DEFINING SENSITIVITY AND SPECIFICITY OF AN IMPERFECT GOLD STANDARD

Dr Direk Limmathurotsakul, MD MSc PhD

Faculty of Tropical Medicine
MAHIDOL UNIVERSITY Wisdom of the Land

40 years of TropMed

How to evaluate accuracy of a new diagnostic test

- Se is a probability that a test is positive in diseased pt
- Sp is a probability that a test is negative in non-diseased pt
- Compare the new diagnostic test with a gold standard

New test	Gold standard	
	Positive	Negative
Positive	a	c
Negative	b	d

$$Se = \frac{a}{a+b}$$

$$Sp = \frac{d}{c+d}$$

- We assume that Se and Sp of the gold standard are 100%

What would happen if the gold standard is imperfect ?

- If the gold standard has true Se of 60% and true Sp of 100%
- Hypothetically tested in **1,000 infected** and **1,000 non-infected** pts

Gold standard	
Positive	Negative
600	400 + 1,000

Se = 95% (570/600)

Sp = 69% (970/1,400)

Prevalence = 30% (600/2,000)

SEVERELY BIASED!!

How to evaluate test accuracy when the gold standard is imperfect

- Assume that Se and/or Sp of gold standard may not be 100%
- Because the gold standard is imperfect, we consider **the true prevalence** instead
- Need at least 3 tests in a single population

How to evaluate test accuracy when the gold standard is imperfect

A probability that

How to evaluate test accuracy when the gold standard is imperfect

Profile	Number
111	.
110	.
101	.
011	.
100	.
010	.
001	.
000	.
TOTAL	N

- We observed **8** total numbers of patients having each profile
- We can estimate **7** unknown parameters

Prevalence = xx%

Test 1 Se = xx% Sp = xx%

Test 2 Se = xx% Sp = xx%

Test 3 Se = xx% Sp = xx%

How to evaluate test accuracy when the gold standard is imperfect

Profile	Number
111	8
110	1
101	1
011	1
100	4
010	4
001	4
000	77
TOTAL	**100**

- We observed **8** total numbers of patients having each profile
- We can estimate **7** unknown parameters

Prevalence = xx%

Test 1 Se = xx% Sp = xx%

Test 2 Se = xx% Sp = xx%

Test 3 Se = xx% Sp = xx%

How to evaluate test accuracy when the gold standard is imperfect

Profile	Number
111	8
110	1
101	1
011	1
100	4
010	4
001	4
000	77
TOTAL	**100**

- We observed **8** total numbers of patients having each profile
- We can estimate **7** unknown parameters

