Taking Costs and Diagnostic Test Accuracy into Account when Designing Prevalence Studies: An Application to Childhood Tuberculosis Prevalence

Zhuoyu Wang, MSc, Nandini Dendukuri, PhD, Madhukar Pai, MD, PhD, Lawrence Joseph, PhD

Background. *When planning a study to estimate disease prevalence to a pre-specified precision, it is of interest to minimize total testing cost. This is particularly challenging in the absence of a perfect reference test for the disease because different combinations of imperfect tests need to be considered. We illustrate the problem and a solution by designing a study to estimate the prevalence of childhood tuberculosis in a hospital setting.* ***Methods.*** *All possible combinations of 3 commonly used tuberculosis tests, including chest X-ray, tuberculin skin test, and a sputum-based test, either culture or Xpert, are considered. For each of the 11 possible test combinations, 3 Bayesian sample size criteria, including average coverage criterion, average length criterion and modified worst outcome criterion, are used to determine the required sample size and total testing cost, taking into consideration prior knowledge about the accuracy of the tests.* ***Results.*** *In some cases, the required sample sizes and total testing costs were both reduced when more tests were used, whereas, in other examples, lower costs are achieved with fewer tests.* ***Conclusion.*** *Total testing cost should be formally considered when designing a prevalence study.* ***Key words****: sample size; cost; diagnostic test accuracy; prevalence studies; diagnostic studies; Bayesian methods.* ***(Med Decis Making XXXX;XX:xx–xx)***

Received 6 July 2016 from Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, QC, Canada (ZW, ND, MP, LJ); Technology Assessment Unit, McGill University Health Center, Montreal, QC, Canada (ND); and Division of Clinical Epidemiology, Department of Medicine, McGill University Health Center, Montreal, QC, Canada (ND, LJ). Financial support was provided by grants from National Science and Engineering Research Council of Canada (NSERC) (RGPIN 238593-2013) and Canadian Institutes of Health Research (CIHR) (MOP 89857). The funding agreement ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report. Revision accepted for publication 8 May 2017.

Address correspondence to Nandini Dendukuri, PhD, Centre for Outcomes Research and Evaluation, 5252 Boulevard de Maisonneuve, 3F.50, Montreal, QC H4A 3S5 Canada; e-mail: nandini.dendukuri@mcgill.ca.

DOI: 10.1177/0272989X17713456

When designing a study to estimate the prevalence of a disease for which no perfect diagnostic test is available, multiple imperfect tests may be employed. These tests are likely to differ from each other not only in terms of sensitivity and specificity but also in terms of cost and ease of implementation. Previous research has provided methods for sample size determination in the absence of a perfect test, demonstrating that use of multiple tests can increase the testing accuracy and decrease the sample size required in such prevalence studies.[1–3] However, test costs have not been formally considered in these studies. The budget for a prevalence study is an important component of study design, and the total testing cost, which is affected by both the particular choice of tests used and the required sample size, is a key component of the total budget. Therefore, it is of interest to develop methods for designing an optimal prevalence study such that a desired precision of the prevalence estimate is achieved with the smallest total testing cost. We will provide a general framework for such

situations, and illustrate the methods for estimating the prevalence of childhood tuberculosis (TB) in a hospital setting in South Africa. We assume that all participants in this hospital-setting survey are suspected to have TB based on clinical symptoms and will be evaluated with the same tests.

The Challenge of Estimating the Prevalence of Childhood TB

Childhood TB is an important health concern worldwide with an estimated half to one million incident cases per year and 140,000 TB deaths in 2015.[4,5] Estimating the prevalence of childhood TB in different settings (including a hospital setting for high-risk cases) is challenging but important for understanding and curtailing the rising trend of childhood TB. One of the major challenges is the lack of a gold standard test for childhood TB.[6–9] In clinical practice, combinations of culture, Xpert MTB/RIF (Xpert), smear microscopy (smear), chest X-ray (CXR) and tuberculin skin test (TST) are currently recommended.[10] However, none of these tests or any combination of them has perfect sensitivity or specificity.

