A Statistician is ________.

- To data what a doctor is to a patient
- The wizard behind the curtain
- An oasis in the desert
- Somebody who is wrong 5% of the time
- Either a pain in the behind or an unveiler of grand secrets
- My best friend when writing a grant proposal
- Someone who answers my questions with more questions
- An unlikely bedfellow that destroys dreams
A Statistician is __________.

- The gateway to understanding
- A scientist who is able to transform data into knowledge
- My best stormy whether friend
- An angel of god
- A friend for life if you can afford the FTE
- The rate-limiting step in manuscript writing
- A person that makes me wish that I knew more statistics
- A person that makes me wish that I knew less statistics
- Wicked smart
“Clinical research has drifted from its early public health orientation … toward RCTs as a business…

…trial methodologies, including statistical methods, QC standards, and data monitoring and analysis procedures, are now largely shaped by imperatives to develop new approved products (or increase sales of existing products) while meeting regulatory requirements.”

DeMets and Califf, JAMA 2011
SWOT Analyses: Clinical Trials

- **Strengths**
  - Randomization (the foundation for statistical inference)
  - Blinding
  - Control groups
  - Prospective observation
  - ITT (protects the benefits of randomization and provides pragmatic analyses)

- **Weaknesses**
  - Expensive
  - Time-consuming
SWOT Analyses: Clinical Trials

- **Opportunities**
  - Pragmatism: more relevant answers for clinical practice

- **Threats**
  - Innate desire to do things faster and cheaper, magnified by today’s business and political pressures
  - Though understandable, such desires can be dangerous threatening our objectivity and ability to reason, resulting in studies with lower integrity, reproducibility, and applicability
  - Susceptibility to sales pitches for approaches labelled as “innovations” that effectively lower the evidentiary standard and introduce greater uncertainty
An Objective Objective

- Typical example of trial objective:
  “To demonstrate that treatment A is superior to treatment B.”

- Incorrect

- The goal of the trial is to get the right answer, a fair contrast between A vs. B.
  - The marketing objective / company goal is to show A is better than B.

- We should at least be objective about the objective
Negative Trial?
Must be something wrong with the trial.

“The greatest obstacle to discovery is not ignorance, it is the illusion of knowledge.”
Daniel Boorstin

“It is not what the man of science believes that distinguishes him, but how and why he believes it. His beliefs are tentative, not dogmatic; they are based on evidence, not on authority or intuition.”
Bertrand Russell
Modern “Innovations”:

Progress ... or regress?
Innovations = Compromise in Rigor?

- Non-randomized rather than randomized evidence rationalized by the increasing access to real world data and the belief that modeling can replace randomization
- Surrogate outcomes
- Surrogate diseases
- PP analyses instead of ITT
- Uncontrolled studies
- Unblinded studies
- Assuming treatment effects rather than collecting data to estimate those effects
- Adaptive designs that promote efficiencies but are inefficient and threaten integrity
A Troubling Trend

Figure 1. Number of NIH-funded, registered trials by phase and year of registration.
Closed-minded refusal to use real world data (RWD) would be an act of foolishness ... 

...foolishness only surpassed by using RWD to subvert randomization, the foundation for statistical inference.
Randomization: The Most Powerful Tool in Clinical Trials

- Foundation for statistical inference (with ITT)
  - Intervention assignment is independent of outcome risk

- Expectation of balance between groups with respect to
  - known factors
  - UNKNOWN factors
    - Protects us from our own ignorance and knowledge limitations
    - Factors that cannot be measured (and thus cannot be controlled)

- Eliminates many biases / confounding that plague observational studies and the need for untestable assumptions
  - E.g., confounding by indication from physician/patient selection

- Now treated as a luxury rather than foundation
The Story of Patulin

- Patulin is a compound from mold *Penicillium patulinum*

- Studied as a potential treatment for the common cold in an early non-randomized, double-blinded concurrently-controlled clinical trial
  - N=180

- Improvement at 48 hours
  - Patulin in buffer = 55/95 (58%)
  - Buffer alone = 8/85 (9.4%)
  - Difference = 48%; CI=(35%, 60%); p <0.002
The Story of Patulin

- A randomized double-blind trial was then conducted in 1449 factory and postal workers

- Cured at 48 hours:
  - Patulin in buffer = 87/668 (13%)
  - Buffer solution alone = 88/680 (13%)
    - Difference = 0%; 95% CI (-3.6%, 3.8%); p = 0.96
Comparison of evidence on harms of medical interventions in randomized and nonrandomized studies

Panagiotis N. Papanikolaou, Georgia D. Christidi, John P.A. Ioannidis

Interpretation: Nonrandomized studies are often conservative in estimating absolute risks of harms. It would be useful to compare and scrutinize the evidence on harms obtained from both randomized and nonrandomized studies.

