the concrete practices of healthcare or technology workers, both from the North and the South. Indeed, our study shows the need to consider the perspective and concrete work of fieldworkers, like other research⁹ have pointed out, at the forefront of detection. While they use these new tools daily, the question of maintaining and caring for the technologies constitute other, less visible operations alongside the determined choice of threshold. Indeed, CAD is truly at the interface between technology and care, in a shifting boundary that is defined according to context. Conceived as global and universal, they turn out to be locally more diverse depending on socio-political factors. While GeneXpert has shown its limits in its implementation, CAD is the promise of a more efficient use of Xpert tests, and CAD represents for global health actors customization systems for GeneXpert. Ultimately, however, their use is determined by the availability of GeneXpert. They therefore represent a paradigmatic

example of how a given technology can be less disruptive than it appears, but rather adapt to a system and its conditions, without making it possible to rediscuss the distribution of funding or the political priorities for public health. By piling on technical tools (GeneXpert + CAD) and focusing on them only with a technical perspective, public health and care priorities as well as political and social solutions in response to TB epidemics might be blurred under a form of care for technologies and their management.

Calibration thus raises the question of the care of technologies that come to compete with traditional care procedures. The tweaking of algorithms takes place in relation to specific populations within the global South: marginalized populations, sometimes more vulnerable to certain risks, far removed from the healthcare system. This tweaking represents both procedures left to actors in the field and processes for re-appropriating the technique. As we saw in the previous section, tweaking is both a technical and a social practice. It depends on external factors and is far removed from the rigor of large-scale epidemiological studies. At each stage, human workers make choices in the context of doubt. Calibration is therefore a key stage in the integration of the large-scale CAD deployment in global health.

Tweaking practices are thus manifold: comparing results on several devices, increasing or decreasing the threshold, experimenting with players in the field, worrying about the future of people identified as positive or not, etc. By opening the black box of calibration, we gain access to the social aspects of learning machines. With the spread of CAD also comes the spread of all these practices, which are necessary to the functioning of CAD, and which may also in some ways constitute diversions from care or else generate new concerns for care, such as the question of what happens to patients identified "abnormal" but non-tuberculous in the detection process. Will they lead to new types of care? Or will they be limited to diagnostic and treatment dead-ends? Tweaking algorithms as a practice reveals how the technology is situated but also the many concerns of the users and workers (insertion in care, control over infrastructure, and political ownership). This should be better considered into truly make CAD innovative tools for TB management in global health.

Conclusions

Tweaking practices provide a better understanding of the tensions between standardization and localization processes in the diffusion of algorithmic technologies in a global health context. While the WHO advises large-scale epidemiological studies with the aim of standardizing the tool, field use involves tweaking both triage thresholds and infrastructural devices. These practices are practically driven by economic considerations associated, for example, with the availability of GeneXpert tests, but also by political

ownership strategies in relation to technologies potentially incorporating a loss of control. We can thus see how the issues of calibration and choice of thresholds constitute a series of reappropriations that play out at other, more contextual levels, beyond the intrinsic performance of CADs. Calibration tweaking is therefore a form of political ownership, with choices being made in relation to the local contexts of the different sites. It is important to take these political re-appropriations seriously, as the risk here is precisely one of management by scarcity. Far from the revolutionary promises of CAD, these tools could facilitate adaptation to situations of scarcity without calling into question the more structural aspects determining disease detection, and more generally the dynamics of TB infection.

Acknowledgements: We thank all the participants for their time and contributions.

Authors' contributions: PMD and JO contributed to conceptualization, investigation, methodology, formal analysis, and writing—original draft preparation. PMD and FAK contributed to the supervision and validation. FAK, PMD, JO, and JP contributed to the writing—reviewing and editing.

Declaration of conflicting interests: The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethical approval: This research has been approved by the Comité d’Ethique de la Recherche Clinique (CERC) of the University of Montréal (Project Number: 2021-1270).

Funding: The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: JO declares operating grants and salary support from publicly funded research agencies to study CAD of TB on chest x-rays (Observatoire international sur les impacts sociétaux de l’IA et du numérique (from the Fonds de Recherche de Québec). PMD declares operating grants from publicly funded research agencies to study CAD of TB on chest x-rays (the Canadian Institutes of Health Research, Fonds de Recherche Québec Santé, and Observatoire international sur les impacts sociétaux de l’IA et du numérique [from the Fonds de Recherche de Québec]). FAK declares operating grants and salary support from publicly funded research agencies to study CAD of TB on chest x-rays (the Canadian Institutes of Health Research, Fonds de Recherche Québec Santé, and Observatoire international sur les impacts sociétaux de l’IA et du numérique [from the Fonds de Recherche de Québec]). FAK reports participating in a technical consultation on WHO prequalification requirements for CAD software for TB. FAK reports that the following developers of CAD software provided his research group with either free or reduced pricing access to their software for evaluative research, and that the groups did not have any role in the study design, analysis, result interpretation, or decision to publish previous research and the submitted work: Delft

(Netherlands, makers of CAD4TB), qure.ai (India, makers of qXR), and Lunit (South Korea, makers of LUNIT INSIGHT). This research has been funded by the OBVIA (Observatoire international sur les impacts sociétaux de l’IA et du numérique, from the Fonds de Recherche de Québec) (Project number: FRQSC, 268938).