Sample Size Determination for a TB Prevalence Study

Although a handbook providing advice on how to design and carry out a prevalence survey of pulmonary TB in adults has been published by the World Health Organization,[11] no guidelines are available specifically for sample size determination when no perfect reference test is available, including for childhood TB. According to the WHO handbook, when a perfect reference test is available, the sample size for a TB prevalence survey based on a simple random sample can be calculated as

$$N = 1.96^2 \left[\frac{1 - \pi_g}{d^2 \pi_g} \right], \qquad (1)$$

where 1.96 is the Z score from a normal density required for 95% coverage, π_g is the ''prior guess'' of the true population prevalence of TB, and d is the relative precision of the prevalence study, defined as $d = \frac{w/2}{\pi}$, where w is the width of the confidence interval and π is the true population prevalence (replaced by π_g at the design stage). The value of d is recommended to be between 0.2 and 0.25 for a prevalence survey for TB in adults. For example, d = 0.2 means that the 95% confidence interval (CI) around the estimated prevalence p is (p-0.2π, p+0.2π).

This method of sample size calculation has important deficiencies, particularly when there is no perfect reference test. First, this approach assumes that the testing strategy has perfect sensitivity and specificity, which is not reasonable for childhood TB. In turn, assuming perfect accuracy in the absence of truly perfect tests can lead to poor sample size suggestions.[1] Second, the uncertainties about the sensitivities and specificities of these tests are ignored, even though these contribute additional uncertainty to the prevalence estimates. In the case when these uncertainties are too large or the desired precision of the prevalence estimate is too small, the desired precision can be unachievable even with an infinite sample size.[2,3] Third, the ''prior guess'' of the true population prevalence of TB used in the calculation is only a point estimate. However, there is always some uncertainty in our knowledge about TB prevalence in the population of interest, otherwise a prevalence study would not be needed. A slight change in the ''prior guess'' of the TB prevalence could lead to a considerable change in the sample size. For example, according to (1) above, when d = 0.2, changing π_g from 100 to 80 per 100,000 population increases the sample size from 95,944 to 119,944; an increase of 25%.

In this paper, we will use a Bayesian approach to design a childhood TB prevalence study in a hospital setting; this has several advantages over other sample size methods in designing prevalence studies. For example, Bayesian sample size methods consider the inaccuracies as well as the uncertainties in the accuracies of the diagnostic tests, leading to more realistic sample sizes. In addition, the uncertainty about the prevalence is also acknowledged. We will also take the testing costs into account to find the optimal testing strategy, suggesting which tests to use and how large the sample size should be, such that the target precision is attained at minimum cost.

SUGGESTED TESTS FOR TB AND THEIR PROPERTIES

We will follow the recommendations of Graham and others[10] to use culture, Xpert, CXR and TST as the potential diagnostic tests to design a childhood TB prevalence study within a hospital setting in South Africa. Though smear microscopy is widely used in the survey of TB in adults, it has a poor sensitivity rate of 16% to 30% in children[12] and is not recommended. Since both culture and Xpert are

based on sputum and are considered to be highly correlated, to minimize the collection of redundant information, they will not be used together. With this constraint, any combination of one of the sputum-based tests, the image-based CXR and the immune-based TST can reasonably be assumed to be independent conditional on the true disease state, the so-called "conditional independence" assumption in the diagnostic testing literature.[13]

The available prior information about the sensitivities and specificities of culture, Xpert, CXR, and TST in detecting childhood TB are given in Table 1. These estimates were obtained from a latent class model fit to data from a cohort of hospitalized children with suspected TB in South Africa.[12] Culture has the highest specificity but the lowest sensitivity in children, and the sensitivities and specificities of CXR and TST are both relatively low. In addition, there are considerable uncertainties in the sensitivities and specificities of these tests, as indicated by the widths of the 95% credible intervals (CrI). In the methods described below, we will show how these uncertainties are quantified into Beta prior distributions over the sensitivity and specificity of each test, to input into the sample size calculations.

The accuracies and unit costs of these tests may vary in different settings. In this illustration, we will use the reported costs of these tests (in US dollars) in South Africa as our test accuracy estimates are based on data from this region,[14–17] as listed in Table 1. Different values of the unit costs of culture and Xpert are available in the literature. We used $5.12 as the unit cost of culture in our main analysis and carried out a sensitivity analysis to increase the unit cost to $14.89. The difference in unit cost arises because solid cultures are cheaper than liquid cultures. We used $14.93 as the unit cost of Xpert in our analysis.

With 4 available tests and the constraint that 2 of them cannot be used together, there are 11 possible combinations: four single-test combinations, five 2-test combinations, and two 3-test combinations. For each combination, the required sample size and total testing costs will be calculated, and the optimal combination, sample size, and total cost will be determined.

