LATENT CLASS ANALYSIS:

AN INDISPENSABLE METHOD FOR DIAGNOSTIC ACCURACY RESEARCH

Nandini Dendukuri

Depart of Medicine,
Department of Epidemiology, Biostatistics and Occupational Health,
McGill University

Technology Assessment Unit, McGill University Health Centre

Outline

- What are latent class models? Why are they necessary?
- Why are they not more widely applied in diagnostic research?
- How can we make them more accessible?

An example from health technology assessment

- Should the MUHC approve purchase of a urinary antigen (UA) test to diagnose *streptococcus* pneumonia?

- Pneumonia commonly suspected in hospitalized patients, but rarely confirmed
  - *Standard culture test has poor sensitivity, takes time*

- Most cases treated empirically with antibiotics
  - *Concern for increased risk of C. difficile diarrhea, antibiotic resistance*

An example from health technology assessment

- An urinary antigen test (UA) with improved sensitivity, better turn around time could aid in choosing targeted antibiotics

- Questions of interest
  - *What is the expected increase in true positives? In false positives?*
  - *Is the addition of the UA test to the routine work-up cost-effective?*

- To answer these questions we carried out a systematic review of studies that estimated the sensitivity and specificity of the urinary antigen test

Results of systematic review

- 27 studies identified

- Statistical analysis involves comparing UA test to assorted reference standards
  - *Not possible to compare results across studies.*
  - *Typical of problems where no perfect reference exists*

- Most common reference standard is a composite of culture tests

Reference class (# of studies)	Reference definition	Plausible range of sensitivity	Plausible range of specificity
A (12 studies)	Blood OR sputum OR respiratory culture positive	40-70%	80-100%
B (11 studies)	Blood OR sputum culture positive	30-60%	80-100%
C (4 studies)	Blood culture positive	10-40%	90-100%

Closer look at one study

- Traditional statistical analysis

UA Sensitivity = $\frac{55}{78}$ = 70.5%

UA Specificity = $\frac{224}{305}$ = 73.4%

		Urinary antigen test		Total
		+	-	
Composite Reference Standard	+	55	23	78
	-	81	224	305

Composite reference standard assumed perfect

⇒ *UA cannot improve over it*

⇒ *Cost-effectiveness analysis would never conclude in favor of UA as it is more expensive and less accurate*

From Sordé et al., Archives of Int Med, 2011

Closer look at one study

- Traditional statistical analysis

$$\text{UA Sensitivity} = \frac{55}{78} = 70.5\%$$

$$\text{UA Specificity} = \frac{224}{305} = 73.4\%$$

		Urinary antigen test		Total
		+	-	
Composite Reference Standard	+	55	23	78
	-	81	224	305

- Since culture has poor sensitivity, some of the 81 may be true positives

  ⇒ *UA Specificity possibly under-estimated*

- The sensitivity could be over- or under-estimated

  ⇒ *We cannot use this estimate as an upper or lower bound*

From Sordé et al., Archives of Int Med, 2011

How can we improve over the traditional analysis?

- Clearly, we need to acknowledge sensitivity and specificity of the composite reference standard are not perfect and need to be estimated,
  - *i.e. we need a latent class analysis*

- Latent class analysis allows us to
  - *estimate increase in true positives detected when using UA*
  - *compare trade-off between true vs false positives on UA*

- It includes the prevalence as an unknown parameter
  - *Therefore, it allows for comparison and pooling of results across studies identified by the systematic review*

A simple latent class model for two tests

		New test (T_1)	
		+	-
Reference (T_2)	+	n_{11}	n_{01}
	-	n_{10}	n_{00}

Assumes each cell is a mixture of disease positive (D+) and disease negative (D-) patients

- Likelihood: $L \propto \prod_{i=0}^{1} \prod_{j=0}^{1} p_{ij}^{n_{ij}}$

- Let D denote the latent disease status. The multinomial probabilities can be expressed as

$$p_{ij} = P(T_1, T_2) = P(T_1, T_2 \mid \mathrm{D}+)\ \mathrm{P}(\mathrm{D}+) + P(T_1, T_2 | D-)P(D-)$$

The terms $P(T_1, T_2 \mid \mathrm{D})$ can be expressed in different ways leading to different types of latent class models

Modeling $P(T_1, T_2 \mid \mathrm{D})$

- **Conditional independence (CI) model:**

Assumes T_1 and T_2 are independent conditional on D, e.g.

$$
\begin{aligned}
&P(T_1 = 1, T_2 = 1) \\
&= P(T_1 = 1, T_2 = 1|\mathrm{D}+)\ \mathrm{P}(\mathrm{D}+) + \mathrm{P}(T_1 = 1, T_2 = 1|\mathrm{D}-)\ \mathrm{P}(\mathrm{D}-) \\
&= P(T_1 = 1|\mathrm{D}+)P(T_2 = 1|\mathrm{D}+)\mathrm{P}(\mathrm{D}+) + P(T_1 = 1|\mathrm{D}-)P(T_2 = 1|\mathrm{D}-)\mathrm{P}(\mathrm{D}-) \\
&= \ S_1 S_2 \pi + (1 - C_1)(1 - C_2)(1 - \pi)
\end{aligned}
$$

where S_1 and S_2 are sensitivities, C_1 and C_2 are the specificities and π is the prevalence

- **Alternatives:**

As the CI model has been criticized for being unrealistic, different approaches have been proposed for allowing T_1 and T_2 to be dependent.