Guarantorship: Pierre-Marie is responsible for the finished article.

Informed consent: Written consent has been obtained from all the participants cited in this paper, according to the ethics approval mentioned.

ORCID iD: Pierre-Marie David  <https://orcid.org/0000-0001-7441-0633>

References

1. World Health Organization. Global Tuberculosis Report 2023 [Internet]. 2023 [cité 1 févr 2024]. Disponible sur: <https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2023>
2. World Health Organization. WHO consolidated guidelines on tuberculosis: module 2: screening: systematic screening for tuberculosis disease. World Health Organization; 2021.
3. Engel N, Hoyweghen IV, Krumeich A, et al. *Making global health care innovation work: standardization and localization*. New-York: Springer; 2014. p.178.
4. Soler L, Wieber F, Allamel-Raffin C, et al. Calibration: a conceptual framework applied to scientific practices which investigate natural phenomena by means of standardized instruments. *J Gen Philos Sci* 2013; 44: 263–317.
5. Lock M and Nguyen VK. *An anthropology of biomedicine*. Hoboken, NJ: John Wiley & Sons, 2018.
6. Faunce TA. Nanotechnology in global medicine and human biosecurity: private interests, policy dilemmas, and the calibration of public health law. *J Law, Med & Ethics* 2007; 35: 629–642.
7. Benjamin R. Assessing risk, automating racism. *Science* 2019; 366: 421–422.
8. Casilli AA. *En attendant les robots: enquête sur le travail du clic*. Paris: Le Seuil, 2019, p.486.
9. Biruk C (Crystal) *Cooking data: culture and politics in an African research world*. Durham: Duke University Press, 2018, p.296. (Critical Global Health: Evidence, Efficacy, Ethnography).
10. Browne JT, Bakker S, Yu B, et al. Trust in Clinical AI: Expanding the Unit of Analysis. In: Schlobach S, Pérez-Ortiz M and Tielman M, éditeurs. *Frontiers in Artificial Intelligence and Applications [Internet]*. IOS Press, 2022 [cité 6 oct 2022]. <https://ebooks.iospress.nl/doi/10.3233/FAIA220192>, pp.93–113.
11. Engström E, Strand F and Strimling P. Human-AI interactions in a trial of AI breast cancer diagnostics in a real-world clinical setting. In: Conference: Online workshop held in conjunction with CHI 2021, 2021.
12. Gerdon F, Bach RL, Kern C, et al. Social impacts of algorithmic decision-making: A research agenda for the social

- sciences. *Big Data and Soc* 2022; 9. DOI: 10.1177/20539517221089305.
13. World Health Organization, UNICEF/UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases. *Determining the local calibration of computer-assisted detection (CAD) thresholds and other parameters: a toolkit to support the effective use of CAD for TB screening* [Internet]. Geneva: World Health Organization; 2021 [cité 17 janv 2022]. Disponible sur: <https://apps.who.int/iris/handle/10665/345925>
 14. Kelly AH, Lezaun J and Street A. Global health, accelerated: rapid diagnostics and the fragile solidarities of ‘emergency R&D’. *Econ Soc* 3 avr 2022; 51: 187–210.
 15. Fehr J, Gunda R, Siedner MJ, et al. CAD4TB Software updates: different triaging thresholds require caution by users and regulation by authorities. *Int J Tuberc Lung Dis*. 1 févr 2023; 27: 157–160.
 16. Lantenois C and Coriat B. La « préqualification » OMS : origines, déploiement et impacts sur la disponibilité des antirétroviraux dans les pays du Sud. *Sci Soc et Santé* 2014; 32: 71–99.
 17. Onno J, Ahmad Khan F, Daftary A, et al. Artificial intelligence-based computer aided detection (AI-CAD) in the fight against tuberculosis: effects of moving health technologies in global health. *Soc Sci Med*. 3 may 2023; 327: 115949.
 18. Png MT. At the tensions of south and north: critical roles of global south stakeholders in AI governance. In: 2022 ACM Conference on Fairness, Accountability, and Transparency [Internet]. New York, NY, USA: Association for Computing Machinery; 2022 [cité 9 févr 2023]. pp.1434–1445. (FAccT ‘22). Disponible sur:
 19. Timmermans S and Berg M. The practice of medical technology. *Sociol Health Illn* 2003; 25: 97–114.
 20. Engel N. *Innovating tuberculosis diagnostics for the point of care*. In: Macdonald and Harper, *Understanding Tuberculosis and its Control*. London: Routledge; 2019, pp.166–184.
 21. Amore L. *Cloud ethics: Algorithms and the attributes of ourselves and others*. Durham: Duke University Press, 2020, p. 1.
 22. Geissler P. The archipelago of public health. Comments on the landscape of medical research in 21st century Africa. In: *Making and unmaking public health in Africa ethnographic and historical perspectives*. Athens: Ohio University Press; 2013, pp.231–256.
-