

Real examples of when to use Bayesian analysis I: animal health



Julio Alvarez

VISAVET Health Surveillance Centre, Animal Health Department
Universidad Complutense

Quick outline

- The problem of evaluation of diagnostic tests: the gold standard dilemma
- Use of Bayesian statistics for evaluation of diagnostic tests
- Practical example: using BLCMs for estimating test performance of bovine tuberculosis diagnostic tests

Diagnostic test evaluation

	Reference +	Reference -	Total
Test +			
Test -			
Total			N

Pos
Sensitivity is the ability of the test to correctly identify diseased individuals

Neg
Specificity is the ability of the test to correctly identify healthy individuals

Healthy

$$Se = \frac{TP}{Sick}$$

$$PPV = \frac{TP}{pos}$$

$$AP = \frac{Pos}{N}$$

$$Sp = \frac{TN}{Healthy}$$

$$NPV = \frac{TN}{Neg}$$

$$TrP = \frac{Sick}{N}$$

The problem of diagnostic test evaluation

- Typically new/alternative tests are compared with well established/standardized tests (“gold standard”/ “reference test”)
- This only informs of the performance of the new test **relative** to the old one!...
... unless we assume the gold standard is perfect (what we typically do)

Perfect test???

Perfect test????



Case example: mycobacterial diseases

- No perfect test (not even close...)
- Chronic slow diseases → delayed (non-protective) (humoral) immune response
- All known tests lack sensitivity and/or specificity
- Usual gold standard is culture... with a sensitivity <50% in multiple cases

Alternatives?

- Choose individuals of known status and see proportion of test positive or test negative animals:
 - Negative: never saw the pathogen (never: different location/epidemiological settings)
 - Positive: experimental infections
- External validity???



Be careful when using certain terms!

Highly Accurate Antibody Assays for Early and Rapid Detection of Tuberculosis in African and Asian Elephants[▽]

Rena Greenwald,¹ Olena Lyashchenko,¹ Javan Esfandiari,¹ Michele Miller,² Susan Mikota,³ John H. Olsen,⁴ Ray Ball,⁴ Genevieve Dumonceaux,⁴ Dennis Schmitt,⁵ Torsten Moller,⁶ Janet B. Payeur,⁷ Beth Harris,⁷ Denise Sofranko,⁸ W. Ray Waters,⁹ and Konstantin P. Lyashchenko^{1*}

recognized by elephant antibodies during disease. The serologic assays demonstrated 100% sensitivity and 95 to 100% specificity. Rapid and accurate antibody tests to identify infected elephants will likely allow

on disease status and history of exposure (Table 1). The TB-infected group included 26 animals from 17 herds with culture-confirmed TB due to *M. tuberculosis* ($n = 25$) or *M. bovis* ($n = 1$). Of the 26 elephants, 7 died and 11 were humanely euthanized. TB was not necessarily the cause of death or the reason for euthanasia. Disease was diagnosed antemortem by trunk wash culture ($n = 15$; 58%) or only postmortem by isolating *M. tuberculosis* or *M. bovis* from various tissues ($n = 11$; 42%). Ten elephants were treated with first-line anti-TB drugs

Development and Evaluation of an Enzyme-Linked Immunosorbent Assay for Use in the Detection of Bovine Tuberculosis in Cattle^{▽†}

W. R. Waters,^{1*} B. M. Buddle,² H. M. Vordermeier,³ E. Gormley,⁴ M. V. Palmer,¹ T. C. Thacker,¹ J. P. Bannantine,¹ J. R. Stabel,¹ R. Linscott,⁵ E. Martel,⁵ F. Milian,⁶ W. Foshaug,⁷ and J. C. Lawrence⁵

TABLE 1. Sensitivity of IDEXX *M. bovis* antibody ELISA with sera collected from naturally infected cattle

Source	ID or characterization ^c	n^d	No. of herds	Sensitivity (%)		
				Lot 1	Lot 2	Lot 3
Great Britain	AHVLA-2 ^a	134	31	74.6	72.4	73.1
	AHVLA-1 ^b	50	>5	86	88	86
Ireland	No visible lesions ^b	50	>5	48	44	46
	With visible lesions ^b	50	>5	72	70	70
	Skin test positive, Bovigam positive, with visible lesions ^b	30	22	96.7	86.7	90.0
New Zealand	AgResearch ^b	42	7	42.9	40.5	35.7
USA	Colorado ^a	81	1	44.4	45.7	49.4
	NVSL serum bank ^a	31	12	48.4	48.4	48.4
	Michigan ^a	10	1	30	30	30
Overall value		478	>89	63.6	61.9	62.6

^a Infection status determined by histopathology with IS6110 PCR and/or culture.

^b Infection status determined by culture or presence of gross lesions and/or from a tuberculosis-affected herd.

^c ID, identification. NVSL, National Veterinary Service Laboratory.

^d n , number of animals.

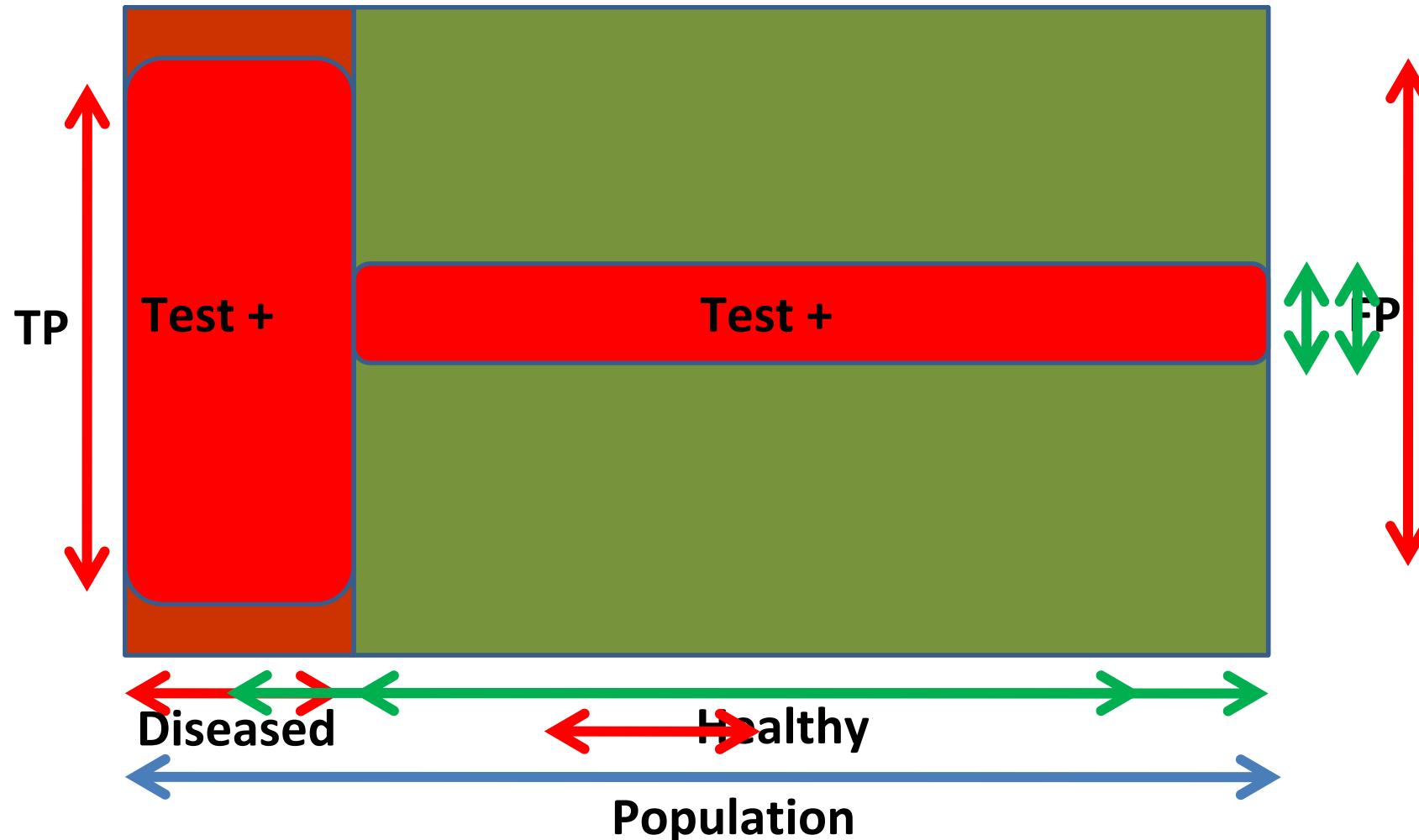
TABLE 2. Specificity of IDEXX *M. bovis* antibody ELISA with sera collected from noninfected cattle from various geographic regions

