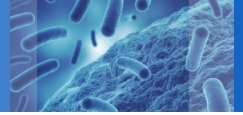
ARLG
Antibacterial Resistance Leadership Group

Response Adjusted for Duration of Antibiotic Risk (RADAR)

Scott Evans, Ph.D., M.S.
Harvard University

CTTI Statistical Think Tank Expert Meeting
November 19, 2014



Special Thank You

- Kunal Merchant
- Dan Rubin
- CTTI
- FDA

Acknowledgements

People making valuable contributions to this work (p<0.001)

Dan Rubin, FDA

Dean Follman, NIH

Gene Penello, FDA

David Schoenfeld, Harvard

John Powers, GWU, NIH

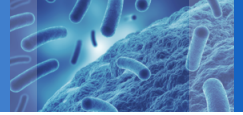
Chip Chambers, UCSF

Vance Fowler, Duke

Ebb Lautenbach, UPENN

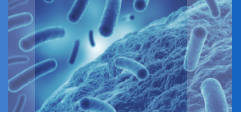
Charles Huskins, Mayo Clinic

Research reported in this publication was supported by the National Institute of Allergy And Infectious Diseases of the National Institutes of Health under Award Number UM1AI104681. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.



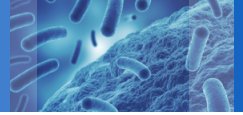
Motivation

- Interest in improving benefit:risk evaluation strategies
- Flawed methodologies for antibiotic stewardship trials



A Simple Test: Question 1

- Setting: analyses of a clinical trial
- Efficacy population = ITT or mITT
- Safety population = those with e.g., > 1 dose of drug
- Efficacy population ≠ Safety population
- What is the “benefit:risk population”?



A Simple Test: Question 2

- Suppose the person that you care about most in the world, has just been diagnosed with a terrible infection
- 3 treatment options: A, B, and C
- Both treatment efficacy (i.e., the benefit) and toxicity (i.e., harm) are binary (w/ similar importance)

Recently Completed RCT Comparing A, B, and C

A (N=100)	B (N=100)	C (N=100)
Benefit: 50%	Benefit: 50%	Benefit: 50%
Toxicity: 20%	Toxicity: 50%	Toxicity: 50%

Which treatment would you choose?

The answer of course as any reasonable researcher would tell you is…

Recently Completed RCT Comparing A, B, and C

A (N=100)	B (N=100)	C (N=100)
Benefit: 50%	Benefit: 50%	Benefit: 50%
Toxicity: 20%	Toxicity: 50%	Toxicity: 50%

Which treatment would you choose?

The answer of course as any reasonable researcher would tell you is…

C

Analysis of Patients: 4 Possible Outcomes

A (N=100)

Benefit: 50%

Toxicity: 20%

		Benefit +	Benefit -
Tox	+	10	10
	-	40	40

B (N=100)

Benefit: 50%

Toxicity: 50%

		Benefit +	Benefit -
Tox	+	50	0
	-	0	50

C (N=100)

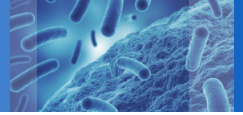
Benefit: 50%

Toxicity: 50%

		Benefit +	Benefit -
Tox	+	0	50
	-	50	0

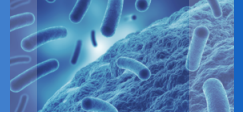
Rate of saving your loved one (benefit without toxicity)

40% **0%** **50%**



Our culture is to collect data on patients to analyze the endpoints.

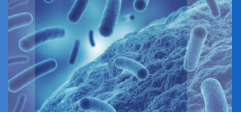
Shouldn't we use endpoint data to analyze the patients?



Good News!

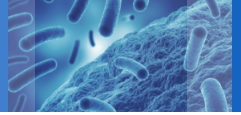
- It turns out that if we analyze patients rather than endpoints, then many of our statistical problems are greatly lessened

- We gain:
  - More informative benefit:risk evaluation
  - Patient-level interpretation (these are the patient outcomes)
  - Alleviation from competing risk problems
  - Clarity with respect to the research questions
  - And more…



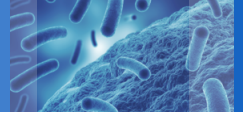
Stewardship Trials

- Thus there is considerable interest in evaluating (stewardship) strategies to see if they result in less antibiotic use but w/o compromising clinical outcomes
- Stewardship trials often compare a new strategy of antibiotic use vs. a standard (control) strategy with respect to clinical outcomes and antibiotic use
- Many issues in these trials
- Current designs do not adequately address the issues