CMAJ 2006;174(5):635-41

Agreement of treatment effects for mortality from routinely collected data and subsequent randomized trials: meta-epidemiological survey

Lars G Hemkens,1,2 Despina G Contopoulos-Ioannidis,3,4 John P A Ioannidis1,4,6

Cite this as: BMJ 2016;352:i493
http://dx.doi.org/10.1136/bmj.i493

RCD studies systematically and substantially overestimate mortality benefits of medical treatments compared with subsequent trials investigating the same question
Comparing Adherers to Non-adherers

In the Coronary Drug Project (CDP), patients randomized to Clofibrate and Placebo were stratified according to adherence:

<table>
<thead>
<tr>
<th>Clofibrate Group</th>
<th>Adherers</th>
<th>Nonadherers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Died</td>
<td>106 (15%)</td>
<td>88 (25%)</td>
</tr>
<tr>
<td>Survived</td>
<td>602</td>
<td>269</td>
</tr>
</tbody>
</table>

Relative risk = 1.39

Does this imply a positive effect of Clofibrate?

Let’s look at the placebo group…
Placebo Group

<table>
<thead>
<tr>
<th></th>
<th>Adherers</th>
<th>Nonadherers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Died</td>
<td>274 (15%)</td>
<td>249 (28%)</td>
</tr>
<tr>
<td>Survived</td>
<td>1539</td>
<td>633</td>
</tr>
</tbody>
</table>

Relative risk = 1.87

Nonadherence predicts poor outcome
Clinically Meaningful Endpoint

A direct measure of how a patient “functions, feels or survives”
Surrogate Endpoint

- A measure that is predictive of clinical outcome but takes a shorter time to observe or is less expensive or invasive.
Validation

- Correlation does not imply surrogacy

- Results in the same conclusions if the clinical endpoint was used
  - If it is more sensitive, then it is not a surrogate!

  - Intervention affects the surrogate
  - Intervention affects the clinical endpoint
  - The association between the surrogate and the clinical endpoint is independent of intervention
  - The null hypothesis for the clinical endpoints implies the null hypothesis for the surrogate
Avastin and Breast Cancer

- In 2007, the NEJM published an open-label ECOG study comparing paclitaxel to paclitaxel plus avastin for first-line treatment of metastatic breast cancer.

- The avastin arm had prolonged progression-free survival (PFS) (11.8 vs. 5.9 mos., HR = 0.60, P < 0.001).

- Median survival was similar (26.7 vs. 25.2 mos.).

- No differences seen in quality of life.

- After considerable discussion with their advisory committee, the FDA granted accelerated approval to Avastin for this indication.
Avastin and Breast Cancer

- With accelerated approval, Genentech was required to conduct additional studies.

- In July 2010, the FDA Advisory Committee reviewed two additional studies, AVADO and RIBBON-1.

- Neither study showed large differences in PFS, overall survival was not improved, and the Avastin group experienced significantly more severe adverse events.

- In December 2010, the FDA withdrew its approval of Avastin for treatment of metastatic breast cancer.
More Questions

- 18 of the 36 cancer drugs that were approved by the FDA from 2008 to 2012 on the basis of a surrogate endpoint, typically tumor shrinkage or PFS. Post-marketing studies did not indicate a survival benefit.
  
BELLINI

- Double-blind, randomised, placebo-controlled trial comparing venetoclax, bortezomib, and dexamethasone vs. placebo, bortezomib, and dexamethasone for treatment of relapsed, refractory multiple myeloma

- Superiority of venetoclax (N=194) vs. placebo (N=97)
  - PFS: 22.4 vs. 11.5 months
  - Response rate: 82% vs 68%
  - Minimal residual disease negative rate: 13.4% vs. 1%

- Mortality
  - Venetoclax: 41/194 (21.1%)
  - Placebo: 11/97 (11.3%)
  - HR 2.03 (1.04-3.94)
The Story of Tredaptive

- Tredaptive increases HDL (good cholesterol) in patients at risk for heart disease with low HDL

- Approved in 70 countries including the EU in 2008 based on trials that showed significant increases in HDL

- Not approved by FDA. Wanted a clinical outcome trial.

