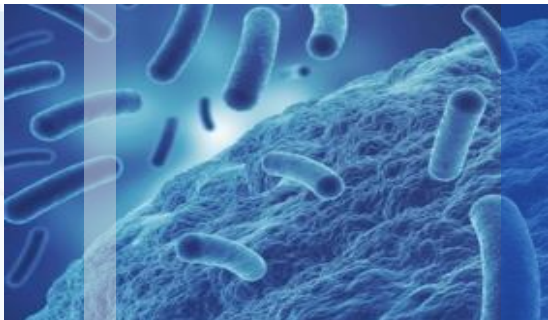
Pragmatism:
Finding and Answering the Important Question

Scott Evans, PhD, MS
Director, The Biostatistics Center
Professor, Department of Epidemiology and Biostatistics
George Washington University

Modern Study Designs for Pragmatic Translational Research
June 5, 2019





A Leaky Roof…

- Created a water bubble in my wall
- In addition to a new roof, I had to re-paper the wall
- I asked my neighbor, who recently papered a similar-sized room in his house:

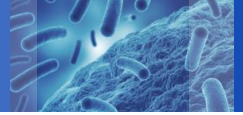
  “How much paper did you buy?”

- He replied: “Six rolls.”

Upon finishing the papering of the wall…

- I had only used only 4 rolls
- I told my neighbor that I had 2 rolls left
- He replied:

"Oh. That happened to you too?"



Two Things I've Learned about Traditional Trials and Diagnostic Studies

1. They are rigorously conducted by experts closely adhering to the highest standards and fundamental principles of randomized clinical trials and diagnostic studies

2. They are essentially useless for helping clinicians make treatment and diagnostic decisions

Most clinical trials fail to provide the evidence needed to inform medical decision-making. However, the serious implications of this deficit are largely absent from public discourse.

DeMets and Califf, *JAMA*, 2011

Outline

- Pragmatism and Pragmatic Clinical Trials
- Pragmatic Benefit:risk Evaluation
  - DOOR
  - Partial Credit
- SMART COMPASS
- Pragmatic Diagnostic Studies

PRAGMATIC CLINICAL TRIALS

Clinical Trials Today

- 18,000 RCTs published each year but reviews often conclude that more evidence is needed to inform clinical decision-making
- Research often not directly relevant to clinical practice
- Typical trial setting is a parallel universe, e.g.,
  - Selective enrollment criteria
  - Surrogate endpoints instead of clinical outcomes
  - Limited use of concomitant therapies
- Analyses not designed to evaluate global effects on patients

Pragmatic Clinical Trials

- Purpose to inform decisions about practice and policy
- Improve relevance and applicability
- Focus on effectiveness rather than efficacy
  - Evaluation under usual (vs. ideal) conditions
  - Extraneous variation, patient biases, and clinician inexperience is not to be controlled but part of the game
- Address questions about strategies for treating patients in practice vs. biology (mechanisms of action; causal pathways)

Characteristics

- Diverse and representative populations
- Multiple heterogeneous real-world settings, i.e., generalizable
  - Settings of everyday care (community clinics, hospitals, and health systems)
- Comparison to real-world alternatives rather than e.g., placebo
- Flexible study protocols
- Important patient-centered outcomes

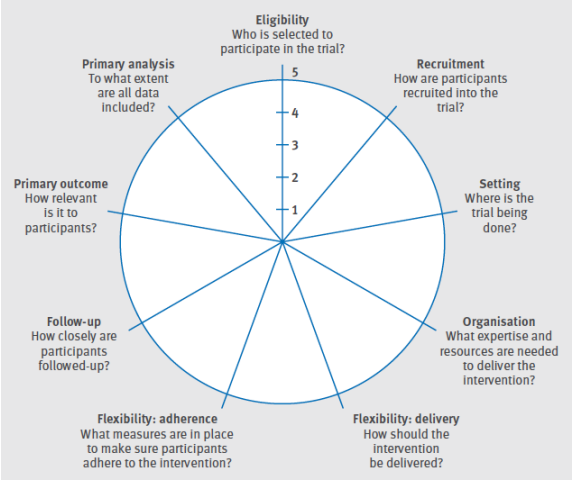
Explanatory vs. Pragmatic: A Continuum

Terminology:

Explanatory trials are often referred to as *efficacy trials*.

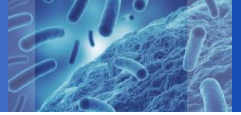
Pragmatic trials are often referred to as *effectiveness trials*.

Characteristic	Explanatory	Pragmatic
Question	Efficacy: can it work?	Effectiveness: Will it work?
Goal	Evaluate biological mechanisms.	Improve practice and policy.
Relevance to Practice	Indirect and rare. Generally little effort made to link design to practical decision-making in setting where intervention will be applied.	Direct. Useful for everyday decision-making.
Setting	Well-resourced. Parallel universe. Standardized.	Normal practice. Real world. Variable.
Participants	Selected with highly-defined entry criteria. Exclude patients unlikely to comply, with confounding conditions, and likely to have complications. Include people likely to respond.	Representative. Few restrictions or entry criteria. Patients as seen in practice.
Protocol	Rigid.	Flexible.
Control group	Placebo.	Real world alternative. Best available therapy.
Variation	Minimized. Standardization encouraged.	Local customization allowed.
Intervention Flexibility	Strict instructions for use.	Flexible.
Practitioner Expertise	Requirements for experience. Training may be required.	Full range of practitioners. Training not required.
Participant Adherence	Monitored and enforced.	May not be monitored.
Practitioner Adherence	Monitored. Poor adherers may be dropped.	May not be monitored.
Patient Follow-up	Formal. Frequently scheduled visits.	Informal.
Outcomes	Surrogates and process measures. May require training regarding methods of measurement.	Relevant to patients / practitioners. No formal training required.
Data Collection	Extensive. Requires data outside of usual care.	Brief. Data collected in usual care. May use administrative databases.
Analyses	Often ITT however frequently supplemented with PP analyses of e.g., compliers.	ITT. Includes all patients.



PRECIS-2 Tool

Loudon K, Treweek S, Sullivan F, Donnan P, Thorpe KE, Zwarenstein M. *BMJ* 2015;350:h 2147 doi: 10.1136



Most late-phase trials are somewhat pragmatic, or should be – we almost always evaluate an intervention strategy.

Concerns with Aspects of Pragmatic Trials

- Lack of blinding in many cases
  - Treatment crossover (but part of the game too)
  - Subjective and patient centered evaluations in particular could be biased
  - Objective evaluations are not entirely immune to biases
    - E.g., patients may selectively drop-out causing a distortion of the estimated effects

- Suggestion to consider blinding

Partially Blinded Trial

Active Albuterol or Placebo, Sham Acupuncture, or No Intervention in Asthma

Michael E. Wechsler, M.D., John M. Kelley, Ph.D., Ingrid O.E. Boyd, M.P.H., Stefanie Dutile, B.S., Gautham Marigowda, M.B., Irving Kirsch, Ph.D., Elliot Israel, M.D., and Ted J. Kaptchuk

N Engl J Med 2011;365:119-26.

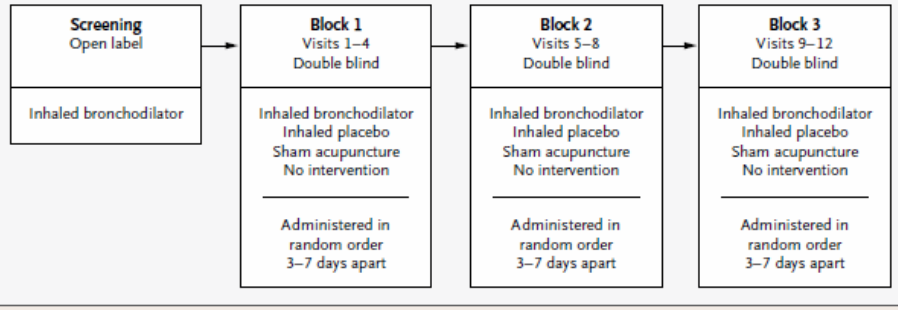


Figure 1. Schema for Study Interventions.
The time between blocks varied, but was generally 3 to 7 days.

Objective vs. Subjective Endpoint

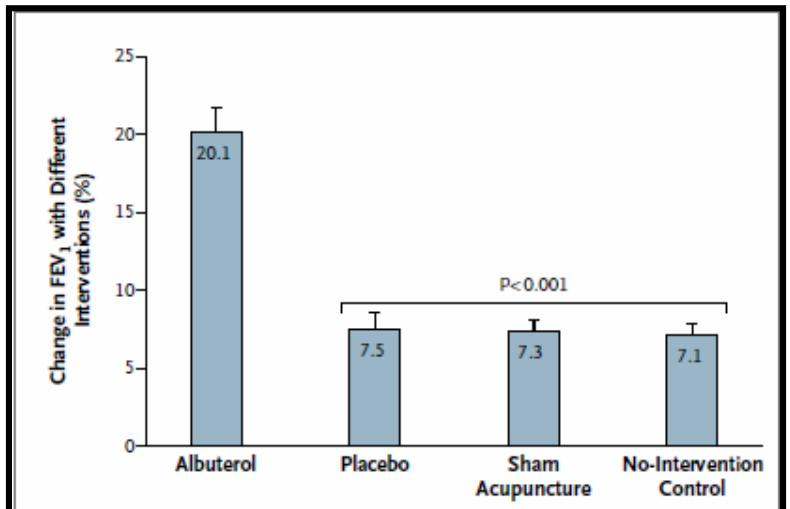


Figure 3. Percent Change in Maximum Forced Expiratory Volume in 1 Second (FEV_1) with Each of the Four Interventions.
The relative improvement in FEV_1 achieved with albuterol was significantly greater than that achieved with each of the other three interventions (P<0.001). No other differences among the four experimental conditions were significant. T bars indicate standard errors.

