Diagnostics for tuberculosis: what new knowledge did we gain through The International Journal of Tuberculosis and Lung Disease in 2008?

J. Minion,*† A. Zwerling,*† M. Pai*†

* Department of Epidemiology & Biostatistics, McGill University, Montreal, † Respiratory Epidemiology & Clinical Research Unit, Montreal Chest Institute, Montreal, Quebec, Canada

THERE IS GROWING EXCITEMENT in the tuberculosis (TB) community over the introduction of new tools, especially new TB diagnostics. Thanks to the involvement of partners such as the World Health Organization (WHO) Stop TB Partnership’s New Diagnostics Working Group and the Foundation for Innovative New Diagnostics (FIND), the new TB diagnostics pipeline is steadily expanding.1,2 The Journal captured some of this excitement in 2008, with 34 articles focused primarily on TB diagnostics.

Before focusing on the new, it is worth remembering the past. Through a series of articles published in the Journal, we revisited the contributions of pioneers such as Koch, Ehrlich, Löwenstein and Jensen, men who made great strides in the field of TB diagnosis.3–7 Although substantial progress has been made since their days, we are still struggling towards the relatively inadequate goal of 70% TB case detection. The complexity of such a simple-seeming goal stems from the fact that ‘TB can be one of the easiest diseases to diagnose, and also one of the most difficult’.8

SMEAR MICROSCOPY

Smear microscopy continues to be the most widely used test for active TB. Two studies evaluated technologies developed to bring fluorescent microscopy (FM) to the front lines. Although known to improve sensitivity and reduce the time required to read a smear,9 FM has not been widely implemented in low-resource settings, mainly due to costs and practicality issues stemming from the use of mercury vapour lamps in conventional fluorescent microscopes. Torrea et al. evaluated the FluoreslenS system (Portable Medical Laboratories Inc, Solana Beach, CA, USA), which grants a standard light microscope fluorescent capability using an external high intensity halogen light source connected to an objective with in-built filters.10 Busy routine laboratories in Nairobi alternating using Ziehl-Neelsen (ZN) stained smears with light microscopy and auramine stained smears with a FluoreslenS converted microscope and compared the percentage of positive smears detected. FM detected on average 18% more positives than ZN; however, there was inconsistent performance at one of the three centres and several issues with practicality and user-friendliness were raised.

Van Deun et al. evaluated an attachment that uses light-emitting diode (LED) technology to convert standard light microscopes for fluorescent microscopy: the FluLED Easy (Fraen Corporation Srl, Settimo Milanese, Italy).11 Comparing the LED attachment with a conventional fluorescent microscope in a reference centre setting, concordance was 99%. Implementation in a low-income field setting in Tanzania was met with enthusiastic approval by laboratory staff, who cited the comfortable background lighting of the smears, lack of need for a dark room and lack of ultraviolet (UV) radiation as reasons for high user acceptance. Other studies have also yielded promising results using LED technology for this purpose.12,13

The remaining publications dealt with perfecting the art of the acid-fast bacilli (AFB) smear. Affolabi et al. looked at bulk staining of sputum smears, a process that is both cheap and rapid.14 Although there has traditionally been some concern with the risk of transferring bacilli from positive to negative smears with bulk staining, Affolabi et al. found no evidence of this phenomenon and called for the re-evaluation of this potentially resource-saving technique.

Sadaphal et al. demonstrated proof of principle of an automated detection system for identifying ‘TB objects’ in ZN-stained smears.15 Using a computational algorithm applied to digital images of ZN smears, they were able to successfully identify AFB in sputum smears. The development of such a technology is posited to have a role in centres with few skilled technicians and high workloads, where improved turnaround
times of smear results could reduce diagnostic delays. Along the same lines of automation, Kim et al. evaluated the 7720 Aerospray AFB slide stainer (Wescor, Logan, UT, USA) for automated auramine staining of respiratory specimens. Among clinical specimens, the assay yielded sensitivity and specificity estimates of respectively 83.8% and 94.2%, with high throughput capability. Using smear-positive specimens that were negative for Mycobacterium tuberculosis by direct DNA probe assay, their multiplex polymerase chain reaction (PCR) restriction analysis was able to amplify mycobacterial DNA in 60% of specimens and correctly speciate NTM in 94% of those successfully amplified specimens.