- HPS2-THRIVE (Heart Protection Study 2-Treatment of HDL to Reduce the Incidence of Vascular Events)
  - 4-year trial conducted by Merck with 26,000 participants
  - Compared statin + Tredaptive vs. statin alone
  - Endpoint: time to heart attack or coronary death, stroke, or need for arterial bypass
Effects of Extended-Release Niacin with Laropiprant in High-Risk Patients

The HPS2-THRIVE Collaborative Group

CONCLUSIONS

Among participants with atherosclerotic vascular disease, the addition of extended-release niacin–laropiprant to statin-based LDL cholesterol–lowering therapy did not significantly reduce the risk of major vascular events but did increase the risk of serious adverse events. (Funded by Merck and others; HPS2-THRIVE ClinicalTrials.gov number, NCT00461630.)

![Table 2](image)

**Table 2. Effects of Niacin–Laropiprant on Selected Serious Adverse Events and Diabetes.**

<table>
<thead>
<tr>
<th>Event Type</th>
<th>Niacin–Laropiprant (N = 12,838)</th>
<th>Placebo (N = 12,835)</th>
<th>Rate Ratio (95% CI)</th>
<th>Absolute Excess with Niacin–Laropiprant percentage points</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Serious adverse event — no. (%)</td>
<td>620 (4.8)</td>
<td>491 (3.8)</td>
<td>1.28 (1.13–1.44)</td>
<td>1.0±0.3</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Gastrointestinal event</td>
<td>86 (0.7)</td>
<td>51 (0.4)</td>
<td>1.67 (1.20–2.34)</td>
<td>0.3±0.1</td>
<td>0.003</td>
</tr>
<tr>
<td>Musculoskeletal event</td>
<td>481 (3.7)</td>
<td>385 (3.0)</td>
<td>1.26 (1.16–1.44)</td>
<td>0.7±0.2</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Skin-related event</td>
<td>362 (2.5)</td>
<td>238 (1.9)</td>
<td>1.38 (1.17–1.62)</td>
<td>0.7±0.2</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Infection event</td>
<td>1031 (8.0)</td>
<td>853 (6.6)</td>
<td>1.22 (1.12–1.34)</td>
<td>1.4±0.3</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Bleeding event</td>
<td>326 (2.5)</td>
<td>238 (1.9)</td>
<td>1.38 (1.17–1.62)</td>
<td>0.7±0.2</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Diabetes mellitus — no./total no. (%)</td>
<td>494/8704 (5.7)</td>
<td>376/8670 (4.3)</td>
<td>1.32 (1.16–1.51)</td>
<td>1.3±0.3</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>New-onset diabetes in participants without diabetes at baseline</td>
<td>460/4134 (11.1)</td>
<td>311/4165 (7.5)</td>
<td>1.55 (1.34–1.78)</td>
<td>3.7±0.6</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Disturbed diabetes control in participants with diabetes at baseline</td>
<td>460/4134 (11.1)</td>
<td>311/4165 (7.5)</td>
<td>1.55 (1.34–1.78)</td>
<td>3.7±0.6</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Figure 1. First Major Vascular Event during Follow-up.

Shown are Kaplan–Meier plots of the first major vascular event during the 4 years of follow-up. The inset shows the same data on an expanded y axis. The numbers of participants at risk for a first postrandomization major vascular event at the start of each year of follow-up are also shown, along with the benefit, which is shown as the absolute differences (with standard errors) in incidence rates between participants assigned to niacin–laropiprant and those assigned to placebo.
Surrogate Diseases

- Today there are proposals to use surrogate diseases.

- For example, infections are typically defined by an infection site (e.g., skin, lung) and the offending pathogen (e.g., *Pseudomonas aeruginosa*).

- Trials evaluating interventions for different infection sites use different endpoints and e.g., mortality is more common in some sites (e.g., bloodstream) than others (e.g., urinary tract).

- If a drug that targets a specific pathogen is effective in one infection site, can’t we use that as evidence of effectiveness in another site?
Daptomycin

- Approved for skin and other infections
- Does not work in respiratory infections
- Deactivation by pulmonary surfactant was only discovered in animal models after community-acquired pneumonia trials failed in humans.

<table>
<thead>
<tr>
<th>CABP Studies 05+08</th>
<th>Clinical Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>ITT Population</td>
<td>Ceftriaxone</td>
</tr>
<tr>
<td></td>
<td>326/431 (77.4)</td>
</tr>
</tbody>
</table>
Doripenem

- FDA-approved for several indications such as abdominal infections
- Not approved in 2008 for ventilator-associated pneumonia
- Post-marketing trial halted early from excess mortality

