Radical Pragmatism:
DOOR and SMART COMPASS for the Evaluation of Antibiotics

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

FDA
November 19, 2019
A Leaky Roof...

- Created a water bubble in my wall

- In addition to a new roof, I had to re-paper the wall

- My neighbor recently papered a similar-sized room in his house. I asked:

  “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 Antibiotic Clinical Trials

1. They are rigorously conducted by experts closely adhering to the highest standards and fundamental principles of RCTs

2. They are essentially useless for helping clinicians make treatment 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
Example Issues

- Drugs are compared in susceptible disease, but susceptibility status is unknown at the time of treatment initiation

- Patients are considered failures when they change therapy, though they may not fail

- We lose interest in patients that change therapy, despite therapeutic adjustments that can effectively treat the patient

- Population studied ≠ population applied
  - E.g., noninferiority trials exclude patients with recent prior therapy. Then these drugs are used in these patients, possibly representing a majority.
Question 1

- We define analysis populations
  - Efficacy: ITT population
  - Safety: safety population

- Efficacy population ≠ safety population

- We combine these two analyses into benefit:risk analyses

- To whom does this analysis apply?
Question 2

- We measure the duration of hospitalization
- Shorter duration is better … or is it?
- The faster the patient dies, the shorter the duration
- Outcome interpretation needs context of other outcomes
Question 3

- Trials typically use binary endpoints

- E.g., “cure”: patient survives, symptoms resolve, microbiological eradication, no changes to therapy

- Consider the following:
  - One patient fails because they die
  - Another patient fails because of lack of micro eradication
  - Primary analyses treats these patients equivalently (failure)

- Shouldn’t primary analysis recognize the difference?
## FDA Advisory Committee Evaluation of Plazomicin in cUTI

<table>
<thead>
<tr>
<th>Drug</th>
<th>Composite Cure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plazomicin</td>
<td>81.7%</td>
</tr>
<tr>
<td>Meropenem</td>
<td>70.1%</td>
</tr>
</tbody>
</table>
## FDA Advisory Committee Evaluation of Plazomicin in cUTI

<table>
<thead>
<tr>
<th>Drug</th>
<th>Composite Cure</th>
<th>Clinical Cure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plazomicin</td>
<td>81.7%</td>
<td>89.0%</td>
</tr>
<tr>
<td>Meropenem</td>
<td>70.1%</td>
<td>90.4%</td>
</tr>
</tbody>
</table>
## FDA Advisory Committee Evaluation of Plazomicin in cUTI

<table>
<thead>
<tr>
<th>Drug</th>
<th>Composite Cure</th>
<th>Clinical Cure</th>
<th>Micro Eradication</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plazomicin</td>
<td>81.7%</td>
<td>89.0%</td>
<td>89.5%</td>
</tr>
<tr>
<td>Meropenem</td>
<td>70.1%</td>
<td>90.4%</td>
<td>74.6%</td>
</tr>
</tbody>
</table>
### FDA Advisory Committee Evaluation of Plazomicin in cUTI

<table>
<thead>
<tr>
<th>Drug</th>
<th>Composite Cure</th>
<th>Clinical Cure</th>
<th>Micro Eradication</th>
<th>1-level Decrease in Creatinine Clearance</th>
<th>Last Serum Creatinine Increased ≥ 0.5 mg/dL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plazomicin</td>
<td>81.7%</td>
<td>89.0%</td>
<td>89.5%</td>
<td>13.7%</td>
<td>3%</td>
</tr>
<tr>
<td>Meropenem</td>
<td>70.1%</td>
<td>90.4%</td>
<td>74.6%</td>
<td>5.7%</td>
<td>1%</td>
</tr>
</tbody>
</table>
$1 + 2 \times 3 \neq 9$

Children in grade school have learned this.

Clinical trialists missed this class.
Question 4

- 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

<table>
<thead>
<tr>
<th></th>
<th>A (N=100)</th>
<th>B (N=100)</th>
<th>C (N=100)</th>
</tr>
</thead>
</table>

# RCT Comparing A, B, and C

## Analysis of Outcomes

<table>
<thead>
<tr>
<th></th>
<th>A (N=100)</th>
<th>B (N=100)</th>
<th>C (N=100)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Success</td>
<td>50%</td>
<td>50%</td>
<td>50%</td>
</tr>
<tr>
<td></td>
<td>A (N=100)</td>
<td>B (N=100)</td>
<td>C (N=100)</td>
</tr>
<tr>
<td>---</td>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
</tr>
<tr>
<td>Success</td>
<td>50%</td>
<td>50%</td>
<td>50%</td>
</tr>
<tr>
<td>Safety event</td>
<td>30%</td>
<td>50%</td>
<td>50%</td>
</tr>
</tbody>
</table>
## RCT Comparing A, B, and C

