Wellness on Wheels (WoW): Iterative evaluation and refinement of mobile computer-assisted chest x-ray screening for TB improves efficiency, yield, and outcomes in Nigeria

Ellen Mitchell (emitchell@itg.be)
Institute of Tropical Medicine Antwerp
Hussien Abdur-Razzaq
Lagos State Ministry of Health
Victor Anyebe
KNCV Nigeria
Adebola Lawanson
National Tuberculosis and Leprosy Program Nigeria
Simeon Onyemaechi
National Tuberculosis and Leprosy Program Nigeria
Nkemdilim Chukwueme
New York University
Jerod Scholten
KNCV TB Foundation
Debby Nongo
United States Agency for International Development
Chidubem Ogbudebe
KNCV Nigeria
Danjuma Aboki
KNCV Nigeria
Chimezie Dimkpa
KNCV Nigeria
Daniel Egbule
KNCV Nigeria
Bassey Nsa
McMaster University
Rupert Eneogu
United States Agency for International Development
Emmy van der Grinten
KNCV TB Foundation
Sani Useni
KNCV Nigeria
Festus, O. Soyinka
Ogun State Ministry of Health
Peter Nwadike
KNCV Nigeria
Mustapha Gidado
KNCV TB Foundation
Emperor Ubochioma
National Tuberculosis and Leprosy Program Nigeria

Research Article

Keywords: CAD4TB, screening, radiography, feasibility, acceptability, quality, yield

Posted Date: March 8th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1343644/v1

Abstract

Background: Wellness on Wheels (WoW) is a model of mobile systematic TB screening of high-risk populations combining digital chest radiography with computer-aided automated interpretation and chronic cough screening to identify presumptive TB in communities, health facilities and prisons in Nigeria. Understanding how models are designed and refined over time helps others to anticipate technical and political challenges, replicate successful strategies, and avoid common mistakes.

Methods: We piloted and refined approaches in phased evaluations, recalibrating CAD4TB thresholds to balance TB yield and feasibility. Iterative data monitoring of screening volumes, risk mix, number needed to screen (NNS), number needed to test (NNT), sample loss, TB treatment initiation and outcomes. Risk factors for loss along the diagnostic cascade were identified and mitigation plans were implemented. Participants with high likelihood on CAD4TB (≥80) who tested negative on a single spot GeneXpert were followed-up.

Results: Gradual improvements included: achieving screening targets (64.0% to 70.5%), risk group inclusion (91.5% to 92.9%), on-site sample processing (84.3% to 86.1%), treatment initiation (86.7% to 90.8%), treatment success (70.6% to 83.2%), and NNT (8.2 to 7.6). However, expectoration by asymptomatic presumptive participants (≈85%) and HIV testing coverage (64.9%) remained suboptimal.

Conclusion: Mobile computer-assisted digital chest x-ray and chronic cough screening with GeneXpert MTB/RIF testing is feasible, acceptable, efficient and high-yield when highest risk groups and key stakeholders are engaged, and operations evolve in real time to fix problems. CAD4TB scores should be used to identify people who need clinical diagnosis and/or longer-term follow-up for progression to TB disease.

Background

The important role of computer-assisted digital chest x-ray screening as a triage tool for identifying people who would benefit from molecular TB diagnostic testing is well documented.[1–3] Although the prospects for a biomarker test in the medium term are good, at present there are no robust triage alternatives to x-ray for most settings.[4–6] While there is now consensus that digital chest x-ray followed by rapid PCR-based testing is the only option in many settings, there are still uncertainties about how best to operationalize this algorithm in low resource settings.[7] Many high burden settings lack enough radiology personnel in the public sector to rapidly interpret the high volumes of chest x-rays.[7, 8] Most images in TB screening show no anomalies or very early disease and are thus time-consuming to read and rule-out.[9] Computer-assisted digital chest x-ray is a technology with potential to rapidly identify those with a high pre-test probability of TB from among a large volume of healthy persons so that a feasible sub-set can be tested.

We report the results of an iterative process to develop and refine TB triage, testing, and treatment algorithm for urban residents of four Nigerian states, in both the southern (Ogun, Lagos) and northern (Kano and Nasarawa) regions, using computer-aided detection (CAD) software program. We describe the operational steps to engage community leaders, define beneficiary groups and participant risk mix, and identify acceptable, feasible, and high transmission settings for high-volume active case finding. Planning processes to ensure yield, flow, confidentiality, and quality of care are summarized. The challenges, miscalculations, and debates during the pilot phase are included to highlight operational lessons useful to implementers elsewhere.

STUDY RATIONALE:

Policy makers, donors, and community advocates are hesitant to invest in the steep infrastructure costs for mobile chest x-ray and GeneXpert MTB/RIF laboratories without a better understanding of how to maximize their impact. A precise assessment of the contribution of mobile TB screening has been challenging because publication bias has limited access to many active case finding projects with poor risk group targeting, scant community engagement, mediocre results, or initial missteps [10–13]. Few authors fully disclose losses along the diagnostic cascade, or report TB treatment outcomes. Finally, few studies have been conducted independently from CAD software developers[14]. To move the field forward it is essential to have a fuller understanding of the demands, constraints, and choices facing implementers and guidance on continuous quality assurance.

STUDY POPULATION

Nigeria is a high TB, TB/HIV, and MDR-TB burden country with an estimated TB treatment coverage of 24%. Home to 9% of the world's missing TB patients, Nigeria has experimented with many models of TB active case-finding (ACF) and results have varied widely.[12, 15–20] Urban Nigeria has an estimated prevalence of bacteriologically-confirmed pulmonary TB between 441-884 cases per 100,000 persons 15 years or older.[21] Prior screening suggested considerable heterogeneity, with high TB yields obtained only with accurate tools and careful attention to participant mix.[16–18, 22, 23]

Nigeria has tried to increase the effectiveness of screening efforts, moving from a three-week chronic cough threshold policy to a two-week cough threshold in 2014, but pulmonary TB (PTB) case notification did not increase significantly.[24] Community referral of chronic coughers is often feasible due to low frequency and high specificity (94%)of chronic cough for TB, the sensitivity of the classical chronic cough screen has recently been downgraded from 56 to 38%.[25]

The Wellness on Wheels (WoW) Intervention

Infrastructure & Staff development

Two large container trucks housing a lead-lined chest-x-ray suite, reading area, and mobile GeneXpert laboratory were sourced via competitive tendering. Solar panels, shade canopies, anti-theft features were added.

