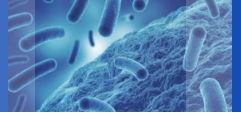
Benefit-risk Evaluation for Diagnostics: A Framework (BED-FRAME)

Scott Evans, PhD, MS
Director, The Biostatistics Center
Founding Chair and Professor, Department of Biostatistics and Bioinformatics
Milken Institute School of Public Health
George Washington University

Liver Forum
Paris, France
2023

Outline

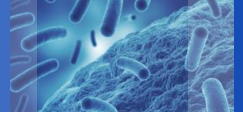
- **Why we need something more to assess diagnostics**
  - **Pragmatism and benefit:risk apply here too**
- **BED-FRAME to address this need**
- **Example**



Suppose there is a choice between 2 diagnostic tests:

one has higher sensitivity and the other has higher specificity.

Which test should be selected to optimize clinical outcomes?



I can say 2 things about the methods I learned in college for the analysis of diagnostic tests.

1. They closely reflect the methods used in the literature.

2. They are useless for informing clinical decision-making.

Sensitivity and Specificity

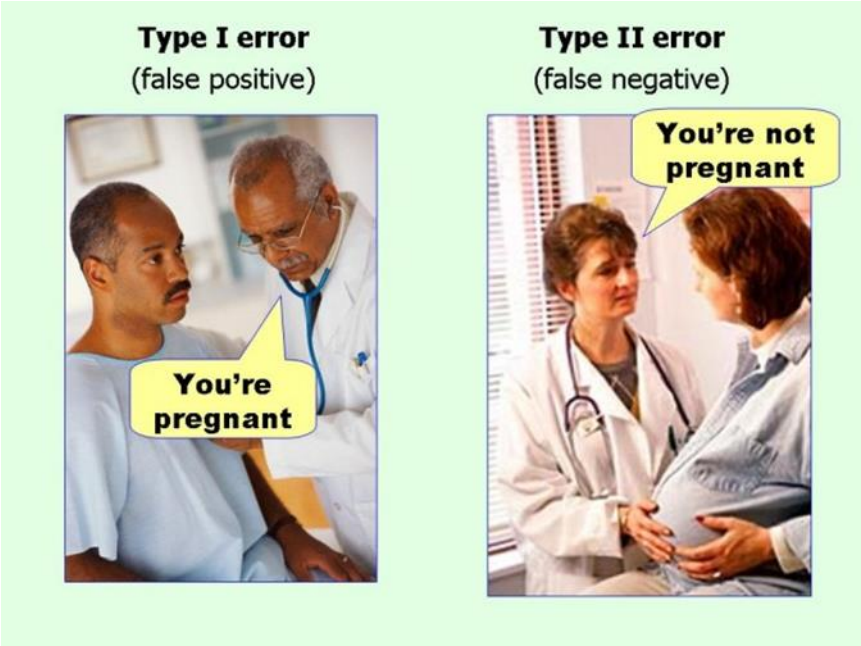
- Sensitivity and specificity are foundations for diagnostic evaluation
- Arbitrary goals are typically defined
  - E.g., sensitivity and specificity must be above 90%
- However…

Important Consideration: Integrating Sensitivity and Specificity

- If sensitivity is very high, then one may be willing to sacrifice specificity to a degree
- If sensitivity was marginal, then … less willing to sacrifice
- Consider
  - A test with a sensitivity and specificity of 99% and 89%, respectively, would not satisfy the arbitrary 90% goals
  - A test with a sensitivity and specificity of 91% would satisfy the goal
  - However, if a false negative is much more important than a false positive, then the first test may have more utility than the second
- Need to integrate sensitivity and specificity while considering the relative importance of false negative and false positive results

Important Consideration: Relative Importance of Errors

Not all errors are equivalent.



Non-invasive Tests for NASH

- Likely a panel of markers e.g., radiology and serum
- Consequence of errors depend on the stage of disease.
- Earlier stages: no treatment and unlikely for a while
  - False positive: patient advised to adapt lifestyle and is monitored for disease progression
  - False negative: patient may progress to advanced fibrosis without monitoring
- Advanced fibrosis
  - False positive: patient may undergo unnecessary diagnostics and maybe put on contra-indicated treatment putting them at risk for harmful toxicity and cost.
  - False negative: patient could progress to compensated cirrhosis and the decompensated cirrhosis without being monitored and treated

Important Consideration: Prevalence

- Consider application of a test with sensitivity = 90% and specificity = 80%, to 1000 patients when the prevalence is 20%

- Expected diagnostic yield
  - 20 false positives
  - 160 false negatives

- If the prevalence was instead 2%, then the expected yield is:
  - 2 false positives
  - 196 false negatives

