

Surrogate endpoints in global health research: still searching for killer apps and silver bullets?

Madhukar Pai,¹ Samuel G Schumacher,² Seye Abimbola^{3,4}

To cite: Pai M, Schumacher SG, Abimbola S. Surrogate endpoints in global health research: still searching for killer apps and silver bullets? *BMJ Glob Health* 2018;**3**:e000755. doi:10.1136/bmjgh-2018-000755

Received 2 February 2018
Accepted 20 February 2018

In clinical research, there is widespread acceptance that surrogate endpoints may not translate to long-term benefits.^{1–3} Clinical epidemiologists highlight the hazards of surrogate measures (eg, biomarkers, laboratory test results and short-term improvements in health) that substitute for outcomes which are important for patients (eg, avoiding premature death or severe disability). For example, in cardiovascular research, improvements in parameters such as blood pressure or cholesterol may not improve outcomes such as deaths. Improvements in surrogate endpoints may not correlate with real outcomes of interest (and may even increase the risk of death, in some cases). And there are many examples and case studies in the literature that illustrate the hazards of using surrogates in clinical epidemiology.^{1–3}

In comparison, in global health, we are often stunned when interventions that showed improvements in surrogate endpoints do not lead to lives being saved. Take, for example, the new tuberculosis (TB) detection technology, Xpert MTB/RIF(R) (Cepheid Inc, Sunnyvale, California, USA), an automated, molecular test for TB and drug resistance. Xpert MTB/RIF was first endorsed by WHO in 2010⁴ and has since been rolled out in many countries with over 23 million tests conducted in the past 6 years.⁵ While the test is rapid, accurate and much superior to tests that have been in use for decades,⁶ some pragmatic randomised controlled trials (RCTs) did not show improvements in long-term outcomes such as reduction in mortality.^{7,8} These results have prompted media headlines such as ‘improved diagnostics fail to halt the rise of tuberculosis.’⁹

The recent RCT in India of the WHO Safe Childbirth Checklist presents another example. The WHO Safe Childbirth Checklist is a quality-improvement tool to promote systematic adherence to practices that have been associated with improved childbirth

outcomes.¹⁰ In a large-scale study in 24 districts in India, adherence of birth attendants to essential birth practices was higher in facilities that participated in the coaching-based WHO Safe Childbirth Checklist programme than in those that did not. But maternal and perinatal mortality and maternal morbidity did not differ significantly between the two groups.¹⁰ Again, this prompted media headlines such as ‘a birth checklist fails to reduce deaths in rural India’¹¹ and ‘a lifesaving childbirth tool was successfully introduced in India—but saved no lives’.¹²

There are many more such examples in global health, from complex water and sanitation interventions, to TB vaccine trials, where surrogate endpoints do not align well with long-term outcomes.^{13–14} But given the weak health systems in many low-income and middle-income countries, it is surprising that global health researchers and journalists have great expectations that new tools, widgets, drones and checklists will save lives and are then stunned and disappointed when they do not. These ‘technological’ innovations often improve surrogate endpoints but may fail to meaningfully improve clinical outcomes in part because such outcomes improve only when a series of causal events are improved or completed. Often, the entire cascade of events in healthcare needs to improve; merely improving one or two steps (eg, diagnosis or process of care) may not lead to improvements in overall outcomes or result in sustained benefit.

In addition, there are innovations for which the expectation of improved health outcomes may not be necessary; especially innovations that aim to facilitate the patient–provider interface through improved coordination and integration of care (eg, using text message reminders, video consultations, remote monitoring and medication adherence technologies).^{15–17} For example, while a patient’s health may not improve simply because they



¹McGill Global Health Programs and McGill International TB Centre, McGill University, Montreal, Quebec, Canada

²Foundation for Innovative New Diagnostics, Geneva, Switzerland

³The George Institute for Global Health, University of New South Wales, Sydney, New South Wales, Australia

⁴School of Public Health, University of Sydney, Sydney, New South Wales, Australia