Computer-aided detection for tuberculosis (CAD4TB) version 4 by Delft Imaging interprets a digital image in less than a minute and can do so consistently during day-long screening events.[1–3] The software generates a likelihood score between 0 and 100, indicating the extent of lung abnormalities. CAD4TB displays a heatmap pinpointing the size and location affected. The scores are used to set a threshold above which persons are invited for TB testing. To determine the threshold, the software is calibrated in each population.

Each team had a radiology technologist, laboratory technologists, data clerks, truck driver, and a clinical supervisor. Training was conducted on standard operating procedures (SOPs) drawn from procedural manuals of successful digital CXR TB screening programs.[4] Experts in social mobilization crafted logos and messaging. Staff exposure to radiation was measured through dosimetry as mandated by the Nuclear Regulatory Agency of Nigeria.

Stakeholder Engagement

Prior to TB screening activity, preparatory visits were made to engage TB stakeholders; sensitize and seek the cooperation of local authorities; assess the feasibility of conducting an effective screening; ensure that all necessary logistics were put in place. The community is mobilized using locally appropriate means such as town announcers, handbills, posters, leaflets, community drama and radio.

Local government TB supervisors, State TB program managers, international and local partners and civil society organizations (CSOs) participated in the inaugural events. Advocacy visits were paid by KNCV active-case finding teams to key stakeholders in Nasarawa state including the governor, local government chairmen and the traditional chiefs particularly the Emir of Lafia detailing the intervention, its objectives and benefits as well as expectations from the selected communities. The initiative was flagged off at the national level in one of the four pilot states, Ogun by the governor supported by the Minister of Health. The flag off brought together different stakeholders including community and religious leaders, community-based organizations, ex-TB patients, different cadre of health workers, security agencies, and political leaders to pledge their support for the intervention. Community leaders, TB program staff, policy makers, and technical experts helped to select the locations for mobile outreach through a participatory mapping process (Table 1).

Strategy and Prioritization

Two week-long planning workshops were held involving state, national, community leadership to plan risk group mix, hot spots, testing algorithms, participant flows and crowd management including retention before screening, performance monitoring and targets. Stakeholders voiced diverse assumptions about the effectiveness of computer-assisted chest x-ray interpretation versus cough screening. To reach consensus on a viable and sustainable approach, sensitivity versus specificity of the screening algorithm and screening quality vs quantity trade-offs were debated. Equity concerns were raised when epidemiologists urged a narrow focus on adult men, older adults, alcohol users, urban poor, prisoners and groups with a higher pre-test probability of TB. Debates ensued about the competing demands of reaching high daily screening volumes versus screening those at highest risk, who would be fewer in number. Modelling diverse yield scenarios helped to make strategic choices and manage expectations of donors, TB program managers, and community leaders (See Supplemental data). The mapping process entailed the identification of groups at higher risk of developing TB disease. These factors included persons sharing similar risk factors for TB such as persons living with HIV and close contacts of pulmonary TB patients or a group of persons living in a specific geographical location associated with high burden of TB e.g. people living in an urban slum or a prison. For each risk group, an estimate of the population and the proportion that could be reached with the screening service was calculated and the number needed to screen (NNS), which itself is a function of the prevalence of TB in the risk group and the screening & diagnostic algorithm, was also estimated. The community mobilization strategy was geared toward recruitment of men and persons over 30 years of age, because of their elevated

vulnerability. Men tend to be less likely to participate in community TB screening, so dedicated efforts are often required to attain a high risk participant mix.[5, 6] Aiming for a high-risk pool with a pulmonary TB prevalence of 1,000/100,000 per population with an estimated 85% sensitivity of the algorithm, we expected an average daily yield of 1.7 persons with bacteriologically-confirmed TB.

The intervention was carried out in three phases. Findings from each phase informed the design of the subsequent phase. Iterative modification of procedures, strategies, test thresholds and targets occurred after review of the results of each phase.

First, a "Calibration phase" was undertaken to assess the viability of the case finding strategy and to identify a feasible TB testing threshold that would ensure a reasonable TB yield given specific micro-epidemic conditions and equipment. It was not possible to conduct the type of calibration that includes precision measurement of algorithm accuracy study via universal testing with a reference standard due to stakeholder opposition and resource limitations. Rigorous calibration would have required TB testing of approximately 30,000 people at low risk over a six month period, at cartridge cost of 300,000 USD. Instead, a sensitive algorithm comprised of a low testing threshold (CAD4TB score ≥ 40) and a classic symptom screen (cough of ≥two weeks) were trialed over 8 days (n=1875). Persons with CAD4TB scores ≥ 80 and negative spot TB test results were followed-up three to six months later to identify missed TB from GeneXpert MTB/RIF testing on a single spot sample. Emphasis was placed on implementation of a simple algorithm to facilitate the highest volume screening of highest risk adults while minimizing the participation burden and risk/benefit balance. Pilots were executed in two regions (North, South), to field test the approach. The third phase ("Scale-up") leveraged learning from the calibration and pilots to refine strategy. Persons classified as presumptive for TB followed the national guidelines. Before treatment initiation, a risk factor and clinical interview of bacteriologically confirmed PTB was added to preclude over-diagnosis.