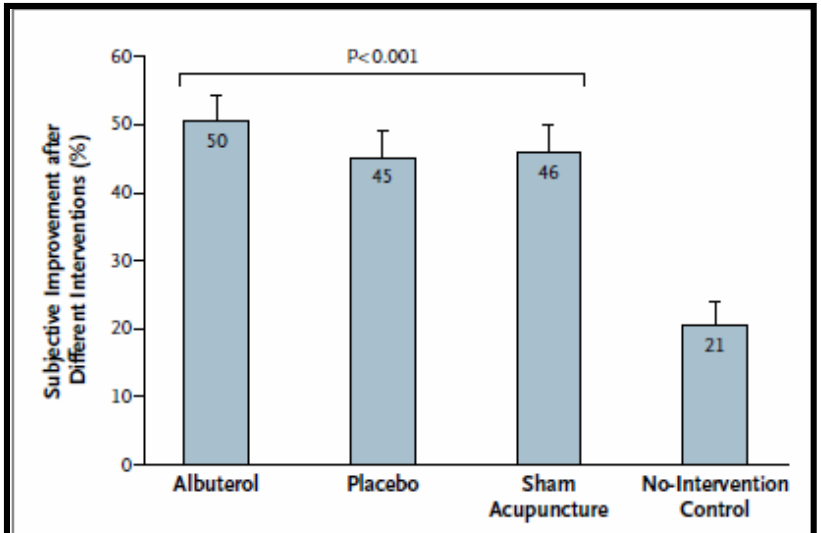


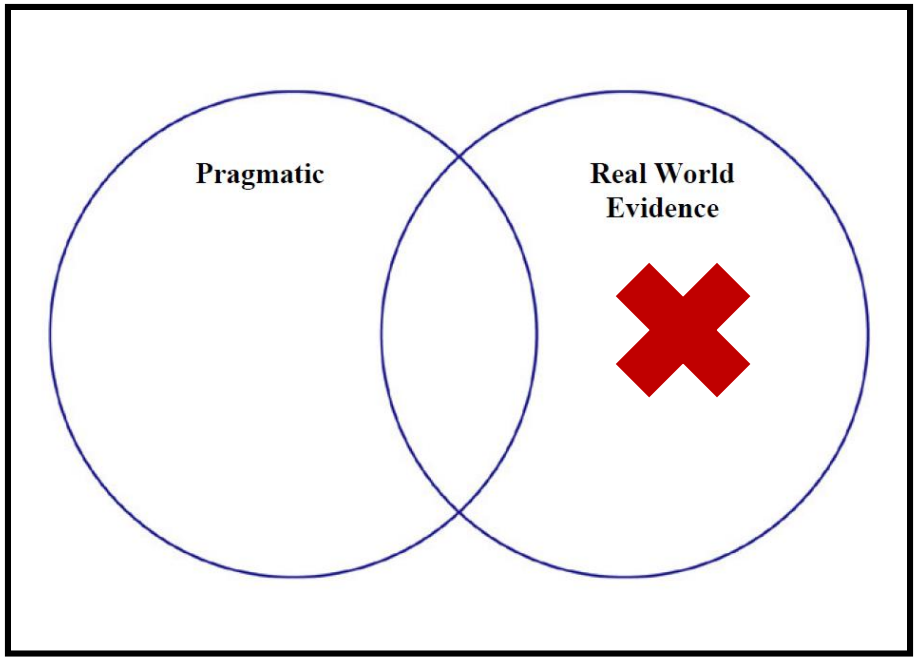
Figure 4. Percent Change in Subjective Improvement with Each of the Four Interventions.
The relative improvement in subjective outcomes, assessed with the use of a visual-analogue scale (with 0 indicating no improvement and 10 indicating complete improvement), was significantly greater with the albuterol inhaler, placebo inhaler, and sham acupuncture interventions than with the no-intervention control (P<0.001). No other differences among the four experimental conditions were significant. T bars indicate standard errors.

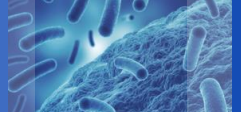
Pragmatism vs. RWE

- Real world evidence (RWE) concerns the data source i.e., evidence acquired using non-traditional sources e.g., EHR
- Pragmatism concerns the question
- One does not necessarily imply the other
- To answer important questions for clinical practice, conduct pragmatic studies
- To gain the cost and resource efficiencies of existing data, then consider utilizing real world data

What is the Motivation?

- Many want the resource efficiencies of RWD but do not want the dilution of treatment effects associated with pragmatic trials





How Pragmatic Are You?

- Suppose an RCT is conducted comparing A and B
- Efficacy is evaluated as a binary response at time T from randomization
- Safety is evaluated as the occurrence of SAEs
- Patient is randomized to A
- Prior to T the patient is changed to treatment C
- The patient subsequently experiences an SAE commonly associated with C
- Patient meets criteria of a responder at time T
- How is this patient evaluated for efficacy?
- For safety?
- Pragmatic evaluation consists of the evaluation of the strategy of application of treatment A
- The positive response and the SAE are considered downstream consequences to the initial assignment to A and are thus attributed to the strategy of application of A

- We should be more interested in downstream effects even after modification of therapy. If a patient is struggling and therapeutic adjustments can recover them, then the strategy of use worked! If a patient cannot recover with adjustments then the strategy of application failed.

DE GRUYTER Statistical Communications in Infectious Diseases. 2018; 20170002

Scott Evans[1] / Daniel B. Rubin[2] / John H. Powers[3,4] / Dean Follmann[5]

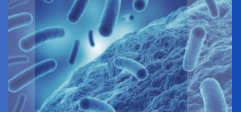
Analysis Populations in Anti-Infective Clinical Trials: Whom to Analyze?

[1] Department of Epidemiology and Biostatistics, Milken Institute School of Public Health of The George Washington University, Innovations in Design, Education, and Analysis (IDEA) Committee ,The Biostatistics Center, 6610 Executive Blvd., Suite 750 Rockville, United States of America, E-mail: SCID@degruyter.com
[2] Center for Drug Evaluation and Research, Food and Drug Administration, Silver Spring, MD, USA
[3] Clinical Research Directorate/Clinical Monitoring Research Program,Leidos Biomedical Research, Inc., NCI Campus at Frederick, Frederick, MD 21702, USA
[4] George Washington University School of Medicine, Washington DC, USA
[5] National Institute of Allergy and Infectious Diseases of the National Institutes of Health, Rockville, United States of America

Table 1: Recommended analysis populations for various research foci in clinical trials.

Research Focus	Primarily Useful to	Whom To Analyze	Clinical Utility/Pragmatism for Empiric Therapy	Clinical Utility/Pragmatism for Confirmatory Therapy	Preserves Integrity of Randomization
Evaluate strategy to treat clinical disease, i. e. empiric therapy in real-world setting	Todays' clinicians and patients	ITT (or mITT if blinded and uniformly assessed and small difference in number of participants with ITT)**	High	High	Yes***
Evaluate strategy to treat clinical disease caused by specific pathogen(s), i. e. confirmatory therapy*	Future clinicians and patients (i. e. with development of rapid point of care diagnostics)	mmITT**	Today: Variable (indirect) Future: possibly high with development of rapid diagnostics assuming affects do not change over time	High	Yes***
Understanding biological mechanisms of action, or evaluating potential for use if therapy can be tolerated/adhered to	Biologists, chemists	PP	Variable	Variable	No. Subject to the biases of observational studies

How pragmatic are typical trials?

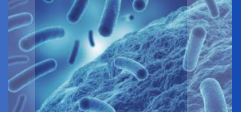


We are drowning in data but starving for knowledge.

Many of our wounds are self-inflicted.

What is the Question?

- We define analysis populations
  - Efficacy: ITT population
  - Safety: safety population
- Efficacy population ≠ safety population
- We combine these analyses into benefit:risk analyses. To whom does this analysis apply? What is the estimand?
- How do we do personalized medicine if we do not evaluate associations between outcomes?
- Is this what we need to inform clinical practice?



Example: Infectious Disease Trial

- Suppose we measure the duration of hospitalization
- Shorter duration is better … or is it?
- The faster the patient dies, the shorter the duration
- Interpretation of an outcome needs context of other clinical outcomes for the same patient
- So why do we analyze them separately?

Example: Cardiovascular Event Prevention Trial

- Evaluate time-to-first event (e.g., death, MI, stroke)
  - But there can be multiple events
- Fail to distinguish differential importance of events
  - Death > non-fatal event
  - Disabling > non-disabling event
  - Permanent sequelae > transient sequelae
- In deciding how to treat patients, shouldn't we consider this information?
- If so, why are we not designing and analyzing trials in this way?

Example: Cardiovascular Event Prevention Trial

- Competing risk challenge: death informatively censors time to stroke
- Decision analysis approach: summarize the marginal effects
  - Double-counting: Fatal bleed counted as a death and a major bleed
  - How do we interpret this?

Quiz

- Suppose a loved one is diagnosed with a serious disease
- You are selecting treatment
- 3 treatment options: A, B, and C
- 2 outcomes, equally important
  - Treatment success: yes/no
  - Safety event: yes/no

RCT Comparing A, B, and C
Analysis of Outcomes

A (N=100)	B (N=100)	C (N=100)

RCT Comparing A, B, and C
Analysis of Outcomes

A (N=100)	B (N=100)	C (N=100)
Success: 50%	Success: 50%	Success: 50%

RCT Comparing A, B, and C
Analysis of Outcomes

A (N=100)	B (N=100)	C (N=100)
Success: 50%	Success: 50%	Success: 50%
Safety event: 30%	Safety event: 50%	Safety event: 50%

RCT Comparing A, B, and C
Analysis of Outcomes

A (N=100)	B (N=100)	C (N=100)
Success: 50%	Success: 50%	Success: 50%
Safety event: 30%	Safety event: 50%	Safety event: 50%

Which treatment would you choose?

RCT Comparing A, B, and C
Analysis of Outcomes

A (N=100)	B (N=100)	C (N=100)
Success: 50%	Success: 50%	Success: 50%
Safety event: 30%	Safety event: 50%	Safety event: 50%

Which treatment would you choose?

They all have the same success rate.

RCT Comparing A, B, and C
Analysis of Outcomes

A (N=100)	B (N=100)	C (N=100)
Success: 50%	Success: 50%	Success: 50%
Safety event: 30%	Safety event: 50%	Safety event: 50%

Which treatment would you choose?

They all have the same success rate.

A has the lowest safety event rate.

RCT Comparing A, B, and C
Analysis of Outcomes

A (N=100)	B (N=100)	C (N=100)
Success: 50%	Success: 50%	Success: 50%
Safety event: 30%	Safety event: 50%	Safety event: 50%

Which treatment would you choose?

They all have the same success rate.

A has the lowest safety event rate.

B and C are indistinguishable.

RCT Comparing A, B, and C
Analysis of Outcomes

A (N=100)	B (N=100)	C (N=100)
Success: 50%	Success: 50%	Success: 50%
Safety event: 30%	Safety event: 50%	Safety event: 50%

Which treatment would you choose?

They all have the same success rate.

A has the lowest safety event rate.

B and C are indistinguishable.

Choose A…right?