Source of noninfected sera ^a	n^b	No. of herds	Specificity (%)		
			Lot 1	Lot 2	Lot 3
Maine	126	2	99.2	98.4	99.2
Maine	126	2	99	98.4	99.2
Pennsylvania	79	1	88.6	92.4	98.7
Arkansas	39	1	100	100	97.4
New York	84	1	98.8	100	98.8
North Dakota	110	1	96.3	97.2	99.1
Washington	84	2	98.8	98.8	98.8
South Dakota	84	1	98.8	100	100
Missouri	92	>2	94.5	95.7	98.9
Texas	96	>2	93.7	93.8	95.8
Michigan	92	2	100	100	100
Iowa	8	1	100	100	100
Colorado	121	11	99.2	100	99.2
Great Britain (AHVLA)	50	>5	94	98	96
Ireland (UCD/DAFF)	92	16	100	100	95.7
Austria	316	>10	97.5	99.1	98.1
Overall value	1,473	>58	97.4	98.2	98.4

^a Samples obtained from cattle from tuberculosis-free herds. AHVLA, Animal Health and Veterinary Laboratories Agency; UCD, University College Dublin; DAFF, Department of Agriculture, Fisheries and Food.

^b n , number of animals.

Gold-standard vs. Bayesian



So in reality...

	Test 2 +	Test 2 -	Total
Test 1 +	a	b	a+b
Test 1 -	c	d	c+d
Total	a+c	b+d	N

Sensitivity 1 and Sensitivity 2???

Specificity 1 and Specificity 2???

Alternatives to the “gold standard”-based approach

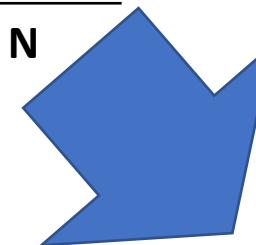
- “Hui and Walter model”(Hui and Walter, 1980)
“If two tests are applied simultaneously to the same individuals from two populations with different disease prevalences, then assuming conditional independence of the errors of the two tests, the error rates of both tests and the true prevalences in both populations can be estimated by a maximum likelihood procedure”.
- Not originally Bayesian: can be solved using ML methods...
but require 1) large sample size, 2) assume tests are independent and 3) their accuracy is assumed to be constant across populations

The Hui-Walter paradigm

- Hui-Walter model implementation to be further discussed in the next session

	Test 2 +	Test 2 -	Total
Test 1 +	a	b	a+b
Test 1 -	c	d	c+d
Total	a+c	b+d	N

Diseased population (P)



Non-diseased population (1-P)

	Test 2 +	Test 2 -
Test 1 +	$p \times Se_1 \times Se_2$	$p \times Se_1 \times (1 - Se_2)$
Test 1 -	$p \times (1 - Se_1) \times Se_2$	$(1 - p) \times (1 - Se_1) \times (1 - Se_2)$

	Test 2 +	Test 2 -
Test 1 +	$(1 - p) \times (1 - Sp_1) \times (1 - Sp_2)$	$(1 - p) \times (1 - Sp_1 \times Sp_2)$
Test 1 -	$(1 - p) \times Sp_1 \times (1 - Sp_2)$	$(1 - p) \times Sp_1 \times Sp_2$