METHODS

Bayesian Sample Size Criteria

For each test combination, we will use Bayesian methods proposed by Dendukuri et al.[3] to determine the required sample size for the TB prevalence study. This method assumes that the diagnostic tests are conditionally independent. In our case, we anticipate that culture and Xpert are conditionally dependent as both are microbiological tests aimed at detecting the presence of TB bacteria in the test sample. However, both these tests can reasonably be assumed to be independent of CXR, which is based on imaging, and TST, which is based on the child's immune response to TB. Further, these assumptions were satisfied in the earlier latent class analysis.[12] To implement our methods, the sensitivities and specificities of the tests in Table 1 were used as the prior information. In addition, a uniform prior (*Beta*(1,1)) was used for the prevalence of TB.

We applied 3 different Bayesian sample size criteria[3,18,19] to illustrate the range of sample sizes resulting from using a less strict versus a stricter sample size criterion. In our application, all 3 criteria provide the sample size required to estimate childhood TB prevalence with a 95% posterior CrI of length l =0.1, but differ on the probability of this occurring across the sample space of possible data sets. The 3 criteria are:

- Average Coverage Criterion (ACC): This criterion ensures that the coverage of the posterior CrI of fixed length 0.1 is at least 95% when averaged across all data sets. To implement this criterion, given prior information, 4,000 random data sets of test results were generated using the prior information. For each data set, a Bayesian latent class model[13] was used to obtain the posterior distribution of the prevalence as well as the coverage of the highest posterior density (HPD) interval with a length of 0.1. The algorithm searches for the minimal sample size to ensure that the average coverage across the 4,000 data sets is no smaller than 95%.
- Average Length Criterion (ALC): Similar to the ACC, this criterion ensures that the expected length of the posterior CrI of fixed coverage 95% is at most 0.1 when averaged across all data sets. It is implemented the same way as the ACC, except that the length of the HPD interval with 95% coverage is calculated for each sample. The minimal sample size is searched to ensure that the average of the lengths is no wider than 0.1.
- Modified Worst Outcome Criterion (MWOC): Because both the ACC and ALC return sample sizes that guarantee desired lengths and coverages only on average, a stricter criterion is also of interest. The criterion ensures that at least 95% of the posterior credible intervals of coverage 95% are no wider than 0.1, which is also equivalent to at least 95% of the posterior CrI of length 0.1 have coverage

Table 1 Accuracy Properties[12] and Unit Cost (USD)[13–16] of Tests for Childhood Tuberculosis Prevalence Survey

Test	Sensitivity, Median (95% CrI)	Prior Distribution of Sensitivity	Specificity, Median (95% CrI)	Prior Distribution of Specificity	Unit Cost, $ Different Sources
Culture	0.60 (0.46, 0.76)	Beta (24.2, 15.4)	1.00 (0.99, 1.00)	Beta (524.1, 2.8)	5.12 12.10 14.89
Xpert	0.49 (0.38, 0.62)	Beta (31.2, 31.3)	0.99 (0.97, 1.00)	Beta (413.9, 5.8)	14.93 16.90 33.50
CXR	0.64 (0.55, 0.73)	Beta (69.8, 39.1)	0.78 (0.734, 0.83)	Beta (201.7, 54.9)	14.43
TST	0.75 (0.61, 0.84)	Beta (42.0, 15.3)	0.69 (0.63, 0.76)	Beta (139.1, 60.4)	10.06

CrI, credible interval; CXR, chest x-ray; TST, tuberculosis skin test; USD, United States dollars ($).

no less than 95%. We thus seek the minimum sample size such that at least 95% of predicted data sets have HPD coverage of at least 95% for intervals of length at most 0.1.

All priors were converted into the closest fitting *Beta* distributions, as listed in Table 1, by matching the plausible range of the prior information to the 2.5% and 97.5% quantiles of the Beta distribution. We assumed sample sizes larger than 100,000 would be infeasible. The methods were implemented using the freely available *PropMisclassSample Size* program.[20]