- Prevalence must be carefully incorporated

Prevalence and Relative Importance are Dynamic

- Prevalence and relative importance are dynamic, varying temporally and geographically
- E.g., changes in COVID infection prevalence over time or geographic diversity of prevalence
- The relative importance of false negative vs. false positive errors can vary depending on e.g.,
  - The availability and costs of effective interventions
  - Ability to manage toxicities associated with the interventions
  - Disease virulence
  - Contagiousness

Towards Improved Clinical Decision-making

- The “real world” consequences of diagnostics application can be evaluated using the expected diagnostic yield
- Diagnostic yield: the distribution of true positives, true negatives, false positives and false negatives
  - Determined by sensitivity, specificity, and prevalence
- The desirability of the resulting diagnostic yield is affected by the relative importance of false negative vs. false positive errors
- Can we make diagnostic studies more pragmatic by comprehensively considering this information?
- Yes …

Keys to Improved Decision-making

- Evaluations with greater pragmatism based on diagnostic yield
  - Consider sensitivity and specificity simultaneously
  - Incorporate prevalence
  - Incorporate relative importance of different errors
  - Analyze overall utility in light of the dynamic nature of the prevalence and relative importance

Clinical Infectious Diseases

INVITED ARTICLE

HEALTHCARE EPIDEMIOLOGY: Robert A. Weinstein, Section Editor

Benefit-risk Evaluation for Diagnostics: A Framework (BED-FRAME)

Scott R. Evans,[1,2] Gene Pennello,[3] Norberto Pantoja-Galicia,[3] Hongyu Jiang,[2] Andrea M. Hujer,[4] Kristine M. Hujer,[4] Claudia Manca,[5] Carol Hill,[6] Michael R. Jacobs,[4] Liang Chen,[5] Robin Patel,[7] Barry N. Kreiswirth,[5] and Robert A. Bonomo[4]; for the Antibacterial Resistance Leadership Group

[1]Department of Biostatistics, and [2]Center for Biostatistics in AIDS Research, Harvard T. H. Chan School of Public Health, Boston, Massachusetts; [3]Division of Biostatistics, Office of Surveillance and Biometrics, Center for Devices and Radiological Health, US Food and Drug Administration, Silver Spring, Maryland; [4]Louis Stokes Cleveland Veterans Affairs Medical Center, Case Western Reserve University School of Medicine, Ohio; [5]Public Health Research Institute, New Jersey Medical School, Rutgers University, Newark; [6]Duke Clinical Research Institute, Duke University, Durham, North Carolina; and [7]Mayo Clinic, Rochester, Minnesota

The medical community needs systematic and pragmatic approaches for evaluating the benefit-risk trade-offs of diagnostics that assist in medical decision making. Benefit-Risk Evaluation of Diagnostics: A Framework (BED-FRAME) is a strategy for pragmatic evaluation of diagnostics designed to supplement traditional approaches. BED-FRAME evaluates diagnostic yield and addresses 2 key issues: (1) that diagnostic yield depends on prevalence, and (2) that different diagnostic errors carry different clinical consequences. As such, evaluating and comparing diagnostics depends on prevalence and the relative importance of potential errors. BED-FRAME provides a tool for communicating the expected clinical impact of diagnostic application and the expected trade-offs of diagnostic alternatives. BED-FRAME is a useful fundamental supplement to the standard analysis of diagnostic studies that will aid in clinical decision making.

Keywords. benefit-risk; diagnostics; diagnostic yield; pragmatism.

JOURNAL OF BIOPHARMACEUTICAL STATISTICS
2016, VOL. 26, NO. 6, 1083–1097
http://dx.doi.org/10.1080/10543406.2016.1226335

Comparing diagnostic tests on benefit-risk

Gene Pennello[a], Norberto Pantoja-Galicia[a], and Scott Evans[b]

[a]Center for Devices and Radiological Health, Food and Drug Administration, Silver Spring, Maryland, USA; [b]Center for Biostatistics in AIDS Research and the Department of Biostatistics, Harvard T. H. Chan School of Public Health, Boston, Massachusetts, USA

Welcome to Nerd Nirvana

AAP News

May 31, 2016

I'll Be Sleeping Well with This BED-FRAME for Diagnostic Tests

Dr Bud Wiedermann, MD,MA, Evidence eMended Editor, Grand Rounds

As much as I poke fun at contrived acronyms, I confess to favor this one. I felt like I was in Nerd Nirvana after reading this early release article:

Evans SR, Pennello G, Pantoja-Galicia N, et al. Benefit-risk evaluation for diagnostics: a framework (BED-FRAME). Clin Infect Dis 2016; May 18. pii:ciw239; Epub ahead of print.