Issues in Stewardship Trials

- Benefit:risk
- Question the question
- Noninferiority design issues
- Competing risks
- Standardization / correction



Noninferiority (NI) Complexities

- Many stewardship trials utilize NI trial designs
- Lower scientific integrity than superiority trials as they are more prone to biases and manipulation
  - Antibiotics have characteristics that exacerbate concerns
- Potentially large and impractical sample sizes that jeopardize feasibility and strain resources
- The validity of NI trials relies upon several foundational requirements during design, conduct, analyses, and reporting
- Avoid NI when possible

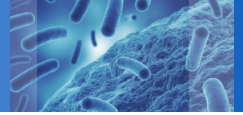
NI Complexities

- Assay sensitivity: reduced (intentionally or unintentionally) by diluting effects through subtle choices in design and conduct
  - E.g., inclusion of subgroups where treatment effects may be small (e.g., participants with skin abscesses in skin infection trials, where placebo-controlled trials have demonstrated minimal drug effects vs. drainage)
- Constancy assumption in doubt in a setting of evolving resistance and thus decreasing effectiveness of antibiotics
- Analysis issues
  - ITT jeopardized by assay sensitivity issues
  - PP vulnerable to all biases of observational studies

NI Complexities: Historical Data

- Often reliable historical data to justify the NI margin does not exist or is no longer applicable due to the evolution of medical practice or the development of resistance
  - Many studies justify selections based on studies from the pre-antibiotic era (1930s - 1950s), often non-randomized
  - But conditions (e.g., the availability / quality of supportive care) and populations have changed

- Biocreep concern
  - From 2002-2009, 43 NME approval packages submitted to FDA with about half for antimicrobials



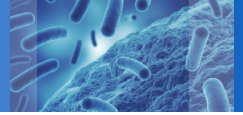
NI Complexities: Ethical Dilemmas

- Null hypothesis is inferiority (assumed to be true)
  - Is this equipoise?
  - Are patients told this in informed consent?

- Why will patients volunteer to risk being randomized to a strategy that might be as good (but unproven as of yet) as a proven existing medical alternative but is not hypothesized to be better?
  - Why not simply opt for the proven alternative?

NI Complexities: Ethical Dilemmas

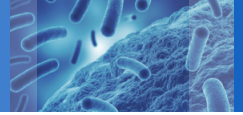
- Concerningly little attention is paid to M2 (clinical acceptability) when defining a margin
  - Most margins are selected based on a preservation of a fraction of the effect criteria (e.g., via meta-analyses of prior trials) as well as cost and feasibility issues associated with sample size, despite ICH-E10 recommending otherwise
- Data regarding what is “clinically acceptable” is often lacking
  - Even when a margin that would ensure effect retention can be identified, the selected margin is often larger than what is acceptable and thus unconvincing to the medical community
- Should we be surveying for this information?



NI Complexities: Troubling Dichotomy

- Troubling dichotomy in the way in which differences between interventions are interpreted in noninferiority vs. superiority studies

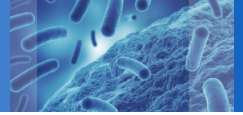
- A typical NI margin is 10% (risk difference)
  - This should mean that inferiority of < 10% is clinically acceptable
  - Some argue that this is too stringent
  - But ask a patient / clinician if they are willing to take a new therapy that may be up to 10% worse than the standard, many will decline
  - Bu had a superiority trial been conducted and showed a 9% improvement, it is unlikely that this difference would be dismissed as clinically irrelevant… such differences would likely be claimed as an advantage



Question the Question

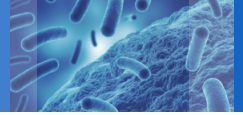
If a new stewardship strategy is not better than the existing strategy, then what is it's value?

We need superior strategies.