	Test 2 +	Test 2 -	Total
Test 1 +	a	b	a+b
Test 1 -	c	d	c+d
Total	a+c	b+d	N

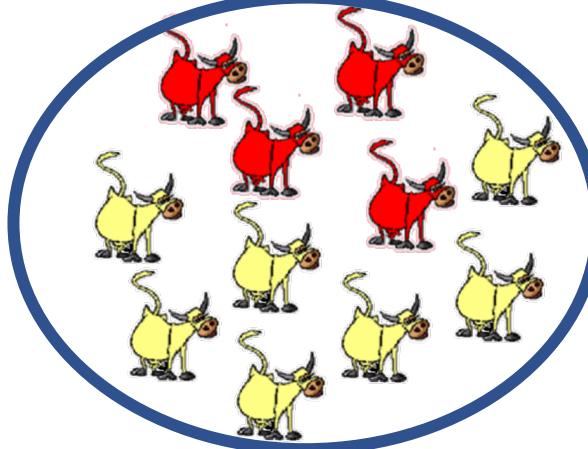
$$P(T1+, T2+) = a/N =$$

$$P \times Se_{T1} \times Se_{T2}$$

+

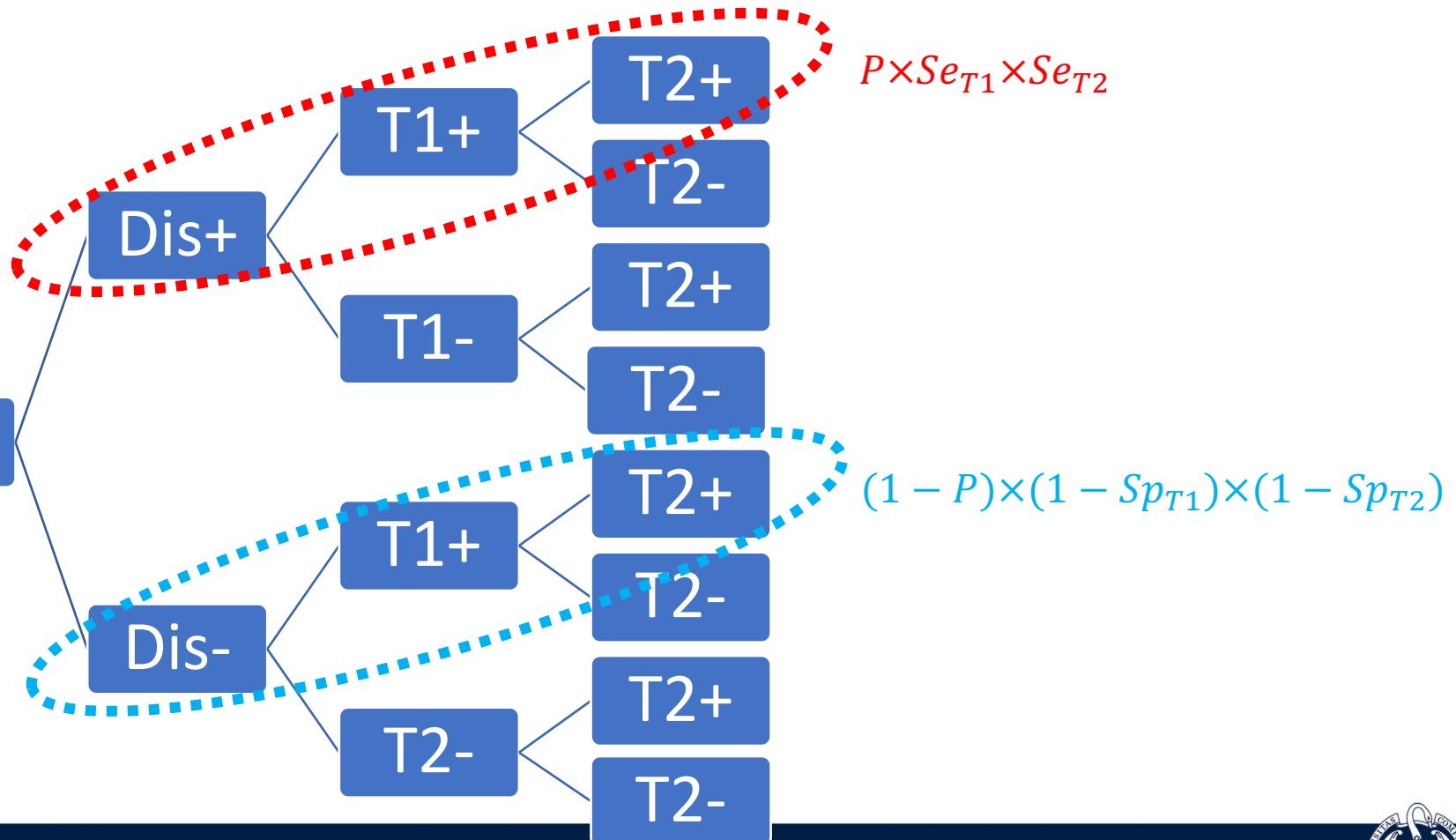
$$(1 - P) \times (1 - Sp_{T1}) \times (1 - Sp_{T2})$$

Diseased population (P)



Non-diseased population (1-P)

Pop



The Hui-Walter paradigm

- $T1+T2+= P*Se1*Se2 + (1-P)*(1-Sp1)*(1-Sp2)$
- $T1+T2-= P*Se1*(1-Se2) + (1-P)*(1-Sp1)*Sp2$
- $T1-T2+= P*(1-Se1)*Se2 + (1-P)*Sp1*(1-Sp2)$
- $T1-T2-= P*(1-Se1)*(1-Se2) + (1-P)*Sp1*Sp2$

5 parameters to estimate ($P, Se1, Se2, Sp1, Sp2$) vs. 3 degrees of freedom
Non-identifiable model ☹

The Hui-Walter paradigm

“If two tests are applied simultaneously to the same individuals from two populations with different disease prevalences, then assuming conditional independence of the errors of the two tests, the error rates of both tests and the true prevalences in both populations can be estimated by a maximum likelihood procedure”

- Population 1

$$T1+T2+: P1*Se1*Se2+(1-P1)*(1-Sp1)*(1-Sp2)$$

$$T1+T2-: P1*Se1*(1-Se2)+(1-P1)*(1-Sp1)*Sp2$$

$$T1-T2+: P1*(1-Se1)*Se2+(1-P1)*Sp1*(1-Sp2)$$

$$T1-T2-: P1*(1-Se1)*(1-Se2)+(1-P1)*Sp1*Sp2$$

- Population 2

$$T1+T2+: P2*Se1*Se2+(1-P2)*(1-Sp1)*(1-Sp2)$$

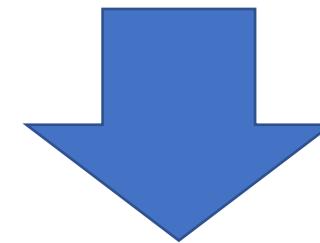
$$T1+T2-: P2*Se1*(1-Se2)+(1-P2)*(1-Sp1)*Sp2$$

$$T1-T2+: P2*(1-Se1)*Se2+(1-P2)*Sp1*(1-Sp2)$$

$$T1-T2-: P2*(1-Se1)*(1-Se2)+(1-P2)*Sp1*Sp2$$

6 parameters to estimate ($P1, P2, Se1, Se2, Sp1, Sp2$)
vs.

6 degrees of freedom
Identifiable model ☺



But is based on (strong) assumptions:

- 2 populations with different prevalences
- Se y Sp constant across populations
- Conditional Independence between tests

Conditional Independence?

- The result of one test is conditionally independent from the other (knowing if one diseased/non-diseased animal is positive/negative in one test gives no information on whether it will be positive/negative in the other (e.g., knowing that a card is a spades gives no information on whether it is a figure))
- Often not true!! (similar tests/tests based on similar principles)
- Requires adding additional terms to equations (Vacek, 1985)

Considering conditional dependence

Number of parameters to estimate increases rapidly as the number of potentially dependent tests increases

TABLE 2. Maximum Number of Estimable Parameters and Number of Parameters to Be Estimated in the Absence of Conditional Independence and Under Conditional Independence as a Function of the Number of Tests per Subject

Number of Tests	Maximum Number of Estimable Parameters	Parameters to be Estimated Under Conditional Dependence	Parameters to Be Estimated Under Conditional Independence
1	1	3	3
2	3	7	5
3	7	15	7
4	15	31	9
5	31	63	11
h	$2^h - 1$	$2^{h+1} - 1$	$2h + 1$

Berkvens D et al. (2006) Estimating Disease Prevalence in a Bayesian Framework Using Probabilistic Constraints.
doi: 10.1097/01.ede.0000198422.64801.8d

Alternatives to the “gold standard”-based approach

- If we could only incorporate some prior knowledge... we could get rid of large sample theory assumptions and “guide” our models
 - We typically have some information on diagnostic performance (i.e., test sensitivity is $>20\%$)
 - We typically have some information on disease prevalence (i.e., proportion of infected is $<50\%$)



Bayesians to the rescue!