### Analysis of Outcomes

<table>
<thead>
<tr>
<th></th>
<th>A (N=100)</th>
<th>B (N=100)</th>
<th>C (N=100)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Success: 50%</td>
<td>Success: 50%</td>
<td>Success: 50%</td>
</tr>
<tr>
<td></td>
<td>Safety event: 30%</td>
<td>Safety event: 50%</td>
<td>Safety event: 50%</td>
</tr>
</tbody>
</table>

**Which treatment would you choose?**
## RCT Comparing A, B, and C

### Analysis of Outcomes

<table>
<thead>
<tr>
<th></th>
<th>A (N=100)</th>
<th>B (N=100)</th>
<th>C (N=100)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Success: 50%</td>
<td>Success: 50%</td>
<td>Success: 50%</td>
</tr>
<tr>
<td></td>
<td>Safety event: 30%</td>
<td>Safety event: 50%</td>
<td>Safety event: 50%</td>
</tr>
</tbody>
</table>

Which treatment would you choose?

They all have the same success rate.
RCT Comparing A, B, and C
Analysis of Outcomes

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

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

C (N=100)
Success: 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**

<table>
<thead>
<tr>
<th></th>
<th>A (N=100)</th>
<th>B (N=100)</th>
<th>C (N=100)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Success: 50%</td>
<td>Success: 50%</td>
<td>Success: 50%</td>
</tr>
<tr>
<td></td>
<td>Safety event: 30%</td>
<td>Safety event: 50%</td>
<td>Safety event: 50%</td>
</tr>
</tbody>
</table>

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)
Success: 50%
Safety event: 30%

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

C (N=100)
Success: 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

<table>
<thead>
<tr>
<th></th>
<th>A (N=100)</th>
<th>B (N=100)</th>
<th>C (N=100)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Success: 50%</td>
<td>Success: 50%</td>
<td>Success: 50%</td>
</tr>
<tr>
<td></td>
<td>Safety event: 30%</td>
<td>Safety event: 50%</td>
<td>Safety event: 50%</td>
</tr>
<tr>
<td>SE</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>15</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>35</td>
<td>35</td>
<td>0</td>
</tr>
</tbody>
</table>
## Analysis of Patients: 4 Possible Outcomes

### A (N=100)
- **Success:** 50%
- **Safety event:** 30%

<table>
<thead>
<tr>
<th>SE</th>
<th>Success+</th>
<th>Success-</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>-</td>
<td>35</td>
<td>35</td>
</tr>
</tbody>
</table>

### B (N=100)
- **Success:** 50%
- **Safety event:** 50%

<table>
<thead>
<tr>
<th>SE</th>
<th>Success+</th>
<th>Success-</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>50</td>
<td>0</td>
</tr>
<tr>
<td>-</td>
<td>0</td>
<td>50</td>
</tr>
</tbody>
</table>

### C (N=100)
- **Success:** 50%
- **Safety event:** 50%

<table>
<thead>
<tr>
<th>SE</th>
<th>Success+</th>
<th>Success-</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>0</td>
<td>50</td>
</tr>
<tr>
<td>-</td>
<td>50</td>
<td>0</td>
</tr>
</tbody>
</table>
## Analysis of Patients: 4 Possible Outcomes

<table>
<thead>
<tr>
<th></th>
<th>Success (+)</th>
<th>Success (-)</th>
<th>SE (+)</th>
<th>SE (-)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A (N=100)</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>35</td>
</tr>
<tr>
<td>Success: 50%</td>
<td>Safety event: 30%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B (N=100)</td>
<td>50</td>
<td>0</td>
<td>0</td>
<td>50</td>
</tr>
<tr>
<td>Success: 50%</td>
<td>Safety event: 50%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C (N=100)</td>
<td>0</td>
<td>50</td>
<td>50</td>
<td>0</td>
</tr>
<tr>
<td>Success: 50%</td>
<td>Safety event: 50%</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Analysis of Patients: 4 Possible Outcomes

<table>
<thead>
<tr>
<th></th>
<th>Success</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Success (A: N=100)</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>Success (B: N=100)</td>
<td>50</td>
<td>0</td>
</tr>
<tr>
<td>Success (C: N=100)</td>
<td>0</td>
<td>50</td>
</tr>
</tbody>
</table>

- **A (N=100)**
  - Success: 50%
  - Safety event: 30%

- **B (N=100)**
  - Success: 50%
  - Safety event: 50%

- **C (N=100)**
  - Success: 50%
  - Safety event: 50%
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…”
“The good physician treats the disease.
The great physician treats the patient.”