<table>
<thead>
<tr>
<th>Mortality</th>
<th>Doripenem</th>
<th>Zosyn</th>
<th>P-value (Fisher’s exact test)</th>
</tr>
</thead>
<tbody>
<tr>
<td>During IV therapy</td>
<td>21/223 (0.09)</td>
<td>9/223 (0.04)</td>
<td>0.04</td>
</tr>
<tr>
<td>Due to pneumonia</td>
<td>9/221 (0.04)</td>
<td>1/221 (&lt;0.01)</td>
<td>0.02</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Analysis Population</th>
<th>Doripenem Group %</th>
<th>Imipenem Group %</th>
<th>Difference %</th>
<th>2-sided 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical Cure Rates</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MITT</td>
<td>45.6</td>
<td>56.8</td>
<td>-11.2</td>
<td>-26.3 to 3.8</td>
</tr>
<tr>
<td>ME</td>
<td>49.1</td>
<td>66.1</td>
<td>-17</td>
<td>-34.7 to 0.8</td>
</tr>
<tr>
<td>All Cause 28-day Mortality Rate (MITT)</td>
<td>21.5</td>
<td>14.8</td>
<td>6.7</td>
<td>-5.0 to 18.5</td>
</tr>
</tbody>
</table>
P-value: one of our greatest tools. Often misused and misinterpreted.

The hammer is a great tool. If someone uses it to wash windows, and breaks the window, do you throw out the hammer?

The American Statistician

Moving to a World Beyond "$p < 0.05$"

Ronald L. Wasserstein, Allen L. Schirm & Nicole A. Lazar

It’s Not the P-Values’ Fault

Yoav Benjamini

I argue that ASA board statement about the $p$-values may be read as discouraging the use of $p$-values because they can be misused, while the other approaches offered there might be misused in much the same way. In particular, ignoring the effect of selection on statistical inferences is common yet potentially very harmful to the replicability of research results.

KEY WORDS: ASA board; Industrialized science; Selective inference.
Innovations are often presented with a degree of commercialism rather than scientific objectivity.
Adaptive Designs: Time for a Market Correction?

ADAPTIVE METHODS: TELLING “THE REST OF THE STORY”

Scott S. Emerson and Thomas R. Fleming
Department of Biostatistics, University of Washington, Seattle, Washington, USA

The Food and Drug Administration (FDA) draft guidance on adaptive design randomized clinical trials provides in-depth consideration of the difficulties that unblinded adaptation of clinical trial design might introduce. We provide extended discussion of these difficulties, with focus on the problems that the adaptive designs pose in the scientific interpretation of randomized clinical trial results, for regulatory authorities as well as for patients and caregivers who wish to make evidence-based decisions regarding the choice of treatment. We consider implications in adequate and well-controlled studies of the use of unblinded measures of treatment effect to make adaptive selection/ modification of treatments, adaptive selection of primary endpoints, adaptive modification of maximal sample size, adaptive modification of randomization ratios, and adaptive modification of target populations (adaptive enrichment), and then we consider the special topic of seamless phase 2–3 designs. We examine the extent to which the adaptive designs do not meet the goals of having greater efficiency, being more likely to identify truly effective treatments, being more informative, and providing greater flexibility. We fully support the FDA’s continued requirement of adequate and well-controlled confirmatory studies, complete with prospective, detailed specification of the entire randomized clinical trial design in a way that allows accurate and precise estimation of treatment effectiveness.

Issues in the use of adaptive clinical trial designs

Scott S. Emerson∗,†,‡
Department of Biostatistics, University of Washington, Box 357232, Seattle, Washington 98195, U.S.A.

SUMMARY
Sequential sampling plans are often used in the monitoring of clinical trials in order to address the ethical and efficiency issues inherent in human testing of a new treatment or preventive agent for disease. Group sequential stopping rules are perhaps the most commonly used approaches, but in recent years, a number of authors have proposed adaptive methods of choosing a stopping rule. In general, such adaptive approaches come at a price of inefficiency (almost always) and clouding of the scientific question (sometimes). In this paper, I review the degree of adaptation possible within the largely prespecified group sequential stopping rules, and discuss the operating characteristics that can be characterized fully prior to collection of the data. I then discuss the greater flexibility possible when using several of the adaptive approaches receiving the greatest attention in the statistical literature and conclude with a discussion of the scientific and statistical issues raised by their use. Copyright © 2006 John Wiley & Sons, Ltd.
Outcome-Adaptive Randomization: Is It Useful?
Edward L. Korn and Boris Freidlin
See accompanying editorial on page 606