Methods

Acceptability of WOW among stakeholders was pre-defined as willingness to engage by TB program staff, community members, local leaders, and technical stakeholders. Acceptability of chest- x-ray screening, cough screening, sputum production, and treatment initiation were all assessed through cascade analysis to measure drop-out or loss of participants at each phase.[26]

Feasibility was defined as a) the ability to screen 1,000 persons per week (200 adults per day over five days), and b) to test those eligible and provide results within 48 hours. To assess the feasibility of different screening and testing algorithms, performance data were collected on daily total screened, proportion screened representing high risk groups, and proportion receiving timely feedback on TB status. Testing frequencies, cartridge cost-per-case, and pseudo receiver operating curves were constructed to model the relative proportion of TB cases detected at each CAD4TB score threshold to inform the choice of score thresholds to balance algorithm sensitivity, feasibility of lab throughput. Sub-group comparisons were conducted to identify recruitment and testing problems and yield in specific groups.

Fidelity to the intervention design was designed as adherence to the standard operating procedures (SOPS) and field manuals after training. Fidelity was measured via monitoring visits and interim data queries.

The yield of the intervention was defined as the prevalence of bacteriologically-confirmed PTB among the population screened, as well as the prevalence of rifampin-resistant TB among the population screened. The efficiency of the intervention was defined as the number needed to screen (NNS) and number needed to test (NNT) to diagnose one bacteriologically-confirmed PTB case.[27] Pseudo ROC curves and cost estimates of CAD4TB thresholds were generated using an online tool developed by Takuya Yamanaka for TDR.[7]

Participants in the calibration and southern pilot with CAD4TB score ≥ 80 and a negative GeneXpert MTB/RIF result were followed up two to six months after screening to assess the volume and pace of TB progression.

Results

The outputs and outcomes of TB screening are described in three phases to highlight the impact of the evolving strategy and procedures. The distribution of screening sites is described in Table 1 and key metrics are summarized in Table 2.

Calibration Phase

Urban slums and health facilities in three local government areas of Ogun state were selected to test the suitability and viability of the algorithm, a test threshold of ≥60, and effectiveness of training on standard operating procedures (SOP). In the Calibration phase recruitment targets were met, with an average of 203 (101%) persons ≥ 15 years screened per day (Table 2). As planned, 86% of those screened were over thirty years of age. However, the calibration cohort mix skewed female (62%). Acceptability of chest x-ray screening was high (99.5%). As intended, during calibration a large proportion of participants (41.0%) were flagged for testing for having a CAD4TB score of ≥40. Cough of

two weeks duration or more flagged 14.7%, but 97.6% of chronic coughers also had qualifying CAD4TB scores. The high sensitivity of the calibration screen yielded 43.4% eligibility for TB testing, The 769 presumptive clients, an average of 96 per day, exceeded the capacity of the two 4-module GeneXpert machines (32 tests/day). Local labs and DOTS facilities were engaged to help handle the overflow of samples and presumptive clients. Over 35% of samples had to be tested off-site or outside the target window of 48 hours. Moreover, 15% of presumptive clients could not spontaneously expectorate a spot sputum sample to test. Electrical interruptions gave high error rates (6%) and obliged re-testing of samples. Fidelity to the data management SOPs was low and data linkage was manual. Two persons with B+ PTB were identified among 1,875 adults screened, yielding an NNS of 958, and a NNT of 385. GeneXpert cartridge costs per confirmed TB case were US$3,842 (385*$9.98) which was deemed unsustainably high. The TB patients identified had high CAD4TB scores (>95). More than a quarter (27%) of the general population had CAD4TB scores between 40 and 59, but testing in this group yielded no cases. So stakeholders urged an increase in specificity to a level that matched the on-site TB testing capacity (max of 32 tests/day) of the trucks. Calibration was viewed as a poor use of limited resources and a threat to project reputation by some stakeholders who urged a more traditional focus on symptomatic individuals. Approaches to address each challenge are summarized in Table 3.

Pilot Phase

Piloting of a revised threshold was conducted over 98 screening days (Table 2). Participation in screening averaged 64.0% of the 200 per day target. A greater proportion of adult men (58.1%) participated in better alignment with Nigeria's disproportionately male TB epidemic.[21] Figure 1 describes the testing algorithm in the Pilot Phase. The CAD4TB score threshold for TB testing was raised from ≥40 to ≥60 resulting in a lower proportion of presumptive TB clients (41.0% to 13.7%). Those with a chronic cough and/or a negative CAD4TB score remained eligible for TB testing and contributed 2.6% to testing volumes. Construction of a quasi 'receiver operating curve' combining the calibration and pilot data suggested possible gains of ≈1-3% in sensitivity at a cost of ≈2-4% specificity and increased testing volume of 3% by lowering the testing threshold to ≥56. As the proportion selected for TB testing (11% or 22 tests/day) in the North was well below that of the laboratory capacity (16% or 32 tests/day) a more sensitive screen seemed viable.

As the volume of people to be tested declined, cascade retention improved; missing or delayed results declined from 58% to 15% (Table 2) Bacteriologically-confirmed TB yield increased eight-fold from the calibration from 0.2%(2/1875) to 1.7%(196/ 11,279) of participants.

Fidelity to data management and laboratory SOPs improved and error rates declined from 6% during calibration to 0.5% in the pilot. To sharpen staff focus on completeness of sample collection, yield, and turnaround time, personal identifiers were only collected if a person screened positive. In all, 86.7% of those diagnosed started treatment,77% were tested for HIV, and 70.6% had a documented successful outcome.

There were 225 people who scored ≥ 80 on CAD4TB yet tested negative on a single spot GeneXpert and were eligible for follow-up. A total of 173 (77%) were reached by phone or home visit. They were asked about TB symptoms and invited to re-test for TB an average of 242 days (range 52-358) after the original screen. Eighty-nine (51%) agreed to re-test, could produce a sample, and four (4.4%) tested positive for TB. Fifty-one (23%) could not produce a sample. Four (2.3%) refused. Three (1.7%) had died, including one from TB (0.6%). Six (3.4%) had been diagnosed with TB in the interim and begun treatment. In total 11 TB cases (6.3%) were identified during 121 person-years of follow-up. The mean time to TB detection was 296 days (95%CI 286-306). Six of 11 cases (55%) detected during follow-up had CAD4TB scores of 99 or 100 during screening. The elevated risk of disease found via follow-up led to a policy of on-site clinical diagnostic review of all participants with CAD4TB scores ≥ 80 regardless of molecular test outcome.