- “How to revise our beliefs in the light of evidence” → probability of an event based on prior knowledge of conditions related to the event
- In its more general form:

$$Prior \times Likelihood = Posterior$$

- In its “primitive” definition: Bayes’ rule

$$P(\theta|Y) = \frac{P(\theta) \times P(Y|\theta)}{P(Y)}$$

Bayesian applied to diagnostic test evaluation

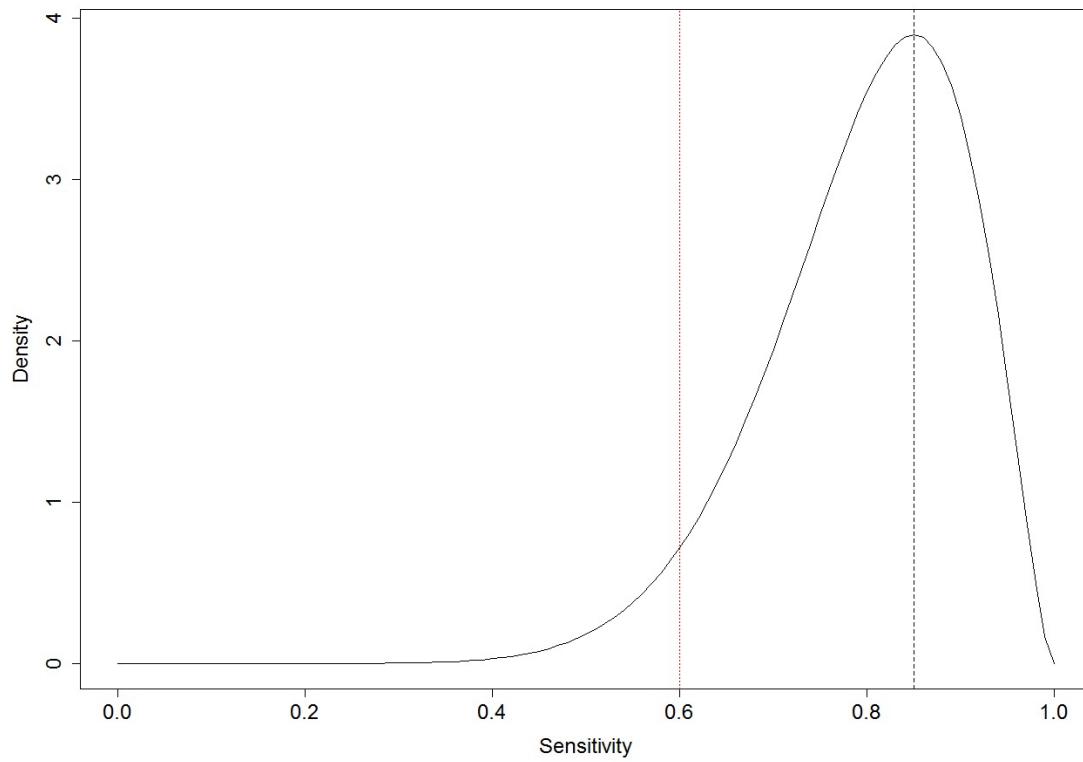
- Joseph et al. (1995) **Bayesian** estimation of disease prevalence and diagnostic test evaluation in the absence of a gold standard

$$P(\theta|Y) = \frac{P(\theta) \times P(Y|\theta)}{P(Y)}$$

- Targets
 - Prevalence $\pi = P(D+)$
 - Sensitivity $Se_i = P(T_i+ | D+)$
 - Specificity $Sp_i = P(T_i- | D-)$
- Prior knowledge (beta distributions)
 - $\pi \sim Beta(a_\pi, b_\pi)$
 - $Se_i \sim Beta(a_{sei}, b_{sei})$
 - $Sp_i \sim Beta(a_{spi}, b_{spi})$

What is the sensitivity of test X?

- Option 1: 85%
- Option B:
 - Typically ~85%
 - Usually >60%

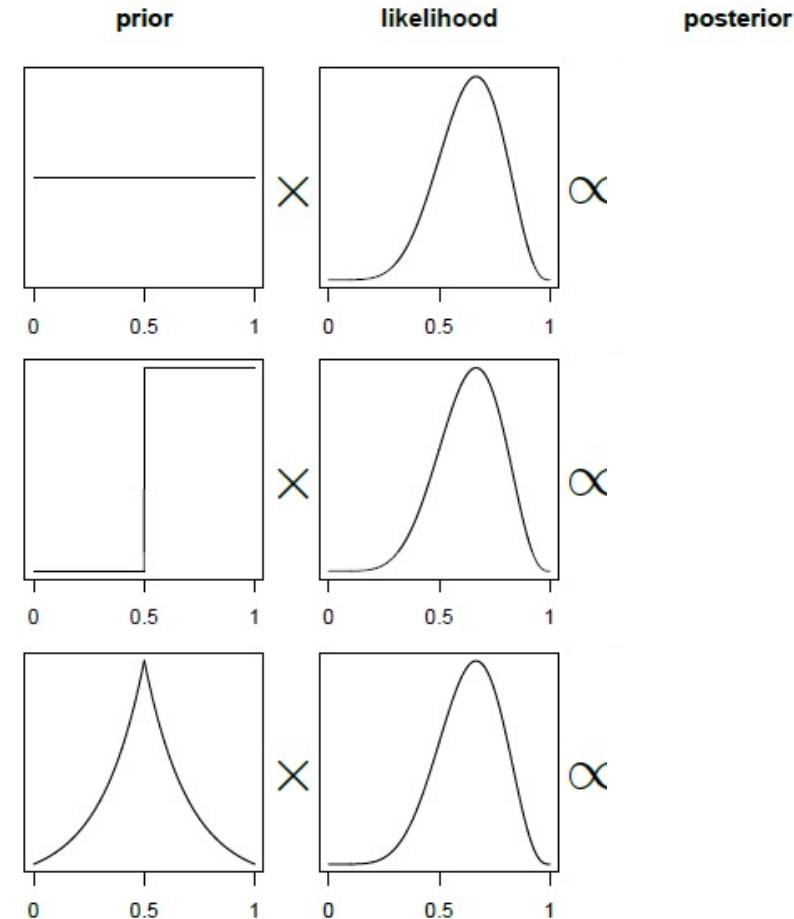


So it is a perfect match!!

- Bayesian statistics allows us to obtain a posterior distribution that can incorporate some prior belief (how sensitive do I think the test is) and some data (what result did I have in population X?)

Posterior is a function of the prior and the likelihood

- The more weight on the prior, the lower the impact of the data (and viceversa)



So how do we estimate the posterior?