Prevalence = 10%

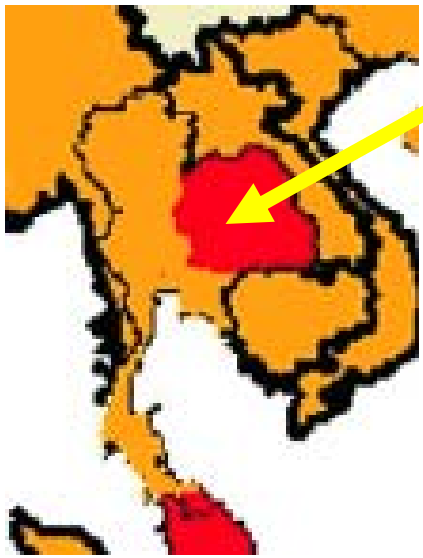
Test 1 Se = 95% Sp = 95%

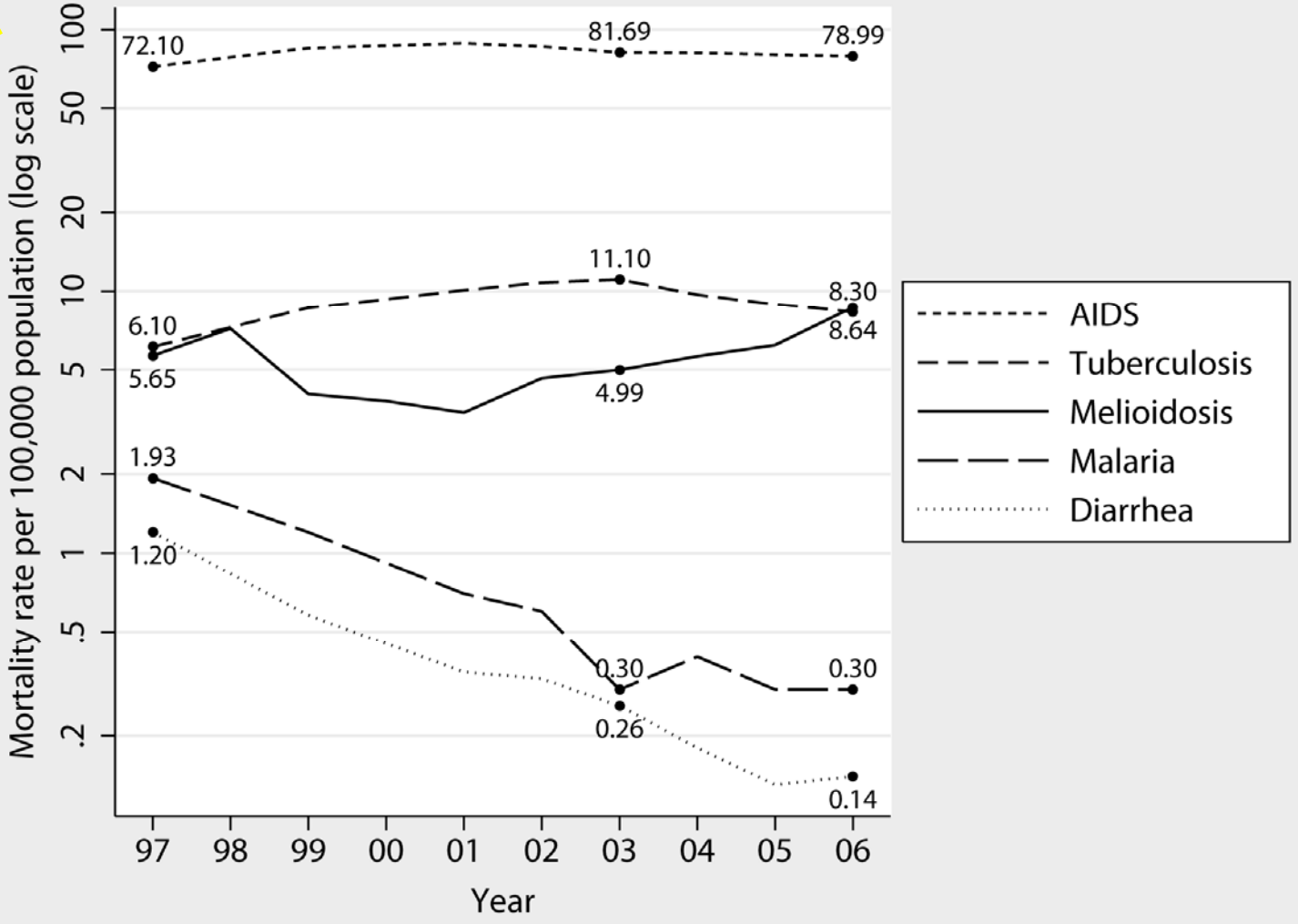
Test 2 Se = 95% Sp = 95%

Test 3 Se = 95% Sp = 95%

An example from melioidosis

An example from melioidosis: Introduction





An example from melioidosis: Introduction

- Melioidosis is a life-threatening infection caused by Gram negative bacilli, ***B. pseudomallei***
- Current gold standard is culture
- **Sp of culture is 100%** because *B. pseudomallei* does not colonize in healthy individuals
- Se of culture seems to be low as clinicians commonly make a **clinical diagnosis of melioidosis in culture-negative patients** based on all clinical data

An example from melioidosis: Introduction

- A number of **serological tests have low specificity** after evaluation by comparing with the culture
- This could be due to

  (1) **high antibody level in the healthy individuals** after exposure to *B. pseudomallei* in the environment

  (2) **misclassification of culture**

An example from melioidosis: Methods

- Data from published studies of **six diagnostic tests** in 320 patients suspected melioidosis were re-analyzed

- Six tests were performed **on admission**

  (1) culture

  (2) clinical criteria

  (3) IHA (indirect hemagglutination test)

  (4) IgM ICT (immunochromogenic cassette test)

  (5) IgG ICT

  (6) ELISA

An example from melioidosis: Methods

- **Bayesian latent class model (LCM) with conditional dependence between diagnostic tests** was used
- Result of the final model was **compared with conventional method** (culture as a perfect gold standard)
- Accuracy of the model was evaluated by **post-hoc model validation;** all clinical data after admission (USS, treatment and progression) was used to categorize each patient into 4 categories (definite, probable, possible and unlikely)

An example from melioidosis: Data

Response profile	Observed frequency
111111	54
111110	3
111101	0
111100	0
111011	8
111010	0
111001	0
…	…
000000	23

An example from melioidosis: Results

Parameters	Culture as a gold standard	Final Bayesian LCM
Prevalence	**37 %**	**71 %**
Culture		
Se	**100 %**	**52 %**
Sp	**100 %**	**100 %**