Scale-up

The scale up period lasted 117 days. The average daily screening volume increased from 64.0% to 70.5% of the 200/day/team target (Table 2). Demographic outreach remained appropriate overall, with 92.9% considered elevated risk(65.9% male, 68.4% over 30 years of age). Acceptability and interpretability of chest x-ray remained high (>99%). The lowered testing threshold in the North (CAD4TB score ≥56) increased the proportion of presumptive clients by 1.4%(n=240).However, the overall proportion of those screening positive declined precipitously from 16.3% to 10.6% (1,527/16,636) although there was no change in the testing threshold.

The yield of TB declined during scale up from 1.7% to 1.0% despite a consistent site mix of roughly a quarter (22-28%) of days spent at prisons, half of days (62-42%) spent in poor urban centers, and the remained spent in hotspots such as health centers (6%-7%), markets(6-8%), and motor parks (2%-4%) (See Table 1). The number needed to screen (NNS) to diagnose one case almost doubled from 58 to 98 during scale up, while NNT declined slightly from 7.8 from the pilot 8.2, suggesting improved efficiency. In the scale-up period the southern truck moved from Ogun state to Lagos state. The trucks were too heavy and wide to navigate the narrow paths of informal settlements built on land reclaimed from the sea. Lack of access to the poorest neighborhoods may explain the increase in NNS from 42 to 86 in the south. Another

possibility is that the proportion of days screening near schools and factories increased slightly, and these locations had low yield. HIV co-infection among TB patients declined from 3.5% to 2.5%.

There were 171 persons over 15 years of age diagnosed with PTB in the scale-up phase. The CAD4TB threshold ≥60 identified 92.9% (159/171) of adult TB patients diagnosed with Xpert. Twelve people (7.0%) with B+PTB screened positive via chronic cough alone. Forty-four people diagnosed with TB (26.0%) screened positive for both chronic cough and chest x-ray abnormalities. Two (2.2%) additional asymptomatic people with TB were detected via the lowered test threshold of ≥56 score in the north. Neither had chronic cough, so the lowered threshold had additionality (See Supplemental data 3).Three people were diagnosed clinically. Approximately 66% (115/174) of PTB identified by Xpert was not accompanied by traditional chronic cough; however in-depth clinical interviews of those bacteriologically positive prior to treatment initiation typically revealed non pathognomonic symptoms compatible with pulmonary TB.

Leveraging lessons from the earlier phases, data and sample management practices were refined (Table 3). New monitoring and evaluation tools, variable standards and linkage methods were introduced to facilitate data-driven decision making in real-time. Barcodes for unique identifiers and auto-fill were introduced to expedite the work of busy screeners and help prevent sample loss. Staff were retrained on revised standard operating procedures (SOPs) and databases revised to expedite sample and client tracking. Fidelity to intervention design, data management and SOPs increased with missing samples declining from 15.7% to 13.9% during scale-up. Linkage to treatment increased from 86.7% to 90.8% and successful treatment outcomes increased from 70.6% to 83.2%. However HIV test coverage decreased to 64.9% from 77.0% due to occasional stock outs of rapid tests and completeness of expectoration remained suboptimal.

Discussion

Over the course of three phases, the WOW project tested models, iteratively identified and overcame challenges, refined strategies in collaboration with stakeholders and tracked outcomes. Monitoring the key metrics to limit losses along the screening cascade helped gain efficiencies with NNT, cartridge utilization, treatment initiation and outcomes over time. Target groups and CAD4TB score testing thresholds should be regularly adjusted to maximize case-finding. We found that challenges were often team -specific; with operational issues such as crowd control, unprocessed sputum sample backlog, difficult terrain, and highly variable proportions of presumptive TB over time. Varying screening criteria and thresholds permitted the comparison of the performance of individual screening methods (chronic cough vs CAD4TB vs combined) and thresholds (≥ 40 vs. ≥ 60). CAD4TB score thresholds for TB testing with Xpert should be set high enough so that sample throughput can be managed on-site to preclude diagnostic delay. Calibration indicated that a test threshold of ≥ 60 (or negative) and/or cough of 2 week duration was a sustainable algorithm in the South, whereas ≥ 56 (or negative) and/or cough of 2 week duration was viable in the North.

We show that high-yield TB case finding campaigns using expensive equipment requires attention to the socio-demographic and risk-factor mix in the intended screening pool. Establishing participant mix goals reflective of the gender and age distribution of the national epidemic proved valuable. Equally crucial was detailed community mapping and early engagement of stakeholders in prisons, health facilities, and workplaces. Close collaboration with the TB program helped minimize the treatment initiation challenges reported in some NGO-driven ACF.[6] A rich site mix of prisons, out-patient departments, dense urban low-income communities favored a high risk pool, but did not guarantee it. The high specificity of the screening algorithm was strategic and reduced excess testing in low burden sites. Further research is needed to explain spatial variation in yield.

Despite conscious community engagement strategy, daily screening targets were not routinely achieved. Our study highlights the challenge of sustaining ambitious daily screening volumes year-round in risk groups of finite size and spatial dispersion.