Question the Question: Another Example

- Colistin
  - Last resort antibiotic
  - Nephrotoxicity
- NI to colistin?
- When considering all information (including toxicity and QOL), we want to know if an alternative is better than colistin
  - Figure out how to construct this evaluation



Competing Risks

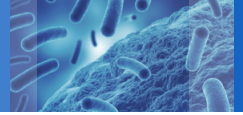
- Common endpoints are distorted / challenging to interpret
  - Days in the hospital
  - Days in the ICU
  - Days of antibiotic use
- Fewer days is interpreted as a better outcome
- Really? A sepsis trial may expect 30% mortality. The faster they die, the fewer days…
- Without clinical context of other outcomes (e.g., survival) for the same patient, interpretation of these endpoints is challenging

Standardizing / Correcting Outcomes Using Variables that Can also be Affected by the Intervention

- (Days of antibiotic use / days in the hospital)… lower is better … or is it?
- What if the effect is to increase the denominator?
- Which would you prefer?
  - Case 1: hospitalization for 10 days with 5 days of antibiotics (50%)
  - Case 2: hospitalization for 15 days with 5 days of antibiotics (33%)
  - Clinically Case #1 is preferred (fewer hospital days)
- Which would you prefer?
  - Case 1: 4 days in the hospital with 2 days of antibiotics
  - Case 2: 20 days in the hospital with 10 days of antibiotics
  - Both have 50% ratio

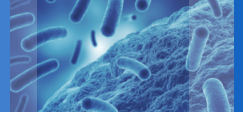
Scientific Rationale?

- "This is the way that we've always done it before…"
- "My advisor did it this way"
- "There's unmet medical need"
- "The FDA said it was okay"

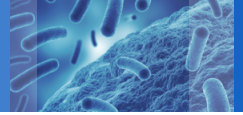


We cannot solve problems using the same thinking that we used to create them.

Albert Einstein



Response Adjusted for Days of Antibiotic Risk (RADAR)



RADAR: Conceptual Framework

- Desire to know if new strategies are BETTER than the standard strategies when we consider the interventions in their TOTALITY
  - Considering all important clinical outcomes (benefits, harms, QOL) and antibiotic use

- The question becomes how to logically put together the important outcomes

Step 1: Generic Examples of ALL_OUT

3 Levels

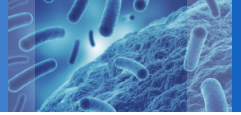
- Survive without toxicity
- Survive with toxicity
- Death

5 Levels

- Benefit w/o toxicity
- Benefit w/ toxicity
- Survive, no benefit w/o toxicity
- Survive, no benefit w toxicity
- Death

Step 2: Desirability Of Outcome Ranking (DOOR)

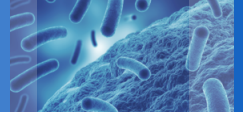
- All trial patients receive DOOR
- DOOR is constructed using 2 rules:
  1. When comparing 2 patients with different clinical outcomes
     - The patient with the better clinical outcome receives a higher rank
  2. When comparing 2 patients with the same clinical outcome
     - The patient with a shorter duration of antibiotic use receives a higher rank
- DOOR is consistent with “reduce use w/o clinical compromise”



STEP 3: Evaluate Superiority of DOOR

- Estimation (using confidence intervals)
  - Probability that a randomly selected patient will have a better DOOR if assigned the new strategy relative to the control

- Hypothesis Testing
  - Null: the probability that a patient assigned to the new strategy will have a better DOOR than if assigned to the control is 50%
  - Alternative: the probability that a patient assigned to the new strategy will have a better DOOR than if assigned to the control is X % (where X is greater than 50%).
  - Sample size for 90% power using a 2-sided alpha=0.05 Wilcoxon Mann-Whitney test
    - If p=60%, then N=360 (180 per arm)
  -

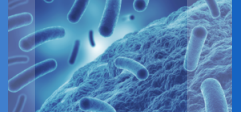


Example: ARLG SCOUT-CAPG

- RCT comparing 5-day vs. standard 10-day course of outpatient antibiotics in children with community-acquired pneumonia (CAP)

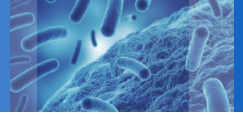
- Original design
  - Debate over appropriate NI margin
  - Questionable feasibility w/ N=800 required for 90% power

- RADAR design
  - Superiority trial (avoiding NI)
  - N=360 (>50% reduction in the required N)



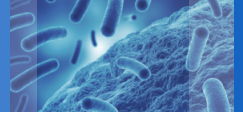
Analyses

- DOOR is a composite endpoint
  - Fundamental to also analyze each component too
- Evaluate ordinal clinical outcome
  - Cumulative difference plot with confidence bands
  - Tests for ordinal outcome: M-H chi-square
  - Components of ordinal outcome
- Sensitivity analyses developing