Analysis of Patients: 4 Possible Outcomes

A (N=100)

Success: 50%

Safety event: 30%

		Success +	Success -
SE	+	15	15
	-	35	35

B (N=100)

Success: 50%

Safety event: 50%

		Success +	Success -
SE	+	50	0
	-	0	50

C (N=100)

Success: 50%

Safety event: 50%

		Success +	Success -
SE	+	0	50
	-	50	0

Analysis of Patients: 4 Possible Outcomes

A (N=100)

Success: 50%

Safety event: 30%

		Success +	Success -
SE	+	15	15
	-	35	35

B (N=100)

Success: 50%

Safety event: 50%

		Success +	Success -
SE	+	50	0
	-	0	50

C (N=100)

Success: 50%

Safety event: 50%

		Success +	Success -
SE	+	0	50
	-	50	0

Analysis of Patients: 4 Possible Outcomes

A (N=100)

Success: 50%

Safety event: 30%

		Success +	Success -
SE	+	15	15
	-	35	35

B (N=100)

Success: 50%

Safety event: 50%

		Success +	Success -
SE	+	50	0
	-	0	50

C (N=100)

Success: 50%

Safety event: 50%

		Success +	Success -
SE	+	0	50
	-	50	0

Analysis of Patients: 4 Possible Outcomes

A (N=100)

Success: 50%

Safety event: 30%

		Success +	Success -
SE	+	15	15
	-	35	35

B (N=100)

Success: 50%

Safety event: 50%

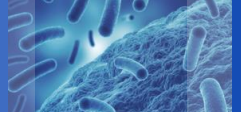
		Success +	Success -
SE	+	50	0
	-	0	50

C (N=100)

Success: 50%

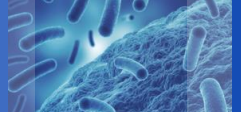
Safety event: 50%

		Success +	Success -
SE	+	0	50
	-	50	0



Our culture is to use patients to analyze the outcomes.

Shouldn’t we use outcomes to analyze the patients?



Scott's father (a math teacher) to his confused son many years ago:

"The order of operations is important…"

A Vision

STATISTICS IN BIOPHARMACEUTICAL RESEARCH
2016, VOL. 8, NO. 4, 386–393
http://dx.doi.org/10.1080/19466315.2016.1207561

Taylor & Francis
Taylor & Francis Group

Using Outcomes to Analyze Patients Rather than Patients to Analyze Outcomes: A Step Toward Pragmatism in Benefit:Risk Evaluation

Scott R. Evans[a,b] and Dean Follmann[c]

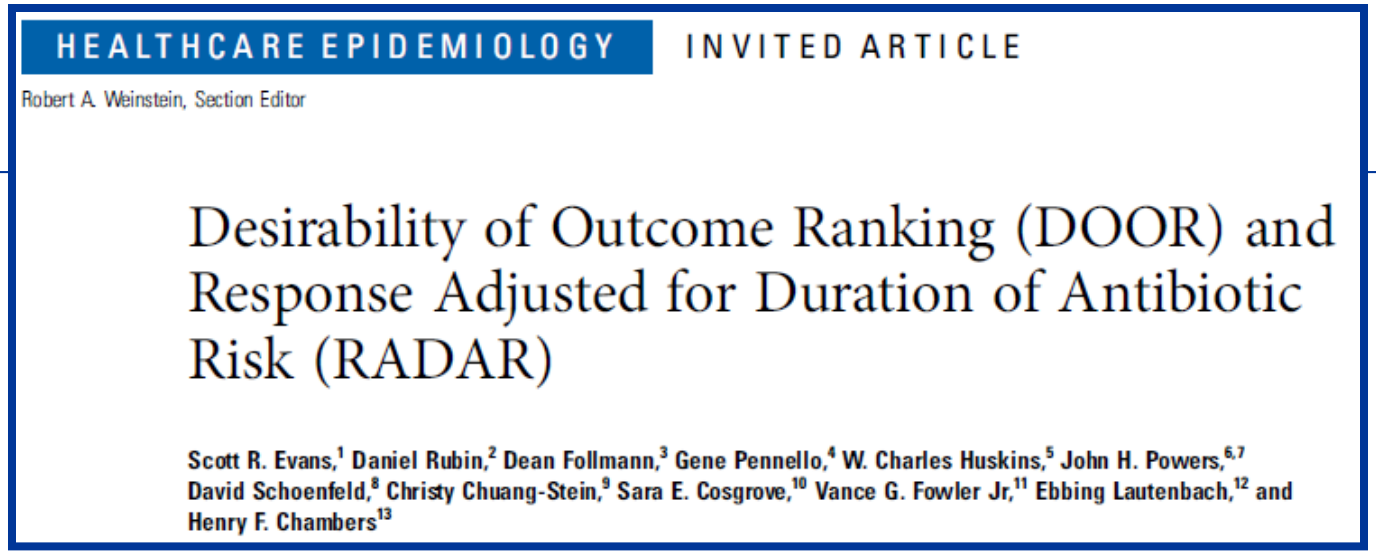
[a]Department of Biostatistics, Harvard University, Boston, MA, USA; [b]Center for Biostatistics in AIDS Research, Harvard University, Boston, MA, USA; [c]National Institute of Allergy and Infectious Diseases (NIAID) of the National Institutes of Health (NIH), Bethesda, MD, USA.

The good physician treats the disease.

The great physician treats the patient.

William Osler

Perhaps we should analyze the patient.



HEALTHCARE EPIDEMIOLOGY INVITED ARTICLE

Robert A. Weinstein, Section Editor

Desirability of Outcome Ranking (DOOR) and Response Adjusted for Duration of Antibiotic Risk (RADAR)

Scott R. Evans,[1] Daniel Rubin,[2] Dean Follmann,[3] Gene Pennello,[4] W. Charles Huskins,[5] John H. Powers,[6,7] David Schoenfeld,[8] Christy Chuang-Stein,[9] Sara E. Cosgrove,[10] Vance G. Fowler Jr,[11] Ebbing Lautenbach,[12] and Henry F. Chambers[13]

Before we analyze several hundred patients,

we must understand how to analyze one.

- *The patient journey*: “exit examination” or “discharge review” based on a synthesis of benefits, harms, QOL

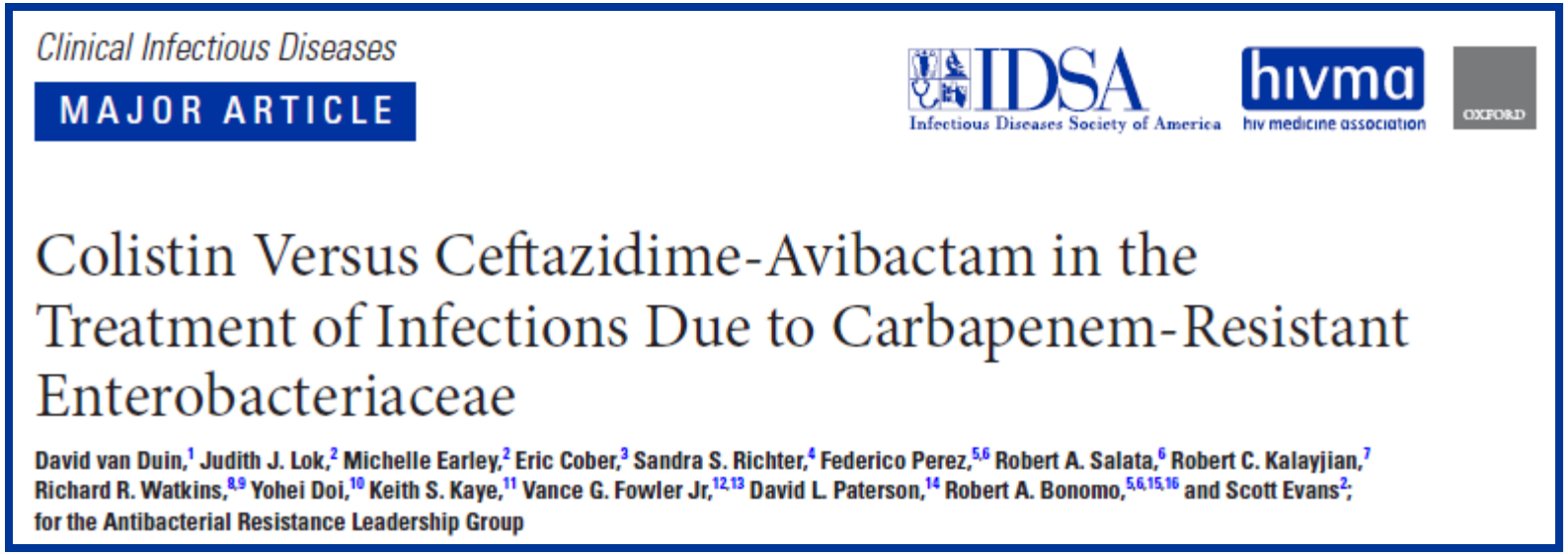
- DOOR probability: probability of a more desirable global outcome when assigned to the new vs. the control treatment

Example

CRACKLE
Antibacterial Resistance Leadership Group

Motivating question:

Should we use ceftazidime-avibactam or colistin for the initial treatment of CRE infection?



Clinical Infectious Diseases

MAJOR ARTICLE

IDSA Infectious Diseases Society of America

hivma hiv medicine association

OXFORD

Colistin Versus Ceftazidime-Avibactam in the Treatment of Infections Due to Carbapenem-Resistant Enterobacteriaceae

David van Duin,[1] Judith J. Lok,[2] Michelle Earley,[2] Eric Cober,[3] Sandra S. Richter,[4] Federico Perez,[5,6] Robert A. Salata,[6] Robert C. Kalayjian,[7] Richard R. Watkins,[8,9] Yohei Doi,[10] Keith S. Kaye,[11] Vance G. Fowler Jr,[12,13] David L. Paterson,[14] Robert A. Bonomo,[5,6,15,16] and Scott Evans[2]; for the Antibacterial Resistance Leadership Group

DOOR

- DOOR with 4 levels
  - Alive; discharged home
  - Alive; not discharged home; no renal failure
  - Alive; not discharged home; renal failure
  - Death

- Looking for northward migration of patients in these categories

DOOR

	Colistin (N=46)	Caz-Avi (N=26)
Discharged home	4 (9%)	6 (23%)
Alive; not discharged home; no renal failure	25 (54%)	17 (65%)
Alive; not discharged home; renal failure	5 (11%)	1 (4%)
Death	12 (26%)	2 (8%)

- IPTW-adjusted DOOR Probability: 64% (53%, 75%)
- IPTW-adjusted Win Ratio: 3.0 (1.32, 9.72)

IPTW adjustments: Pitt score, infection type (BSI vs. UTI), and creatinine (sensitivity analyses only)

Challenges

- Cultural change
- Composites
  - Are tricky and require great care
    - Several very good references (e.g., Neaton et.al., J Cardiac Failure, 2005)
  - Commonly used
    - E.g., PFS in oncology, MACE in cardiovascular disease
    - Though the motive is often to reduce the sample size in event-time trials

