The Risk of Falsely Declaring Non-Inferiority of Novel Latent Tuberculosis

Treatment in Large Trials

Ibrahim Abubakar[1], Frank GJ Cobelens[2] and Molebogeng X Rangaka[1,3]

1. UCL Institute for Global Health, London, UK

2. Department of Global health and Amsterdam Institute for Global Health and Development, Amsterdam University Medical Centers, Amsterdam, The Netherlands

3. Institute of Infectious Disease and Molecular Medicine & School of Public Health, University of Cape Town, South Africa

AJRCCM Articles in Press. Published on 09-December-2019 as 10.1164/rccm.201911-2258ED

Page 2 of 7

Addressing the global burden of latent tuberculosis (TB) infection is critical to eliminate TB and will require a much-improved diagnostic test, a much shorter treatment or both. It is an exciting time for research to shorten latent TB treatment with ongoing and recently completed studies holding promise of ultra-short, safer, more sterilizing regimens that would be easier to deliver to large populations.(1-3) As with other treatment interventions where there is an existing effective treatment, latent TB trials often opt for a non-inferiority design which can contain trial size and cost. A particular challenge for these trials is whom to enrol in the absence of a robust "gold-standard" diagnostic test for latent TB infection (LTBI). The modelling analysis by Jason Stout and colleagues, here (4), examined the factors that would lead to a false positive outcome in a non-inferiority trial comparing new vs. established treatments for LTBI in the absence of a perfect test. Sensitivity analyses around key assumptions concluded that their model findings were valid under certain alternate scenarios.

The authors examined the impact of LTBI prevalence, the sensitivity and specificity of currently available proxy tests for LTBI, choice of non-inferiority margins and other parameters on the design and interpretation of non-inferiority trials. There is much debate about what constitutes "true" latent infection. A particular concern in non-inferiority trials relates to the specificity of a test for LTBI and the prevalence of true LTBI in the study population.

A low prevalence of latent TB would mean that many individuals in the trial are not infected, which increases the risk of falsely declaring non-inferiority. This modelling analysis suggests that without testing for LTBI that risk is substantial when the latent TB prevalence is below 45%. Where latent TB prevalence is less than 45%, it is still better to 'enrich' the trial population for latent TB by enrolment based on LTBI tests. However, as low specificity of the LTBI test would again result in low prevalence, more specific tests such as interferon-gamma release assays (IGRA) should be used. Indeed, more broadly, the authors concluded that non-inferiority trials evaluating regimens for treating latent TB should enrol participants based on IGRA rather than on the PPD tuberculin skin test (TST), to decrease the risk of misclassifying ineffective regimens as non-inferior. The conclusion that IGRAs are better than TST was

predetermined by the assumption that TST has a considerably lower specificity (59-70% for TST versus 93-99% for the IGRAs) in "standard practice" compared to IGRA. These specificity estimates are derived from studies that used a TST cut-off of 5 mm for children and HIV-infected individuals in accordance with CDC recommendations; a sensible public health decision to ensure that few cases of LTBI are missed by maximizing sensitivity at the expense of specificity.(5) However, trials do not have to use standard TST cut-offs. Previous studies suggest that it is possible to achieve very high specificity levels with TST using higher thresholds.(6) Indeed, at cut-offs of 15 mm or more TST's predictive ability for subsequent TB is similar to that for IGRA.(7-9)

The authors also used their model to examine the impact of the enrolment strategy used in the BRIEF-TB trial, an important study that is shifting the paradigm of latent TB treatment by demonstrating the non-inferiority of a one-month rifapentine plus isoniazid regimen compared to 9 months of isoniazid alone.(3) BRIEF-TB enrolled a high-risk population with HIV (median CD4 470 cells/mm^3) residing in high-burden settings and prescribed anti-retroviral therapy, including many whose LTBI tests were unreactive, that is "negative", at baseline. Based on the Stout et al model, inclusion of individuals with that high a proportion of negative LTBI tests would essentially reduce the proportion who have true LTBI in the enrolled population, and consequently fewer with incident TB outcomes thus increasing the risk of falsely declaring non-inferiority. A false positive non-inferiority result would be unhelpful as it may result in assuming, at the very least, the wrong magnitude of effect and at worse, the use of an ineffective regimen. There may, however, be other that are not sufficiently addressed. The model does not consider the impact of ART-mediated immune-reconstitution; an initial negative LTBI result or presumed estimate of burden may be a poor indicator of who progresses to disease or benefits from TB preventive therapy.(10) Another potential limitation is that in settings with intensive exposure, and with the recognised high probability of progression soon after infection, some of the observed effects of latent TB treatment will be due to a reduction of new infections during the trial. The model places a greater weight on

AJRCCM Articles in Press. Published on 09-December-2019 as 10.1164/rccm.201911-2258ED

Page 4 of 7

reactivation rather than reinfection being a driver of active disease in high burden areas, and decreases the impact treatment has on progression of incident infections(11, 12). The impact of new infection would be most pronounced in very high burden settings, although, the absolute effect of this is uncertain and likely modest. The ultrashort regimen does have important advantages including greater treatment completion which at a population level may justify the use of a less effective regimen, for example, by allowing the use of a larger non-inferiority margin in future trials. It is important to conduct further trials of this, or other short regimens particularly in HIV-uninfected individuals and child contacts where assumptions regarding the performance of LTBI tests and prevalence of LTBI are different.