  - *These approaches add more unknown parameters to the model*

Model identifiability

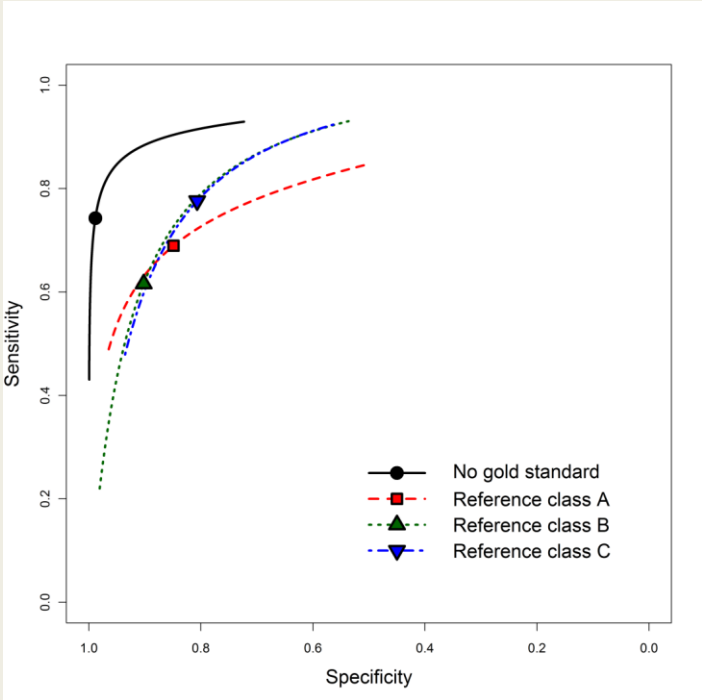
- When tests are dichotomous, it is not uncommon to encounter a situation where we have inadequate degrees of freedom
- Clearly, modeling dependence means we will encounter non-identifiability even when higher numbers of tests are available
- When the model is non-identifiable, external information will be needed in terms of constraints or prior information
  - *This makes Bayesian estimation a natural choice for these models*

# of tests	# of degrees of freedom	# of parameters in CI model
1	1	3
2	3	5
3	7	7
4	15	9
5	31	11

CI model: Conditional independence model

Returning to the health technology assessment question

Summary ROC curves for urinary antigen test



- We can see that the specificity estimate is higher under the latent class analysis, the sensitivity lies in between
- Further, we found that the sensitivities of the imperfect reference standards ranged from about 50-60% and specificities were 98-99%
- These results permitted us to carry out a cost-effectiveness analysis comparing UA to a composite of culture tests

Sinclair et al, J Clin Micro, 2013; Xie et al, Res Synth Meth, 2017

Brief history of the use of latent class (LC) modeling

1968	First introduced by Lazarsfeld and Henry
1974	Maximum likelihood solution proposed by Goodman in Biometrika
1980	First application in diagnostic research by Hui and Walter, Biometrics Proposed a method for dealing with non-identifiability
1985	Model for conditional dependence between a pair of tests proposed by Vacek, Biometrics
1995	Bayesian approach for non-identifiable models, Joseph et al, Am J Epi
1996	Modeling conditional dependence between multiple tests using random effects proposed by Qu et al., Biometrics
2000	Modeling conditional dependence in the presence of non-identifiability using a Bayesian approach, Dendukuri & Joseph, Biometrics
1990s onwards	In reputed journals: • models for conditional dependence • checking model assumptions • sample size estimation • correcting verification bias • meta-analysis

A systematic review of LC models in diagnostic research

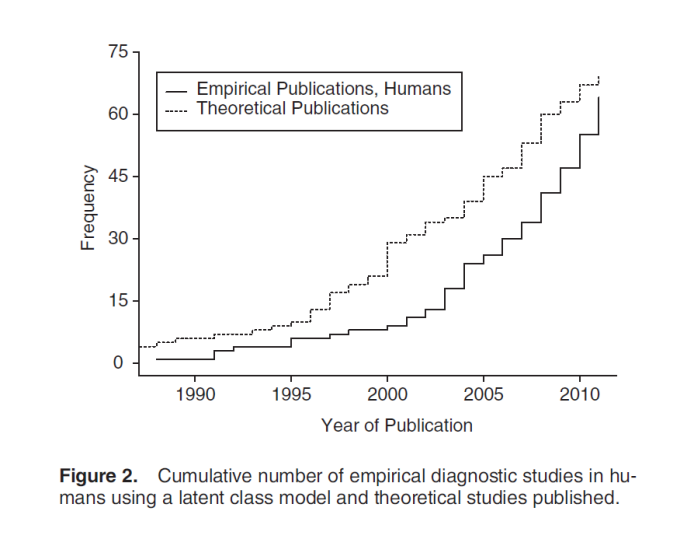


Figure 2. Cumulative number of empirical diagnostic studies in humans using a latent class model and theoretical studies published.

- van Smeden et al., *Am J Epi*, 2013 identified
  - *69 theoretical papers*
  - *64 applied papers in human research + 47 in veterinary sciences*
- Shows that applications of LC models are still not common in human diagnostic research even after 3 decades since the publication by Hui & Walter

Beliefs about LC models in the statistics literature

- "It requires that a minimum of three (imperfect) diagnostic tests be measured on every specimen",

  Alonzo & Pepe, *Stats in Med*, 1999

- "… the CI assumption often fails in practice … considerable bias can occur when the CI assumption is violated …"

  Pepe & Janes, *Biostatistics*, 2007

Beliefs about LC models in the statistics literature

- "The approach yields estimates that are derived from a black box and are not intuitively well connected with the data",

  Pepe, *The statistical evaluation of medical tests for classification and prediction,* 2011

- "… even in cases when models are identifiable, their estimators may not be robust to the assumed dependence structure, and it may be impossible to distinguish between competing conditional dependence models"

  Albert & Dodd, *Biometrics*, 2004

Beliefs about LC models in the medical literature

- "LCA is not designed for hypothesis testing and therefore cannot estimate differences in performance among the three methods, if any exist. Thus, the use of the PIS for comparison purposes is warranted.",

  van der Pol et al., *J Clin Micro*, 2012

- "We recommend using the composite reference standard method [*over latent class analysis*] both for its statistical properties and its relative ease of use."