Challenges

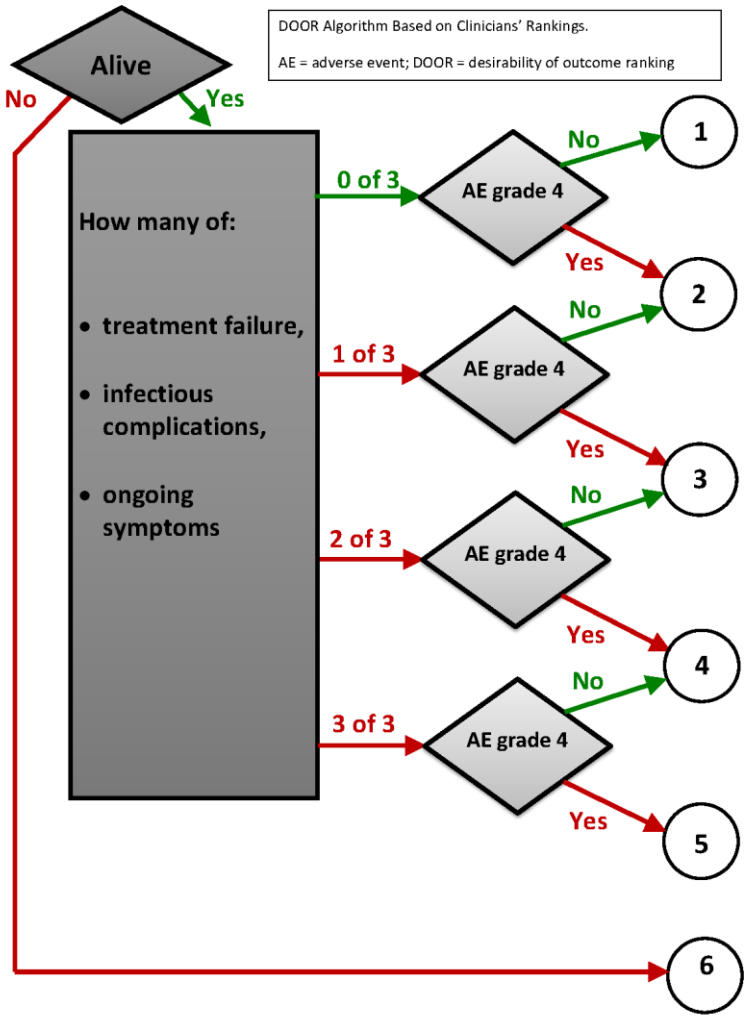
- Construction of ordinal DOOR is novel and challenging
- Careful deliberation is essential to synthesize the outcomes
- An example strategy …

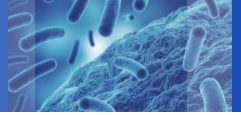
BAC DOOR

- ARLG conducted a pre-trial sub-study to develop DOOR in *Staphylococcus aureus* bacteremia
- 20 representative patient profiles (benefits, harms, and QoL) constructed based on experiences observed in prior trials
- Profiles sent to 43 expert clinicians. They were asked to rank the patient profiles by desirability of outcome.
- Examined clinician consensus and component outcomes that drive clinician rankings

Decision Tree Algorithm

- Things that we learned
  - Cumulative effect
  - Symptoms important
  - Major non-fatal outcomes had similar importance





Can we account for:

1. **Potential unequal steps between categories?**

2. **Varying perspectives among patients / clinicians regarding the desirability of the categories?**

PARTIAL CREDIT

	Score
Discharged home	100
Alive; not discharged home; no renal failure	Partial credit
Alive; not discharged home; renal failure	Partial credit
Death	0

Partial Credit: How Much?

A clinical trials doctrine:

Transparency and pre-specification are the law …

except when it comes to defining the relative importance of different outcomes… in which case it is shunned.

But once study conclusions have been drawn, we have made a decision about the value of the outcomes without transparency…

and the decision-makers may not consciously know what those values are.

Partial Credit: How Much?

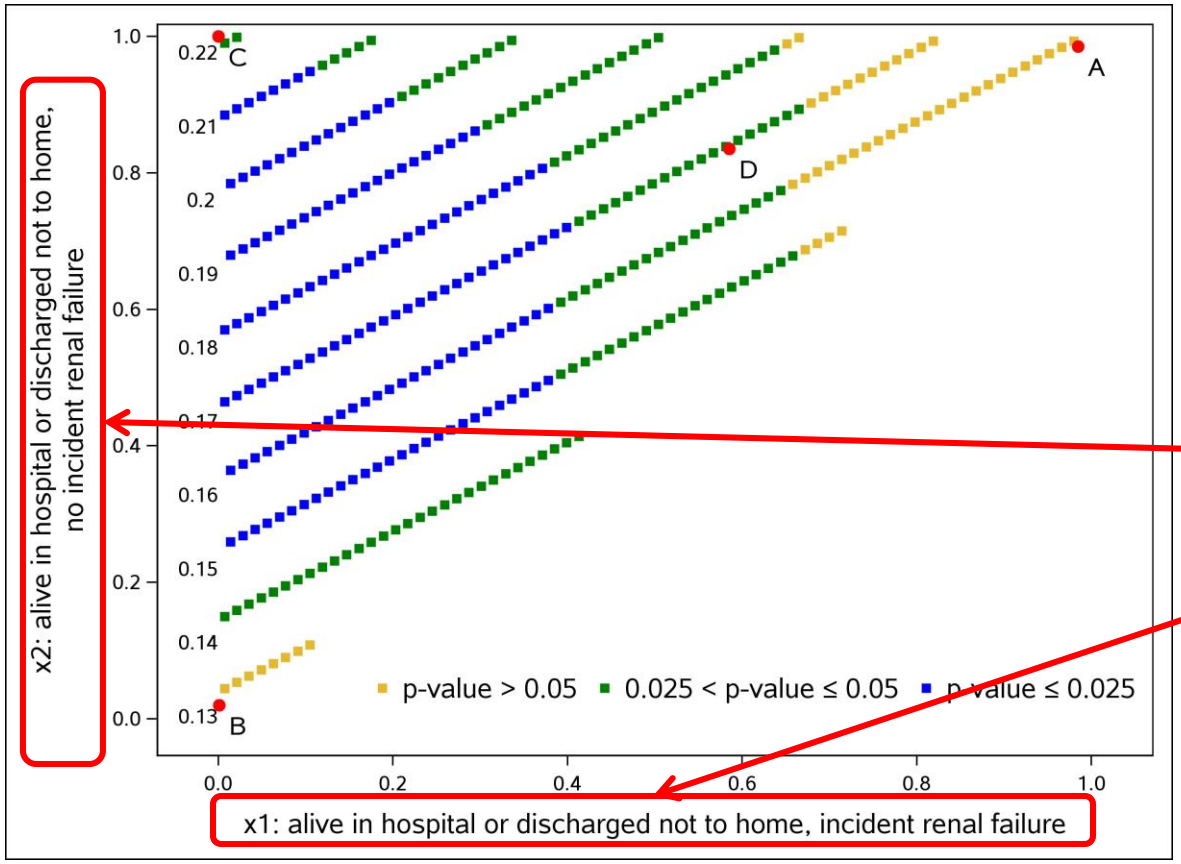
- Strategies
  - Survey expert clinicians for grading key
  - Patient-guided using QOL

Partial Credit

People have different perspectives.

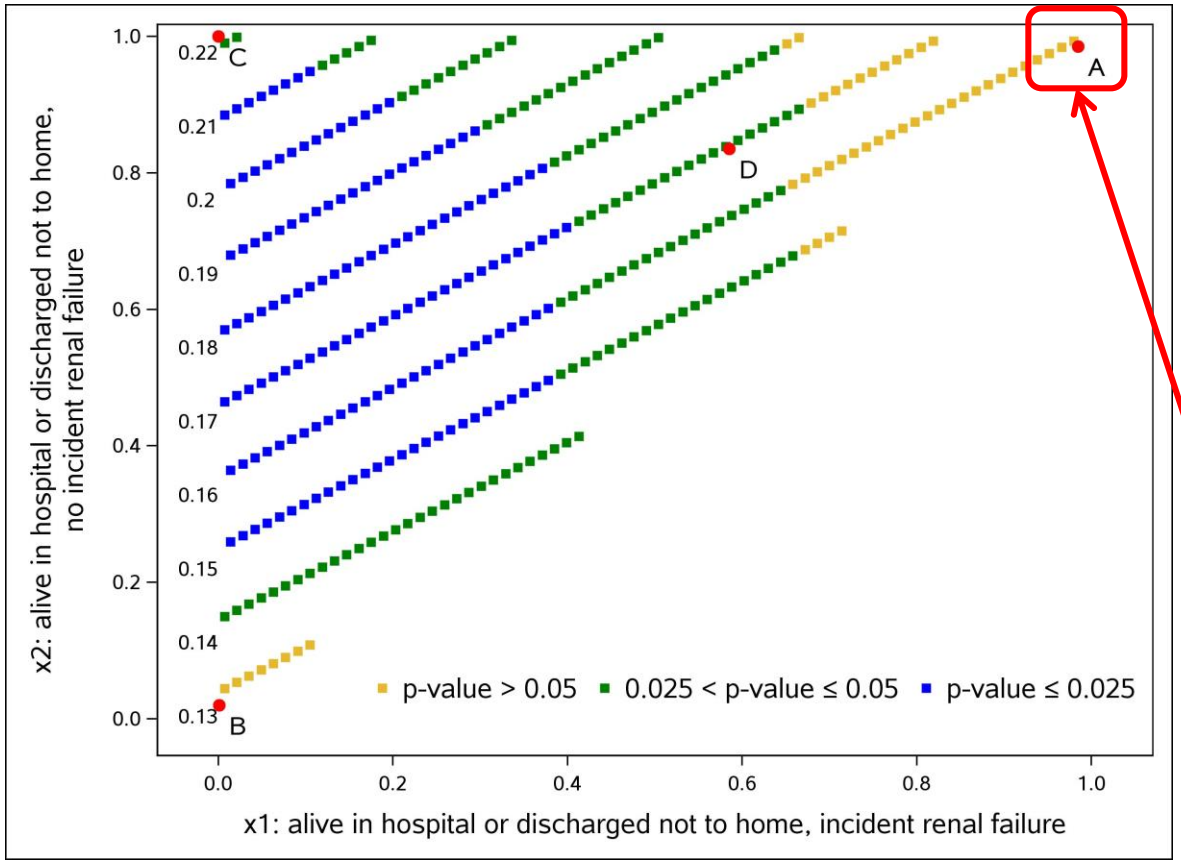
Display treatment contrast as partial credit varies, allowing people to make their own choices based on their own value system.