I struggled whether to use this article for my precious 5th Tuesday posting, where I've freed myself from the confines of AAP Grand Rounds to comment on any article I want. I finally decided that I loved this article too much, so I'm indulging myself.

The article will appeal only to true EBM nerds. I promise not to bore you with the mathematical minutiae, but I really think these authors' approach, or something similar to it, represent a leap forward in how we use diagnostic tests.

We all know that no diagnostic test is perfect, but beyond that fact lies the dilemma of how these inaccuracies impact clinical outcomes in different patient scenarios. BED-FRAME is an attempt at a graphical display to understand how to use test results, based on the tests' diagnostic performance, incorporating all those delightful terms like sensitivity, specificity, likelihood ratios, and disease prevalence.

Weighted Accuracy (WA) and Average Weighted Accuracy (AWA)

Clinical Infectious Diseases

SPECIAL SECTION/INVITED ARTICLE

IDSA Infectious Diseases Society of America · hivma hiv medicine association · OXFORD

Average Weighted Accuracy: Pragmatic Analysis for a Rapid Diagnostics in Categorizing Acute Lung Infections (RADICAL) Study

Ying Liu,[1] Ephraim L. Tsalik,[2,3] Yunyun Jiang,[4] Emily R. Ko,[2] Christopher W. Woods,[2,5] Ricardo Henao,[2] and Scott R. Evans[4]

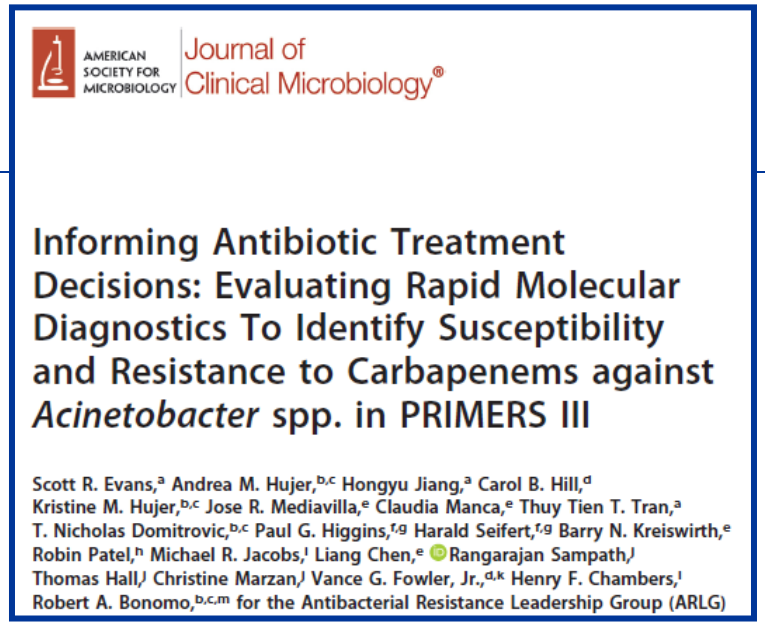
[1]Biogen, Inc., Boston, MA, USA; [2]Center for Applied Genomics and Precision Medicine, Department of Medicine, Duke University, Durham, NC, USA; [3]Emergency Department Service, Durham VA Health Care System, Durham, NC, USA; [4]Biostatistics Center, George Washington Milken Institute School of Public Health, Rockville, MD, USA; and [5]Medicine Service, Durham VA Health Care System, Durham, NC, USA

$$c_1 = \frac{1}{1-r} - \frac{r}{(b-a)(1-r)^2} ln\left(\frac{r+b(1-r)}{r+a(1-r)}\right),$$

$$c_2 = \frac{r}{(1-r)(b-a)} ln\left(\frac{r+b(1-r)}{r+a(1-r)}\right) - \frac{r}{1-r} + \frac{r^2}{(1-r)^2(b-a)} ln\left(\frac{r+b(1-r)}{r+a(1-r)}\right),$$