Correspondence to

Dr Madhukar Pai;
madhukar.pai@mcgill.ca

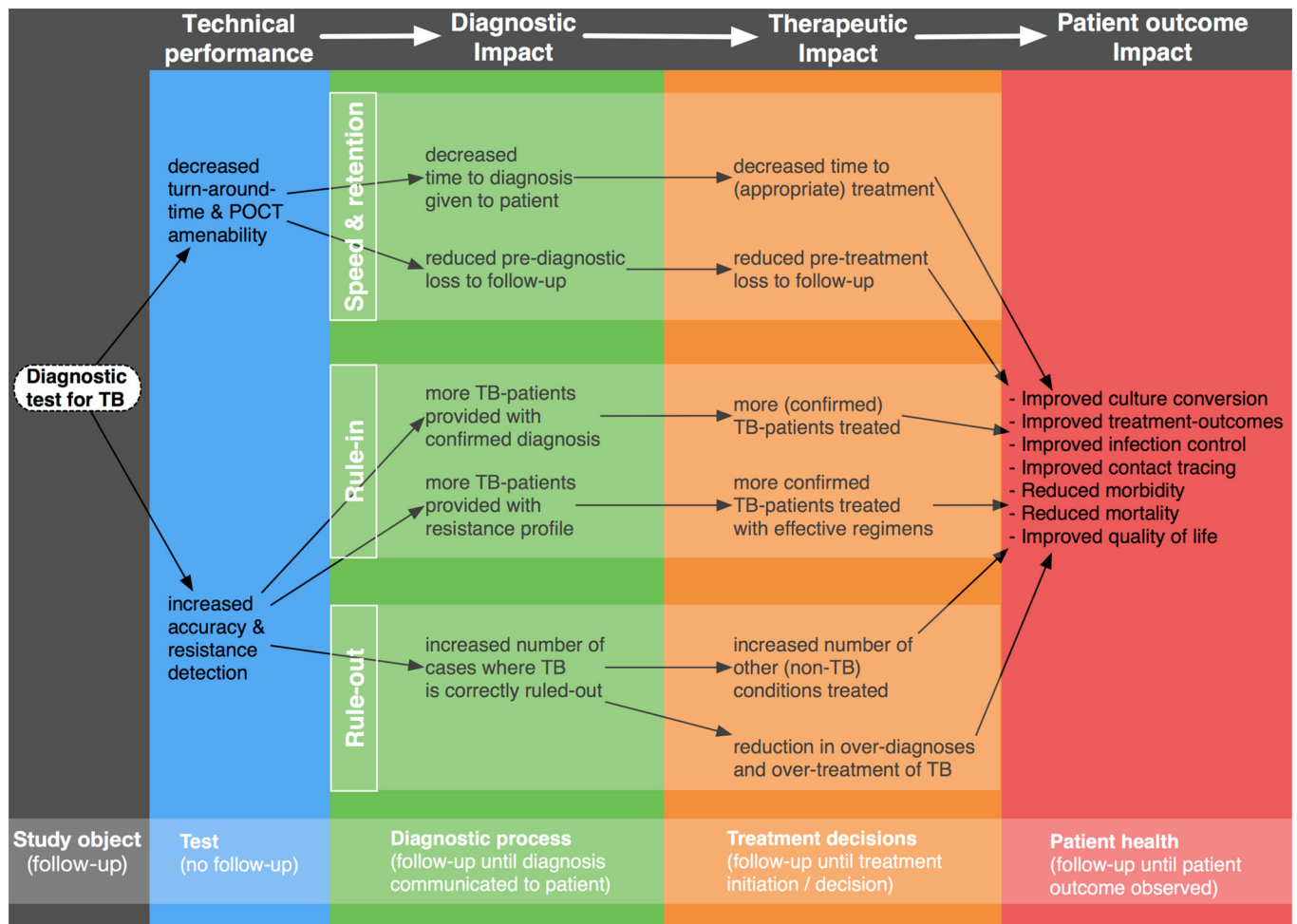


Figure 1 A framework for outlining the pathways through which new tuberculosis (TB) tests can result in improved patient outcomes. Source: Schumacher *et al*¹⁸ *PLoS ONE* 2016 (open access under Creative Commons license).

are able to consult their general practitioner via Skype, such innovations may make the process and experience of care more convenient, save the costs of travel and forfeited work and reduce care-seeking delays. But again, important as they are, these benefits are only points in the causal cascade that link innovations to improved health outcomes, and indicators of these benefits (rather than health outcomes) may be sufficient to determine whether an innovation is effective.