Contours of Effects as Partial Credit Varies



Category	Credit
Discharged home	100
Alive; Not discharged home; No renal failure	Partial credit
Alive; Not discharged home; Renal failure	Partial credit
Death	0

Survival



Category	Credit
Discharged home	100
Alive; Not discharged home; No renal failure	100
Alive; Not discharged home; Renal failure	100
Death	0

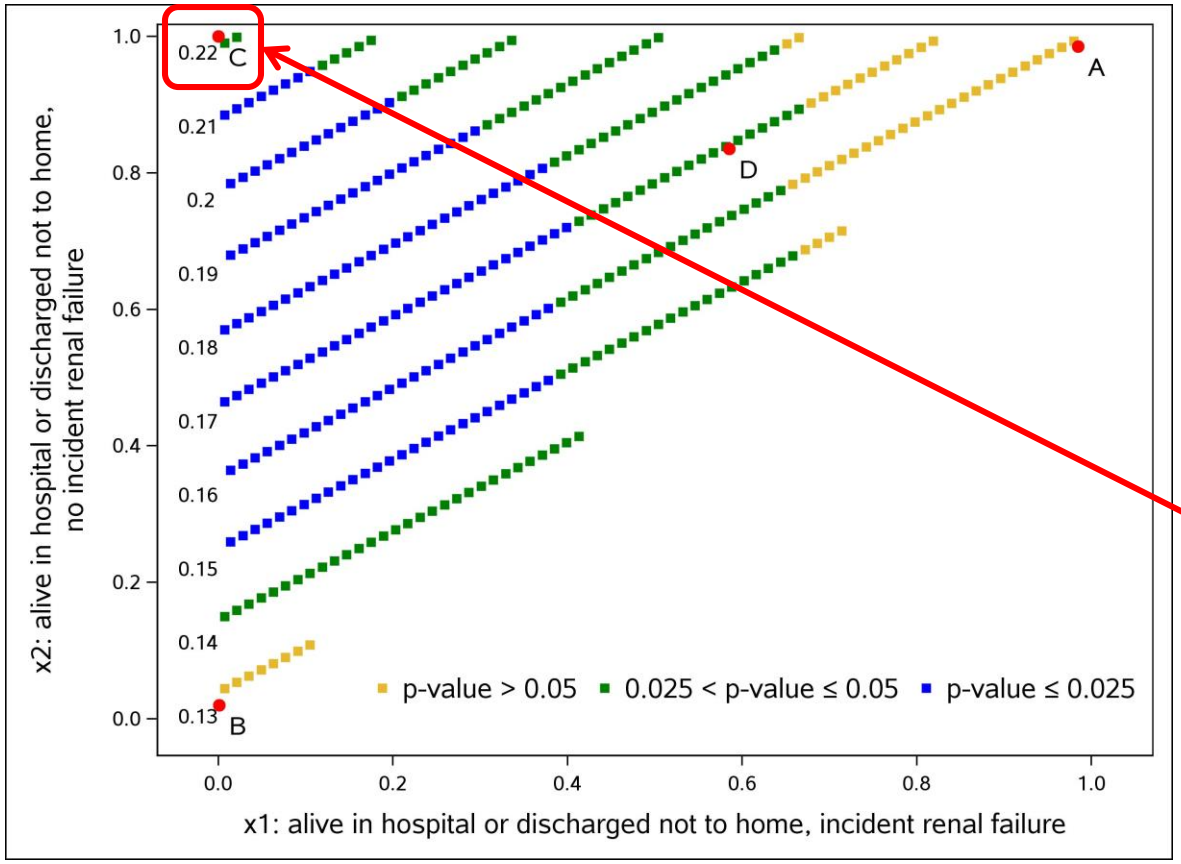
Caz-avi advantage: 0.16 (-0.04, 0.32), p = 0.10

Discharged Home

Category	Credit
Discharged home	100
Alive; Not discharged home; No renal failure	0
Alive; Not discharged home; Renal failure	0
Death	0

Caz-avi advantage: 0.13 (-0.03, 0.31), p = 0.12

Alive without Renal Failure



Category	Credit
Discharged home	100
Alive; Not discharged home; No renal failure	100
Alive; Not discharged home; Renal failure	0
Death	0

Caz-avi advantage: 0.22 (0.02, 0.40), p = 0.03

Compromise

Category	Credit
Discharged home	100
Alive; Not discharged home; No renal failure	80
Alive; Not discharged home; Renal failure	60
Death	0

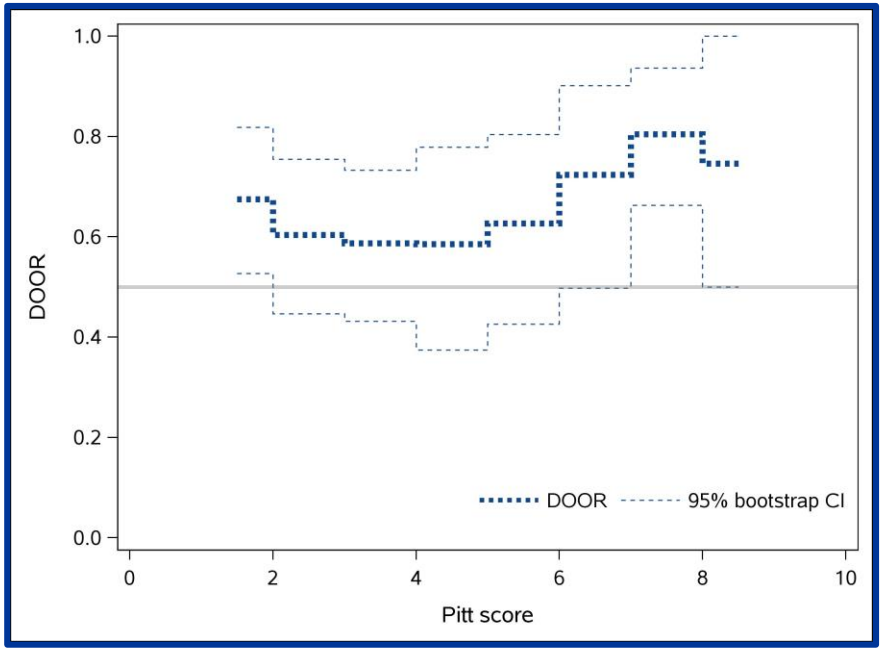
Caz-avi advantage: 0.17 (0.01, 0.30), p = 0.04

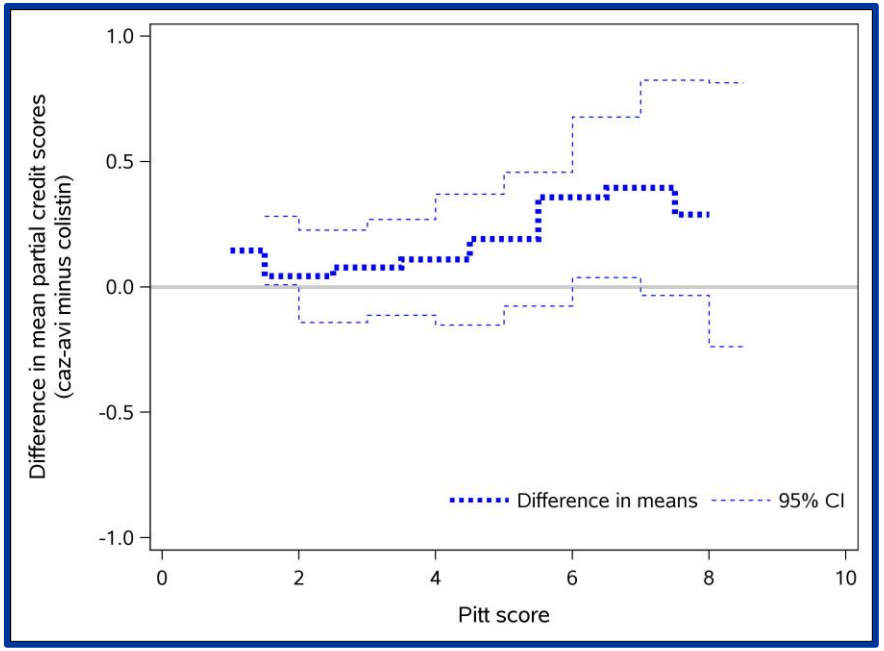
Tailoring Medicine

Who benefits from this new therapy?

DOOR STEPP

Caz-Avi-Colistin Contrast as a Function of Disease Severity





DOOR Probability

Partial Credit (80/60)

Largest differences are in the most severe patients.

DOOR STEPP

PROVIDE

- Prospective multi-center observational evaluation among adult hospitalized patients with MRSA bloodstream infections
- Research Question
  - What is the vancomycin pharmacodynamic exposure target associated with optimal treatment outcome?
- N=265