Abstract
Outcome-adaptive randomization is one of the possible elements of an adaptive trial design in which the ratio of patients randomly assigned to the experimental treatment arm versus the control treatment arm changes from 1:1 over time to randomly assigning a higher proportion of patients to the arm that is doing better. Outcome-adaptive randomization has intuitive appeal in that, on average, a higher proportion of patients will be treated on the better treatment arm (if there is one). In both the randomized phase II and phase III settings with a short-term binary outcome, we compare outcome-adaptive randomization with designs that use 1:1 and 2:1 fixed-ratio randomizations (in the latter, twice as many patients are randomly assigned to the experimental treatment arm). The comparisons are done in terms of required sample sizes, the numbers and proportions of patients having an inferior outcome, and we restrict attention to the situation in which one treatment arm is a control treatment (rather than the less common situation of two experimental treatments without a control treatment). With no differential patient accrual rates because of the trial design, we find no benefits to outcome-adaptive randomization over 1:1 randomization, and we recommend the latter. If it is thought that the patient accrual rates will be substantially higher because of the possibility of a higher proportion of patients being randomly assigned to the experimental treatment (because the trial will be more attractive to patients and clinicians), we recommend using a fixed 2:1 randomization instead of an outcome-adaptive randomization.
A Statistician is ________.

- A coupon … gets me 15% off on sample size
  
- Gullible
  
- Salesperson
A Statistician is ________.

- Formerly the gatekeeper of trial integrity but is now often complicit in lowering it.

- Egotistical… my modeling skills can replace randomization.

- Powerless? We remain silent while the foundation for statistical inference is shown the door.
Traditional approaches are in need in improvements.

Rather than approaches that compromise scientific rigor, can we redirect our motivation to find BETTER answers to the most important questions for patients and clinicians?

Yes!

Perhaps there are even efficiencies in doing so.
One concept consistent with these goals though often misunderstood, is pragmatism, more thoroughly understanding the effects of interventions as experienced by patients, and the value of diagnostics in real world practice.

Proposal:

Place increased interest on questions of a pragmatic origin to match their clinical importance and utility.
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
• 14 years ago I started annually teaching clinical trials

• ~40 early career MDs beginning research careers in trials

• Major assignment: Develop your own protocol
  – The protocol turned into a real study for most

• The first year: one student wanted to evaluate if doing nothing was better than the common treatment.

• Ten years later: 20-25% of students were doing such trials
Why?

- Off-label use?
- Evolving effectiveness?
- Original studies not very pragmatic?
  - Restricted populations
  - Select settings
  - Limitations on common concomitant therapies
  - Surrogate outcomes
  - Analyses not focused on benefit:risk
We are drowning in data but starving for knowledge.

Many of our wounds are self-inflicted.
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?”
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.
Great work on estimands…we finally get people to recognize that different analysis populations address different questions…

Then they choose the wrong question

Let’s ignore that randomization is the foundation for statistical inference, only ITT analyses preserves the benefits provided by randomization regardless of whether an endpoint is labeled as one of efficacy or safety, and that if you conduct PP/as-treated that you have surrendered the integrity of an RCT instead opting for an observational study … let’s set those small facts aside.

What is the relevant question when evaluating an intervention?
What is Most Important for Trial Participants and Future Patients?

- Suppose an RCT is conducted comparing A vs. B

- A trial participant assigned to A, discontinues A, and begins a new intervention C

- The participant then experiences an AE, adjudicated as related to C but not A

- This leads some to believe that safety is not an issue for A. It is C’s fault.
What is Most Important for Trial Participants and Future Patients?

- Now suppose ten additional trial participants discontinue A, begin C, and experience the AE.

- Again adjudication links the relationship to C but not A.

- There are no such events in Arm B.

- Conclusion?

- I don’t want treatment A.

- Neither should you.
ITT Addresses the Most Important Question

- The assessment of most importance for patients and clinicians making decisions is conducted through a contrast of randomized interventions using ITT.

- Events are experienced by the trial participants and are thus important as downstream consequences to the initial intervention assignment and application.

- Causality ≠ Adjudicated Relationship.

- Vioxx studies: on-treatment analyses led to underestimation of the risks for harm only uncovered with subsequent ITT analyses that included events after treatment discontinuation.
Greater Pragmatism is Needed

- The benefit-risk profile of an intervention within the context of the trial and potentially future use in clinical practice encompasses therapeutic management after intervention withdrawal.

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

- How does this 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
- 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?