Determining the Total Cost of Testing

Though costs from different perspectives could be considered in a disease prevalence survey, we consider only direct costs from the perspective of the research or public health body that is planning the prevalence study. We will not consider indirect costs that may arise from a societal perspective. In a TB prevalence survey, the direct testing cost includes both fixed costs (denoted by I) and variable costs. Fixed costs would include the capital cost of the testing equipment as well as staff training for operating the equipment. We consider 2 types of variable costs. One type of variable cost would be proportional to sample size of the prevalence study, denoted by N, with a fixed unit cost, denoted by m. This would include consumables and reimbursement to subjects. A second type of variable cost could increase non-linearly with the sample size; e.g., via a step function. This kind of variable cost would include additional test kits, equipment and operating staff that become necessary after a certain quantum increase in the sample size.[11] For example, assume the capacity of an equipment during a study period of a fixed duration is N_0, then $[N/N_0]$ units of this equipment are needed, where $[x]$ returns the smallest integer no less than x. Assume the variable cost of the equipment is m_0. In this case, the total testing cost, M, of a TB prevalence survey can be defined as the sum of the fixed and variable costs:

$$M = I + m * N + m_0 * \left\lceil \frac{N}{N_0} \right\rceil. \qquad (2)$$

More generally there could be more than one term in the variable costs in (2) since the variable costs of each test in the test combination may be affected differently by the sample size. As more data on cost-sample size relationships become available, more complex testing cost functions can be defined. For example, the marginal unit cost of a test may decrease with the increase in sample size. Then, a power function with exponent less than one could be used to model the total cost of this test.

In real applications, the fixed cost, the unit cost, and the sample size could all be affected by the choice of tests being used. Usually, when multiple tests are used and/or when the tests used are more accurate, the fixed cost and unit cost become larger, since more test equipment is required and more accurate tests are usually more expensive. However, since multiple tests or more accurate tests usually provide more accurate results, this can reduce the required sample size.

Initially, we will assume that, in our hospital setting, the testing equipment and trained staff are already available. When fixed costs are already paid, the total testing cost simplifies to

$$M = m * N,$$

as I is set to zero. The unit cost m of a testing combination is the sum of the listed prices of the tests included in the combination. For example, the unit cost of carrying out culture plus CXR is $19.55 in our main analysis.

We then illustrate the impact of variable costs by taking the purchase price of additional testing devices into account in an additional analysis. One disadvantage of Xpert is its low throughput. The commonly used model GX4, with 4 modules, allows no more than 20 tests per 8-h shift. Assume the existing Xpert device is mainly used for clinical diagnosis and can only allow 5 research tests per day. To reduce the duration of the study period, additional Xpert devices might be needed. For example, if a researcher desires to complete the study in 6 months or 125 working days, and if the required sample size N exceeds $5 \times 125 = 625$, we will need [(N-625)/(20×125)] additional Xpert devices. The purchase price for a GX4 Xpert device is $17,000.

We will calculate the required sample size and compare the total testing cost for all possible testing combinations. The testing combination with the smallest total testing cost is optimal and should be recommended. In sensitivity analyses, we will examine the impact of using an alternative unit cost for culture and the impact of using a larger required width of the posterior CrI of l=0.2 in our sample size criteria.

RESULTS

Main Results

In our initial analyses, the required sample sizes and total testing costs for each testing combination under the MWOC, the ACC, and the ALC criteria for the desired 95% CrI width of 0.1 are listed in Table 2. When the MWOC is used, only the 2 combinations of 3 tests obtain the desired precision, within a reasonable sample size of 2,407 and a total testing cost of $71,271.27 for the culture, CXR and TST combination, and a sample size of 3,652 and a total testing cost of $143,961.84 for the Xpert, CXR and TST combination. When the ACC is used, the required sample size for the combination of culture and CXR is 4,356 with a total testing cost of $85,159.80, and the required sample size for the combination of Xpert and CXR is 7,459 with a total testing cost of $218,996.24. Adding a third test to these 2-test combinations dramatically reduces the sample size. For example, adding TST to the Xpert plus CXR combination reduces the required sample size to 1,428, or less than one-fifth of the previous size. Though the unit cost is increased from $29.36 to $39.42, the total testing cost is reduced to $56,291.76, or a quarter of the cost. Similarly, adding TST into the culture plus CXR combination reduces the sample size from 4,356 to 1,070 with a total testing cost of $31,682.70. When the ALC is used, the sample sizes are also substantially reduced when adding a third test to any combination of 2 tests, with a minimal sample size of 836 and a total testing cost of $24,753.96. Note that no single test can obtain the desired precision regardless of the criterion used. Across all 3 Bayesian sample size criteria, the two 3-test combinations resulted in not only the minimal sample sizes but also the minimal total testing costs, compared with any of the 2-test or single-test combinations. Therefore, though adding an additional test increases the unit cost, the required sample size and the total testing cost may be decreased dramatically. In such a case, it is worth spending more per subject to reduce the total cost. However, it is not always true that more tests per subject or smaller sample size leads to lower costs. For example, when the ALC is used, the required sample size of the combination of Xpert, CXR, and TST is 1,000 smaller than the required sample size of the combination of culture and X-ray, the cost is about $3,000 less.