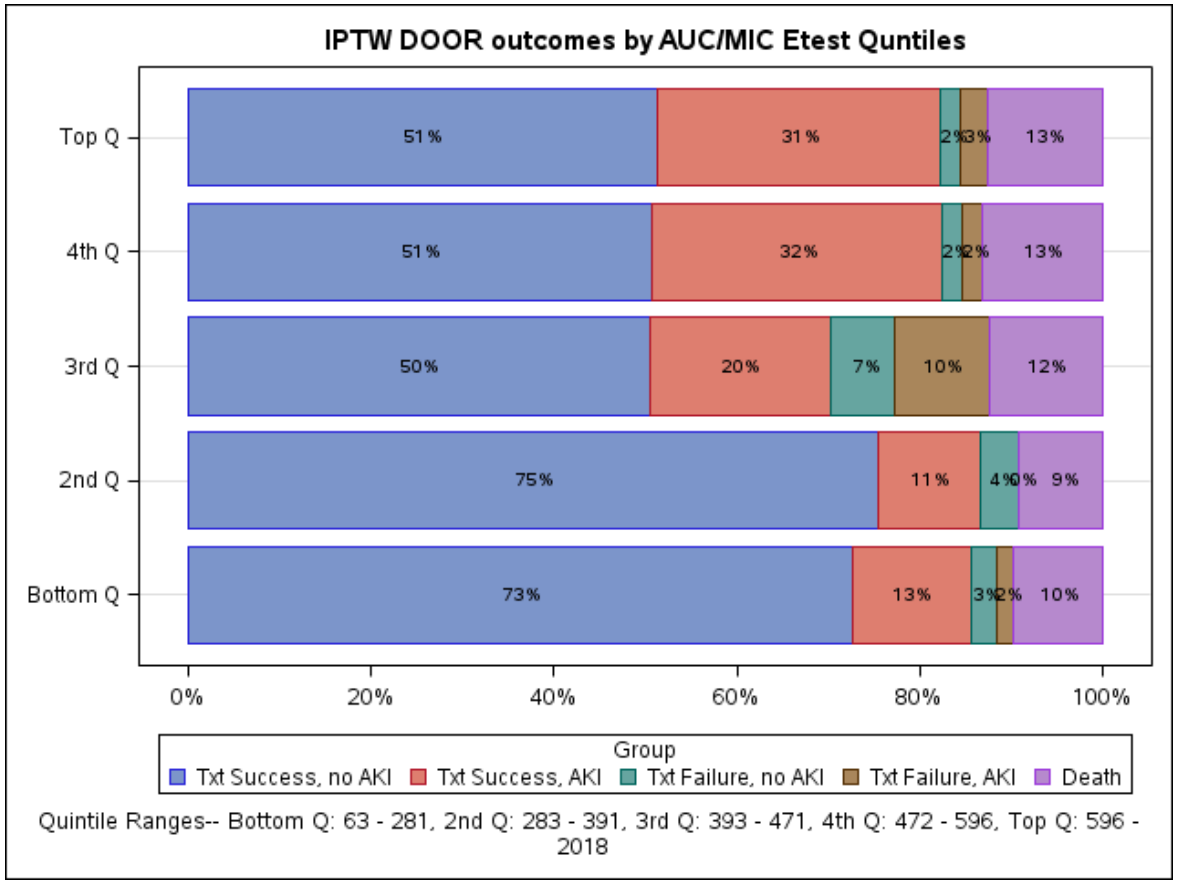
DOOR

Better outcome

Worse outcome

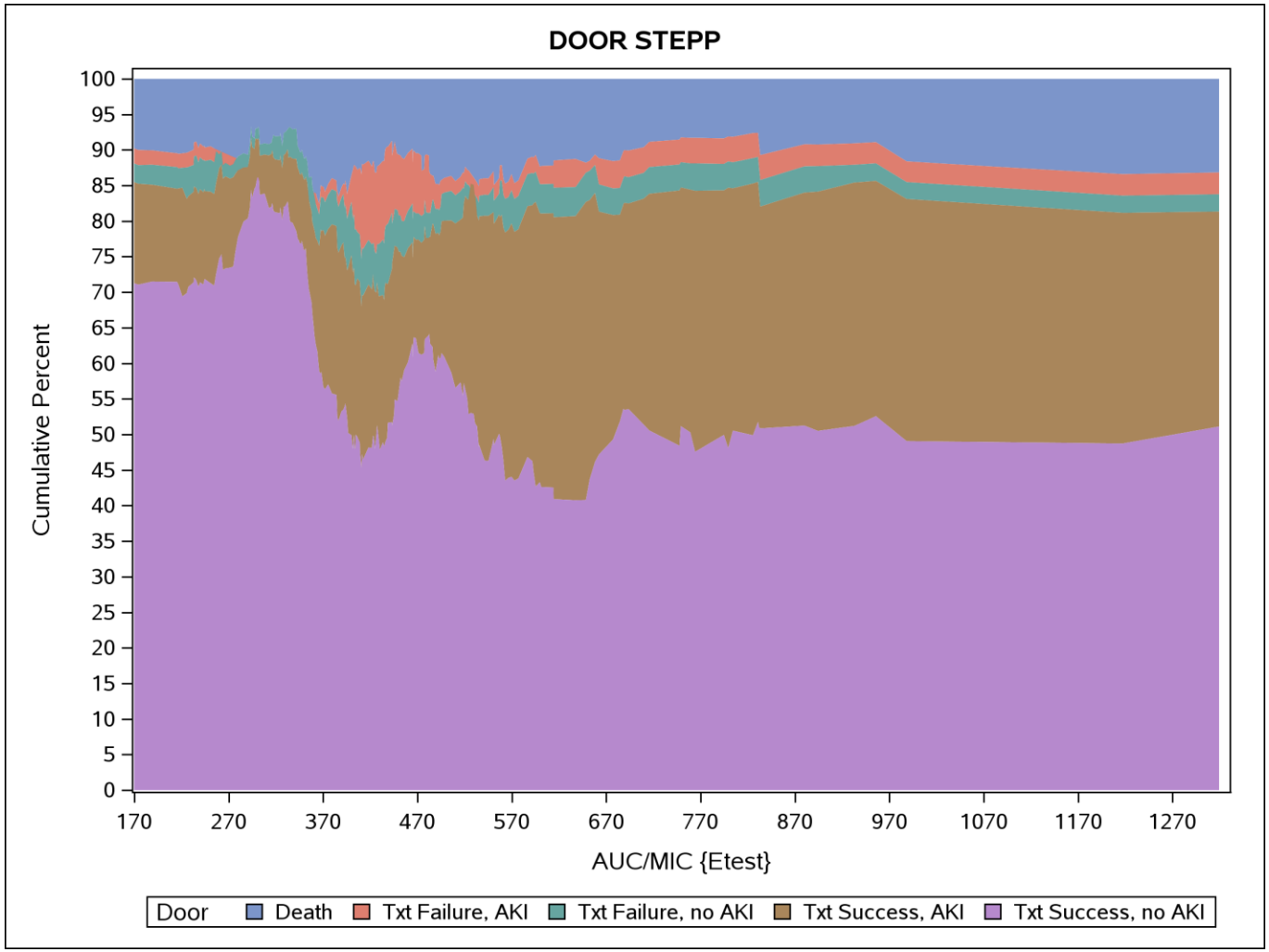
	Treatment success without AKI
	Treatment success with AKI
	Treatment failure (persistent bacteremia) without AKI
	Treatment failure with AKI
	Death

DOOR Outcomes by Dosing Quintiles

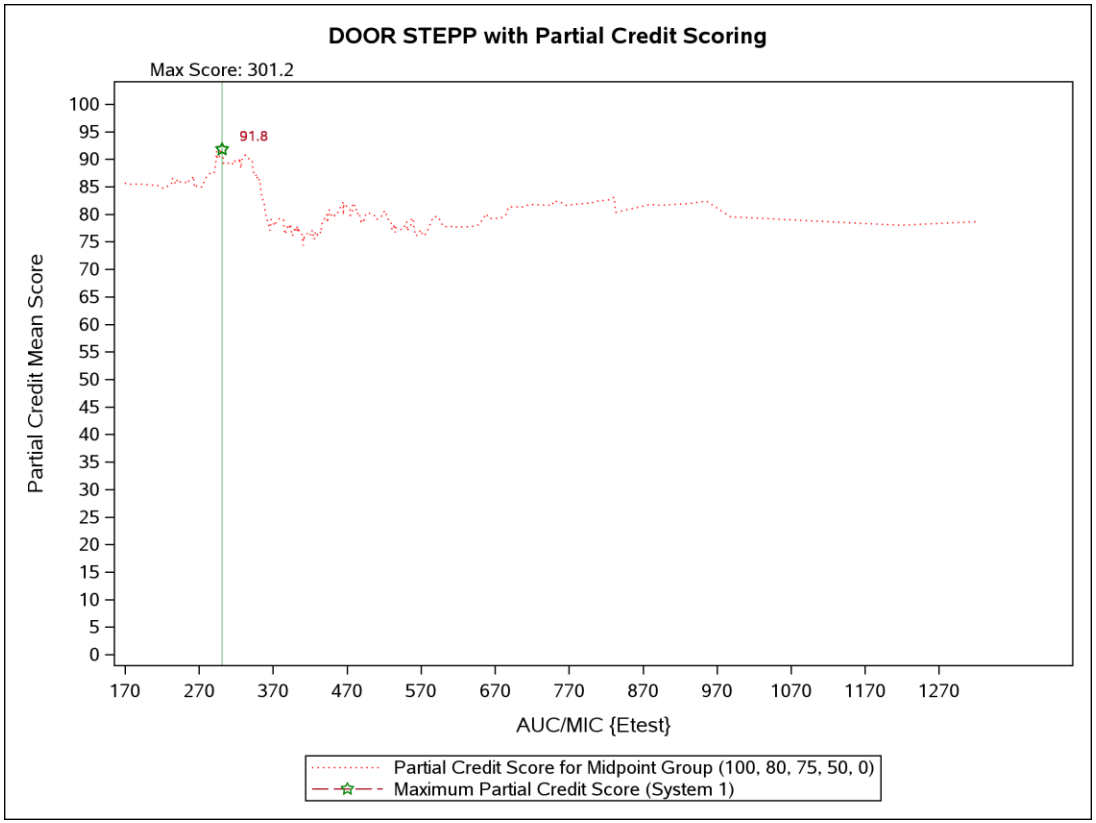


- IPTW adjustments for: presence of infective endocarditis, baseline calculated creatinine clearance, Apache II score, and indicator of any of: prosthetic joint, cardiac prosthetic device, intravascular prosthetic material.

DOOR STEPP


DOOR STEPP
Cumulative Percent
100
95
90
85
80
75
70
65
60
55
50
45
40
35
30
25
20
15
10
5
0
170
270
370
470
570
670
770
870
970
1070
1170
1270
AUC/MIC {Etest}
Door
Death
Txt Failure, AKI
Txt Failure, no AKI
Txt Success, AKI
Txt Success, no AKI

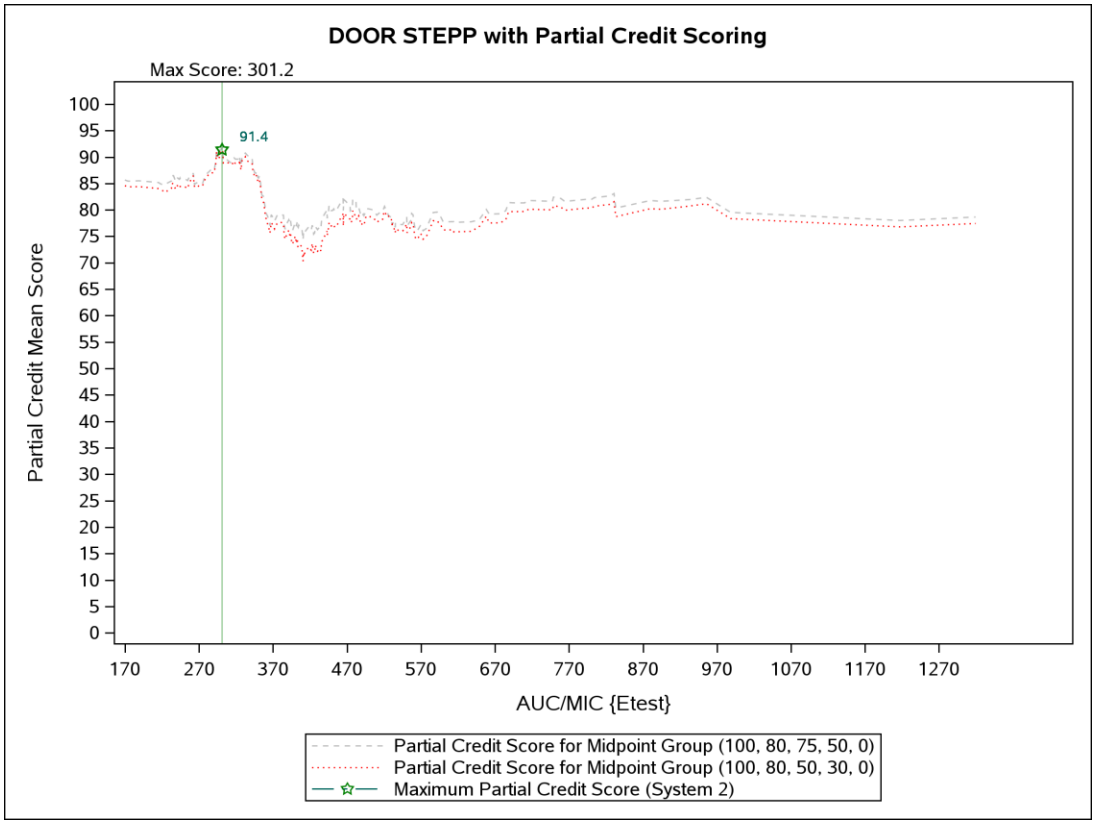
DOOR STEPP: Partial Credit Clinician A



Category	Credit
Treatment Success; No Kidney Injury	100
Treatment Success; Kidney Injury	80
Treatment Failure; No Kidney Injury	75
Treatment Failure; Kidney Injury	50
Death	0