In summary, these studies suggest that further work is needed to produce innovative new NAAT-based diagnostics, such as loop-mediated isothermal amplification (LAMP), which may overcome the limitations of current technologies.

Tuberculin Skin Test

Even after decades of use, tuberculin skin test (TST) cut-off points and the effect of bacille Calmette-Guérin (BCG) on TST are still under debate. In Taiwan, Chan et al. examined the effect of age following BCG vaccination on the size of TST induration. This was then used to evaluate different cut-off values and their ability to discriminate between high-risk and low-risk populations. They found that using a uniform cut-off of 10 mm in human immunodeficiency virus (HIV) negative, BCG-vaccinated children led to an excess of test-positive cases in a low-risk population. They proceeded to develop age-specific cut-offs that better discriminated between populations at high and low risk for latent TB infection (LTBI).

The interpretation of serial TST is complicated, especially when boosting is confused with real conversions. Teixeira et al. performed a survey of Brazilian medical students with repeated TST after 5 weeks for those with initially negative (<10 mm induration) re-
results. They found that students with a history of BCG vaccination after infancy and students with 1–9 mm induration on their first TST were more likely to boost their second test. Depending on the criteria used, the prevalence of boosting in their setting was estimated at between 4.2% and 11.6%, supporting the recommendation that two-step TST be used routinely in the initial evaluation of young adults who will undergo serial TST.

The utility and interpretation of the TST in HIV-infected patients is controversial. Swaminathan et al. presented TST results from HIV-positive patients in India recruited into two clinical trials, either with active TB ruled out or with active TB confirmed. Using 5 mm or 10 mm as the cut-off, the poor sensitivity of TST in diagnosing active TB worsened further in those with low CD4 counts. If treatment of LTBI was determined by positive TST results, as is currently recommended, 42.6% or 36% (using respectively 5 mm or 10 mm cut-offs) of the active TB-negative group would have been given therapy for LTBI. Comparing these rates of LTBI to general population estimates, the authors predicted that using the TST for this purpose would prevent treatment from being offered to a large number of HIV-positive, latently infected individuals.

Recognising that the interpretation of a TST result is complex, Menzies et al. developed an algorithm to calculate the risk of active TB disease for adults with various TST reactions. This web-based tool (http://meakins.mcgill.ca/respepi/homeE.htm) uses basic information such as age, country of origin, size of TST reaction, BCG status and risk factors for progression to disease. It then calculates the probability of the positive TST being a true marker of LTBI, and the person’s annual and lifetime risk of developing active TB.

INTERFERON-GAMMA RELEASE ASSAYS

A major advance in the diagnosis of LTBI in recent times has been the development of interferon-gamma release assays (IGRAs) such as the QuantiFERON®-TB Gold In-Tube (QFT-G) assay (Cellestis Ltd, Carnegie, VIC, Australia), and the T-SPOT®.TB assay (Oxford Immunotec, Oxford, UK). Nishimura et al. conducted a prospective investigation of 194 TB suspects. The overall sensitivity of the QFT-G assay for the diagnosis of active TB was 76.6% (74.5% in pulmonary TB and 80% in extra-pulmonary TB patients), while overall specificity was high (91.2%).

LTBI prevalence surveys have traditionally been conducted with the use of the TST, but the introduction of IGRAs along with novel statistical methods may provide alternative approaches to prevalence estimation. Pai et al. compared conventional and model-based approaches to estimating LTBI prevalence among 719 Indian health care workers tested with both TST and QFT-G. Depending on the model, estimates of prevalence varied from 33.8% to 60.7%. Bayesian latent class analysis was used to estimate prevalence using both test results. These methods, along with the more specific IGRAs, may allow for more realistic and informative estimates of LTBI prevalence.