As is common, this mobile digital chest x-ray and Xpert screening effort required active management of diverse stakeholder expectations.[8, 9] Pressures to demonstrate return on investment in the short term meant little appetite for full calibration due to cost and efficiency concerns. Moreover, the high fixed and running costs of digital chest x-ray and mobile GeneXpert labs made detection of two or three people with TB per day seem disappointing to some stakeholders, even though WOW was among the highest yielding community active case finding efforts ever conducted in Nigeria.[10–13] Digital chest x-ray with CAD4TB version 4 appears to triage better in Nigeria than in contexts with aging populations and higher co-morbidity with HIV, smoking, and diabetes.[1, 3, 14, 15]

A key finding of this study is that CAD4TB scores should be used to pinpoint people who need clinical diagnosis Over-reliance on bacteriological confirmation from a single spot sample misses active TB due to limitations of sample collection and test performance.[16] We found annual risk of progression to TB disease was in the range of 2–6% in those with chest x-ray abnormalities and negative bacteriology. In the small control arm of a 1979 chemotherapy study, trialists in Hong Kong documented active disease in 93 (53%) of 176 bacteriologically negative, CXR-positive participants over 30 months, 75 (43%) during the first year of follow up.[17] In 2008, a population-wide study in

Cambodia showed progression to TB in a similar group was 8.5% (95% CI 6.3–11.2) annually.[18] This suggests that return on investment from active case finding may be doubled by providing high risk participants with appropriate secondary preventive care.

Losses due to inability to expectorate are common in active case-finding with digital chest x-ray but under-reported.[6, 19, 20] One in every 8 participants selected for TB testing had bacteriologically-confirmed PTB in our study, so inability to assess 13.9% of presumptive clients due to inability to provide a sample would have left 26 TB patients without treatment if clinical diagnoses were not availed. The development of solutions for collecting samples from asymptomatic clients with high likelihood of early disease is urgently needed if the full benefits of computer-assisted digital chest x-ray screening are to be realized.[21–24]

This intervention highlights the importance of ensuring that the basics of TB screening, testing, and TB treatment linkage are operating well before adding complementary services. A step-wise approach involving meeting screening targets, ensuring sample collection, quality-assured TB testing, and initiating TB treatment was sufficiently challenging at the onset. Household contact investigation, risk-group specific algorithms, screening for co-morbidities, clinical diagnoses of extrapulmonary TB, and preventive therapy for those at high risk of progression are all recommended, once the quality and timeliness of core TB screening functions is demonstrated.[25]

Innovations in CAD software now leverage accumulated information iteratively to improve prediction in a process called deep learning.[26, 27] Applying that same ‘deep learning’ logic to the entire process of TB screening is important to derive the full societal benefit of these technological innovations.

Conclusion

Implementation research in two settings with an evolving strategy shows the key success factors for mobile CXR/GXP case-finding in urban Nigeria. Striking a sustainable balance between sensitivity and feasibility in mobile CXR/GXP screening requires iterative, data-driven adaptation to respond to both the heterogeneity of TB micro-epidemics and the known socio-demographic correlates of risk. To maximize feasibility, acceptability, quality, and yield of active-case finding over time, continuous review, and re-calibration of outreach strategies, algorithm, and procedures were essential. To build effective screening and testing thresholds and algorithms early pilot efforts require experimentation and regular course correction.

Abbreviations

CAD4TB- Computer-aided detection for tuberculosis

CXR- chest x-ray

GXP- GeneXpert MTB/RIF

NNS- Number needed to screen to detect 1 true TB case

NNT- Number needed to test to detect 1 true TB case

PMVs- patent medicine vendors

TB- Tuberculosis

WHO-World Health Organization

Declarations

Ethics approval and consent to participate

This evaluation was conducted using de-identified secondary analysis of program surveillance data at the request of stakeholders. The absence of identifying information prohibited both consent as well as harms due to deductive disclosure.

Consent for publication

All authors have reviewed and approved this version of the manuscript for submission.

Availability of data and materials

The dataset will be made available on request provided that the National TB Program of Nigeria approves the request.

Competing interests

We have no competing interests to declare.

Funding

The Global Health Bureau, Office of Infectious Disease, US Agency for International Development, financially supported this study through Challenge TB under the terms of Agreement No. AID-OAA-A-14-00029. This publication is made possible by the generous support of the American people through the United States Agency for International Development (USAID). The contents are the responsibility of Challenge TB and do not necessarily reflect the views of USAID or the United States Government.

Authors' contributions

EMHM, RE, AD, VAA, DE, FS, SU, SO, EU, JS, JV, PN, NC, MG, and CO planned the WOW truck intervention. AD, BN, FS, AR,AL,SO,EU,JV,NC, MG conducted advocacy and engagement of strategic partners across ministries. JV, EVdG, BN, and MG contributed resource mobilization and donor engagement. VAA, DE, CD supervised engagement, screening and testing. JS, PN contributed to laboratory design, algorithms, and field manuals. EMHM, RE, NK, and CO designed the monitoring and evaluation system and field manuals. EMHM and CO harmonized datasets, conducted analyses, and RE and EMHM wrote the draft manuscript. All authors contributed to data interpretation and writing.

Acknowledgements

We acknowledge the important contributions of Nigerian community leaders and advocates including Dr. Baba Gana Adams, Mrs Ibiyemi Fakande, and Cecilia Kafran in planning and accountability monitoring. The contributions of World Health organization staff including Omoniyi Fadare and Enang Oyama are similarly appreciated. We are grateful to Amy Piatek and Mukadi Ya Diul and others at the United States Agency for International Development (USAID) for technical and financial support. We acknowledge the contributions and comradery of our departed colleague Bashir P. Zakariyya who worked tirelessly to bring the WOW project to fruition.