Presenting results for all 3 sample size criteria allows us to see how much more we would need to pay if we insist on a more stringent criterion. Whether we pursue the more stringent criterion will depend on whether the cost and sample size are feasible in practice, as well as the additional value of information that could be obtained with larger sizes.

Sensitivity Analysis

In the sensitivity analysis where the unit cost of culture was increased to $14.89, the total testing costs of the combination of culture, CXR, and

Table 2 Required Sample Sizes and Total Testing Costs of All Potential Testing Combinations

	MWOC		ACC		ALC	
Test Combination	**Sample Size Required**	**Cost, USD**	**Sample Size Required**	**Cost, USD**	**Sample Size Required**	**Cost, USD**
Culture + CXR + TST	2,407	71,271.27	1,070	31,682.70	836	24,753.96
Xpert + CXR + TST	3,652	143,961.84	1,428	56,291.76	1,141	44,978.22
Culture + CXR	>100,000	>1,955,000	4,356	85,159.80	2,151	42,052.05
Culture + TST	>100,000	>1,518,000	>100,000	>1,518,000	3,520	53,433.60
CXR + TST	>100,000	>2,449,000	>100,000	>2,449,000	>100,000	>2,449,000
Xpert + CXR	>100,000	>2,936,000	7,459	218,996.24	3,396	99,706.56
Xpert + TST	>100,000	>2,499,000	>100,000	>2,499,000	6,004	150,039.96
Culture	>100,000	>512,000	>100,000	>512,000	>100,000	>512,000
Xpert	>100,000	>1,493,000	>100,000	>1,493,000	>100,000	>1,493,000
CXR	>100,000	>1,443,000	>100,000	>1,443,000	>100,000	>1,443,000
TST	>100,000	>1,006,000	>100,000	>1,006,000	>100,000	>1,006,000

Data presented using 3 Bayesian sample size criteria when the width of the desired 95% credible interval (CrI) is 0.1. ACC, average coverage criterion; ALC, average length criterion; CXR, chest x-ray; MWOC, modified worst outcome criterion; TST, tuberculosis skin test; USD, United States dollars ($).

TST under the MWOC, ACC, and ALC become $94,787.66, $42,136.60 and $32,921.68, respectively.

When the cost of purchasing additional Xpert devices is considered, the optimal testing combination does not change, since Xpert was not included in the optimal combination of tests to use. However, all testing combinations including the Xpert test will need to purchase additional Xpert devices. For example, the combination of Xpert, CXR, and TST will need 2 additional devices under the MWOC, with costs of $34,000, and 1 additional device is required under the ACC or ALC, with a cost of $17,000.

When we doubled the width of the desired CrI from 0.1 to 0.2, and recalculated the sample size required for the ALC criterion, the required sample sizes and total testing costs reduced dramatically. Once again, the combination of culture, CXR, and TST achieved the smallest sample size of 133 with total cost of $10,507, whereas the combination of Xpert, CXR, and TST achieved the lowest total cost of $10,304 with a sample size of 161. In addition, the 2-test combination of Xpert and TST as well as the combination of culture and TST had lower total testing costs than the 3-test combination of smear, CXR, and TST, with the required sample sizes of 264, 201, and 242, respectively, and the total testing costs of $11,088, $11,457, and $14,278, respectively. Thus, here we find another example where a smaller number of tests per subject leads to lower overall cost, similar to earlier examples.

DISCUSSION

Accurately measuring the burden of a disease in a given setting can be challenging when there is no single perfect diagnostic test for the disease. Typically, multiple imperfect tests need to be employed, potentially increasing the cost of the study. The challenge is illustrated here by the problem of designing a study for measuring the prevalence of childhood TB. A recent expert consensus document has encouraged the use of a panel of commonly used, imperfect tests in research studies aimed at measuring disease prevalence and/or diagnostic accuracy[10] related to childhood TB, with the consideration that the diagnosis of TB in children has additional challenges to the diagnosis in adults. In this paper, we have discussed how tests from this panel (culture, Xpert, CXR, and TST) can be used jointly to design a childhood TB prevalence study in a hospital setting.