- 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*

<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>
</tbody>
</table>
## RCT Comparing A, B, and C

### Analysis of Outcomes

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Success (%)</th>
<th>Safety Event (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>100</td>
<td>50</td>
<td>30</td>
</tr>
<tr>
<td>B</td>
<td>100</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>C</td>
<td>100</td>
<td>50</td>
<td>50</td>
</tr>
</tbody>
</table>
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?
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.
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>Treatment</th>
<th>的成功率(N=100)</th>
<th>成功率</th>
<th>安全事件率</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>50%</td>
<td>30%</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>50%</td>
<td>50%</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>50%</td>
<td>50%</td>
<td></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

### 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></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>A (N=100)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Success</td>
<td>15</td>
<td>15</td>
<td>50</td>
<td>0</td>
</tr>
<tr>
<td>Safety event</td>
<td>35</td>
<td>35</td>
<td>0</td>
<td>50</td>
</tr>
</tbody>
</table>

| B (N=100) |         |        |         |        |
| Success   | 50      | 0      |        |        |
| Safety event | 50 | 0      |        |        |

| C (N=100) |         |        |         |        |
| Success   | 0       | 50     |        |        |
| Safety event | 50 | 0      |        |        |
### Analysis of Patients: 4 Possible Outcomes

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

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

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

Perhaps we should analyze the patient.
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
Motivating question:

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

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

- **IPTW-adjusted DOOR Probability:** 64% (53%, 75%)

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 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 …
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

DOOR Algorithm Based on Clinicians’ Rankings.
AE = adverse event; DOOR = desirability of outcome ranking

Alive

How many of:
- treatment failure,
- infectious complications,
- ongoing symptoms

0 of 3

1 of 3

2 of 3

3 of 3

1

2

3

4

5

6
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>Condition</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>
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…

even the decision-makers may not 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

<table>
<thead>
<tr>
<th>Category</th>
<th>Credit</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>
Survival

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

Caz-avi advantage: 0.13 (-0.03, 0.31), p = 0.12
Alive without Renal Failure

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

<table>
<thead>
<tr>
<th>Category</th>
<th>Credit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discharged home</td>
<td>100</td>
</tr>
<tr>
<td>Alive;</td>
<td></td>
</tr>
<tr>
<td>Not discharged home;</td>
<td>80</td>
</tr>
<tr>
<td>No renal failure</td>
<td></td>
</tr>
<tr>
<td>Alive;</td>
<td></td>
</tr>
<tr>
<td>Not discharged home; Renal failure</td>
<td>60</td>
</tr>
<tr>
<td>Death</td>
<td>0</td>
</tr>
</tbody>
</table>

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

Who benefits from this new therapy?
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>Treatment success without AKI</td>
<td>Treatment failure (persistent bacteremia) without AKI</td>
</tr>
<tr>
<td>Treatment success with AKI</td>
<td>Treatment failure with AKI</td>
</tr>
<tr>
<td>Death</td>
<td></td>
</tr>
</tbody>
</table>
**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.

[Bar chart showing IPTW DOOR outcomes by AUC/MIC Etest Quintiles]
DOOR STEPP

![Graph showing cumulative percent over AUC/MIC (Etest) with different categories: Door, Death, Txt Failure, AKI, Txt Failure, no AKI, Txt Success, AKI, Txt Success, no AKI]
## DOOR STEPP: Partial Credit Clinician A

<table>
<thead>
<tr>
<th>Category</th>
<th>Credit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment Success; No Kidney Injury</td>
<td>100</td>
</tr>
<tr>
<td>Treatment Success; Kidney Injury</td>
<td>80</td>
</tr>
<tr>
<td>Treatment Failure; No Kidney Injury</td>
<td>75</td>
</tr>
<tr>
<td>Treatment Failure; Kidney Injury</td>
<td>50</td>
</tr>
<tr>
<td>Death</td>
<td>0</td>
</tr>
</tbody>
</table>

**Optimal Dose:** 301.2
DOOR STEPP: Partial Credit Clinician B

<table>
<thead>
<tr>
<th>Category</th>
<th>Credit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment Success; No Kidney Injury</td>
<td>100</td>
</tr>
<tr>
<td>Treatment Success; Kidney Injury</td>
<td>80</td>
</tr>
<tr>
<td>Treatment Failure; No Kidney Injury</td>
<td>50</td>
</tr>
<tr>
<td>Treatment Failure; Kidney Injury</td>
<td>30</td>
</tr>
<tr>
<td>Death</td>
<td>0</td>
</tr>
</tbody>
</table>

Optimal Dose: 301.2
DOOR STEPP: Partial Credit Clinician C

<table>
<thead>
<tr>
<th>Category</th>
<th>Credit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment Success; No Kidney Injury</td>
<td>100</td>
</tr>
<tr>
<td>Treatment Success; Kidney Injury</td>
<td>50</td>
</tr>
<tr>
<td>Treatment Failure; No Kidney Injury</td>
<td>50</td>
</tr>
<tr>
<td>Treatment Failure; Kidney Injury</td>
<td>25</td>
</tr>
<tr>
<td>Death</td>
<td>0</td>
</tr>
</tbody>
</table>