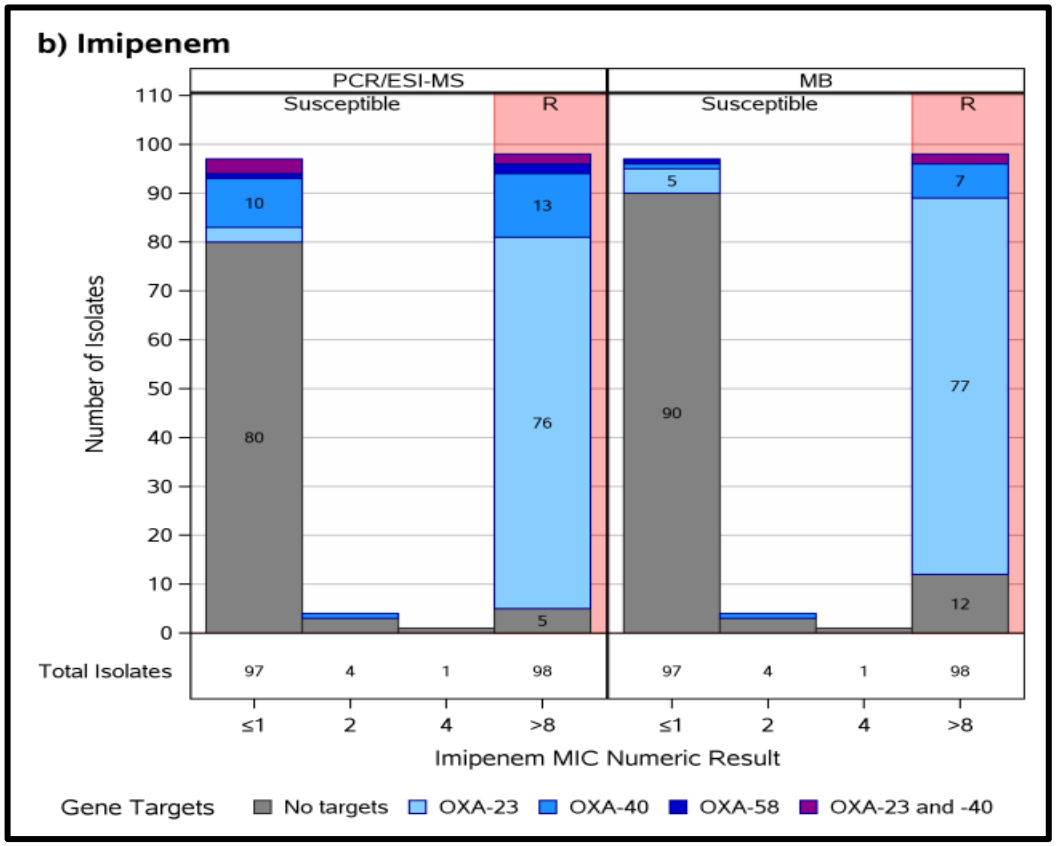
- $SE\left(\widehat{AWA}\right) = \sqrt{\frac{{c_1}^2 \widehat{PPA}(1-\widehat{PPA})}{n_1} + \frac{{c_2}^2 \widehat{NPA}(1-\widehat{NPA})}{n_2}}.$

Example



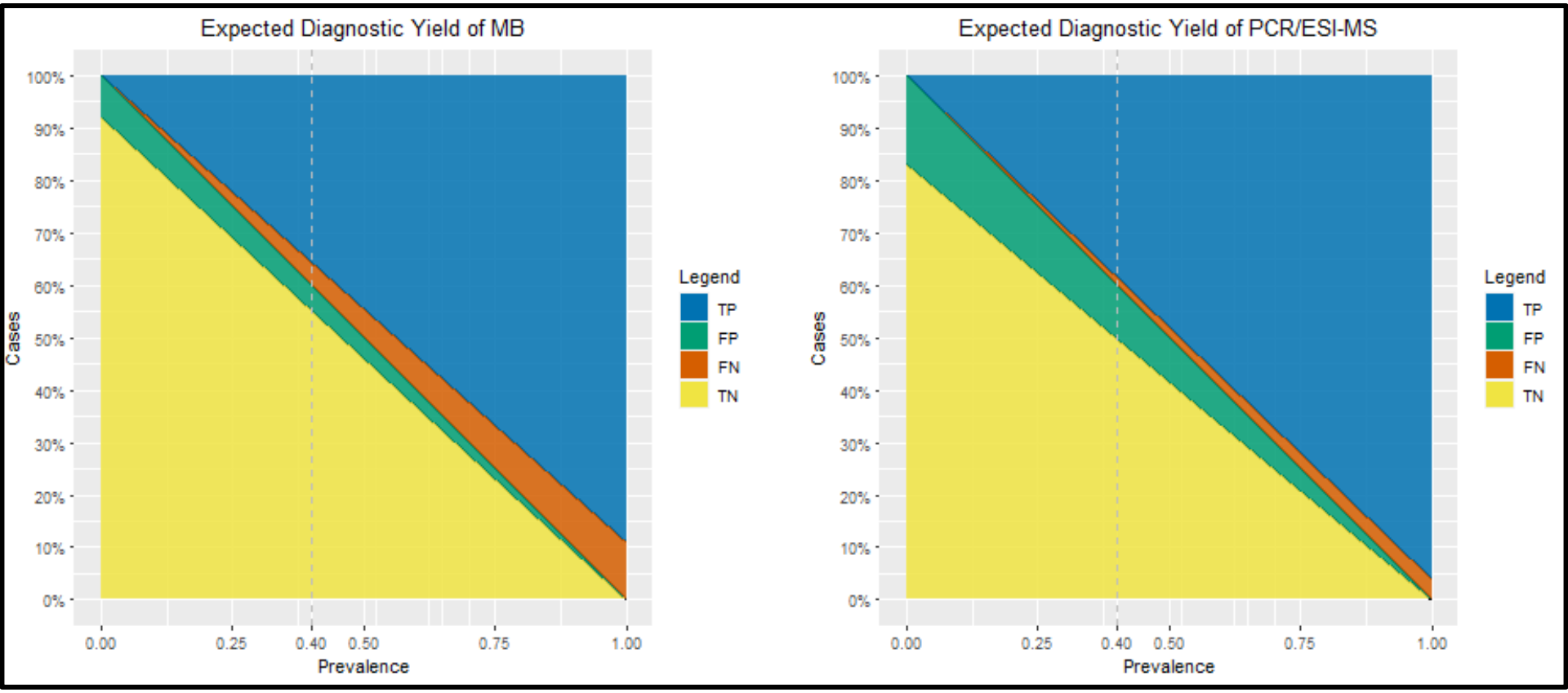
- Evaluation of two platforms (PCR/ESI-MS and MB) for discriminating resistance vs. susceptibility to carbapenems
  - Based on the absence / presence of 7 genetic targets
    - OXA-23, -40, -58, NDM, KPC, VIM, and IMP
- Reference standard: Minimal inhibitory concentration (MIC)
  - Smallest antibiotic concentration sufficient to inhibit bacterial growth when tested in vitro
- 200 strains (~50% susceptible to carbapenems)