  Hess et al, *Eur J Clim Micro Inf Dis*, 2012

Beliefs about LC models in the medical literature

- "… latent class analysis is unlikely to provide more confidence about our understanding of the effectiveness of Xpert MTB/RIF in identifying the presence or absence of true [*childhood*] tuberculosis disease."

  Dodd and Wilkinson, *Lancet*, 2013

- "… there is no consensus on the optimal [statistical] approach to evaluating the performance of NAATs [*for Chlamydia trachomatis*]"

  Centers for Disease Control (CDC), 2014

- "… latent class models, now allow investigators to liberate themselves from the restrictive assumption of a perfect reference test and estimate the accuracy of the candidate tests and the reference standard with the same data."

  World Organisation for Animal Health, 2014

Anticipated advantages of using a composite reference standard (CRS)

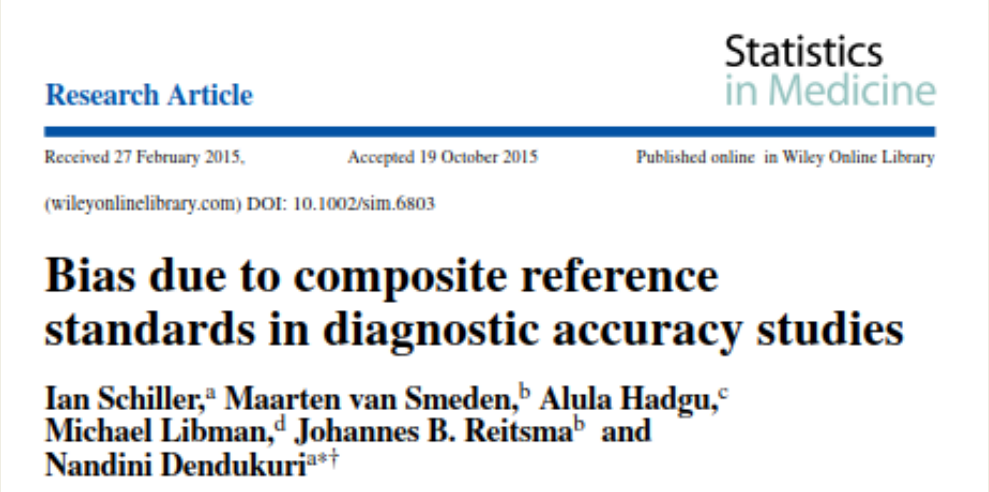
- Increased accuracy in disease classification compared to single imperfect reference test
  - *Therefore, decreased bias in estimated accuracy of test under evaluation and estimated prevalence*

- Avoid incorporation bias because the CRS is independent of the test under evaluation

- Transparency, simplicity
  - *Achieve standardization across studies*

Unanswered questions about composite reference standards

- Does increasing the number of component tests improve the CRS?
- What is the impact of 'conditional dependence' between the test under evaluation and the CRS?
- How do changes in the underlying prevalence affect estimates?



Research Article

Statistics in Medicine

Received 27 February 2015, Accepted 19 October 2015 Published online in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.6803

Bias due to composite reference standards in diagnostic accuracy studies

Ian Schiller,[a] Maarten van Smeden,[b] Alula Hadgu,[c] Michael Libman,[d] Johannes B. Reitsma[b] and Nandini Dendukuri[a*†]

RESEARCH METHODS AND REPORTING

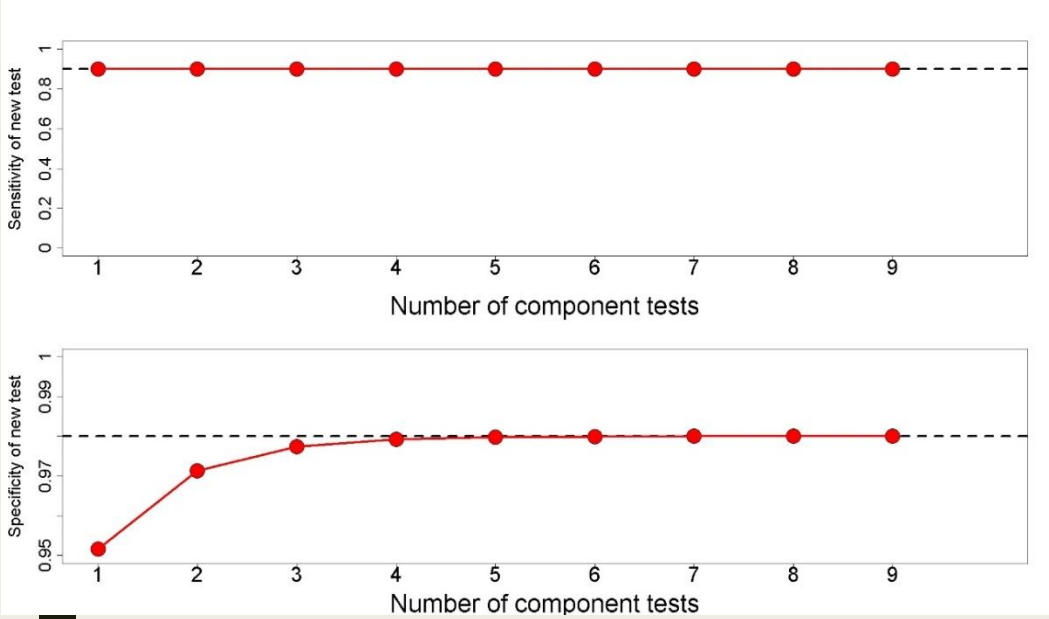
Concerns about composite reference standards in diagnostic research