- Other than analytical mathematics (always unattractive for non-mathematicians, and sometimes non-computable depending on the model) there are several numerical techniques, such as:
 - Grid approximation
 - Quadratic approximation
 - **Markov Chain Monte Carlo (beloved MCMC)** → several tools developed in the last decades that allow its easy implementation, e.g.
 - Bayesian inference Using Gibbs Sampling: BUGS (1997) → WinBUGS, OpenBUGS
 - Just Another Gibbs Sampler: JAGS (2007)
 - Stan (based on Hamiltonian/hybrid Monte Carlo)

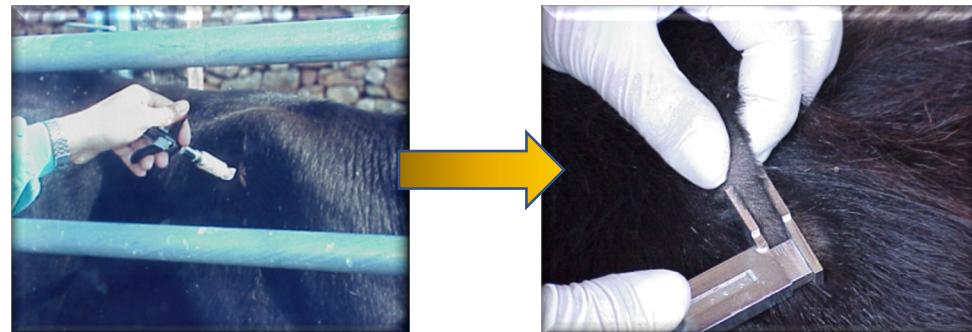
How does this work in practice? Bovine TB

- Imagine you are the manager of a bTB disease eradication program in a (non-OTF) EU region
- There is a (sort of) new test becoming increasingly used (IGRA) that does not agree (so much) with previously established standard (skin test) or with the usual gold standard (bacteriology) so people are concerned (there are more positives now → false positives??)
- Gold standard is now to underperform, but you need to have an (ideally unbiased) estimate on the performance of the new test
- How can you move forward?



How does this work in practice? Bovine TB

- First of all: tests available (and expected performance if needed)
 - Skin test (~Mantoux): single or comparative intradermal intradermotuberculinization



- Interferon-gamma release assay (IGRA): “in-vitro” skin test (sort of!)

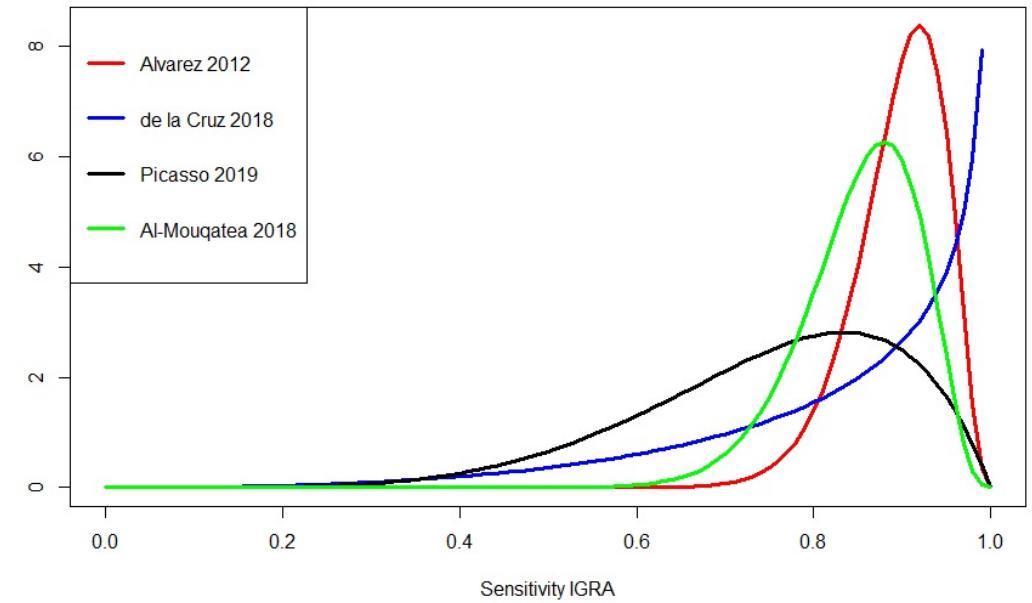
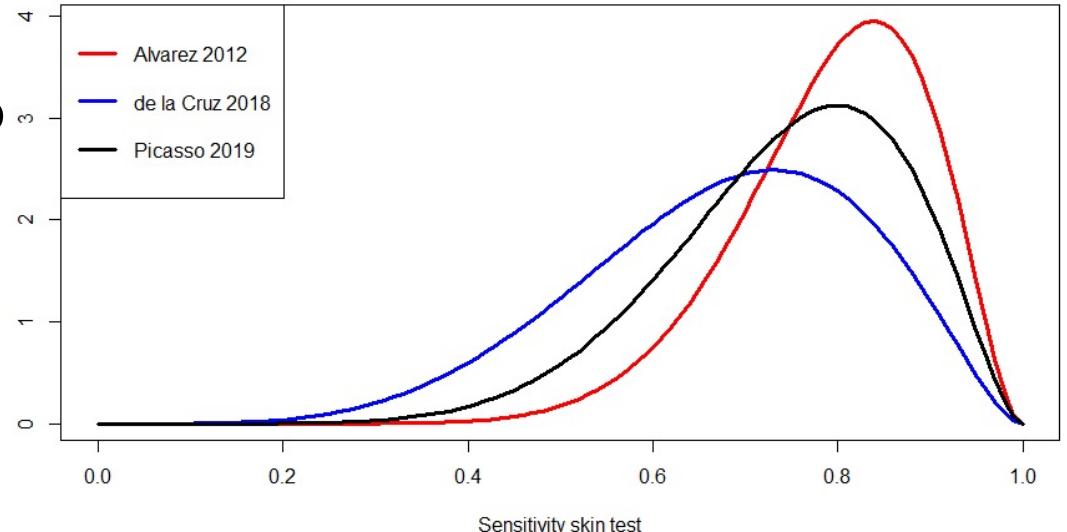


Expected test performance?

- Wide range of values reported in the literature
- (Widely) different methodologies
 - Single skin test:
 - Sensitivity=63.2-100% (83.9%) → metaanalysis (UK, 16 estimates) = 94% (49-100%)
 - Specificity=75.5-99% (96.8%) → metaanalysis (UK, 10 estimates) = 91% (70-100%)
 - IFN- γ :
 - Sensitivity=73-100% (87.6%) → metaanalysis (UK, 166 estimates) = 67% (49-82%)
 - Specificity=85-99.6% (96.6%) → metaanalysis (UK, 137 estimates) = 98% (96-99%)

How to incorporate this into priors?