Genetic Target Identification by MIC Level

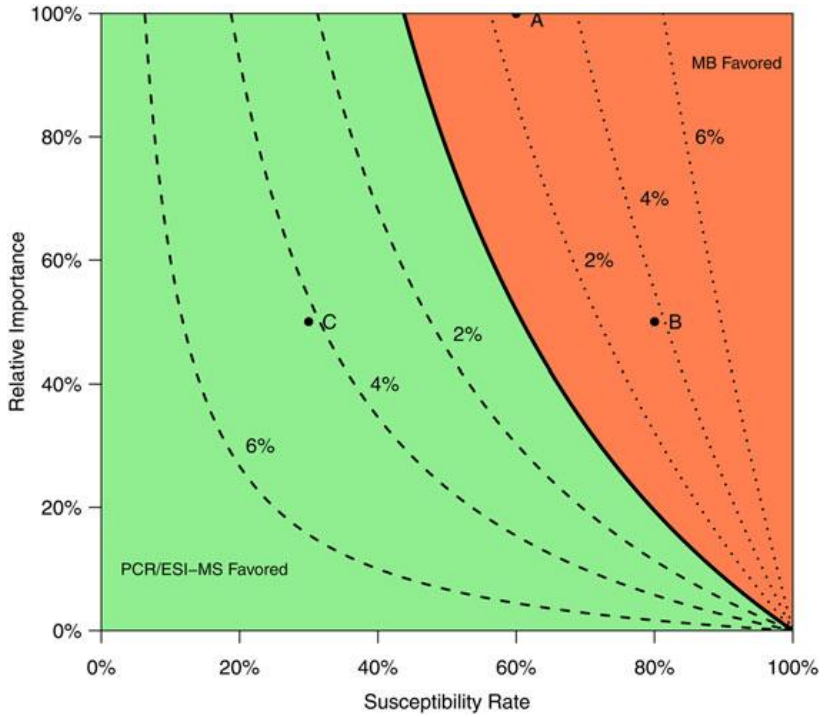


Slide Rule Profile Plot:

Expected Diagnostic Yield as a Function of the Prevalence of Susceptibility



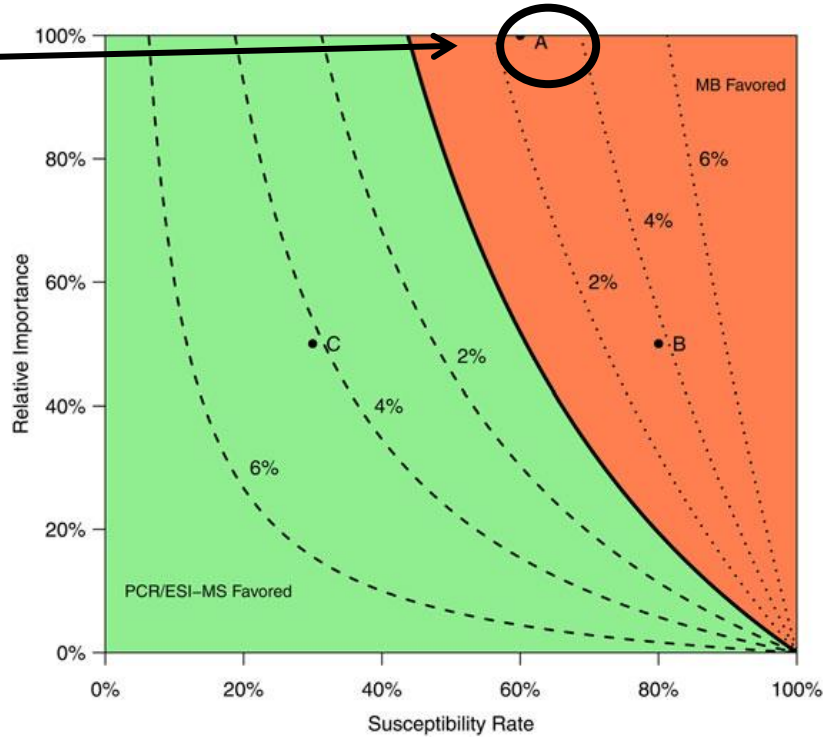
Contours of the Difference in Weighted Accuracy as a Function of Relative Importance and Susceptibility Rate



Red favors MB; Green favors PCR/ESI-MS

Contours of the Difference in Weighted Accuracy as a Function of Relative Importance and Susceptibility Rate

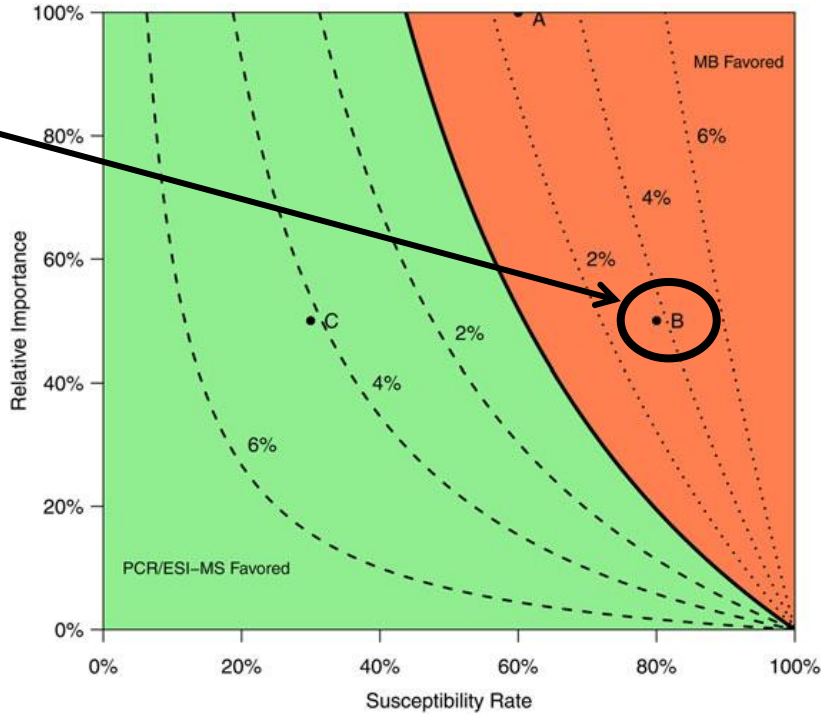
- Consider Point A
  - R =100% [i.e., errors are equivalent)
  - Susceptibility rate=60%
  - ~ 3% higher weighted accuracy for MB



Red favors MB; Green favors PCR/ESI-MS

Contours of the Difference in Weighted Accuracy as a Function of Relative Importance and Susceptibility Rate