The use of IGRAs to diagnose LTBI in the immunocompromised, including HIV patients, is an area of expanding research. Mandalakas et al. evaluated the performance of TST, T-SPOT.TB and QFT-G in HIV-infected adults and children in South Africa. Among 20 adults and 23 children tested, 41% were TST-positive, 28% QFT-G-positive and 61% positive by T-SPOT.TB, suggesting that T-SPOT.TB may have improved sensitivity in this group. Overall agreement between TST and IGRAs was poor to moderate.

Like the TST, the selection and evaluation of an appropriate cut-off value has also been debated for IGRAs. Soysal et al. evaluated a range of cut-off values for TST, QFT-G and the T-SPOT.TB test. Using the manufacturer’s recommended cut-offs they found QFT-G and T-SPOT.TB to be both more sensitive and significantly more specific than the TST. Employing lower cut-offs for TST, QFT-G and T-SPOT.TB resulted in increased sensitivity but also decreased specificity estimates for all but the QFT-G assay. Using a mathematical model of agreement between IGRA and TST results, Davidow et al. also explored the effect of cut-off selection. They found that for a fixed prevalence, the agreement between tests depended on the cut-off point selected, further emphasising the importance of appropriate cut-off selection.

Can IGRAs be potentially useful as biomarkers in monitoring anti-tuberculosis treatment efficacy? To assess treatment efficacy in patients with pulmonary TB, Katiyar et al. performed serial testing by QFT-G at baseline, at 2 months and again at 6 months; results were then compared with sputum culture status. At 2 months, 77.6% of patients showed a decline in interferon-gamma (IFN-γ) production, and this predicted the likelihood of remaining sputum culture-positive at the end of the intensive phase of anti-tuberculosis treatment. An increase of 1 IU/ml in IFN-γ levels was associated with a higher likelihood of failing to convert to negative culture. The authors concluded that there may be potential for the use of IGRAs in monitoring the efficacy of anti-tuberculosis treatment. Other studies from high-endemic settings do not support this approach. Indeed, studies reporting on the use of IGRAs to assess the efficacy of treatment of both active TB and LTBI are highly inconsistent and hard to interpret.

Overall, IGRAs have excellent specificity that is unaffected by BCG vaccination. This characteristic is most useful clinically in BCG-vaccinated individuals, particularly in settings where BCG vaccination is given routinely after infancy and/or on multiple occasions. The sensitivity of IGRAs and TST varies.
across tests and populations, but T-Spot.TB appears to be the most sensitive test for LTBI, with perhaps a slightly lower specificity than QFT-G.\textsuperscript{51,52} However, the diagnosis of active TB rests on microbiological detection of M. tuberculosis. Immune-based tests, such as IGRAs and TST, do not directly detect M. tuberculosis, and they cannot distinguish between LTBI and active TB. Furthermore, a negative IGRA result would not conclusively rule out active disease in an individual suspected to have TB.\textsuperscript{53,54}

Despite the substantial body of literature on IGRAs, almost all the available studies on IGRAs have limitations, namely lack of a gold standard for LTBI, cross-sectional design, use of sensitivity and specificity as surrogates for patient-important outcomes, and lack of adequate data on patient-important outcomes.\textsuperscript{52} Outcomes such as incremental or added value of IGRAs, impact of IGRAs on clinical decision-making and therapeutic choices, and the prognostic ability of IGRAs to accurately identify individuals with LTBI who are at highest risk for progressing to active TB, have not been thoroughly examined.\textsuperscript{52}

**DRUG SUSCEPTIBILITY TESTING**

With the growing threat of multidrug-resistant (MDR) and extensively drug-resistant (XDR) TB, there is a widespread need for greater access to drug susceptibility testing (DST). Nine studies this year looked specifically at ways to detect the resistance patterns of M. tuberculosis. Microscopic detection of M. tuberculosis growth is a simple idea that has captured the attention of several investigators. Robledo et al. evaluated the thin layer agar (TLA) method, which involves microscopic examination of growth on solid media, using 7H11 Middlebrook agar in quadrant petri plates with one compartment containing para-nitrobenzoic acid, two containing isoniazid (INH) and rifampicin (RMP), respectively, and one without additives.\textsuperscript{55} Using decontaminated sputum specimens of confirmed TB patients, they found the sensitivity of TLA to be 91.3%, compared to 84.7% for Löwenstein-Jensen (LJ) and 96.7% for automated Mycobacteria Growth Indicator Tube (MGIT) 960 (Becton Dickinson, Sparks, MD, USA) culture. Contamination rates for TLA, LJ and MGIT were respectively 4.1%, 10.5% and 2.2%, and times to detection were respectively 10, 22 and 7.1 days. Where TLA distinguished itself was in detecting resistance, where they found 100% concordance with the standard proportion method for both RMP and INH, but with time to result being about 11 days compared to 21 days.

