References

1. Castro-Añón O, Pérez de Llano LA, De la Fuente Sánchez S, Golpe R, Méndez Marote L, Castro-Castro J, *et al*. Obesity-hypoventilation syndrome: increased risk of death over sleep apnea syndrome. *PLoS One* 2015;10:e0117808.
2. Berg G, Delaive K, Manfreda J, Walld R, Kryger MH. The use of health-care resources in obesity-hypoventilation syndrome. *Chest* 2001;120: 377–383.
3. Jennum P, Ibsen R, Kjellberg J. Social consequences of sleep disordered breathing on patients and their partners: a controlled national study. *Eur Respir J* 2014;43:134–144.
4. Alawami M, Mustafa A, Whyte K, Alkhater M, Bhikoo Z, Pemberton J. Echocardiographic and electrocardiographic findings in patients with obesity hypoventilation syndrome. *Intern Med J* 2015;45:68–73.
5. Corral J, Mogollon MV, Sánchez-Quiroga MA, Gómez de Terreros J, Romero A, Caballero C, *et al*.; Spanish Sleep Network. Echocardiographic changes with non-invasive ventilation and CPAP in obesity hypoventilation syndrome. *Thorax* 2018;73:361–368.
6. Howard ME, Piper AJ, Stevens B, Holland AE, Yee BJ, Dabscheck E, *et al*. A randomised controlled trial of CPAP versus non-invasive ventilation for initial treatment of obesity hypoventilation syndrome. *Thorax* 2017;72:437–444.
7. Masa JF, Mokhlesi B, Benítez I, Gomez de Terreros FJ, Sánchez-Quiroga MA, Romero A, *et al*.; Spanish Sleep Network. Long-term clinical effectiveness of continuous positive airway pressure therapy versus non-invasive ventilation therapy in patients with obesity hypoventilation syndrome: a multicentre, open-label, randomised controlled trial. *Lancet* 2019;393:1721–1732.
8. Piper AJ, Wang D, Yee BJ, Barnes DJ, Grunstein RR. Randomised trial of CPAP vs bilevel support in the treatment of obesity hypoventilation syndrome without severe nocturnal desaturation. *Thorax* 2008;63:395–401.
9. Castro-Añón O, Golpe R, Pérez-de-Llano LA, López González MJ, Escalona Velasquez EJ, Pérez Fernández R, *et al*. Haemodynamic effects of non-invasive ventilation in patients with obesity-hypoventilation syndrome. *Respirology* 2012;17:1269–1274.
10. Held M, Walthelm J, Baron S, Roth C, Jany B. Functional impact of pulmonary hypertension due to hypoventilation and changes under noninvasive ventilation. *Eur Respir J* 2014;43: 156–165.
11. Masa JF, Corral J, Alonso ML, Ordax E, Troncoso MF, Gonzalez M, *et al*.; Spanish Sleep Network. Efficacy of different treatment alternatives for obesity hypoventilation syndrome: Pickwick Study. *Am J Respir Crit Care Med* 2015;192:86–95.
12. Masa JF, Mokhlesi B, Benítez I, Mogollon MV, Gomez de Terreros FJ, Sánchez-Quiroga MÁ, *et al*.; Spanish Sleep Network. Echocardiographic changes with positive airway pressure therapy in obesity hypoventilation syndrome: long-term Pickwick randomized controlled trial. *Am J Respir Crit Care Med* 2020;201:586–597.
13. Galiè N, Humbert M, Vachiery JL, Gibbs S, Lang I, Torbicki A, *et al*. 2015 ESC/ERS guidelines for the diagnosis and treatment of pulmonary hypertension: The Joint Task Force for the Diagnosis and Treatment of Pulmonary Hypertension of the European Society of Cardiology (ESC) and the European Respiratory Society (ERS). Endorsed by: Association for European Paediatric and Congenital Cardiology (AEPC), International Society for Heart and Lung Transplantation (ISHLT). *Eur Respir J* 2015;46:903–975.
14. Kauppert CA, Dvorak I, Kollert F, Heinemann F, Jörres RA, Pfeifer M, *et al*. Pulmonary hypertension in obesity-hypoventilation syndrome. *Respir Med* 2013;107:2061–2070.
15. Mokhlesi B, Masa JF, Brozek JL, Gurubhagavatula I, Murphy PB, Piper AJ, *et al*. Evaluation and management of obesity hypoventilation syndrome: an official American Thoracic Society clinical practice guideline. *Am J Respir Crit Care Med* 2019;200:e6–e24.