Cite this as: *BMJ* 2018;360:j5779
http://dx.doi.org/10.1136/bmj.j5779
Accepted: 20 November 2017

Nandini Dendukuri,[1] Ian Schiller,[1] Joris de Groot,[2] Michael Libman,[3] Karel Moons,[2] Johannes Reitsma,[2] Maarten van Smeden[2]

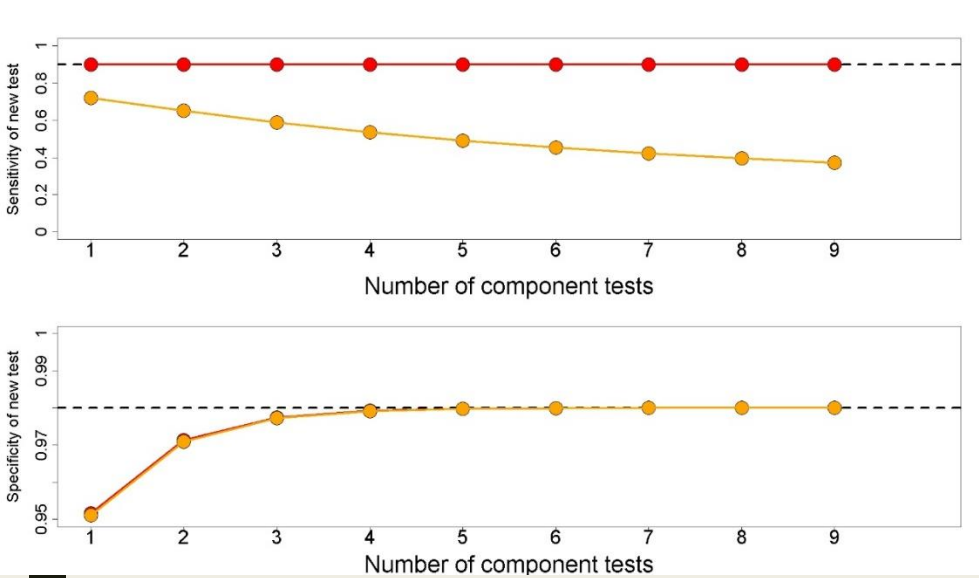
Schiller et al., Stats in Med, 2016, Dendukuri et al., BMJ, 2018

Bias due to OR-rule composite reference standard



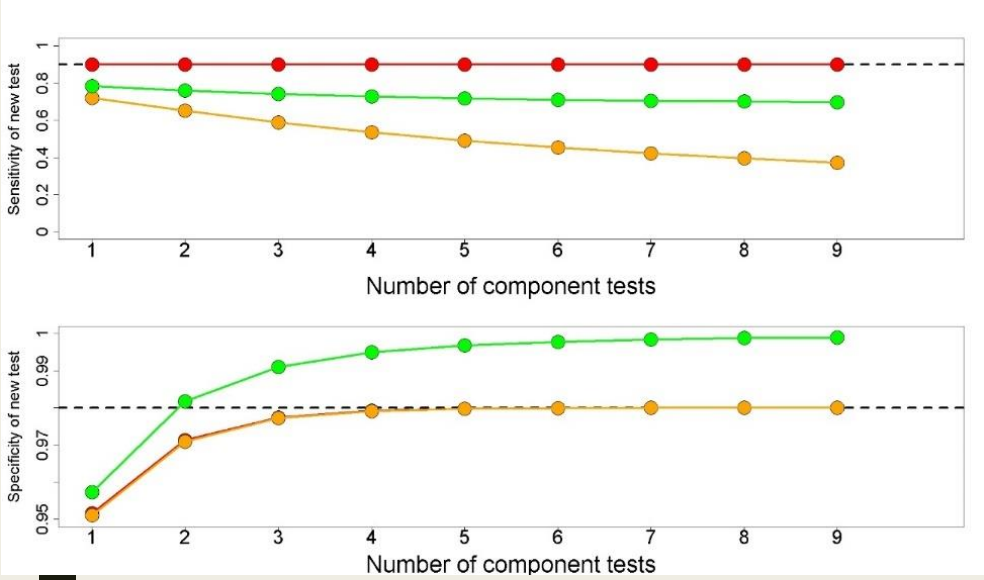
- When component tests have perfect specificity line
  - *Estimate of new test's sensitivity unbiased (i.e. red line falls on the dashed line)*
  - *Bias in estimate of new test's specificity decreases with each added test, eventually becoming unbiased*