Ejigu et al. reported their results using the microscopic-observation drug-susceptibility (MODS) assay for direct detection of INH and RMP resistance in an urban Ethiopian setting.\textsuperscript{56} Similar in principle to TLA, MODS uses liquid culture in 24-well plates and involves examination with an inverted microscope. The turnaround time for DST results using MODS was 9 days using direct specimens, compared to 8 days with MGIT indirect DST using primary isolates from LJ culture. Concordance of the two methods was high.

Nic Fhogartaigh et al. looked at the impact of MODS results on patient-oriented outcomes.\textsuperscript{57} Using physician-selected patients where MODS testing was requested, information was extracted from request forms, health records and interviews to determine if MODS results provided important diagnostic information or resulted in initiation or change of treatment. In 24 of 58 (41.4%) reviewed cases, MODS results should have provoked a change in patient care, supporting the practical utility of MODS within a clinical context. However, the gain in turnaround time, often cited as a central benefit of MODS, was not realised at the patient care level, where health system obstacles led to lengthy delays between availability of results and treatment initiation or change.

Asencios et al. assessed the programmatic implementation of rapid DST testing using the Griess method, a colorimetric assay that uses a nitrate reductase reaction to indicate growth of M. tuberculosis on modified LJ medium.\textsuperscript{58} Detailing the steps taken and sharing the lessons they learned during the decentralisation of first-line DST in two populous districts of Lima, Peru, they illustrated the complexities involved beyond simple markers of test performance. Used as a rapid screen to test for MDR-TB, the Griess method was successfully incorporated into their diagnostic algorithms for only US$4.80 per sample (not including labour or capital costs), resulting in a shortened time to result of 31 days compared to 99 days using conventional DST.

Rapid DST using microscopic examination of cultured growth on a novel porous ceramic (Anopore or PAO, PamGene International, Hertogenbosch, The Netherlands) was described by Ingham et al.\textsuperscript{59} Previously investigated for rapid susceptibility testing of gram-negative bacilli,\textsuperscript{60} this artificial matrix allows nutrients to be provided from beneath the surface through pores and allows microcolonies to be manipulated in situ. Taking advantage of these characteristics, growth of M. tuberculosis was visualised after only 3 days by heat killing, fluorescent staining and using digitally processed images of the microcolonies.

Line-probe assays are rapid molecular tests for direct DST on clinical specimens. In a policy statement released in June 2008, the WHO endorsed the line-probe assay technology as a well-validated mechanism of rapid screening of patients at risk of MDR-TB.\textsuperscript{61} Although the details of how and where this technology will best fit into diagnostic algorithms are yet to be worked out, systematic reviews of the technology have shown consistently high accuracy for RMP resistance detection, with variable sensitivity but good specificity for detection of INH resistance.\textsuperscript{62,63}

The second generation GenoType® MTBDRplus
assay (Hain Lifescience GmbH, Nehren, Germany) detects not only RMP resistance due to *rpoB* mutations and INH resistance due to *katG* mutations, but also INH resistance conferred from changes in the *inhA* promoter region. Causse et al. evaluated this assay using patient samples and cultured isolates from a population of TB suspects in Spain. They found 100% concordance with phenotypic resistance testing for RMP resistance, and 96.5% sensitivity and 95.8% specificity for INH resistance overall.