The Risk of Falsely Declaring Noninferiority of Novel Latent Tuberculosis Treatment in Large Trials

Addressing the global burden of latent tuberculosis infection (LTBI) 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 LTBI treatment, with ongoing and recently completed studies holding promise of ultrashort, safer, more sterilizing regimens that would be easier to deliver to large populations (1–3). As in trials involving other diseases for which there is an existing effective treatment, developers of LTBI trials often opt for a noninferiority design to minimize sample sizes and costs. A particular challenge for investigators in these trials is deciding which subjects to enroll in the absence of a robust "gold-standard" diagnostic test for LTBI. In a modeling analysis presented in this issue of the *Journal*, Stout and colleagues (pp. 598–605) examined the factors that would lead to a false-positive outcome in a noninferiority trial comparing new versus established treatments for LTBI in the absence of a perfect test (4). After performing sensitivity analyses of key assumptions, the authors 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, and the choice of noninferiority margins and other parameters on the design and interpretation of noninferiority trials. There is much debate about what constitutes "true" latent infection. A particular concern in noninferiority trials relates to the specificity of a test for LTBI and the prevalence of true LTBI in the study population.

A low prevalence of LTBI would mean that many individuals in the trial are not infected, which increases the risk of falsely declaring noninferiority. This modeling analysis suggests that without testing for LTBI, that risk is substantial when the LTBI prevalence is below 45%. When LTBI prevalence is less than 45%, it is still better to "enrich" the trial population for LTBI by enrollment based on LTBI tests. However, as the low specificity of the LTBI test would again result in low prevalence, more specific tests, such as IFN-γ release assays (IGRAs), should be used. Indeed, more broadly, the authors conclude that noninferiority trials evaluating regimens for treating LTBI should enroll participants based on IGRAs rather than on the PPD tuberculin skin test (TST), to decrease the risk of misclassifying ineffective regimens as noninferior. The conclusion

Originally Published in Press as DOI: 10.1164/rccm.201911-2258ED on December 9, 2019

that IGRAs are better than TSTs was predetermined by the assumption that the TST has a considerably lower specificity (59–70% for TSTs vs. 93–99% for IGRAs) in "standard practice" than IGRAs. These specificity estimates are derived from studies that used a TST cutoff 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 cutoffs. Previous studies suggested that it is possible to achieve very high specificity levels with the TST using higher thresholds (6). Indeed, at cutoffs of $\geqslant$15 mm, the TST's ability to predict subsequent TB is similar to that of IGRAs (7–9).

The authors also used their model to examine the impact of the enrollment strategy used in the BRIEF-TB (Brief Rifapentine-Isoniazid Evaluation for TB Prevention) trial, an important study that is shifting the paradigm of LTBI treatment by demonstrating the noninferiority of a 1-month rifapentine plus isoniazid regimen compared with 9 months of isoniazid alone (3). BRIEF-TB enrolled a high-risk population of individuals with HIV (median CD4 cell count, 470 cells/mm^3) who resided in high-burden settings and had been prescribed antiretroviral therapy, including many whose LTBI tests were unreactive (i.e., "negative") at baseline. Based on Stout and colleagues' model, inclusion of individuals with that high a proportion of negative LTBI tests would essentially reduce the proportion of subjects with true LTBI in the enrolled population, resulting in fewer subjects with incident TB outcomes, thus increasing the risk of falsely declaring noninferiority. A false-positive noninferiority result would be unhelpful because it may result in assuming, at the very least, the wrong magnitude of effect or, at the worst, the use of an ineffective regimen. There may, however, be other issues that are not sufficiently addressed in the model. The model does not consider the impact of antiretroviral therapy–mediated immune reconstitution; an initial negative LTBI result or presumed estimate of burden may be a poor indicator of who will progress to disease or benefit from TB-preventive therapy (10). Another potential limitation is that in settings with intensive exposure, and with the recognized high probability of progression soon after infection, some of the observed effects of LTBI treatment will be due to a reduction of new infections during the trial. The model places a greater weight on reactivation than on reinfection as 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 infections 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 rates, which at a population level may justify the use of a less effective regimen, for example, by allowing the use of a larger noninferiority margin in future trials. It is important to conduct further trials of this and 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.