Bias due to OR-rule composite reference standard



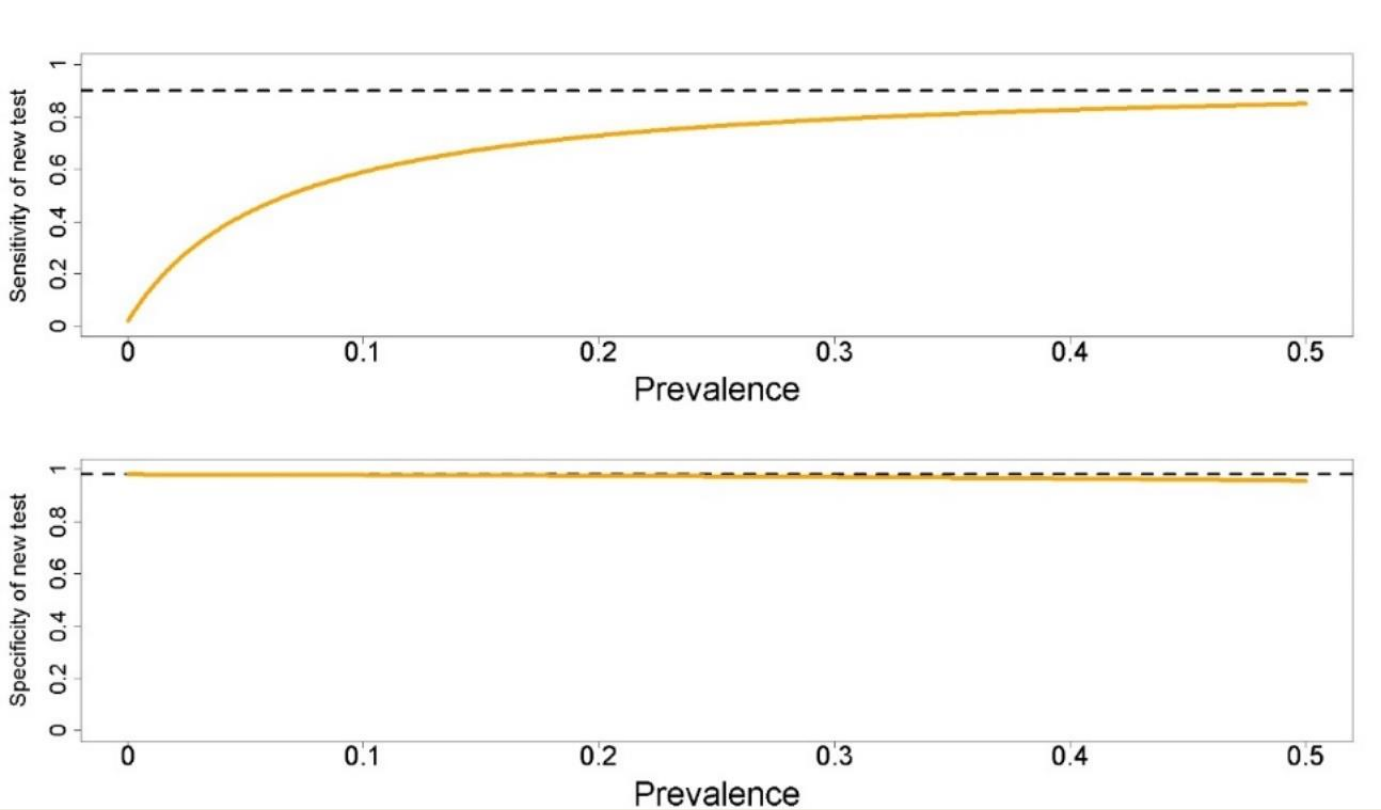
- However, if component tests have 98% specificity
  - *Sensitivity estimate of new test is biased (yellow line), with bias increasing with every component test*

Bias due to OR-rule composite reference standard



- If the component tests have 98% specificity and also make the same errors as the test under evaluation
  - *The specificity of the new test is over-estimated (green line)*

Further, sensitivity and specificity estimates can vary across settings because they depend on the disease prevalence



In summary:

- Problems with composite reference standard more apparent when we examine the impact of increasing the number of tests
  - *Unless specificity of component tests is perfect, new test's sensitivity is underestimated*
  - *When conditional dependence is present, new test's specificity is overestimated*
  - *Bias worsens with increasing number of component tests!*
- Not what is expected of a sound statistical method

Bias due to composite reference standards

- Other types of composite reference standards (e.g. based on an AND rule) also have similar problems

- We also found that CRS based estimates are not comparable across studies, because they are functions of the underlying disease prevalence

- Poor performance of the CRS can be explained by the fact that it makes sub-optimal use of the data
  - *It makes a simplistic classification*
  - *And then it ignores the uncertainty in that classification*

How can we improve over the CRS?

- We need an approach that
  - *Uses the complete cross-tabulation between all imperfect tests without simplifying it*
  - *An approach that models conditional dependence*
  - *An approach that includes the prevalence as an unknown parameter*

i.e. we need latent class analysis!

Returning to latent class models

- What are the challenges in estimating latent class models?
  - *Interpreting the latent disease status*
  - *Selecting the appropriate conditional dependence structure*
  - *Dealing with non-identifiability*

An illustrative example: Childhood Pulmonary Tuberculosis (TB)

- Diagnosis of childhood pulmonary TB relies on multiple tests/signs as no single measure is considered adequate:
  - *Microbiological tests (e.g. Culture, Xpert)*
  - *Symptoms/signs of TB*
  - *Chest radiograph*
  - *Immunologic evidence of TB (e.g. tuberculin skin test (TST))*
  - *Contact with TB patient*

- A consequence is that there are no reliable estimates for the burden of childhood pulmonary TB, despite it being a major public health problem

Clinical Case Definitions for Classification of Intrathoracic Tuberculosis in Children: An Update

Stephen M. Graham,[1,2,3] Luis E. Cuevas,[4] Patrick Jean-Philippe,[5] Renee Browning,[6] Martina Casenghi,[7] Anne K. Detjen,[2] Devasena Gnanashanmugam,[6] Anneke C. Hesseling,[8] Beate Kampmann,[9,10] Anna Mandalakas,[11] Ben J. Marais,[12] Marco Schito,[5,a] Hans M. L. Spiegel,[5] Jeffrey R. Starke,[11] Carol Worrell,[13,b] and Heather J. Zar[14]

Goal: "… enhance harmonized classification … across studies, resulting in greater comparability and the much-needed ability to pool study results."