It is important that global health researchers are realistic when choosing indicators of effectiveness—an innovation designed to reduce costs or improve convenience should be evaluated primarily based on those indicators. For example, the purpose of a TB diagnostic test is to rapidly and accurately identify patients with TB. Once this is done, other factors become more prominent (see figure 1),¹⁸ for example, what treatment is initiated and why (empirical vs test and treat), how quickly, treatment completion rates and treatment of comorbidities. These steps in the care cascade are often weak in many settings.^{7 19–21} In that case, is it fair to expect a TB test to save lives? Likewise, it is not fair to expect that adherence to a childbirth checklist would save lives. The purpose of a checklist is to ensure that essential tasks are done

during childbirth. But what if pregnant women do not come to health facilities on time, or when referred for urgent hospital care they are unable to reach hospitals, which may even lack facilities for Caesarean section or blood transfusion?¹²

We need to be more strategic about using surrogate endpoints in global health. First, because some innovations are developed essentially to influence such surrogate endpoints; second, because health system factors may predictably intervene in the care cascade and third because waiting for long-term outcomes could delay the introduction of useful innovations. On the other hand, we must not use surrogate endpoints naively, given the dangers inherent in such endpoints. We must learn from clinical epidemiologists who argue that, ‘researchers should avoid surrogate endpoints unless they have been validated’² and caution us that ‘the use of surrogate outcomes should be limited to situations where a surrogate has demonstrated robust ability to predict meaningful benefits’.³

Global health researchers should design innovative studies to show if and how surrogate endpoints alter subsequent causal events or influence patient outcomes. If we care about reducing mortality after use of a TB test