All possible combinations of conditionally independent tests were considered. For each test combination, Bayesian sample size criteria were applied to obtain the required sample size. Our results suggest that using more diagnostic tests is a potential strategy to reduce the total sample size and total testing cost of a prevalence study. In our example, the total testing cost was reduced to less than one-half when adding a third test to any 2-test combination. This is an important finding that should encourage TB researchers to gather data on more rather than fewer tests. The availability of

information on more diagnostic tests will in turn allow for the construction of a more informative latent class model. This study also suggests that test costs can sometimes be an important component of study design in diagnostic research. It is necessary to consider and collect cost data in addition to the usual consideration of specificity and sensitivity data of the diagnostic tests used.

The methods in this paper can be applied to settings other than hospitals, such as population prevalence studies, and to diseases other than TB. Test accuracy may vary depending on the setting. For example, the microbiological tests we considered (culture and Xpert) may have lower sensitivity in a community setting where TB cases are likely to be less severe and therefore have a lower bacillary load. In general, these methods are potentially extendable to any setting where costs will be considered along with sample size in the design of a study.

We assumed that the prevalence study would be designed using conditionally independent tests, so as to maximize the incremental value of each test,[21] thus decreasing the total sample size required. While, conditional dependence between any pair of tests cannot be ruled out completely, our recent work has shown that unless there is a strong relation between two tests due to a missing covariate, it is unlikely that the magnitude of conditional dependence has an important impact.[22] Such a high conditional dependence is likely to exist between the culture and Xpert tests, which both depend on the unmeasured bacillary load. Therefore, we reasoned that these 2 tests should not be used in the same study, as they would provide highly correlated information.

It should be noted that we assumed that once a testing strategy is selected, all test results will be gathered on all individuals included in a prevalence study, as it will be carried out within the context of research. This may be different from routine clinical practice, where the goal is to make a decision for an individual patient. In such a setting, additional cost savings may be incurred per patient by carrying out tests in a sequential manner. For example, only patients with an abnormal CXR may be selected for additional testing. Such data, however, is less suitable for research as not all test combinations will be observed. Therefore, in a prevalence study one might prefer to consider all tests regardless of results or sequence to fully use all available information.

In our example with the required 95% CrI length of 0.1, a larger sample size was always associated with greater cost. However, when we doubled the length, some testing combinations resulted in lower costs even though they required larger sample sizes. This occurred because the sample sizes are all reduced dramatically so that the differences in sample sizes become smaller, meaning that unit costs have more impact on the total costs. When some variable costs are step functions of the required sample size, small increases in sample size of a testing combination may require additional equipment, raising the variable cost even if it results in a smaller sample size. In our initial example, more tests always resulted in a reduced sample size but the above case shows that this is not always true. Similar situations may also occur when there is dependence between 2 or more tests. For example, the use of the Xpert test together with culture is likely to increase the cost with little increase in information, as the Xpert test will primarily identify the same subset of the patients identified by culture. More work is needed to extend our methods to models that account for test dependence.

In our example, we only considered the need for additional Xpert devices, which had only minimal effect and did not change our optimal testing combination. In general, however, the impact of the cost of new or additional devices may depend on the specific costs, and results can differ more substantially. When there are no existing devices available for some tests, the initial purchase cost of the devices may dominate the total cost.

Our methods were limited to the consideration of financial costs, but a broader economic analysis might consider other aspects such as the feasibility of carrying out each test or even patient outcomes. Such analyses could consider the tradeoff between using a cheaper test with a long turnaround time v. a more expensive test with a shorter turnaround time. We have provided a general formula for calculating the total costs of a disease prevalence study but the exact formula and costs used will depend on the particulars of each study setting.

Extensions to the latent class model used here that are of interest for future research could include accounting for conditional dependence between the tests, or adding a hierarchical structure suitable for clustered data, which may arise in a population survey with clustered sampling. In this case, when calculating the total testing cost, the impact of the number of sampling clusters on fixed costs could

also be considered. Variations in costs of testing across clusters may also need to be considered.

In conclusion, the total testing cost should be considered when designing a prevalence study in the absence of a perfect reference test. Using more diagnostic tests is a potential strategy to reduce the total testing cost, but as our examples show, this is not guaranteed.