Case Definition	Refined Criteria[a]
Confirmed tuberculosis	Bacteriological confirmation obtained Requires *Mycobacterium tuberculosis* to be confirmed (culture or Xpert MTB/RIF assay) from at least 1 respiratory specimen

Unconfirmed tuberculosis	Bacteriological confirmation NOT obtained AND at least 2 of the following:
	• Symptoms/signs suggestive of tuberculosis (as defined)
	• Chest radiograph consistent with tuberculosis
	• Close tuberculosis exposure or immunologic evidence of *M. tuberculosis* infection
	• Positive response to tuberculosis treatment (requires documented positive clinical response on tuberculosis treatment—no time duration specified)
	- With *M. tuberculosis* infection
	  • Immunological evidence of *M. tuberculosis* infection (TST and/or IGRA positive)
	- Without *M. tuberculosis* infection
	  • No immunological evidence of *M. tuberculosis* infection

Unlikely tuberculosis	Bacteriological confirmation NOT obtained AND Criteria for "unconfirmed tuberculosis" NOT met
	- With *M. tuberculosis* infection
	  • Immunological evidence of *M. tuberculosis* infection (TST and/or IGRA positive)
	- Without *M. tuberculosis* infection
	  • No immunological evidence of *M. tuberculosis* infection

Latent class model for childhood pulmonary TB

American Journal of Epidemiology Advance Access published October 13, 2016

American Journal of Epidemiology

DOI: 10.1093/aje/kww094

Practice of Epidemiology

Diagnostic Test Accuracy in Childhood Pulmonary Tuberculosis: A Bayesian Latent Class Analysis

Samuel G. Schumacher*, Maarten van Smeden*, Nandini Dendukuri, Lawrence Joseph, Mark P. Nicol, Madhukar Pai, and Heather J. Zar

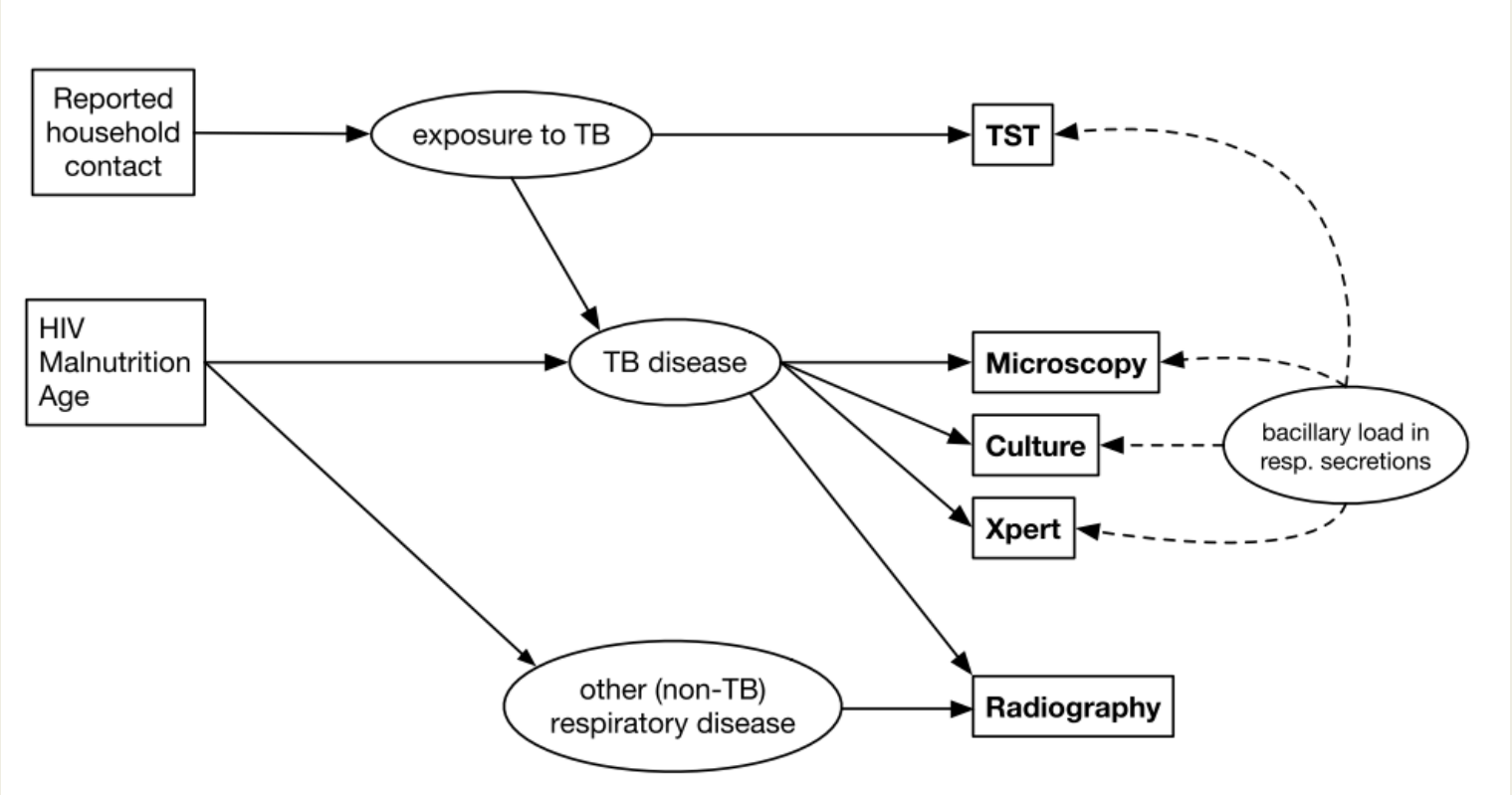
Schumacher et al, Am J Epi, 2016

Latent class model for childhood pulmonary TB

- We had data from a cohort of 749 children hospitalized with suspected pulmonary TB in South Africa
- A heuristic model was set up to explain how the observed data relate to the latent variables
- Importantly, both clinicians and methodologists were involved in this exercise