There are other important issues in designing latent TB treatment non-inferiority trials, such as questions around power and the risk of falsely failing to demonstrate non-inferiority. Small trials which are underpowered may show an intervention arm to be inferior to the standard resulting in a missed opportunity. The true difference between interventions is often unknown at the time of setting up a trial. While the authors used a pragmatic 30% difference, a larger or smaller true difference between the two regimens may be possible. A sensitivity analysis to examine the effect of different scenarios would have been informative for future trials. It is likely that where the true difference is small, it would be even more pertinent to conduct the trial in a high prevalence setting and to enrol participants using the most specific assay available.

The authors also point out a broader issue: that of efficacy versus effectiveness. While efficacy trials measure the effect of treatment under optimized and controlled conditions, effectiveness trials measure this effect in the real world. Latent TB treatment that enrol individuals by approaches that have limited specificity for LTBI but are easily applied, are valuable as they show how much TB disease can be prevented in real-world practice. One must be aware however, that such trials don't establish efficacy of a new regimen, and that a non-inferiority design may provide misleading results. This begs the question whether non-

inferiority designs are at all suitable for effectiveness trials that often apply imperfect methods for selecting patients or for implementing the intervention.

Stout and colleagues have addressed a key issue in conducting non-inferiority trials for latent TB treatment with implications for future studies. The conclusions, which raise valid concerns, should not lead to "throwing the baby with the bath water". The future should be one of multiple, more robust trial designs that include further evaluation of the ultra-short regimen or an even more ambitious programme to demonstrate superiority where the risk-benefit is less clear, rather than a step back due to a hypothetical concern that the result of this elegant trial may be a false positive.

AJRCCM Articles in Press. Published on 09-December-2019 as 10.1164/rccm.201911-2258ED

Page 6 of 7

References

1.   Menzies D, Adjobimey M, Ruslami R, Trajman A, Sow O, Kim H, et al. Four Months of Rifampin or Nine Months of Isoniazid for Latent Tuberculosis in Adults. N Engl J Med. 2018;379(5):440-53.

2.   Sterling TR, Villarino ME, Borisov AS, Shang N, Gordin F, Bliven-Sizemore E, et al. Three months of rifapentine and isoniazid for latent tuberculosis infection. N Engl J Med. 2011;365(23):2155-66.

3.   Swindells S, Ramchandani R, Gupta A, Benson CA, Leon-Cruz J, Mwelase N, et al. One Month of Rifapentine plus Isoniazid to Prevent HIV-Related Tuberculosis. N Engl J Med. 2019;380(11):1001-11.

4.   Stout JE, Turner NA, Belknap R, Horsburgh CR, Sterling TR, Phillips PPJ. Optimizing the design of latent tuberculosis treatment trials: Insights from mathematical modeling. *Am J Respir Crit Care Med* [online ahead of print] 11 November 2019; https://www.atsjournals.org/doi/abs/10.1164/rccm.201908-1606OC.

5.   Cobelens FG, Egwaga SM, van Ginkel T, Muwinge H, Matee MI, Borgdorff MW. Tuberculin skin testing in patients with HIV infection: limited benefit of reduced cutoff values. Clin Infect Dis. 2006;43(5):634-9.

6.   Berkel GM, Cobelens FG, de Vries G, Draayer-Jansen IW, Borgdorff MW. Tuberculin skin test: estimation of positive and negative predictive values from routine data. Int J Tuberc Lung Dis. 2005;9(3):310-6.

7.   Abubakar I, Drobniewski F, Southern J, Sitch AJ, Jackson C, Lipman M, et al. Prognostic value of interferon-gamma release assays and tuberculin skin test in predicting the development of active tuberculosis (UK PREDICT TB): a prospective cohort study. Lancet Infect Dis. 2018;18(10):1077-87.

8.	Pai M, Denkinger CM, Kik SV, Rangaka MX, Zwerling A, Oxlade O, et al. Gamma interferon release assays for detection of Mycobacterium tuberculosis infection. Clin Microbiol Rev. 2014;27(1):3-20.

9.	Rangaka MX, Wilkinson KA, Glynn JR, Ling D, Menzies D, Mwansa-Kambafwile J, et al. Predictive value of interferon-gamma release assays for incident active tuberculosis: a systematic review and meta-analysis. Lancet Infect Dis. 2012;12(1):45-55.

10.	Rangaka MX, Wilkinson RJ, Boulle A, Glynn JR, Fielding K, van Cutsem G, et al. Isoniazid plus antiretroviral therapy to prevent tuberculosis: a randomised double-blind, placebo-controlled trial. Lancet. 2014;384(9944):682-90.

11.	Crampin AC, Mwaungulu JN, Mwaungulu FD, Mwafulirwa DT, Munthali K, Floyd S, et al. Recurrent TB: relapse or reinfection? The effect of HIV in a general population cohort in Malawi. AIDS. 2010;24(3):417-26.

12.	Houben RM, Crampin AC, Ndhlovu R, Sonnenberg P, Godfrey-Faussett P, Haas WH, et al. Human immunodeficiency virus associated tuberculosis more often due to recent infection than reactivation of latent infection. Int J Tuberc Lung Dis. 2011;15(1):24-31.