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
## Benefit-risk category

<table>
<thead>
<tr>
<th></th>
<th>Ticagrelor (N=6589)</th>
<th>Aspirin (N=6610)</th>
<th>Cumulative difference % (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Most desirable</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Survived with no event</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Survived with non-disabling stroke, MI or PLATO major bleeding, 1 event</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Survived with non-disabling stroke, MI or PLATO major bleeding, &gt;1 event</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Least desirable</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Survived with disabling stroke</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Death</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Aspirin results

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

Will people on Ticagrelor migrate to a more desirable outcome?
## Ticagrelor results

<table>
<thead>
<tr>
<th>Benefit-risk category</th>
<th>Ticagrelor (N=6589)</th>
<th>Aspirin (N=6610)</th>
<th>Cumulative difference % (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Survived with no event</td>
<td>6124 (92.9)</td>
<td>6089 (92.1)</td>
<td></td>
</tr>
<tr>
<td>Survived with non-disabling stroke, MI or PLATO major bleeding, 1 event</td>
<td>147 (2.2)</td>
<td>171 (2.6)</td>
<td></td>
</tr>
<tr>
<td>Survived with non-disabling stroke, MI or PLATO major bleeding, &gt;1 event</td>
<td>6 (0.1)</td>
<td>11 (0.2)</td>
<td></td>
</tr>
<tr>
<td>Survived with disabling stroke</td>
<td>244 (3.7)</td>
<td>281 (4.3)</td>
<td></td>
</tr>
<tr>
<td>Death</td>
<td>68 (1.0)</td>
<td>58 (0.9)</td>
<td></td>
</tr>
</tbody>
</table>
# DOOR contrast

<table>
<thead>
<tr>
<th>Benefit-risk category</th>
<th>Ticagrelor (N=6589) n (%)</th>
<th>Aspirin (N=6610) n (%)</th>
<th>Cumulative difference % (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Survived with no event</td>
<td>6124 (92.9)</td>
<td>6089 (92.1)</td>
<td>0.8 (–0.1, 1.7)</td>
</tr>
<tr>
<td>Survived with non-disabling stroke, MI or PLATO major bleeding, 1 event</td>
<td>147 (2.2)</td>
<td>171 (2.6)</td>
<td>0.5 (–0.3, 1.2)</td>
</tr>
<tr>
<td>Survived with non-disabling stroke, MI or PLATO major bleeding, &gt;1 event</td>
<td>6 (0.1)</td>
<td>11 (0.2)</td>
<td>0.4 (–0.3, 1.1)</td>
</tr>
<tr>
<td>Survived with disabling stroke</td>
<td>244 (3.7)</td>
<td>281 (4.3)</td>
<td>–0.2 (–0.5, 0.2)</td>
</tr>
<tr>
<td>Death</td>
<td>68 (1.0)</td>
<td>58 (0.9)</td>
<td>-</td>
</tr>
</tbody>
</table>
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
Quotes by Socrates

*The unexamined life is not worth living.*

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

*Wisdom begins in wonder.*
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 they don’t know, then we should educate them.

If they don’t care, then we should motivate them.
My Plea

- Be motivated to do things better rather than faster and cheaper.

- Place increased interest and importance on questions of a pragmatic origin. These are the most important questions for patients and clinicians.

- Remain objective. Don’t buy new tires for your boat.

- When necessary sacrifice quantity based on feasibility. Don’t sacrifice quality.
Considering all of the medical priorities, if you had $500 million to spend on medical research today, how would you spend it?

Training good biostatisticians.

It would be the most impactful.
Significant Contributors (p<0.001)

- Dean Follmann
- Dan Rubin
- Chip Chambers
- David van Duin
- The Antibacterial Resistance Leadership Group
- The SOCRATESES Steering Committee
I have no doubt that you will enthusiastically applaud now ... because you are so relieved that it is over.