References

1. Zaidi SMA, Habib SS, Van Ginneken B, Ferrand RA, Creswell J, Khowaja S, et al. Evaluation of the diagnostic accuracy of Computer-Aided Detection of tuberculosis on Chest radiography among private sector patients in Pakistan. Scientific Reports. 2018;8:12339.
2. Philipsen RHHM, Sánchez CI, Maduskar P, Melendez J, Peters-Bax L, Peter JG, et al. Automated chest-radiography as a triage for Xpert testing in resource-constrained settings: a prospective study of diagnostic accuracy and costs. Scientific Reports. 2015;5 July:12215.
3. Creswell J, Qin ZZ, Gurung R, Lamichhane B, Yadav DK, Prasai MK, et al. The performance and yield of tuberculosis testing algorithms using microscopy, chest x-ray, and Xpert MTB/RIF. Journal of Clinical Tuberculosis and Other Mycobacterial Diseases. 2019;14 November:1–6.
4. Shazzadur Rahman AAM, Langley I, Galliez R, Kritski A, Tomeny E, Squire SB. Modelling the impact of chest X-ray and alternative triage approaches prior to seeking a tuberculosis diagnosis. BMC Infectious Diseases. 2019;19:93.
5. Kik S V, Schumacher S, Cirillo DM, Churchyard G, Boehme C, Goletti D, et al. An evaluation framework for new tests that predict progression from tuberculosis infection to clinical disease. The European respiratory journal. 2018;52.
6. Penn-Nicholson A, Hraha T, Thompson EG, Sterling D, Mbandi SK, Wall KM, et al. Discovery and validation of a prognostic proteomic signature for tuberculosis progression: A prospective cohort study. PLoS medicine. 2019;16:e1002781.
7. World Health Organization (WHO). WHO consolidated guidelines on tuberculosis. Module 2: screening – systematic screening for tuberculosis disease. second. Geneva, Switzerland; 2021.
8. World Health Organization. Rapid communication on systematic screening for tuberculosis. Geneva, Switzerland; 2020.
9. Henostroza G, Harris JB, Kancheya N, Nhandu V, Besa S, Musopole R, et al. Chest radiograph reading and recording system: evaluation in frontline clinicians in Zambia. BMC Infectious Diseases. 2016;16:136.

10. Kerrigan D, West N, Tudor C, Hanrahan CF, Lebina L, Msandiwa R, et al. Improving active case finding for tuberculosis in South Africa: Informing innovative implementation approaches in the context of the Kharitode trial through formative research. Health Research Policy and Systems. 2017;15:1–8.
11. Biermann O, Atkins S, Lönnroth K, Caws M, Viney K. "Power plays plus push": experts' insights into the development and implementation of active tuberculosis case-finding policies globally, a qualitative study. BMJ open. 2020;10:e036285.
12. Adejumo AO, Azuogu B, Okorie O, Lawal OM, Onazi OJ, Gidado M, et al. Community referral for presumptive TB in Nigeria: a comparison of four models of active case finding. BMC Public Health. 2016;16:177.
13. Onazi O, Adejumo AO, Redwood L, Okorie O, Lawal O, Azuogu B, et al. Community health care workers in pursuit of TB: Discourses and dilemmas. Social Science & Medicine. 2020;246:112756.
14. Harris M, Qi A, Jeagal L, Torabi N, Menzies D, Korobitsyn A, et al. A systematic review of the diagnostic accuracy of artificial intelligence-based computer programs to analyze chest x-rays for pulmonary tuberculosis. PLOS ONE. 2019;14:e0221339.
15. Onazi O, Adejumo AO, Redwood L, Okorie O, Lawal OM, Azuogu B, et al. Nigerian Community Workers and the Search for TB: The Discourses and Dilemmas in Four Active Case finding Models. 2017.
16. Abdurrahman ST, Mbanaso O, Lawson L, Oladimeji O, Blakiston M, Obasanya J, et al. Testing pooled sputum with Xpert MTB/RIF for the diagnosis of pulmonary tuberculosis to increase affordability in low income countries. Journal of Clinical Microbiology. 2015; May:JCM.00864-15.
17. Oshi DC, Chukwu JN, Nwafor CC, Meka AO, Madichie NO, Ogbudebe CL, et al. Does intensified case finding increase tuberculosis case notification among children in resource-poor settings? A report from Nigeria. International journal of mycobacteriology. 2016;5:44–50.
18. Ogbudebe CL, Chukwu JN, Nwafor CC, Meka AO, Ekeke N, Madichie NO, et al. Reaching the underserved: Active tuberculosis case finding in urban slums in southeastern Nigeria. International Journal of Mycobacteriology. 2015;4:18–24.
19. Onazi O, Adejumo AO, Redwood L, Okorie O, Lawal O, Azuogu B, et al. Community health care workers in pursuit of TB: Discourses and dilemmas. Social Science & Medicine. 2020;246:112756.
20. Kusimo OC, Olukolade R, Ogbuji Q, Osho J, Onikan A, Hassan A, et al. Implementation of the Active TB Case Finding in Nigeria ; Processes , Lessons Learnt and Recommendations. 2018;:10–8.
21. Nigeria National Tuberculosis and Leprosy Control Programme. Report of the First National TB Prevalence Survey in Nigeria. Abuja; 2014.
22. Organization WH. Systematic screening for active TB – an operational guide. Geneva; 2013.
23. Lonnroth K, Corbett E, Golub J, Godfrey-Faussett R, Uplekar M, Weil D, et al. Systematic screening for active tuberculosis: rationale, definitions and key considerations. INTERNATIONAL JOURNAL OF TUBERCULOSIS AND LUNG DISEASE. 2013;17:289–98.
24. Gupta R, Espinal MA, Raviglione MC. Tuberculosis as a major global health problem in the 21st century: A WHO perspective. Seminars in Respiratory and Critical Care Medicine. of Publication: Jun 2004:25 (3) (pp 245-253), 2004.
25. Assefa Y, Woldeyohannes S, Gelaw YA, Hamada Y, Getahun H. Screening tools to exclude active pulmonary TB in high TB burden countries: systematic review and meta-analysis. The International Journal of Tuberculosis and Lung Disease. 2019;23:728–34.
26. Subbaraman R, Nathavitharana RR, Mayer KH, Satyanarayana S, Chadha VK, Arinaminpathy N, et al. Constructing care cascades for active tuberculosis: A strategy for program monitoring and identifying gaps in quality of care. PLOS Medicine. 2019;16:e1002754.
27. Van't Hoog AH, Onozaki I, Lonnroth K. Choosing algorithms for TB screening: a modelling study to compare yield, predictive value and diagnostic burden. BMC infectious diseases. 2014;14:532.
28. WHO. Operational handbook on tuberculosis Module 2: Screening. 2nd edition. Geneva, Switzerland: WHO; 2021.