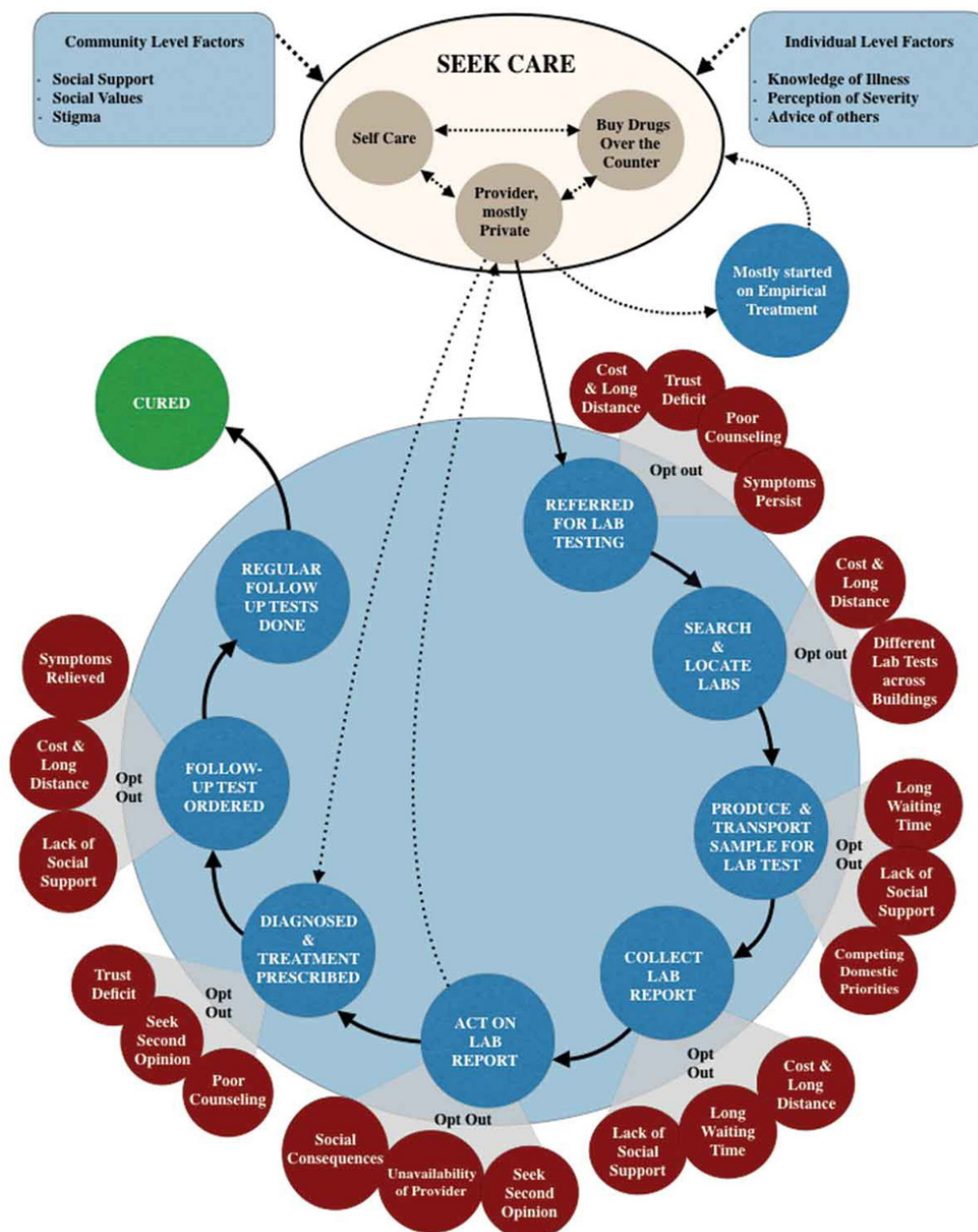


Figure 2 How patients navigate the diagnostic ecosystem in a fragmented health system in India. Source: Yellapa *et al*²⁸ Global Health Action 2017 (open access under Creative Commons license).

or a childbirth checklist, then we should also ensure that health systems are able to deliver subsequent life-saving activities in the care cascade; a shift from a fixation on tools to patient-centred solutions; from trials of stand-alone innovations to evaluations of complex, multisectoral health interventions. Such studies do not have to be large RCTs with mortality as the main outcome, given the methodological challenges of conducting RCTs, when (unlike for drug or vaccines) the effectiveness of the innovation being trialled (eg, a diagnostic, checklist, or text message reminder) depends on events further downstream in the care cascade, which in turn depend on health system context.^{22–24}

RCTs may have little value in evaluating innovations such as the WHO Childbirth Checklist for which there is

already strong and widely accepted evidence for the effectiveness of each of their component interventions.¹⁰ Even if an RCT were to show improved maternal and neonatal outcomes in one setting, it is unclear that the intervention would have had a similar effect elsewhere, given that implementation and health system context vary significantly.²⁵ Indeed, causal pathways in public health interventions are often long and complex, and RCT results are subject to effect modification.²⁶ Unfortunately, when such positive effects are found in RCTs, the result is often promoted as though the findings of the study would be applicable everywhere. And despite the limitations of RCTs or outcomes used in RCTs in global health, donors and guideline development groups (eg, Grading of Recommendations, Assessment, Development and

Evaluations²⁷ often prioritise evidence from RCTs, even when RCTs may not be necessary or appropriate for the innovation being considered for policy.

We therefore propose two ways forward. First, map out the exact point in the cascade of care pathway in which an innovation is inserted and theorise how it may make a difference and what barriers may impede its effects on health outcomes. Using the TB example, while figure 1 shows a conceptual causal pathway through which a diagnostic can have an impact,¹⁸ figure 2 shows an actual, messy pathway that patients navigate within a real world, fragmented health system,²⁸ thus identifying assumptions that must hold, and barriers that must be overcome, for a diagnostic test to fulfil its potential. Second, use theory-driven health systems and implementation research^{29–30} on the adoption of innovations to confirm or refute assumptions of how an innovation might work along the mapped-out care pathway and examine the impact of innovations on the surrogate endpoints along the care cascade. Such implementation research can provide rich insights into how we can optimise the impact and transferability of innovations, depending on context.^{31–33}