Optimal Dose: 301.2

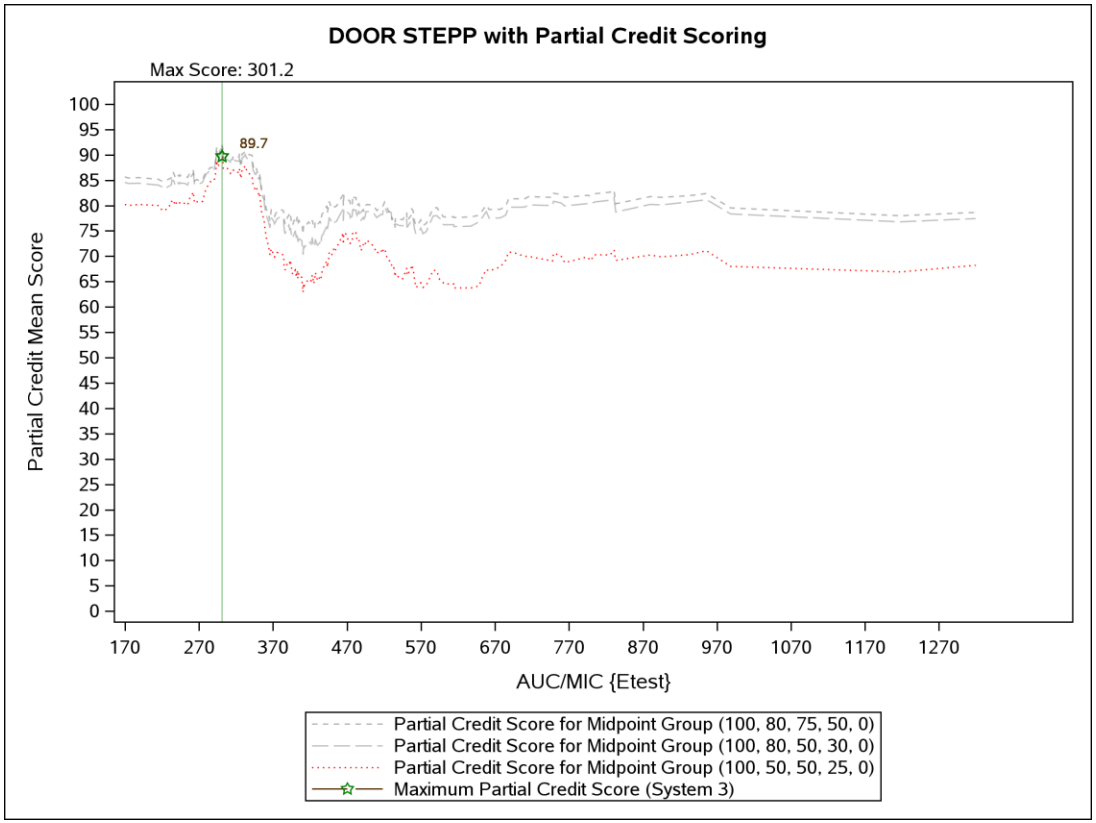
DOOR STEPP: Partial Credit Clinician B



Category	Credit
Treatment Success; No Kidney Injury	100
Treatment Success; Kidney Injury	80
Treatment Failure; No Kidney Injury	50
Treatment Failure; Kidney Injury	30
Death	0

Optimal Dose: 301.2

DOOR STEPP: Partial Credit Clinician C



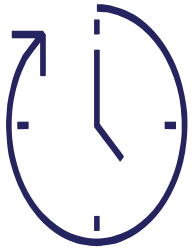
Category	Credit
Treatment Success; No Kidney Injury	100
Treatment Success; Kidney Injury	50
Treatment Failure; No Kidney Injury	50
Treatment Failure; Kidney Injury	25
Death	0

Optimal Dose: 301.2

ANOTHER EXAMPLE

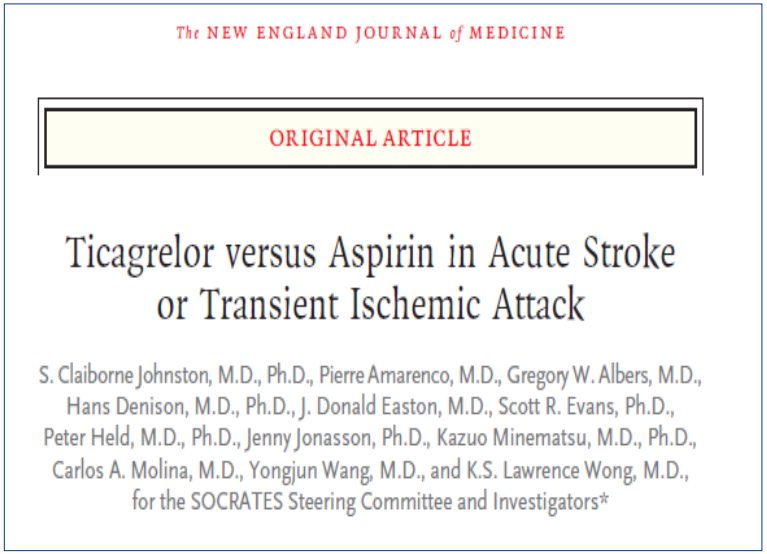
SOCRATES

International (**674** centres in **33** countries), double-blind, randomised controlled trial of 13,199 participants randomised to ticagrelor vs. aspirin in acute stroke or transient ischemic attack (NCT01994720)



Primary end point: **time to stroke**, **MI**, or **death by 90 days**

- **6.7%** event rate in ticagrelor group
- **7.5%** event rate in aspirin group
- **HR=0.89** (0.78, 1.01), p=0.07



The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

Ticagrelor versus Aspirin in Acute Stroke or Transient Ischemic Attack

S. Claiborne Johnston, M.D., Ph.D., Pierre Amarenco, M.D., Gregory W. Albers, M.D., Hans Denison, M.D., Ph.D., J. Donald Easton, M.D., Scott R. Evans, Ph.D., Peter Held, M.D., Ph.D., Jenny Jonasson, Ph.D., Kazuo Minematsu, M.D., Ph.D., Carlos A. Molina, M.D., Yongjun Wang, M.D., and K.S. Lawrence Wong, M.D., for the SOCRATES Steering Committee and Investigators*

SOCRATES Quotes

*The **unexamined life** is **not worth living.***

*Not life, but good life, is to be **chiefly valued.***

Wisdom** begins in **wonder.

DOOR

MOST DESIRABLE

LEAST DESIRABLE

Benefit-risk category	Ticagrelor (N=6589) n (%)	Aspirin (N=6610) n (%)	Cumulative difference % (95% CI)
Survived with no event			
Survived with non-disabling stroke, MI or PLATO major bleeding, 1 event			
Survived with non-disabling stroke, MI or PLATO major bleeding, >1 event			
Survived with disabling stroke			
Death			

Aspirin results

Benefit-risk category	Ticagrelor (N=6589) n (%)	Aspirin (N=6610) n (%)	Cumulative difference % (95% CI)
Survived with no event		6089 (92.1)	
Survived with non-disabling stroke, MI or PLATO major bleeding, 1 event		171 (2.6)	
Survived with non-disabling stroke, MI or PLATO major bleeding, >1 event		11 (0.2)	
Survived with disabling stroke		281 (4.3)	
Death		58 (0.9)	

Will people on Ticagrelor migrate to a more desirable outcome?

Ticagrelor results

Benefit-risk category	Ticagrelor (N=6589) n (%)	Aspirin (N=6610) n (%)	Cumulative difference % (95% CI)
Survived with no event	6124 (92.9)	6089 (92.1)	
Survived with non-disabling stroke, MI or PLATO major bleeding, 1 event	147 (2.2)	171 (2.6)	
Survived with non-disabling stroke, MI or PLATO major bleeding, >1 event	6 (0.1)	11 (0.2)	
Survived with disabling stroke	244 (3.7)	281 (4.3)	
Death	68 (1.0)	58 (0.9)	

Analyses

- DOOR probability = 0.504 (95% CI 0.499–0.508, p=0.096)
  - The probability of a more desirable result with ticagrelor is 50.4%
- Win ratio = 1.11 (95% CI 0.98–1.26, p=0.096)
  - Ticagrelor wins 1.11 times more frequently than it loses
- Partial credit can be applied using QOL instruments

SMART COMPASS

Clinical Infectious Diseases

INVITED ARTICLE

IDEA: Scott R. Evans and Victor De Gruttola, Section Editors

Sequential, Multiple-Assignment, Randomized Trials for COMparing Personalized Antibiotic StrategieS (SMART-COMPASS)

Scott R. Evans,[1] Dean Follmann,[2] Ying Liu,[3] Thomas Holland,[4] Sarah B. Doernberg,[5] Nadine Rouphael,[6] Toshimitsu Hamasaki,[7] Yunyun Jiang,[1] Judith J. Lok,[8] Thuy Tien T. Tran,[1] Anthony D. Harris,[9] Vance G. Fowler Jr,[4] Helen Boucher,[10] Barry N. Kreiswirth,[11] Robert A. Bonomo,[12] David van Duin,[13] David L. Paterson,[14] and Henry Chambers[5]

[1]The Innovations in Design, Education, and Analysis Committee of the Biostatistics Center, George Washington Milken Institute School of Public Health; [2]National Institute of Allergy and Infectious Diseases, Bethesda, Maryland; [3]Biogen, Inc., Cambridge, Massachusetts; [4]Duke University, Durham, North Carolina; [5]University of California at San Francisco; [6]Emory University, Atlanta, Georgia; [7]National Cerebral and Cardiovascular Center, Japan; [8]Boston University, Massachusetts; [9]University of Maryland School of Medicine, Baltimore; [10]Tufts Medical Center and Tufts University School of Medicine, Boston, Massachusetts; [11]New Jersey Medical School-Rutgers University, Newark; [12]Case Western Reserve University, Cleveland, Ohio; [13]University of North Carolina, Chapel Hill; and [14]University of Queensland Centre for Clinical Research, Royal Brisbane and Women's Hospital Campus, Australia.