Schumacher et al, Am J Epi, 2016

Heuristic Model



Heuristic Model

- 3 possible latent variables were identified
  - *Active TB disease*
  - *Exposure to TB (latent TB)*
  - *Other respiratory disease*
- Combinations of these latent variables would lead to four possible latent classes. Of these two (Active TB, Not active TB) were considered relevant and distinguishable with the available tests
- Conditional dependence is anticipated between 4 of the tests
- Covariates Age, HIV and MaInutrition affect model parameters
- The preferred model was defined at the outset rather than by relying on statistical criteria

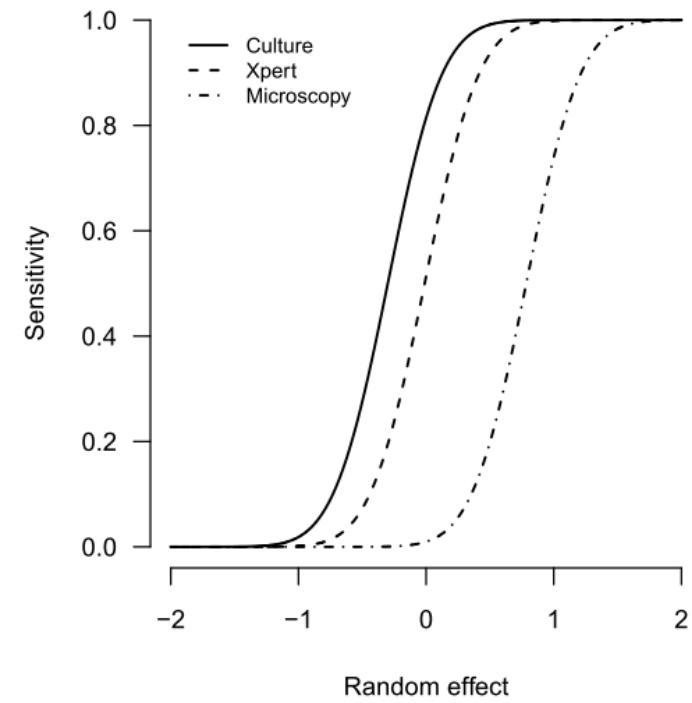
Modeling conditional dependence

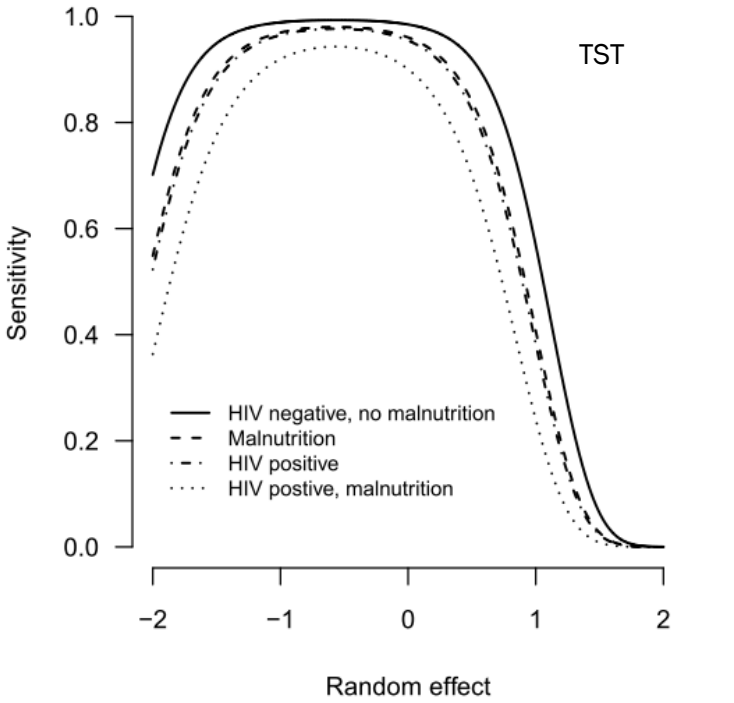
- Culture, Xpert and Smear are all influenced by bacillary load
  - *They could all be false negative for the same group of children who have a low bacillary load, leading to a high positive dependence*

- The TST test is expected to be negatively correlated with the severity of infection (which is also affected by the bacillary load)

Wang et al, Stats in Med, 2016

Random effect used to model conditional dependence





Results

		Conditional Independence Model		Model Adjusting for conditional dependence	
Test and Parameter		Posterior Median Estimate	95% CrI	Posterior Median Estimate	95% CrI
CPTB Prevalence		16.6	15.6 – 18.0	26.7	20.8 – 35.2
Culture					
	Sensitivity	96.7	87.8 – 99.8	60.0	45.7 – 75.5
	Specificity	99.8	98.9 –100.0	99.6	98.7 –100.0
Xpert					
	Sensitivity	74.4	66.0 – 82.2	49.4	37.7 – 62.2
	Specificity	98.3	97.0 – 99.4	98.6	97.3 – 99.5
Microscopy					
	Sensitivity	33.3	25.3 – 42.1	22.3	15.6 – 30.3
	Specificity	99.8	99.2 –100.0	99.7	99.0 –100.0
Radiography					
	Sensitivity	65.4	56.5 – 73.8	64.2	54.9 – 72.8
	Specificity	73.1	69.6 – 76.6	78.0	73.4 – 83.4
TST					
	Sensitivity	69.0	60.5 – 76.7	75.2	61.2 – 83.8
	Specificity	62.4	58.5 – 66.1	69.3	63.2 – 75.9