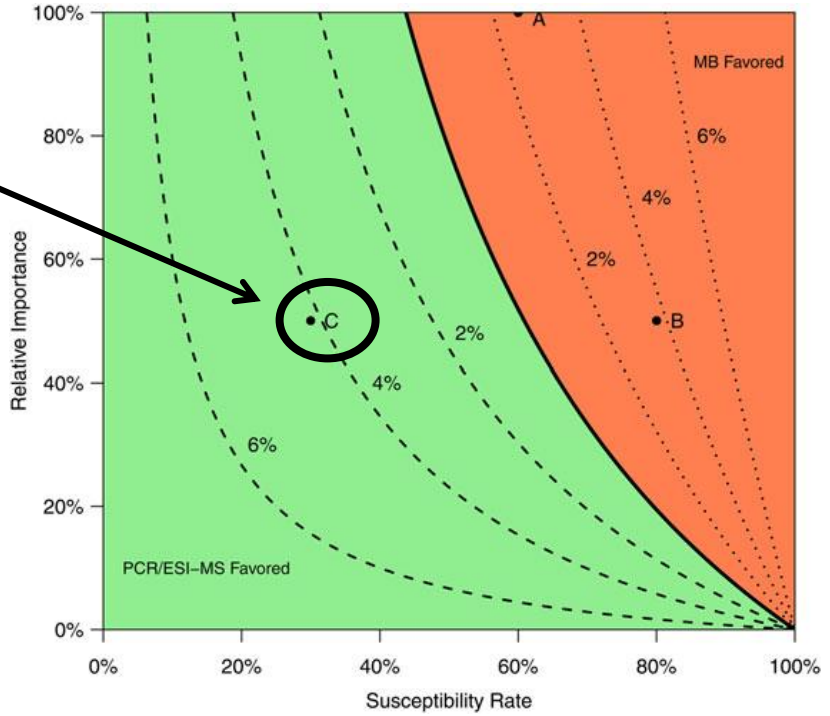
- Consider Point B
  - R=50% (i.e., FR is half as important as FS)
  - Susceptibility rate=80%
  - ~ 4% higher weighted accuracy for MB.



Red favors MB; Green favors PCR/ESI-MS

Contours of the Difference in Weighted Accuracy as a Function of Relative Importance and Susceptibility Rate

- Consider Point C
  - R=50%
  - Susceptibility rate=30% (i.e., resistance outbreak)
  - ~ 4% higher weighted accuracy for PCR/ESI-MS vs MB.



Red favors MB; Green favors PCR/ESI-MS

BED-FRAME Analyses

- Forest plot summary of point and confidence interval estimates for sensitivity, specificity, and the between diagnostic differences in sensitivity and specificity
- Plots of the estimated positive predictive values as a function of prevalence with confidence bands
- Plots of the differences in estimated predictive values as a function of prevalence with confidence bands
- Slide rule profile plot of the expected diagnostic yield as a function of the prevalence
- Estimated between diagnostic difference in FN and TN as a function of prevalence
- The estimated numbers needed to test (NNT)
- A plot of WA as a function of relative importance
- A contour plot of the between diagnostic difference in WA under combinations of prevalence and relative importance
- Plots of AWA and the between diagnostic difference in AWA, as a function of relative importance
- A plot of AWA for random tests as a function of probability ranging from 0 to 1

BED-FRAME Analyses
Free Online Tool in Development

BED-FRAME

- Choosing scenario
  - WA method
  - AWA method comparing two tests
  - AWA method evaluating one test
- Support

Data Input | Summary Plots | Expected Diagnostic Yield | Expected between-diagnostic differences and NNTs | WA comparison | Between-diagnostic WA differences

Prevalence (range from 0 to 1)

.4

Error Score (range from 0 to 100)

35

Note: Suppose the false negative score is 100, then what score would you like to give for a false positive? Lower score means less preference.

The relative importance is 0.35.

Label of Test A

MB

Label of Test B

PCR/ESI-MS

Confidence Interval for two-sided CI

0.5 0.95 0.99

0.5 0.55 0.6 0.65 0.7 0.75 0.8 0.85 0.9 0.95 0.99

Data Input Table

Table 1

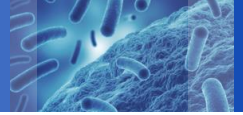
MB	True Results +	True Results -	Total
+	6408	384	7200
-	792	4416	4800
Total	6792	5208	12000

Table 2

PCR/ESI-MS	True Results +	True Results -	Total
+	6912	816	7200
-	288	3984	4800
Total	7728	4272	12000

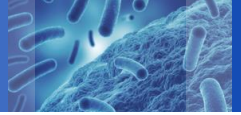
Intention-to-Diagnose (ITD)

- Analog of ITT in interventional studies
- Application carries similar protections, i.e., valid statistical inference (e.g., error control during hypothesis testing and correct coverage probabilities for confidence interval estimates of sensitivity or specificity) and clear generalizability
- Provides a realistic and unbiased answer to the most relevant question for diagnostic tests - that is, how they might perform in clinical practice - by capturing the range of consequences associated with test implementation in an intended-use setting
- However, ITD is rarely recognized or applied



Summary

- Greater pragmatism is needed in diagnostic studies too
  - Connection to clinical practice
  - Intention-to-diagnose
- Benefit:risk assessment in a structured and meaningful way is an important aspect of this
- BED-FRAME: a tool for pragmatic evaluation and comparison of diagnostic alternatives to aid in clinical decision making
  - Free online application for analyses in development