Clinical Patient Management

- Not a single decision
- Dynamic
  - Sequential treatment decisions with tailored (personalized!) adjustments of therapy over time
- Adjustments based on newly available information
  - E.g., AST, early clinical results (e.g., toxicity)

Treatment of Bacterial Infections: Two Major Therapeutic Decisions

- Treat today. Diagnose tomorrow.

- 1st decision: empiric therapy
  - To antibiotic or not to antibiotic?
  - Broad or narrow spectrum?
  - Dual or mono therapy?

- 2nd decision: definitive therapy (48-72 hours later)
  - Keep current therapy or modify?

COMparing Personalized Antibiotic StrategieS

(COMPASS)

Compares decision-making *strategies* consistent with clinical practice rather than specific treatments

Strategies

- A strategy is a *decision-rule guiding patient treatment*
  - Combines empiric and definitive therapy decisions based upon available data at that time
- Strategy ≠ drug(s)
- Distinction between the strategy dictating patient treatment vs. drugs received
  - Patients on the same strategy can have different treatment experiences due to different early responses or AST results

Strategies

- Consider the following strategy for the oral step-down therapy for treatment of cUTI
- Empiric treatment with levofloxacin. For definitive therapy, if AST indicates resistance, then change to alternative. Otherwise continue levofloxacin. This is ONE strategy.
- Suppose Simon and Garfunkel are randomized to this strategy
  - Simon's AST indicates resistance and thus is switched to an alternative
  - Garfunkel's AST indicates susceptibility and thus remains on levofloxacin
- Simon and Garfunkel: different treatment experiences but are part of the same strategy

Tailoring Criterion

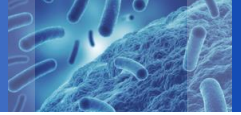
- Here AST is the *tailoring criterion* for directing patient treatment at the definitive stage
- The tailoring criterion can incorporate short-term clinical response, e.g., toxicity requiring therapy adjustment

SMART COMPASS

- If there are multiple definitive therapy options to be investigated, then appropriate trial participants can be re-randomized at the definitive stage
- This allows the opportunity to evaluate which down-stream adjustments would be optimal when we have competing options
- Uses sequential randomization, essentially nesting or embedding sub-trials

SMART COMPASS

- Can address several types of research questions
  - Identification of optimal strategies
  - Evaluate empiric therapies
  - Evaluate definitive therapies (licensure questions)
  - Explore more refined tailoring criterions
- Provides efficiency compared to traditional multi-arm trials
  - Individual patient data can contribute to multiple strategies
- Pragmatic: mirrors clinical decision-making regarding treatment
  - Focus on finding optimal treatment strategies
  - Personalized medicine



PRAGMATIC DIAGNOSTIC STUDIES

Motivating Questions

Why do we set separate goals for sensitivity and specificity when the acceptable level for each depends on the other?

Suppose there is a choice between 2 diagnostics:
one with a higher sensitivity and one with a higher specificity.

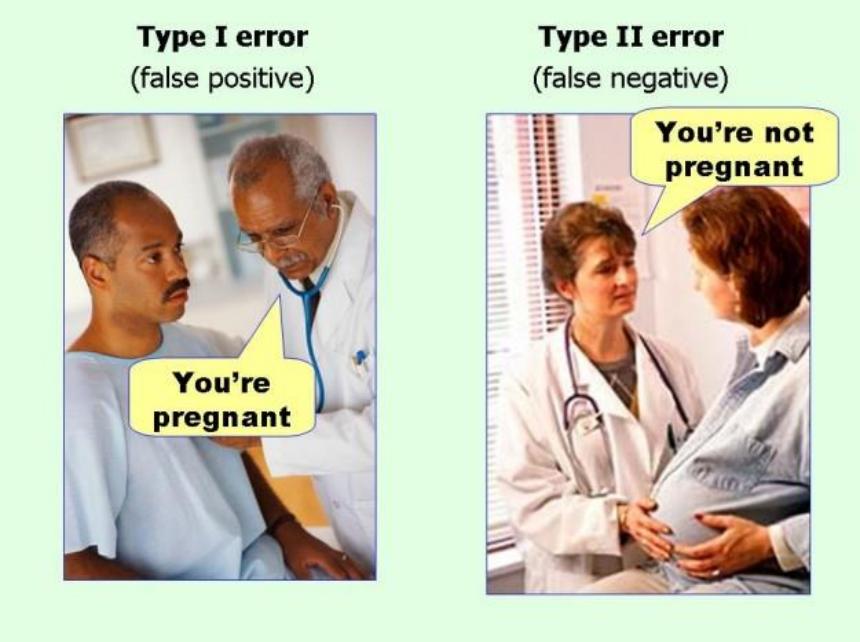
Which test should be selected to optimize clinical outcomes?

Accuracy

- Accuracy = total percent correctly classified
- Two challenges with interpretation
    1. Accuracy treats all errors as if they are equally important.

Accuracy

- Accuracy = total percent correctly classified
- Two challenges with interpretation
  1. Accuracy treats all errors as if they are equally important.



Accuracy

- Accuracy = total percent correctly classified

- Two challenges with interpretation
  1. Accuracy treats all errors as if they are equally important.
  2. It depends on prevalence. Thus accuracy is not generally comparable from study to study, as prevalence rates may differ between studies.

BED-FRAME

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

Taylor & Francis
Taylor & Francis Group

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.

Diagnostic Yield

- The distribution of true positive (TP), true negative (TN), false positive (FP), and false negative (FN) results

- Basis for evaluation considers
  - The magnitude of sensitivity and specificity
  - Prevalence of disease in a given region and time-frame
  - The *relative importance* of false positive vs. negative errors

To appear in CID 2019

Average Weighted Accuracy (AWA): Pragmatic Analysis for a 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), Scott R. Evans (4)

1: Biogen, Inc.; 2: Center for Applied Genomics and Precision Medicine, Department of Medicine, Duke University; 3: Emergency Department Service, Durham VA Health Care System; 4: Biostatistics Center, George Washington Milken Institute School of Public Health; 5: Medicine Service, Durham VA Health Care System

Two Renown Doctors

What people think of as the discovery is really discovery of the question.

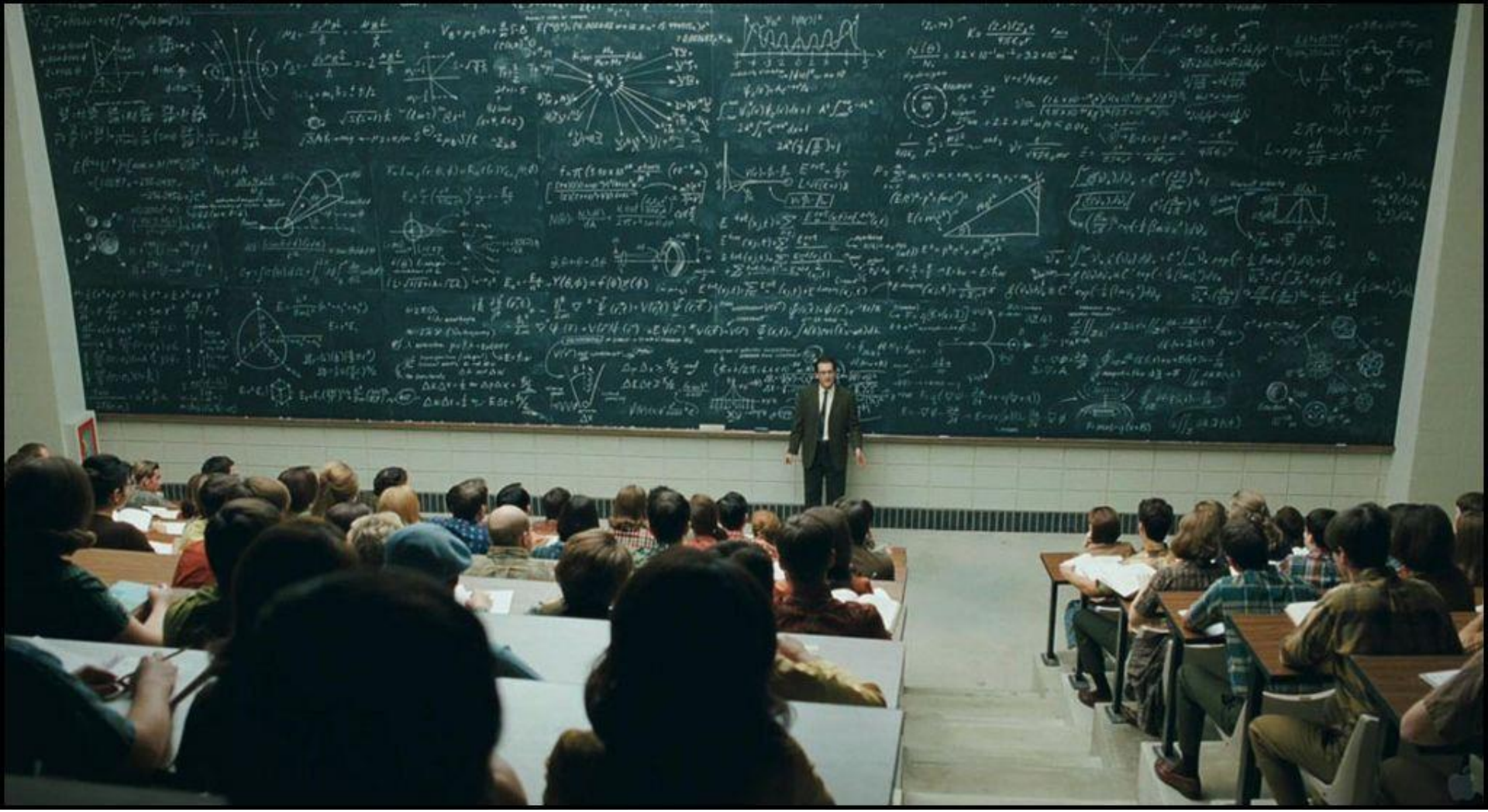
Jonas Salk

Sometimes the questions are complicated and the answers are simple.

Dr. Seuss

Significant Contributors (p<0.001)

- Dean Follmann
- Dan Rubin
- Chip Chambers
- David van Duin
- Gene Pennello
- The Antibacterial Resistance Leadership Group
- The SOCRATES Steering Committee



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

Thank you.