Other important issues should be considered in the design of LTBI treatment noninferiority trials, such as the power of the study and the risk of falsely failing to demonstrate noninferiority. Small trials that 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 a trial is set up. Although Stout and colleagues 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 when the true difference is small, it would be even more pertinent to conduct the trial in a high-prevalence setting and enroll participants using the most specific assay available.

The authors also point out a broader issue: efficacy versus effectiveness. Whereas efficacy trials measure the effect of treatment under optimized and controlled conditions, effectiveness trials measure this effect in the real world. Trials investigating approaches that have limited specificity for LTBI but are easily applied are valuable because they show how much TB disease can be prevented in real-world practice. One must be aware, however, that such trials do not establish the efficacy of a new regimen, and that a noninferiority design may provide misleading results. This raises the question as to whether noninferiority designs are at all suitable for effectiveness trials, which often apply imperfect methods for selecting patients or for implementing the intervention.

Stout and colleagues have addressed a key issue in conducting noninferiority trials for LTBI treatment, and their work has implications for future studies. The authors' conclusions, which raise valid concerns, should not lead us to "throw the baby out with the bath water." In the future, we need to develop multiple, more robust trial designs that include further evaluation of the ultrashort regimen or an even more ambitious program to demonstrate superiority when the risk/benefit ratio is less clear, rather than take a step back because of a hypothetical concern that the result of this elegant trial may be a false positive. ■

Author disclosures are available with the text of this article at www.atsjournals.org.

Ibrahim Abubakar, M.B. B.S., Ph.D.
UCL Institute for Global Health
London, United Kingdom

Frank G. J. Cobelens, M.D., Ph.D.
Department of Global Health
and
Amsterdam Institute for Global Health and Development
Amsterdam University Medical Centers
Amsterdam, the Netherlands

Molebogeng X. Rangaka, M.B. Ch.B., Ph.D.
UCL Institute for Global Health
London, United Kingdom
and
Institute of Infectious Disease and Molecular Medicine
and
School of Public Health
University of Cape Town
Cape Town, South Africa

ORCID IDs: 0000-0002-0370-1430 (I.A.); 0000-0001-9594-188X (M.X.R.).

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;379:440–453.

2. Sterling TR, Villarino ME, Borisov AS, Shang N, Gordin F, Bliven-Sizemore E, *et al*.; TB Trials Consortium PREVENT TB Study Team. Three months of rifapentine and isoniazid for latent tuberculosis infection. *N Engl J Med* 2011;365:2155–2166.
3. Swindells S, Ramchandani R, Gupta A, Benson CA, Leon-Cruz J, Mwelase N, *et al*.; BRIEF TB/A5279 Study Team. One month of rifapentine plus isoniazid to prevent HIV-related tuberculosis. *N Engl J Med* 2019;380:1001–1011.
4. Stout JE, Turner NA, Belknap RW, Horsburgh CR, Sterling TR, Phillips PPJ. Optimizing the design of latent tuberculosis treatment trials: insights from mathematical modeling. *Am J Respir Crit Care Med* 2020;201:598–605.
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: 634–639.
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: 310–316.
7. Abubakar I, Drobniewski F, Southern J, Sitch AJ, Jackson C, Lipman M, *et al*.; PREDICT Study Team. Prognostic value of interferon-γ release assays and tuberculin skin test in predicting the development of active tuberculosis (UK PREDICT TB): a prospective cohort study. *Lancet Infect Dis* 2018;18:1077–1087.
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;27:3–20.
9. Rangaka MX, Wilkinson KA, Glynn JR, Ling D, Menzies D, Mwansa-Kambafwile J, *et al*. Predictive value of interferon-γ release assays for incident active tuberculosis: a systematic review and meta-analysis. *Lancet Infect Dis* 2012;12:45–55.
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: 682–690.
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;24:417–426.
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;15:24–31.