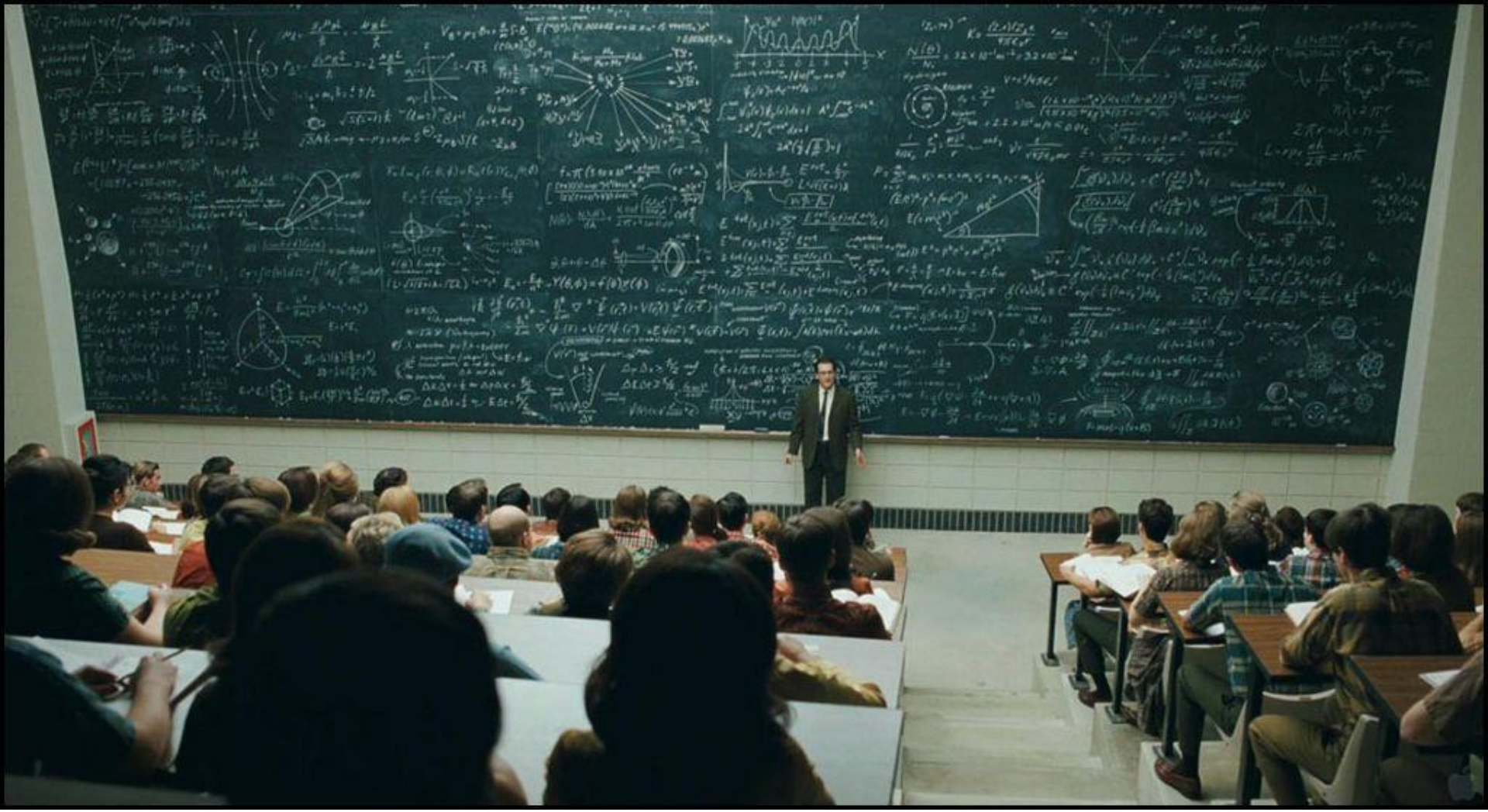


Summary

- Diagnostic yield can be used to evaluate diagnostic value
- Sensitivity, specificity, prevalence and the relative importance of errors are important elements of value assessment
- Prevalence and the relative importance may vary temporally, geographically, and culturally
  - Evaluate the utility of a diagnostic as these factors vary
- Consider studies to inform the prevalence and the relative importance of diagnostic errors

Significant Contributors (p<0.001)

- Gene Pennello
- Shanshan Zhang
- Toshi Hamasaki
- The Antibacterial Resistance Leadership Group



I have no doubt that you will enthusiastically applaud now …
because you are so relieved that it is over.
Thank you.