REFERENCES

1. Rahme E, Joseph L, Gyorkos TW. Bayesian sample size determination for estimating binomial parameters from data subject to misclassification. Appl Stat. 2000;49:119–28.

2. Dendukuri N, Rahme E, Bélisle P, Joseph L. Bayesian sample size determination for prevalence and diagnostic test studies in the absence of a gold standard test. Biometrics. 2004;60:388–97.

3. Dendukuri N, Bélisle P, Joseph L. Bayesian sample size determination for diagnostic test studies in the absence of a gold standard: Comparing identifiable to non-identifiable models. Stat Med. 2010;29:2688–97.

4. Jenkins HE, Tolman AW, Yuen CM, et al. Incidence of multidrug-resistant tuberculosis disease in children: Systematic review and global estimates. Lancet. 2014;383:1572–9.

5. World Health Organization. Global tuberculosis report. 2015. Available from: URL: http://apps.who.int/iris/bitstream/10665/191102/1/9789241565059_eng.pdf.

6. Swaminathan S, Rekha B. Pediatric Tuberculosis: Global Overview and Challenges. Clin Infect Dis. 2010;50:S184–94.

7. Zar HJ, Connell TG, Nicol M. Diagnosis of pulmonary tuberculosis in children: new advances. Expert Rev Anti Infect Ther. 2010;8:277–88.

8. Graham SM, Ahmed T, Amanullah F, et al. Evaluation of tuberculosis diagnostics in children: 1. Proposed clinical case definitions for classification of intrathoracic tuberculosis disease. Consensus from an expert panel. J Infect Dis. 2012;205(Suppl 2): S199–S208.

9. Cuevas LE, Browning R, Bossuyt P, et al. Evaluation of tuberculosis diagnostics in children: 2. Methodological issues for conducting and reporting research evaluations of tuberculosis diagnostics for intrathoracic tuberculosis in children. Consensus from an expert panel. J Infect Dis. 2012;205(Suppl 2):S209–S215.

10. Graham SM, Cuevas LE, Jean-Philipe P, et al. Clinical case definitions for classification of intrathoracic tuberculosis in children: an update. Clin Infect Dis. 2015;61(Suppl 3):S179–S187.

11. World Health Organization. Tuberculosis prevalence surveys: A handbook. 2011.

12. Schumacher S, van Smeden M, Dendukuri N, Joseph L, Nicol P, Pai M, Zar J. Diagnostic test accuracy in childhood pulmonary tuberculosis: a Bayesian latent class analysis. Am J Epidemiol. 2016;184:690–700.

13. Joseph L, Gyorkos T, Coupal L. Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard. Am J Epidemiol. 1995:263–72.

14. Hausler HP, Sinanovic E, Kumaranayake L, et al. Costs of measures to control tuberculosis/HIV in public primary care facilities in Cape Town, South Africa. Bull World Health Organ. 2006;84:528–36.

15. TB Diagnostics Market Analysis Consortium. Market assessment of tuberculosis diagnostics in South Africa in 2012–2013. Int J Tuberc Lung Dis. 2015;19:216–22.

16. Shah M, Chihota V, Coetzee G, Churchyard G, Dorman SE. Comparison of laboratory costs of rapid molecular tests and conventional diagnostics for detection of tuberculosis and drug-resistant tuberculosis in South Africa. BMC Infec Dis. 2013;13: 352.

17. Cunnama L, Sinanovic E, Ramma L, et al. Using top-down and bottom-up costing approaches in LMICs: The case for using both to assess the incremental costs of new technologies at scale. Health Econ. 2016;25:S1.

18. Adcock CJ. A Bayesian approach to calculating sample size. Statistician. 1988:433–9.

19. Joseph L, du Berger R, Bélisle P. Bayesian and mixed Bayesian/likelihood criteria for sample size determination. Stat Med. 1997;16:769–81.

20. PropMisclassSampleSize Version 5.9 [Computer software]. 2016. Available from: URL: http://www.medicine.mcgill.ca/epidemiology/Joseph/software/Bayesian-Sample-Size.html.

21. Ling DI, Pai M, Schiller I, et al. A Bayesian framework for estimating the incremental value of a diagnostic test in the absence of a gold standard. BMC Med Res Methodol. 2014;14:67.

22. Wang Z, Dendukuri N, Joseph L. Understanding the effects of conditional dependence in research studies involving imperfect diagnostic tests. Stat Med. 2017;36:466–80.