An example from melioidosis: Results

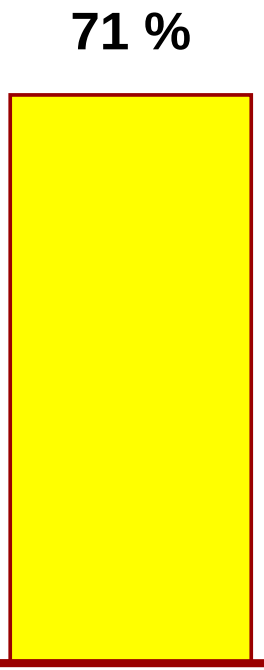
Parameters	Culture as a gold standard	Final Bayesian LCM
ELISA		
Se	**82 %**	**66 %**
Sp	**73 %**	**98 %**

An example from melioidosis: Post-hoc model evaluation

Category	Definition	N (%)
Unlikely	Firm alternative diagnosis *or* Recover without effective antimicrobials	84 (26%)
Possible	Improved after effective antimicrobials *or* Died before improvement observed	83 (26%)
Probable	Specific USS finding *or* Representation with culture +ve in 1 mo	34 (11%)
Definite	Culture +ve	119 (37%)

An example from melioidosis: Post-hoc model evaluation

Category	Definition	N (%)
Unlikely	Firm alternative diagnosis *or* Recover without effective antimicrobials	84 (26%)
Possible	Improved after effective antimicrobials *or* Died before improvement observed	83 (26%)
Probable	Specific USS finding *or* Representation with culture +ve in 1 mo	34 (11%)
Definite	Culture +ve	119 (37%)



Model predicted prevalence

An example from melioidosis: Summary

- Sensitivity of culture is very low
- Previous findings by using culture as a perfect gold standard is inaccurate
- If the Se and Sp of ELISA had been properly estimated, ELISA should have been used in the real clinical setting
- ***** A model for evaluating diagnostic tests with an imperfect gold standard should be used when the accuracy of gold standard is imperfect or unknown *****

Acknowledgement

- Lisa White
- Sue Lee
- Vanaporn Wuthiekanun
- Narisara Chantratita
- Prof Nick Day
- Prof Sharon Peacock
- Kris Jamsen
- Julie Simpson
- Allen Cheng
- Claudio Verzilli

THE END

How to evaluate test accuracy when the gold standard is imperfect

Profile	Probability of obtaining each profile
111	Prev * Se1*Se2*Se3 + (1-Prev) * (1-Sp1)*(1-Sp2)*(1-Sp3)
110	Prev * Se1*Se2*(1-Se3) + (1-Prev) * (1-Sp1)*(1-Sp2)*Sp3
101	Prev * Se1*(1-Se2)*Se3 + (1-Prev) * (1-Sp1)*Sp2*(1-Sp3)
011	Prev * (1-Se1)*Se2*Se3 + (1-Prev) * Sp1*(1-Sp2)*(1-Sp3)
100	Prev * Se1*(1-Se2)*(1-Se3) + (1-Prev) * (1-Sp1)*Sp2*Sp3
010	Prev * (1-Se1)*Se2*(1-Se3) + (1-Prev) * Sp1*(1-Sp2)*Sp3
001	Prev * (1-Se1)*(1-Se2)*Se3 + (1-Prev) * Sp1*Sp2*(1-Sp3)
000	Prev * (1-Se1)*(1-Se2)*(1-Se3) + (1-Prev) * Sp1*Sp2*Sp3

How to evaluate test accuracy when the gold standard is imperfect

Profile	Number
111	8
110	1
101	1
011	1
100	4
010	4
001	4
000	77
TOTAL	**100**

- We observed **8** total numbers of patients having each profile
- We can estimate **7** unknown parameters

Prevalence = 10%

Test 1 Se = 95% Sp = 95%

Test 2 Se = 95% Sp = 95%

Test 3 Se = 95% Sp = 95%

How to evaluate test accuracy when the gold standard is imperfect

Profile	Number
11	.
10	.
01	.
00	.

- We observed **4** total numbers of patients having each profile
- We need to estimate **5** unknown parameters

Prevalence = xx%

Test 1 Se = xx% Sp = xx%

Test 2 Se = xx% Sp = xx%

IMPOSSIBLE TO ESTIMATE !!

An example from melioidosis: Methods

Analysis plan	Assumption that Se of culture is 100%	Assumption that tests are independent
(1) conventional method by using culture as the gold standard	Yes	Yes
(2) Bayesian LCM with conditional independence between diagnostic tests (Model 0)	No	Yes
(3) Bayesian LCM with conditional dependence between diagnostic tests (Final model)	No	No