We need to explicitly lower unreasonable expectations of the impact of innovations, when surrogate endpoints are used, and when findings (including of RCTs) may not be transferable beyond specific and similar context. We need to explain the difference between surrogate endpoints and patient outcomes to policymakers and also to journalists to make sure their reporting is factual and honest. Neither the Xpert MTB/RIF test nor the WHO Safe Childbirth Checklist should be given up just because results of RCTs on their effect on mortality are not favourable. New tools have their place and are urgently needed in global health. Searching for silver bullets and killer apps are worthwhile endeavours, but we must not expect them to be ‘silver’ or ‘killer’ when introduced into systems that are suboptimal. If we care about making a real difference in global health, we also need to work on strengthening health systems to ensure holistic, effective and long-lasting solutions for patients and communities.

Twitter Follow MadhukarPai @paimadhu, Samuel Schumacher @sgschumacher and SeyeAbimbola @seyeabimbola.

Contributors MP wrote the initial draft (published on Nature Microbiology blog). SGS and SA contributed to further developing the ideas in the initial draft. All authors revised and approved the final version.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests SA is Editor in Chief of *BMJ Global Health*.

Patient consent Not required.

Provenance and peer review Not commissioned; internally peer reviewed.

Data sharing statement No additional data are available.

Open Access This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: <http://creativecommons.org/licenses/by/4.0/>

© Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2018. All rights reserved. No commercial use is permitted unless otherwise expressly granted.

REFERENCES

1. Fleming TR, DeMets DL. Surrogate end points in clinical trials: are we being misled? *Ann Intern Med* 1996;125:605–13.
2. Grimes DA, Schulz KF. Surrogate end points in clinical research: hazardous to your health. *Obstet Gynecol* 2005;105(5 Pt 1):1114–8.
3. Kemp R, Prasad V. Surrogate endpoints in oncology: when are they acceptable for regulatory and clinical decisions, and are they currently overused? *BMC Med* 2017;15:134.
4. World Health Organization. Policy statement: automated real-time nucleic acid amplification technology for rapid and simultaneous detection of tuberculosis and rifampicin resistance: Xpert MTB/RIF system. 2011 http://www.who.int/tb/laboratory/roadmap_xpert_mtb_rif_rev23dec2010.pdf
5. Albert H, Nathavitharana RR, Isaacs C, et al. Development, roll-out and impact of Xpert MTB/RIF for tuberculosis: what lessons have we learnt and how can we do better? *Eur Respir J* 2016;48:516–25.
6. Steingart KR, Schiller I, Horne DJ, et al. Xpert® MTB/RIF assay for pulmonary tuberculosis and rifampicin resistance in adults. *Cochrane Database Syst Rev* 2014:CD009593.
7. Theron G, Zijenah L, Chanda D, et al. Feasibility, accuracy, and clinical effect of point-of-care Xpert MTB/RIF testing for tuberculosis in primary-care settings in Africa: a multicentre, randomised, controlled trial. *Lancet* 2014;383:424–35.
8. Churchyard GJ, Stevens WS, Mamejta LD, et al. Xpert MTB/RIF versus sputum microscopy as the initial diagnostic test for tuberculosis: a cluster-randomised trial embedded in South African roll-out of Xpert MTB/RIF. *Lancet Glob Health* 2015;3:e450–57.
9. Callaway E. Improved diagnostics fail to halt the rise of tuberculosis. *Nature* 2017;551:424–5.
10. Semrau KEA, Hirschhorn LR, Marx Delaney M, et al. Outcomes of a coaching-based WHO safe childbirth checklist program in India. *N Engl J Med* 2017;377:2313–24.
11. Ross C. Promise unrealized: a birth checklist fails to reduce deaths in rural India. 2017 <https://www.statnews.com/2017/12/13/who-birth-checklist-fails/>
12. Merelli A. A lifesaving childbirth tool was successfully introduced in India—but saved no lives. 2017 <https://qz.com/1154504/why-a-world-health-organization-checklist-didnt-change-maternal-mortality-rates-in-india/>
13. Luby SP, Rahman M, Arnold BF, et al. Effects of water quality, sanitation, handwashing, and nutritional interventions on diarrhoea and child growth in rural Bangladesh: a cluster randomised controlled trial. *Lancet Glob Health* 2018;6:e302–15.
14. Tameris MD, Hatherill M, Landry BS, et al. Safety and efficacy of MVA85A, a new tuberculosis vaccine, in infants previously vaccinated with BCG: a randomised, placebo-controlled phase 2b trial. *Lancet* 2013;381:1021–8.
15. Flodgren G, Rachas A, Farmer AJ, et al. Interactive telemedicine: effects on professional practice and health care outcomes. *Cochrane Database Syst Rev* 2015;9:CD002098.
16. McWilliams JM. Cost containment and the tale of care coordination. *N Engl J Med* 2016;375:2218–20.
17. McWilliams JM, Hatfield LA, Cherner ME, et al. Early performance of accountable care organizations in medicare. *N Engl J Med* 2016;374:2357–66.
18. Schumacher SG, Sohn H, Qin ZZ, et al. Impact of molecular diagnostics for tuberculosis on patient-important outcomes: a systematic review of study methodologies. *PLoS One* 2016;11:e0151073.
19. Subbaraman R, Nathavitharana RR, Satyanarayana S, et al. The tuberculosis cascade of care in india's public sector: a systematic review and meta-analysis. *PLoS Med* 2016;13:e1002149.
20. Naidoo P, Theron G, Rangaka MX, et al. The South African tuberculosis care cascade: estimated losses and methodological challenges. *J Infect Dis* 2017;216(Suppl 7):S702–13.
21. Cazabon D, Alsdurf H, Satyanarayana S, et al. Quality of tuberculosis care in high burden countries: the urgent need to address gaps in the care cascade. *Int J Infect Dis* 2017;56:111–6.
22. Ferrante di Ruffano L, Hyde CJ, McCaffery KJ, et al. Assessing the value of diagnostic tests: a framework for designing and evaluating trials. *BMJ* 2012;344:e686.
23. Ferrante di Ruffano L, Dinnes J, Sitch AJ, et al. Test-treatment RCTs are susceptible to bias: a review of the methodological quality of randomized trials that evaluate diagnostic tests. *BMC Med Res Methodol* 2017;17:35.