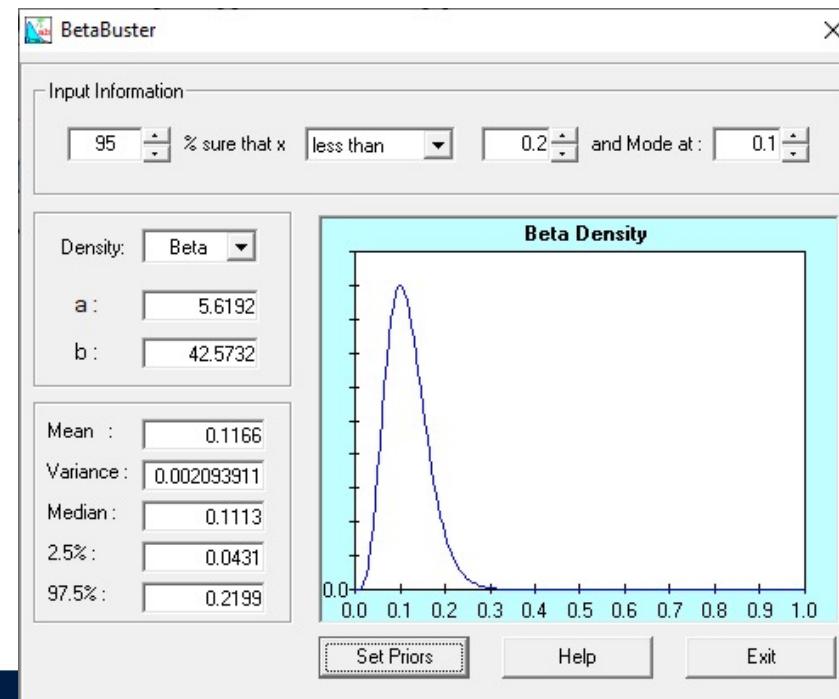
- Skin test:
 - Se
 - 84% (>60%) Alvarez 2012
 - 69% (>40%) de la Cruz 2018
 - 80% (>51%) Picasso-Riso 2019 (**CFT**)
 - Sp
 - 95% (>75%) Alvarez 2012, de la Cruz 2018
 - 90% (>60%) Picasso-Riso 2019 (**CFT**)
- IFN-g:
 - Se
 - 92% (>80%) Alvarez 2012
 - 90% (>50%) de la Cruz 2018
 - 88% (>73%) Al-Mouqatea 2018
 - 83.5% (>48%) Picasso-Riso 2019
 - Sp
 - 90% (>80%) Alvarez 2012, de la Cruz 2018
 - 97% (>85%) Al-Mouqatea 2018
 - 95% (>80%) Picasso-Riso 2019



*Their influence should be always assessed!

Expected prevalence?

- Expert opinion (always challenging!)
 - What is the expected prevalence of bovine tuberculosis in infected herds?
 - <https://shiny.vet.unimelb.edu.au/epi/beta.buster/>
 - Betabuster program



Alvarez et al., 2012

The likelihood (data)

Table 2

Number of reactors to single intradermal tuberculin (SIT) test and interferon-gamma (IFN- γ) assay performed on 6202 cattle in Castilla and Leon (Spain) for each of the combinations of diagnostic tests and interpretation criteria.

SIT test	IFN- γ assay	SIT+/IFN+	SIT+/IFN-	SIT-/IFN+	SIT-/IFN-
Severe	0.05	113	30	894	5165
	0.1	93	50	625	5434
Standard	0.05	87	20	920	5175
	0.1	75	32	643	5452

T1+/T2+

T1+/T2- <<<

T1-/T2+

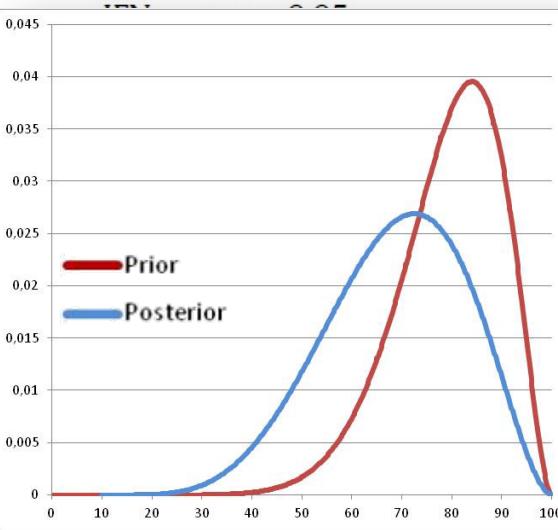
T1-/T2-

The results (posterior): Alvarez et al. 2012

Table 1

Prior (mode and low 95% credibility interval (CI) bound) and posterior estimates (median and 95% CI) for sensitivity, specificity and prevalence of disease (%) obtained for each of the combinations of diagnostic interpretation criteria on 6202 cattle in Leon, Spain.

Diagnostic interpretation/threshold	Skin test priors		IGRA priors		Prevalence	
	SIT test estimates		IFN assay estimates			
	Sensitivity	Specificity	Sensitivity	Specificity		
Prior estimates						
Severe interpretation	83.9 (95% conf. >60)	95 (95% conf. >75)	92 (95% conf. >80)	90 (95% conf. >80)	10 (95% conf. <20)	
Standard interpretation	73.9 (95% conf. >50)	98 (95% conf. >82)	85 (95% conf. >75)	98 (95% conf. >90)		
Posterior estimates						
SIT test: severe	69.4 (40.1–92.2)	99.4 (98.7–99.9)	89.3 (77.5–97.2)	85.7 (84.4–87.6)	2.59 (1.48–4.84)	
	66.1 (35.3–91.3)	99.3 (98.6–99.8)	83.1 (71.9–91.4)	90.3 (89.1–92.5)	2.52 (1.38–5.17)	
	56.6 (29.2–83.2)	99.7 (99.1–100)	90 (78.9–96.7)	85.7 (84.3–88)	2.58 (1.48–5.04)	
	53 (27.3–81.5)	99.6 (99–100)	83.5 (73.6–91.6)	90.4 (89.1–92.7)	2.71 (1.38–5.48)	



Skin test posterior

IGRA posterior

Influence of priors!

- Se of skin test → lower Se
- Prevalence → higher prev

Does the methodology used influence the estimates?

- Frequentist estimates

Test	N [†]	APM and 95% APPI Se Estimates (%) [‡]		APM and 95% APPI Sp Estimates (%) [§]		Species
		Se Estimates [‡]	Sp Estimates [‡]	Se Estimates [‡]	Sp Estimates [‡]	
Skin test (PPD-based)	13	49	66.3 (52.5-74.6)	49	99.1 (98.6-99.5)	
Cervical	10	43	66.1 (49.3-72.0)	43	99.2 (98.7-99.5)	Cattle
SIT standard interpretation	6	12	63.9 (46.6-84.0)	12	96.9 (96.3-97.4)	Cattle
SIT severe interpretation	6	10	69.8 (46.0-91.8)	10	99.4 (98.8-99.9)	Cattle
IFN-γ blood test	12	44	78.1 (70.5-89.7)	44	89.3 (86.9-91.7)	
Bovigam (OD 0.05)	5	10	88.7 (80.1-96.2)	10	88.1 (85.6-91.1)	Cattle
Bovigam (OD 0.1)	7	23	75.1 (69.1-81.3)	23	88.2 (86.6-89.6)	Cattle

Table 4
Meta-analysis results for sensitivity of diagnostic tests for bovine tuberculosis in cattle.