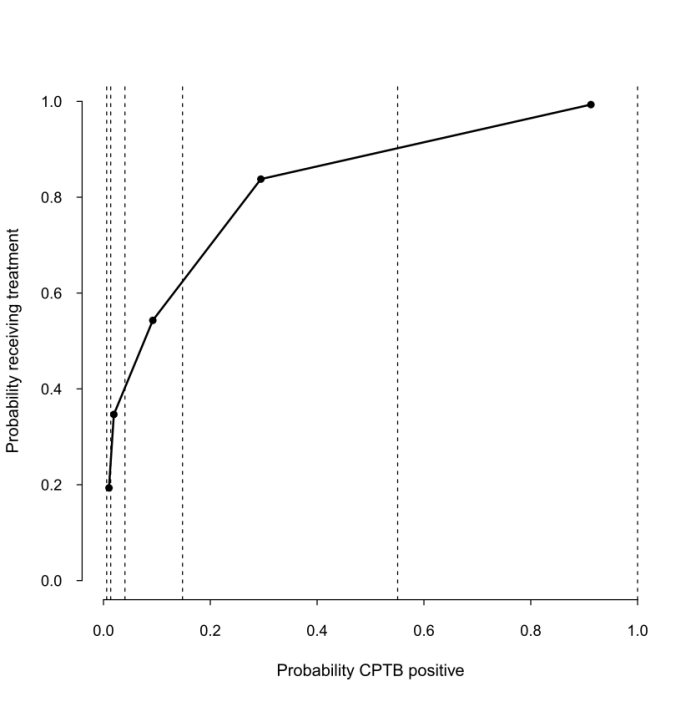
Test outcome pattern					Observed frequency	Predicted probability of TB	
Cu	Xp	Mi	Ra	TS		%	95% CrI
0	0	0	0	0	296	2	0 – 7
0	0	0	0	1	149	16	5 – 33
0	0	0	1	0	87	9	0 – 34
0	0	0	1	1	78	52	26 – 74
0	0	1	0	1	1	11	0 – 100
0	1	0	0	0	5	4	0 – 40
0	1	0	0	1	7	56	0 – 100
0	1	0	1	0	2	12	0 – 100
0	1	0	1	1	2	88	50 – 100
1	0	0	0	0	3	23	0 – 100
1	0	0	0	1	8	93	62 – 100
1	0	0	1	0	1	54	0 –100
1	0	0	1	1	20	99	90 –100
1	1	0	0	0	1	100	100 - 100
1	1	0	0	1	17	100	100 - 100
1	1	0	1	0	4	100	100 - 100
1	1	0	1	1	27	100	100 - 100
1	1	1	0	0	8	100	100 - 100
1	1	1	0	1	5	100	100 - 100
1	1	1	1	0	21	100	100 - 100
1	1	1	1	1	7	100	100 - 100

Predicted probabilities

- Examining the predicted probabilities is another way to see how the observed data relates to the latent disease status

- This is another advantage of latent class analysis over descriptive classification methods like cluster analysis

Probability child was treated increased with probability of CPTB



- Without a perfect reference, we can only evaluate the face validity of our latent class model
- An estimated 95.5% of TB positive children receive anti-TB treatment
- An estimated 45.8% of TB negative children receive anti-TB treatment

Future research

- Fit the CPTB model in other datasets drawn from other settings
  - *Settings where the prevalence of active TB is lower may lead to other choices for latent classes*
  - *Use datasets where more variables are recorded*
- Develop robust latent class models for other disease areas
- Develop prediction models that can help optimize diagnosis?
- Several interesting methodological questions remain to be answered!

References

- Lazarsfeld, P. F., and Henry, N. W. (1968), *Latent Structure Analysis*, Boston: Houghton Mifflin.
- Goodman, L. a. Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika* **61,** 215 (1974).
- Hui, S. & Walter, S. Estimating the error rates of diagnostic tests. *Biometrics* **36**, 167–171 (1980).
- Vacek, P. M. The effect of conditional dependence on the evaluation of diagnostic tests. *Biometrics* **41,** 959–68 (1985).
- Joseph, L., Gyorkos, T. W. & Coupal, L. Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard. *Am. J. Epidemiol.* **141,** 263–273 (1995).
- Qu, Y., Tan, M. & Kutner, M. H. Random effects models in latent class analysis for evaluating accuracy of diagnostic tests. *Biometrics* **52,** 797–810 (1996).
- Dendukuri, N. & Joseph, L. Bayesian approaches to modeling the conditional dependence between multiple diagnostic tests. *Biometrics* 158–167 (2001).
- Wang, Z. Y., Dendukuri, N. & Joseph, L. Understanding the effects of conditional dependence in research studies involving imperfect diagnostic tests. *Stat. Med.* (2016).
- Sinclair, A., Xie, X., Teltscher, M. & Dendukuri, N. Systematic review and meta-analysis of a urine-based pneumococcal antigen test for diagnosis of community-acquired pneumonia caused by Streptococcus pneumoniae. *J. Clin. Microbiol.* **51,** 2303–2310 (2013).
- Schiller, I. *et al.* Bias due to composite reference standards in diagnostic accuracy studies. *Stat. Med.* n/a-n/a (2015).
- Schumacher, S. G. *et al.* Diagnostic Test Accuracy in Childhood Pulmonary Tuberculosis: A Bayesian Latent Class Analysis. *Am. J. Epidemiol.* (2016).
- van Smeden M, Naaktgeboren CA, Reitsma JB, Moons KGM, de Groot JAH. Latent Class Models in Diagnostic Studies When There is No Reference Standard--A Systematic Review. Am J Epidemiol [Internet]. 2014;179(4):423–31.