Tables

Table 1

Comparison of TB yield by locations of TB screening and testing events by phase

Phase:	Calibration	Pilot South		Pilot North		Pilot total		Scale Up South		Scale Up North		scale up total		Total screening events		Mean TB cases per event
Events:	n=8	n=52		n=46		n=98		n=57		n=60		n=117		n=249		
	n	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n
Poor urban communities	6	35	61	31	63	66	62	24	38	30	45	54	42	128	51	1.8
Health facilities	2	4	7	2	4	6	6	0	0	9	13	9	7	17	7	2.1
Factories	1	1	2	1	2	2	2	4	6	2	3	6	5	9	4	0
Markets	0	5	9	1	2	6	6	9	14	1	1	10	8	16	7	1.8
Prisons	0	6	11	11	22	17	16	17	27	19	28	36	28	54	22	3.6
Schools	0	1	2	2	4	3	3	5	8	5	7	10	8	13	5	1.6
Motor parks	1	2	4	0	0	2	2	4	6	1	1	5	4	8	3	2.5
Barracks	0	0	0	1	2	1	1	0	0	0	0	0	0	1	0	0
Missing	0	3	5	0	0	3	3	0	0	0	0	0	0	3	1	1.6
Total sites	10	57	100	49	100	106	100	63	100	67	100	130	100	249	100	

Number of events do not match number of days because sometimes the trucks were in more than one location in a given day.

Table 2

Comparison of Key Performance Metrics by Phase

Phase	Calibration		Pilot South		Pilot North		Pilot Total		Scale-Up South		Scale Up North		Scale-Up Total	
Test threshold	≥ 40 CAD4TB		≥ 60 CAD4TB or neg		≥ 60 CAD4TB or neg		≥ 60 CAD4TB or neg		≥ 60 CAD4TB or neg		≥ 57 CAD4TB or neg		mixed threshold	
sample size	n=1875		n=5,418		n=5,861		n=11,280		n=7,418		n=9,215		n=16,637	
days	8 days		52 days		46 days		98 days		57 days		60 days		118 days	
Acceptability in Population	n	**%**	n	%	n	%	n	%	n	%	n	%	n	%
Average daily screen (SD)	203	**101**	129	65	127	64	128	**64.0**	122	61	154	77	140	**70.5**
Age over 30 years	1625	**87**	5135	95	3636	63	8771	**77.8**	5781	78	5599	61	11380	**68.4**
Male gender	721	**37**	2656	49	3896	67	6552	**58.1**	4155	56	6816	74	10971	**65.9**
Higher risk for TB	1751	**93**	5135	95	5181	90	10316	**91.5**	6945	94	8520	92	15465	**92.9**
Screened for chronic cough	1875	**100**	5418	100	5861	100	11279	**100.0**	7418	100	9215	100	16636	**100.0**
Screened with CXR	1865	**100**	5402	100	5835	100	11237	**99.6**	7396	100	9211	100	16615	**99.9**
Interpretable image	1844	**98**	5324	99	5795	99	11119	**98.9**	7325	99	9211	100	16540	**99.1**
Presumptive by chronic cough only	27	**2**	148	4	147	3	295	**2.6**	198	3	268	3	439	**2.6**
Presumptive by CAD4TB threshold	263	**41**	1133	21	409	7	1542	**13.7**	703	10	388	4	1091	**6.6**
Presumptive: eligible	769	**43**	1281	24	636	11	1917	**17.0**	880	12	647	7	1527	**10.6**
Fidelity to Design and Protocol														
Sample not produced, low quality, not tested	435	**57**	261	20	40	6	301	**15.7**	85	10	128	1	213	**13.9**
Tested for TB with GXP on-site among eligible	334	**43**	1020	80	596	94	1616	**84.3**	795	90	519	80	1314	**86.1**
Bacteriologically confirmed TB	2	**0**	126	2	70	12	196	**1.7**	86	1	83	1	169	**1.0**
RIF resistant	1	**50**	1	1	3	4	4	**2.0**	7	8	0	0	7	**4.1**
Clinically diagnosed TB	0	0	0	0	0	0	0	0	1	1.1	2	2.4	3	**1.7**
Average daily yield	0.25		2.4		1.5		2.0000		1.5		1.4		1.46	
No. needed to screen (NNS) to detect 1 B+ patient	958		42		84		58		85		108		98	
No. needed to test (NNT) detect 1 B+ patient	385		7.9		8.5		8.2		9.2		6.1		7.6	
Cartridge costs	3842		78		85		82		92		62		107	

per TB case (US$)														
HIV tested	2	**100%**	111	97%	40	57%	151	**77.0%**	38	45%	75	90%	113	**64.9%**
TB Treatment initiation	2	**100%**	114	89%	56	80%	170	**86.7%**	83	97%	75	93%	158	**90.8%**
Treatment success	2	**100%**	96	76%	24	62%	120	**70.6%**	63	73%	71	88%	134	**83.2%**