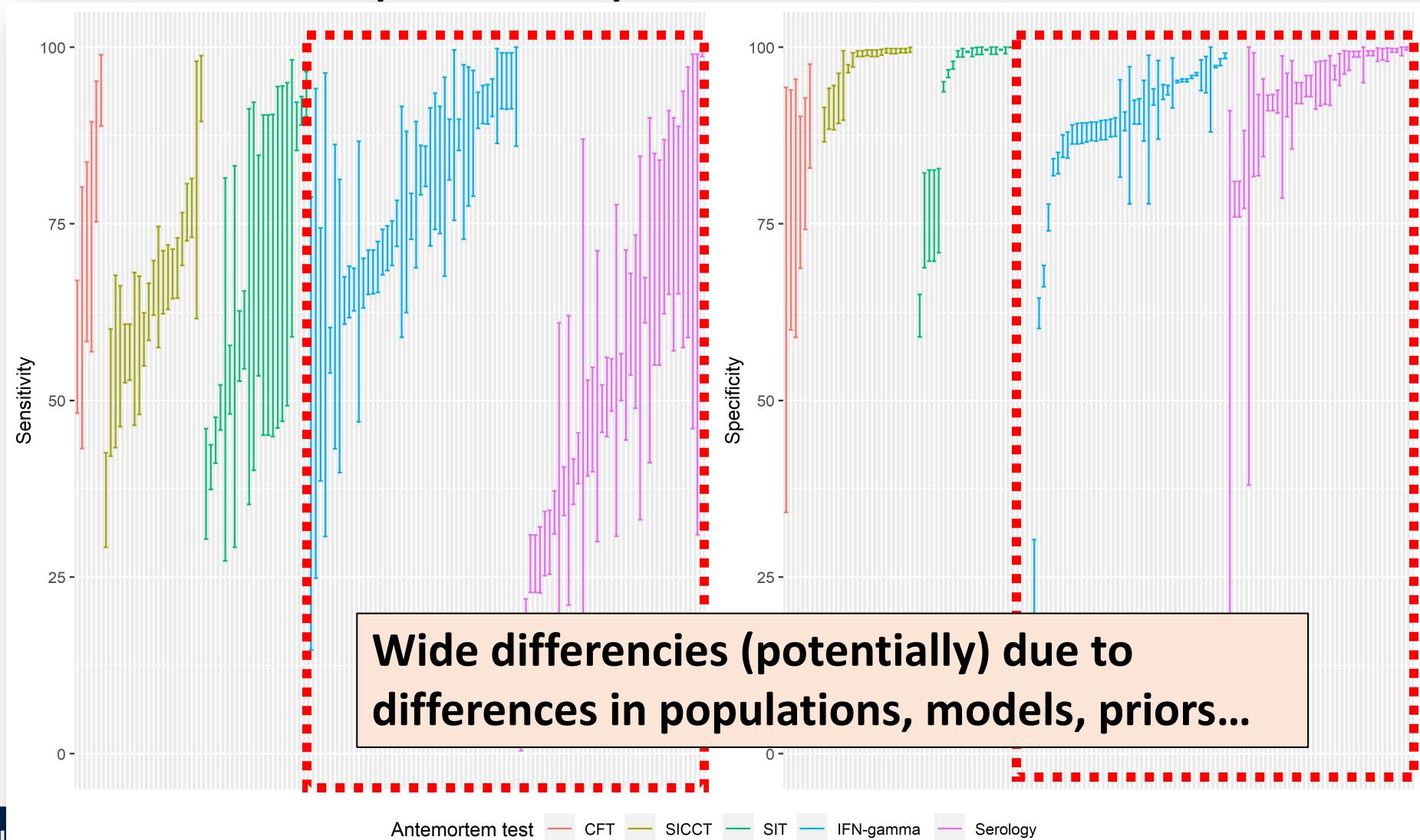
Test Name	References	Sensitivity Estimates	Pooled Sensitivity	Models A & B	Model A			Model B		
					95%			95%		
					n	N	Estimate	Type	Median	Credible Interval
<i>Skin tests with PPD</i>										
Single intradermal skin test	7	16	0.92	REM	0.94	0.49, 1.0	88	0.81	0.53, 0.94	85
SICCT test standard interpretation	14	38	0.78	REM	0.50	0.26, 0.78	159	0.64	0.48, 0.78	156
SICCT test severe interpretation	14	38	0.84	REM	0.63	0.40, 0.84	183	0.75	0.61, 0.86	183
Caudal fold	15	69	0.92	REM	0.76	0.56, 0.89	405	0.96	0.88, 0.98	403
<i>IFN-γ blood tests</i>										
IFN-γ Bovine	27	166	0.84	REM	0.87	0.72, 0.95	796	0.87	0.76, 0.94	790
IFN-γ Bovine PPD-Avian PPD	27	166	0.83	REM	0.67	0.49, 0.82	845	0.70	0.55, 0.92	822

Table 5
Meta-analysis results for specificity of diagnostic tests for bovine tuberculosis in cattle.

Test Name	References	Specificity Estimates	Pooled Specificity	Models A & B	Model A			Model B		
					95%			95%		
					n	N	Estimate	Type	Median	Credible Interval
<i>Skin tests</i>										
Single Intradermal skin test	4	10	0.89	REM	0.91	0.70, 1.00	29	0.91	0.63, 1.00	30
SICCT test standard	7	13	1.00	REM	1.00	0.99, 1.00	^a	1.00	0.99, 1.00	^a
Caudal fold	2	3	0.99	FEM	1.00	0.92, 1.00	^a	^b	^b	^b
<i>IFN-γ blood tests</i>										
IFN-γ Bovine PPD	19	137	0.91	REM	0.97	0.94, 0.98	647	0.94	0.86, 0.98	649
IFN-γ Bovine PPD-Avian PPD	19	137	0.96	REM	0.98	0.96, 0.99	645	0.94	0.88, 0.97	621

- Estimates from BLCMs

Also variability in Bayesian estimates!



In any case... BLCMs are a reliable tool

- BLCMs widely applied for the evaluation of (veterinary) diagnostic tests: endorsed by the WOAH

2019 © OIE - *Manual of Diagnostic Tests for Aquatic Animals* - 14/11/2019

2.2.2. Samples from animals of unknown status

When the so-called reference standard is imperfect, which is the rule with any diagnostic tests, estimates of DSe and DS_p for the candidate assay based on this standard will be flawed. A way to overcome this problem is to perform a latent class analysis of the joint results of the two tests assuming neither test is perfect.

Latent-class models do not rely on the assumption of a perfect reference test but rather estimate the accuracy of the candidate test and the reference standard with the joint test results (Branscum et al., 2005; Enøe et al., 2000; Georgiadis et al., 2003; Hui & Walter, 1980). If a Bayesian latent class analysis is used, prior knowledge about the performance of the reference test and the candidate test can be incorporated into the analysis.

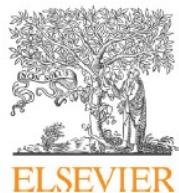
Because these statistical models are complex and require critical assumptions, statistical assistance should be sought to help guide the analysis and describe the sampling from the target population(s), the characteristics of other tests included in the analysis, the appropriate choice of model and the estimation methods based on peer-reviewed literature (see *Terrestrial Manual Chapter 3.6.5* [footnote ¹⁴] for details).