In the interest of developing a more affordable molecular assay for RMP resistance, Tho et al. evaluated a multiplex allele specific (MAS) PCR assay and compared its results with those of the GenoType MTBDR assay and with phenotypic resistance testing. The assay successfully detected 83.7% of 104 phenotypically RMP-resistant isolates with no false-positive resistance detected, compared to 93.3% detected using the MTBDR assay. The MAS-PCR required only 2 hours of post-extraction processing, compared to 6 hours for the MTBDR, and was estimated to cost US$1.60 per sample when batched (excluding labour and equipment costs).

Rodrigues et al. examined the procedures used for second-line DST using the MGIT 960 (Becton Dickinson, Sparks, MD, USA). Comparing their results to those obtained with the Bactec 460 system, Rodrigues et al. found good concordance after testing a large number of resistant strains, including amikacin, kanamycin, capreomycin, ofloxacin, moxifloxacin, ethionamide and para-aminosalicylic acid resistant isolates.

Focusing on one of the most important classes of second-line drugs, van Doorn et al. developed a real time PCR assay to detect *gyrA* mutations associated with fluoroquinolone (FQ) resistance. An estimated 60–70% of FQ-resistant isolates display *gyrA* mutations, and this assay was designed with three probes aimed at detecting the most common FQ-resistance-associated mutations in this gene. Testing of 42 phenotypically FQ-resistant isolates resulted in an overall sensitivity of 71% and correct identification of 40 susceptible isolates, yielding 100% specificity.

The new reality is that empiric anti-tuberculosis treatment is becoming less and less of a viable option in much of the world. Susceptibility testing in all its manifestations still leaves much to be desired. Whether it is improving performance, reducing cost or translating results to clinical action, this area of research will continue to grow in importance as our battle against drug-resistant TB continues.

**COST-EFFECTIVENESS AND IMPACT OF TB DIAGNOSTICS**

If new tools are to be introduced more widely, it behoves us to estimate how and where to get the maximum benefit for our dollars. Marra et al. assessed the cost-effectiveness of three different screening strategies for contacts of active TB patients, including TST alone, QFT-G alone, and TST followed by QFT-G. Screening with QFT-G was cost-effective only in those vaccinated with BCG, while for all other sub-groups TST remained the most cost-effective option. An increase in the prevalence of recent infection, an increase in the risk of reactivation, and a reduction in utility or an increase in adherence to treatment resulted in the QFT-G strategy becoming cost-effective in additional sub-groups.

Mueller et al. performed a cost-effectiveness analysis comparing four culture options using performance data from the National TB Reference Laboratory in Lusaka, Zambia: home-made LJ, commercially acquired LJ, manually read MGIT and automated MGIT culture systems. They found that the superior sensitivity of the liquid-based culture methods led to greater cost-effectiveness than either of the solid LJ culture methods. Despite the high initial investment required for the automated MGIT instrument, this was overwhelmed by the effect of overheads, staff and consumables, and ultimately had a minimal effect on cost per specimen or cost per case detected. Nevertheless, even with maximised throughput estimates, the total cost per positive TB culture ranged from US$97 to US$147, and the cost per additionally identified TB case was at least US$321.

Dowdy et al. analysed the potential cost-effectiveness of adding a hypothetical point-of-care test to the current TB programmes in selected high-burden countries. Using epidemiological data from Kenya, South Africa and Brazil, models were developed to evaluate several diagnostic scenarios under which a new point-of-care test might be used. They found that although increasing sensitivity of the new test was important in maximising its public health impact, gains in specificity as well as the price per test were the biggest drivers of cost-effectiveness. Their analysis also showed that estimated costs and benefits are highly dependent on the setting where the test would be implemented, and that maximal benefit would come from a test that could contribute to case-finding of smear-negative TB, and could be used in areas where the existing diagnostic infrastructure is poor.

**CONCLUSION**

It is not so unexpected that our modern day history of science is devoid of single names that we can point to as ‘great contributors’ to the field. The game is now definitively a team sport: all of us are striving towards the same goal, with gradations of success, slowly transforming the world from one with a deadly, mysterious disease called ‘consumption’ to a better one with easily identifiable persons with tuberculosis. This review highlights the important contributions made by authors publishing their diagnostic research in the
Journal in 2008. We would like to encourage and express appreciation for all those TB researchers continuing to strive for that better world.

References