Table 3

Challenges and Solutions by intervention phase

Phase	Challenge	Remedy implemented
Calibration	Stakeholders dissatisfied with high testing proportion, poor sample management, and perceived low B+TB yield	Practice good communication about goals and limited duration of calibration, expectation management is crucial for ACF
	Low testing threshold (≥40 CAD4TB score), identifies approximately 40% of population as presumptive TB case.	Construct pseudo ROC curves and cost estimation to identify threshold that maximizes yield without exceeding test capacity
	Low testing threshold (≥40 CAD4TB score)identifies many asymptomatic individuals who could not produce a testable sputum sample	Raise the CAD threshold for testing to ≥60, while maintaining the specific symptom (chronic cough of 2 week duration) to avoid under-detection Use an inexpensive, field-safe induction method. Provide sputum coaching using evidence-based methods
	Linking participants' symptom screen, CAD4TB, and testing data is time consuming and error prone due to lack of Unique ID	Use barcoded unique ID to match symptoms, socio-demographics, CAD scores, presumptive registry, GXP results, and treatment outcomes
	Vigorous social mobilization leads to crowd control challenges, delays, skewed participant-mix	Truck driver performs crowd control function Participants receive a time window
	Laboratory sample management, testing suboptimal, GeneXpert machine has 6% error rate	Retrain staff on procedures for GeneXpert MTB/RIF machine use. Stabilize electricity
Pilot	Participant follow-up indicates single spot specimen for GeneXpert MTB/RIF misses some early TB	Add clinical diagnosis of high CAD4TB scores Refer those with CAD4TB >80 for follow-up Follow-up persons with a positive CAD (score >60) and a negative GeneXpert MTB/RIF
	Adolescents (10-17 years) and persons with fibrotic lesions often scored negative CAD4TB vales, but positive on GeneXpert MTB/RIF	Collect sputum for testing among those with negative CAD4TB scores Add clinical diagnosis of negative CAD4TB scores
Scale- up	TB hotspots (prisons 3.6 cases/day, motor parks 2.5/day) are limited in number and size of target population. Poor communities have more potential participants but lower TB prevalence (1.8 cases/day)	Use micro-targeting to influence participant mix: Combine TB screening with services that are more highly valued by highest risk groups (e.g. vision, hearing screenings to increase engagement of older people). Actively monitor yield with weekly dashboards and use to make decisions on when to relocate
	High volume screening is logistically intensive and difficult to sustain. Staff burnout, fatigue, and stress are risks	Rotate staff, provide security, limit shift length Conduct time-motion studies to maximize efficiency, reduce wait times, and limit crowding
	Initiating TB treatment in asymptomatic people based on chest x-ray and GeneXpert MTB/RIF findings can lead to over-treatment	Avoid over-diagnosis of people with previous TB by instituting a clinical exam and history before making a definitive TB diagnosis[28]
	Matching screening and testing data are still not seamless even with barcodes	Develop application programming interface (API)s to integrate data from CAD and GeneXpert machines.

Figures

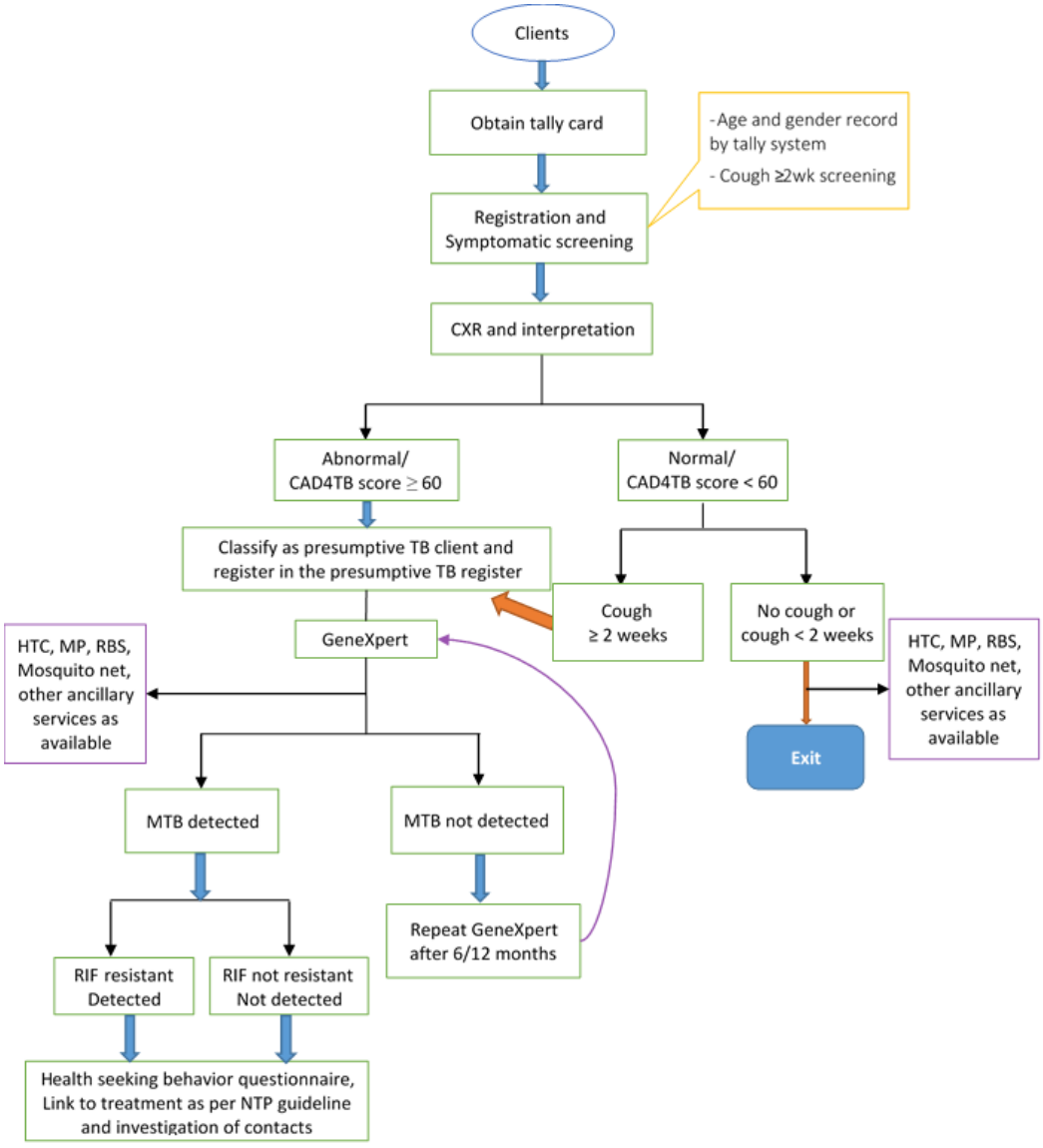

Figure 1

Client flow and testing algorithm in Pilot and Scale Up Phases

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementarydataWOWstudy.pdf