24. Ferrante di Ruffano L, Deeks JJ. Test-treatment RCTs are sheep in wolves' clothing (Letter commenting on: J Clin Epidemiol. 2014;67:612-21). *J Clin Epidemiol* 2016;69:266–7.
25. Deaton A, Cartwright N. Understanding and misunderstanding randomized controlled trials. *Soc Sci Med* 2017.
26. Victora CG, Habicht JP, Bryce J. Evidence-based public health: moving beyond randomized trials. *Am J Public Health* 2004;94:400–5.
27. Balslem H, Helfand M, Schünemann HJ, et al. GRADE guidelines: 3. Rating the quality of evidence. *J Clin Epidemiol* 2011;64:401–6.
28. Yellapa V, Devadasan N, Krumeich A, et al. How patients navigate the diagnostic ecosystem in a fragmented health system: a qualitative study from India. *Glob Health Action* 2017;10:1350452.
29. Ridde V. Need for more and better implementation science in global health. *BMJ Glob Health* 2016;1:e000115.
30. Van Belle S, van de Pas R, Marchal B. Towards an agenda for implementation science in global health: there is nothing more practical than good (social science) theories. *BMJ Glob Health* 2017;2:e000181.
31. Davis J, Katamba A, Vasquez J, et al. Evaluating tuberculosis case detection via real-time monitoring of tuberculosis diagnostic services. *Am J Respir Crit Care Med* 2011;184:362–7.
32. Shete PB, Nalugwa T, Farr K, et al. Feasibility of a streamlined tuberculosis diagnosis and treatment initiation strategy. *Int J Tuberc Lung Dis* 2017;21:746–52.
33. Schumacher SG, Thangakunam B, Denkinger CM, et al. Impact of point-of-care implementation of Xpert® MTB/RIF: product vs. process innovation. *Int J Tuberc Lung Dis* 2015;19:1084–90.