Thank you.
Today's Random Medical News

from the New England Journal of
Panic-Inducing
Gobbledygook

Can cause:
- Hypothermia
- Breast cancer
- Glaucoma
- Rheumatoid arthritis

According to a report released today...
ISIS-2: Second International Study of Infarct Survival

- Randomized Placebo-controlled Factorial Trial of Streptokinase and Aspirin after MI

- Both interventions improved outcomes
  - E.g., Aspirin reduced vascular mortality by 23%, $P<0.0001$

- BUT aspirin increased mortality in 2 subgroups…

  Gemini’s and Libras
Post hoc ergo propter hoc
Knowledge can affect behavior

- Changes in behavior can be unintentional and subtle

- Placebo-controlled trial for cyclophosphamide and plasma exchange for the treatment of MS (Noseworthy et al., Neurology, 1994)
  - Assessments performed by both blinded and unblinded neurologists
  - Benefit suggested when using evaluations from unblinded neurologists; but not for blinded neurologists
Subjective and Objective Outcomes

- Without blinding, subjective evaluations in particular could be biased

- Objective evaluations are not entirely immune to biases induced by a lack of blinding
  - E.g., patients may selectively drop-out or selectively use concomitant therapy causing a distortion of the estimated effects
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

Figure 3. Percent Change in Maximum Forced Expiratory Volume in 1 Second (FEV₁) with Each of the Four Interventions.

The relative improvement in FEV₁ 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.
Blinding

- Particularly important for:
  - Patient reported outcomes (PROs)
    - E.g., pain, depression, anxiety
  - Clinician ratings when there are strong prior beliefs or financial incentives
FDA enforcement actions plummet under Trump

Agency is sending many fewer warning letters

By Charles Piller

From monitoring clinical trials and approving medications and vaccines to ensuring the safety of blood transfusions, medical devices, prescription drugs, and more, the Food and Drug Administration (FDA) is one of the nation's most vital institutions. By several measures, however, FDA's compliance and enforcement actions have plummeted since President Donald Trump took office, science has found.

The agency's "warning letters"—a key tool for keeping dangerous or ineffective drugs, devices, and vaccines off the market—have fallen by one-third, for example. Such letters typically demand swift corrective action to protect public health and safety. FDA records from Trump's inauguration through January 2021 show the agency issued 503B warning letters, compared with 752 for the most recent equivalent period under former President Barack Obama. Compared with the start of the Obama presidency, Trump-era letters dropped by nearly half.

Warnings from the FDA Center for Devices and Radiological Health, which helps ensure the safety and quality of medical devices, and from some of the agency's district offices—such as those in Philadelphia, Florida, and New York—have dropped even more sharply, by more than twofold. Two district offices have not issued a warning in more than 5 years. The numbers don't just reflect a new administration's slow start; FDA's overall, significantly fewer warning letters in the second year of Trump's presidency than in his first.

FDA watchers say they can't pinpoint what's driving the decline, but they are alarmed. "Those who think the Trump administration has not succeeded in its deregulatory efforts ought to look at these data," says Peter Larlee, an FDA executive under Obama and Trump and now executive director of the Center for Science in the Public Interest, a Washington, D.C., advocacy group. "Industry may well take the message that the era is not as tough as it was before.

Several other FDA actions under Trump show similar declines when measured against the end of the Obama administration. FDA inspection reports labeled "Official action indicated"—typically a trigger for warning letters or similar actions—have fallen by about half under Trump and are continuing to trend downward. Even FDA's rare injunctions, a more powerful step than warnings to prevent sales or distribution of unsafe or otherwise illegal products, fell from 23 in the last part of the Obama administration to 6 under Trump. (During a comparable period at the start of the Obama years, FDA issued 31 injunctions.) The agency's "untitled letters"—notes to concerns that fall short of thresholds for formal warnings—also have dropped sharply under Trump.

"FDA's power to enforce its requirements is an important part of how it achieves its public-health mission," says Patricia Zetter, a law professor at Ohio State University College of Law and former FDA commissioner. "If FDA is not using that power, it sends a signal that violations will be tolerated."

FDA did not dispute the enforcement and compliance data, which Science compiled from the agency's own public records in a written statement, the agency responded, "Sometimes the actions we take are visible, like warning letters and recalls. At other times, our actions to protect consumers are less noticeable, but equally vital." Among the efforts that can obviate the need for warning letters, FDA noted, are meetings with companies, follow-up inspections, the threat of mandatory recalls, a new voluntary improvement pilot program for manufacturers, coordination with European regulators, and unspecified "regulatory and compliance measures" conducted "behind the scenes."

Scott Gottlieb, Trump's first FDA commissioner, defended his record after reviewing a summary of Science's data. "We were very aggressive," he wrote in an email. "I don't think you can paint us with a political narrative—just that because we were a Republican administration, somehow we must have relaxed down enforcement activity. We didn't."

Gottlieb, a physician and venture capitalist before being tapped for FDA, resigned in March and is now at the American Enterprise Institute, a conservative think tank in Washington, D.C. On his watch, Gottlieb said later in an interview, FDA made enforcement more efficient and expanded field resources. The agency's budget rose 23% from fiscal years 2015 to 2018. He says that under Trump, FDA issued guidance or beefed-up enforcement