Nowadays

- Guidelines available for adequate reporting of the use of BLCMs in the context of diagnostic test evaluation

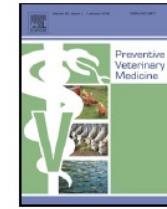
Preventive Veterinary Medicine 138 (2017) 37–47



Contents lists available at ScienceDirect

Preventive Veterinary Medicine

journal homepage: www.elsevier.com/locate/prevetmed



STARD-BLCM: Standards for the Reporting of Diagnostic accuracy studies that use Bayesian Latent Class Models



Polychronis Kostoulas ^{a,*}, Søren S. Nielsen ^b, Adam J. Branscum ^c, Wesley O. Johnson ^d, Nandini Dendukuri ^e, Navneet K. Dhand ^f, Nils Toft ^g, Ian A. Gardner ^h

^a Laboratory of Epidemiology, Biostatistics and Animal Health Economics, Faculty of Veterinary Medicine, University of Thessaly, Karditsa GR43100, Greece

^b Department of Veterinary and Animal Sciences, Faculty of Health and Medical Sciences, University of Copenhagen, Grønnegårdsvej 8, DK-1870 Frederiksberg C, Denmark

^c Biostatistics Program, Oregon State University, Corvallis, OR, 97331, USA

^d Department of Statistics, University of California, Irvine, CA, 92697, USA

^e McGill University Health Centre, McGill University, Montréal, QC, Canada

^f Faculty of Veterinary Science, The University of Sydney, 425 Werombi Road, Camden, 2570 NSW, Australia

^g Technical University of Denmark, National Veterinary Institute, Bülowsvej 27, DK-1870 Frederiksberg C, Denmark

^h Department of Health Management, Atlantic Veterinary College, University of Prince Edward Island, Charlottetown, Prince Edward Island C1A4P3, Canada



VISAVET HEALTH SURVEILLANCE CENTRE

UNIVERSIDAD COMPLUTENSE MADRID



www.visavet.es

Nowadays

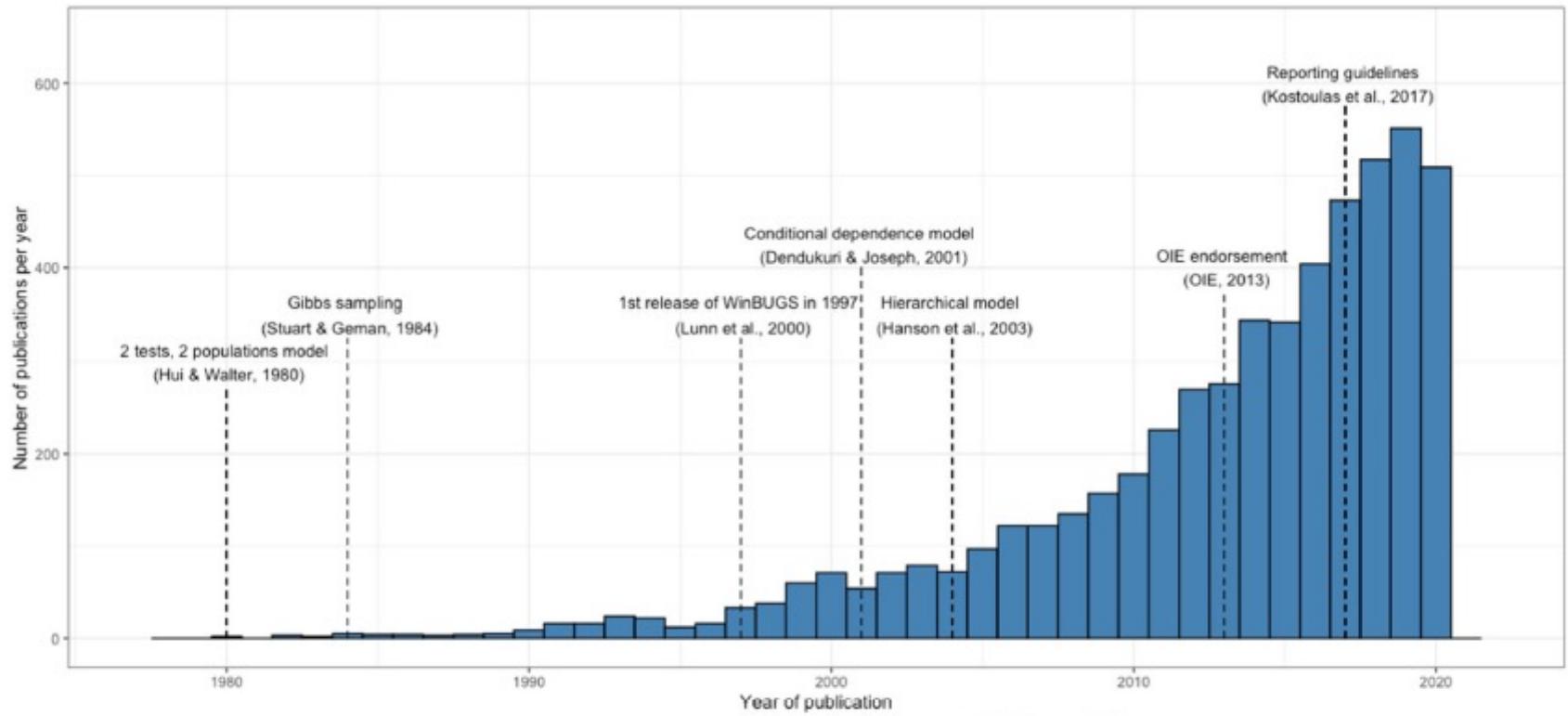


Fig. 2

Frequency histogram of the number of peer-reviewed articles published on latent class analysis when there is an imperfect reference test

References

- Hui, S. L., & Walter, S. D. (1980). Estimating the error rates of diagnostic tests. *Biometrics*, 167-171.
- Vacek, P. M. (1985). The effect of conditional dependence on the evaluation of diagnostic tests. *Biometrics*, 959-968. The effect of conditional dependence on the evaluation of diagnostic tests
- Joseph, L., Gyorkos, T. W., & Coupal, L. (1995). Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard. *American journal of epidemiology*, 141(3), 263-272.
- Spiegelhalter, D. J., Thomas, A., Best, N. G., & Lunn, D. J. (2002). *WinBUGS: Bayesian Inference Using Gibbs Sampling Manual*, Version 1.4. London: Imperial College; Cambridge, UK: MRC Biostatistics Unit.
- Enøe, C., Georgiadis, M. P., & Johnson, W. O. (2000). Estimation of sensitivity and specificity of diagnostic tests and disease prevalence when the true disease state is unknown. *Preventive veterinary medicine*, 45(1-2), 61-81.
- Plummer, M. (2003, March). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In *Proceedings of the 3rd international workshop on distributed statistical computing* (Vol. 124, No. 125.10, pp. 1-10).
- Toft, N., Innocent, G. T., Gettinby, G., & Reid, S. W. (2007). Assessing the convergence of Markov Chain Monte Carlo methods: an example from evaluation of diagnostic tests in absence of a gold standard. *Preventive veterinary medicine*, 79(2-4), 244-256.
- Kostoulas, P., Nielsen, S. S., Branscum, A. J., Johnson, W. O., Dendukuri, N., Dhand, N. K., . . . & Gardner, I. A. (2017). STARD-BLCM: Standards for the Reporting of Diagnostic accuracy studies that use Bayesian Latent Class Models. *Preventive veterinary medicine*, 138, 37-47.