William Osler

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

Scott R. Evans\textsuperscript{a,b} and Dean Follmann\textsuperscript{c}

\textsuperscript{a}Department of Biostatistics, Harvard University, Boston, MA, USA; \textsuperscript{b}Center for Biostatistics in AIDS Research, Harvard University, Boston, MA, USA; \textsuperscript{c}National Institute of Allergy and Infectious Diseases (NIAID) of the National Institutes of Health (NIH), Bethesda, MD, USA.
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. Chambers13

DOOR probability: probability of a more desirable global outcome when assigned to the new vs. the control treatment
Should we use ceftazidime-avibactam or colistin for the initial treatment of CRE infection?
DOOR

- DOOR: 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

<table>
<thead>
<tr>
<th></th>
<th>Colistin (N=46)</th>
<th>Caz-Avi (N=26)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Discharged home</strong></td>
<td>4 (9%)</td>
<td>6 (23%)</td>
</tr>
<tr>
<td><strong>Alive; not discharged home; no renal failure</strong></td>
<td>25 (54%)</td>
<td>17 (65%)</td>
</tr>
<tr>
<td><strong>Alive; not discharged home; renal failure</strong></td>
<td>5 (11%)</td>
<td>1 (4%)</td>
</tr>
<tr>
<td><strong>Death</strong></td>
<td>12 (26%)</td>
<td>2 (8%)</td>
</tr>
</tbody>
</table>

**DOOR Probability:** 64% (53%, 75%)
Summarizing the “Patient Journey”

Before we analyze several hundred patients, we must understand how to analyze one.

An example strategy …

Good Studies Evaluate the Disease While Great Studies Evaluate the Patient: Development and Application of a DOOR Endpoint for *Staphylococcus aureus* Bloodstream Infection
BAC DOOR

- Pre-trial sub-study to develop DOOR in *S. 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 profiles by *desirability of outcome*

- Examined components 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

<table>
<thead>
<tr>
<th></th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discharged home</td>
<td>100</td>
</tr>
<tr>
<td>Alive; not discharged home; no renal failure</td>
<td>Partial credit</td>
</tr>
<tr>
<td>Alive; not discharged home; renal failure</td>
<td>Partial credit</td>
</tr>
<tr>
<td>Death</td>
<td>0</td>
</tr>
</tbody>
</table>
Tailoring Medicine
Who Benefits from Caz-Avi?

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
<table>
<thead>
<tr>
<th>Better outcome</th>
<th>Worse outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Death</td>
</tr>
<tr>
<td>Treatment success without AKI</td>
<td></td>
</tr>
<tr>
<td>Treatment success with AKI</td>
<td></td>
</tr>
<tr>
<td>Treatment failure (persistent bacteremia) without AKI</td>
<td></td>
</tr>
<tr>
<td>Treatment failure with AKI</td>
<td></td>
</tr>
</tbody>
</table>
Higher doses bring toxicity but not greater treatment success.
Presenting Risks and Benefits: Helping the Data Monitoring Committee Do Its Job

Scott R. Evans, PhD; Robert Bigelow, PhD; Christy Chuang-Stein, PhD; Susan S. Ellenberg, PhD; Paul Gallo, PhD; Weili He, PhD; Qi Jiang, PhD; and Frank Rockhold, PhD

Data monitoring committees (DMCs), or data and safety monitoring boards, protect clinical trial participants by conducting benefit-risk assessments during the course of a clinical trial. These evaluations may be improved by broader access to data and more effective analyses and presentation. Data monitoring committees should have access to all data, including efficacy data, at each interim review. The DMC reports should include graphical presentations that summarize benefits and harms, and propose that such summaries become standard in DMC reports.

For author affiliations, see end of text.

This article was published at Annals.org on 19 November 2019.
Development of standardized syndrome-specific DOORs
- ABSSSI
- CABP
- HABP/VABP
- cIAI
- cUTI
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 Tollmann,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,9 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 Chambers5

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

- Addresses several types of research questions
  - Identifies optimal strategies
  - Evaluates empiric therapies
  - Evaluates definitive therapies (licensure questions)

- Provides efficiencies compared to traditional multi-arm trials

- Pragmatic: mirrors clinical decision-making
  - Personalized medicine
NBA Coach Frank Layden

Had a player that was not producing.

Layden asked the player:

“Son, what is it with you? Is it ignorance or apathy?”

The player looked at Layden and said:

“Coach, I don't know and I don't care.”
If people don’t know, then let’s educate them.

If they don’t care, then let’s motivate them.
Significant Contributors (p<0.001)

- Dean Follmann
- Dan Rubin
- Chip Chambers
- Vance Fowler
- 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.