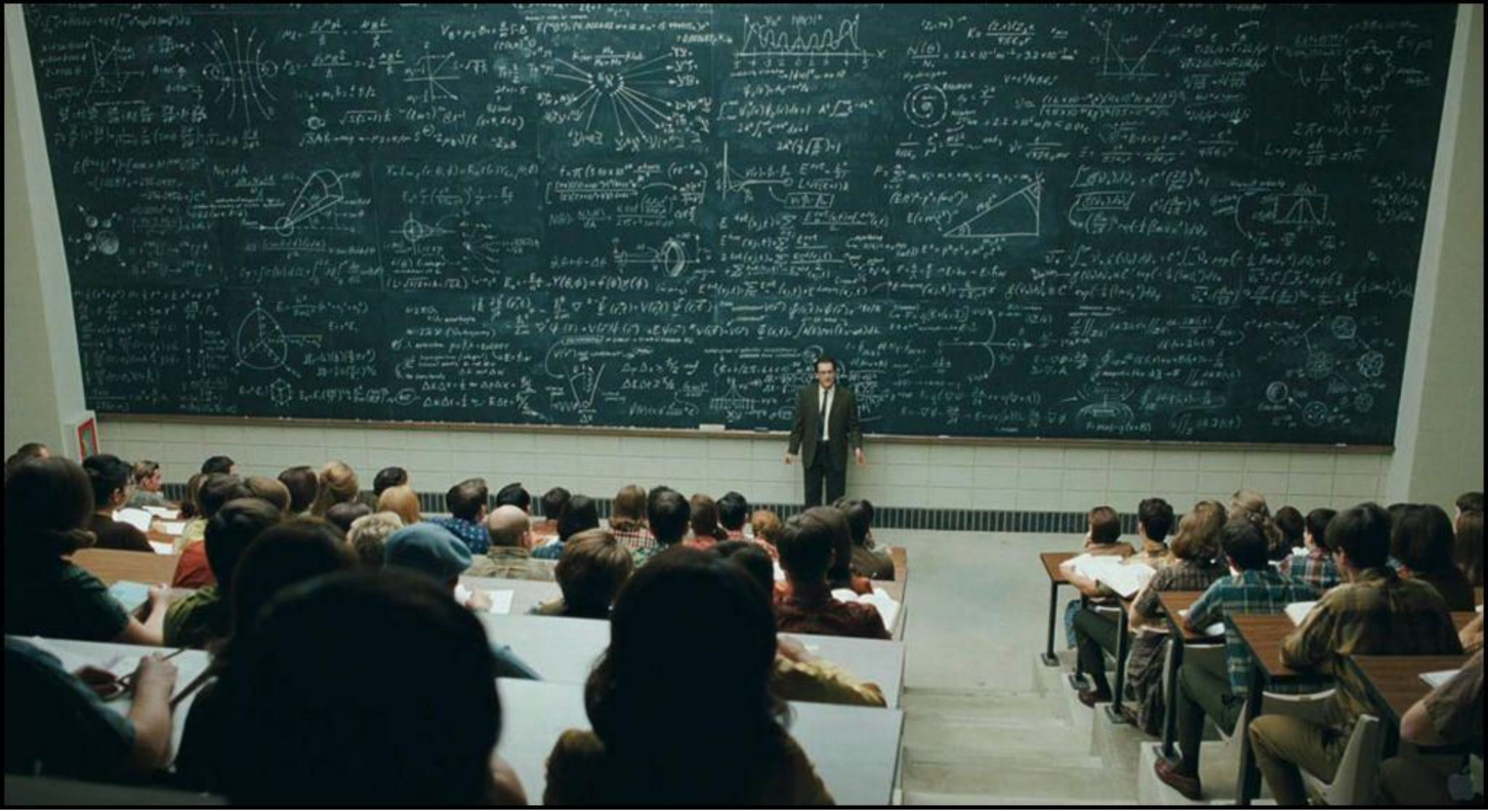
RADAR Advantages

- Superiority design; avoids NI complexities
- Reduction of sample size in many cases
- Alleviates competing risk problems
- More informative benefit:risk analyses
- Patient-level interpretation
- Collaboration between academic, NIH, and regulator researchers



RADAR Challenges

- Culture change
- Creating an ordinal category can be challenging
- Concern that drop in clinical outcome would be trumped by improvement in antibiotic use
  - Evaluate with tipping point analyses and sensitivity analyses
- We avoid weighting categories by using a ranking strategy
  - But ranking equates to weighting
- Ranking is not transitive



I am looking for a post-doc to work on these elementary ideas …
please let me know if you know of good candidates.

Also, SCID is coming to a library near you.

Thank you for your kind attention.