p-Values vs. Patient Values: An Analytic Perspective

Andrew W. Lo, MIT

September 13, 2018

ASA Biopharma Regulatory-Industry Statistics Workshop

FDA Approvals Involve Trade-offs

	Decision: Approve	Decision: Reject
Ineffective Therapy	Type I Error (False Positive), α	Correct
Effective Therapy	Correct	Type II Error (False Negative), β

- Goal: minimize errors by setting the threshold for approval (typically a "*p*-value" of 5%)
- But there's a trade-off between these errors
- "Greatest good for the greatest number" J. Bentham

Why 5%??

"...If one in twenty does not seem high enough odds, we may, if we prefer it, draw the line at one in fifty or one in a hundred. Personally, the writer prefers to set a low standard of significance at the 5 per cent point, and ignore entirely all results which fails to reach this level. A scientific fact should be regarded as experimentally established only if a properly designed experiment rarely fails to give this level of significance..."

– **RA Fisher**, 1926, "The arrangement of field experiments," ***Journal of the Ministry of Agriculture of Great Britain*** 33:503–513.

Why 5%??

THE AMERICAN STATISTICIAN
2016, VOL. 70, NO. 2, 129–133
http://dx.doi.org/10.1080/00031305.2016.1154108

Taylor & Francis
Taylor & Francis Group

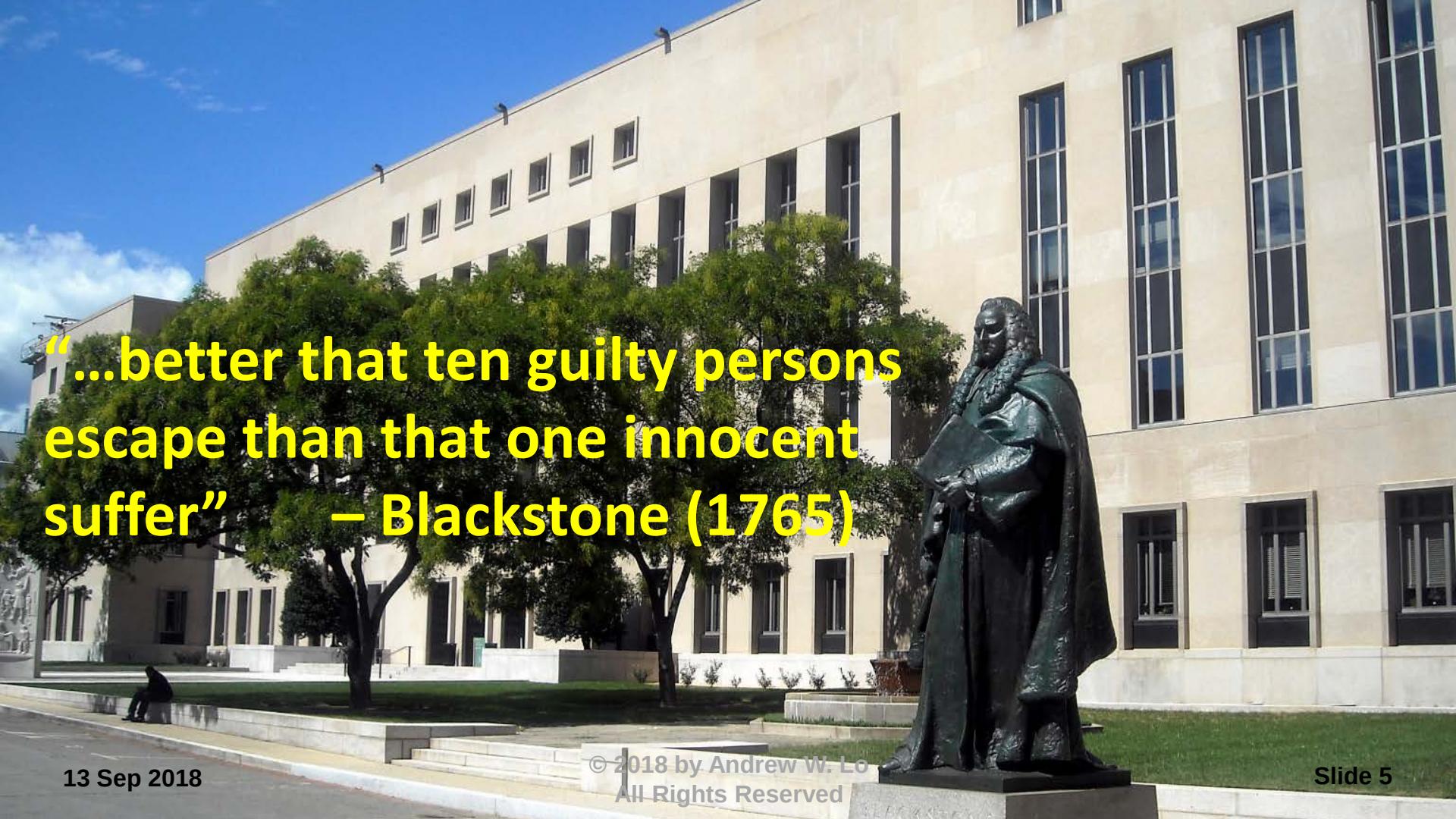
EDITORIAL

The ASA's Statement on *p*-Values: Context, Process, and Purpose

In February 2014, George Cobb, Professor Emeritus of Mathematics and Statistics at Mount Holyoke College, posed these questions to an ASA discussion forum:

> Q: Why do so many colleges and grad schools teach $p = 0.05$?
> A: Because that's still what the scientific community and journal editors use.
> Q: Why do so many people still use $p = 0.05$?
> A: Because that's what they were taught in college or grad school.

2014) and a statement on risk-limiting post-election audits (American Statistical Association 2010). However, these were truly policy-related statements. The VAM statement addressed a key educational policy issue, acknowledging the complexity of the issues involved, citing limitations of VAMs as effective performance models, and urging that they be developed and interpreted with the involvement of statisticians. The statement on election auditing was also in response to a major but specific



Incorporating Patient Preferences

Guidance for Industry & FDA Staff (2012)

"FDA recognizes that patient tolerance for risk and a patient-centric assessment of risk may reveal reasonable patients who are willing to tolerate a very high level of risk to achieve a probable benefit, especially if that benefit results in an improvement in quality of life."

21st Century Cures, Sec. 3002. "Patient-Focused Drug Development Guidance."

"How the FDA plans to use relevant patient experience data and related information when evaluating the risks and benefits of a drug.

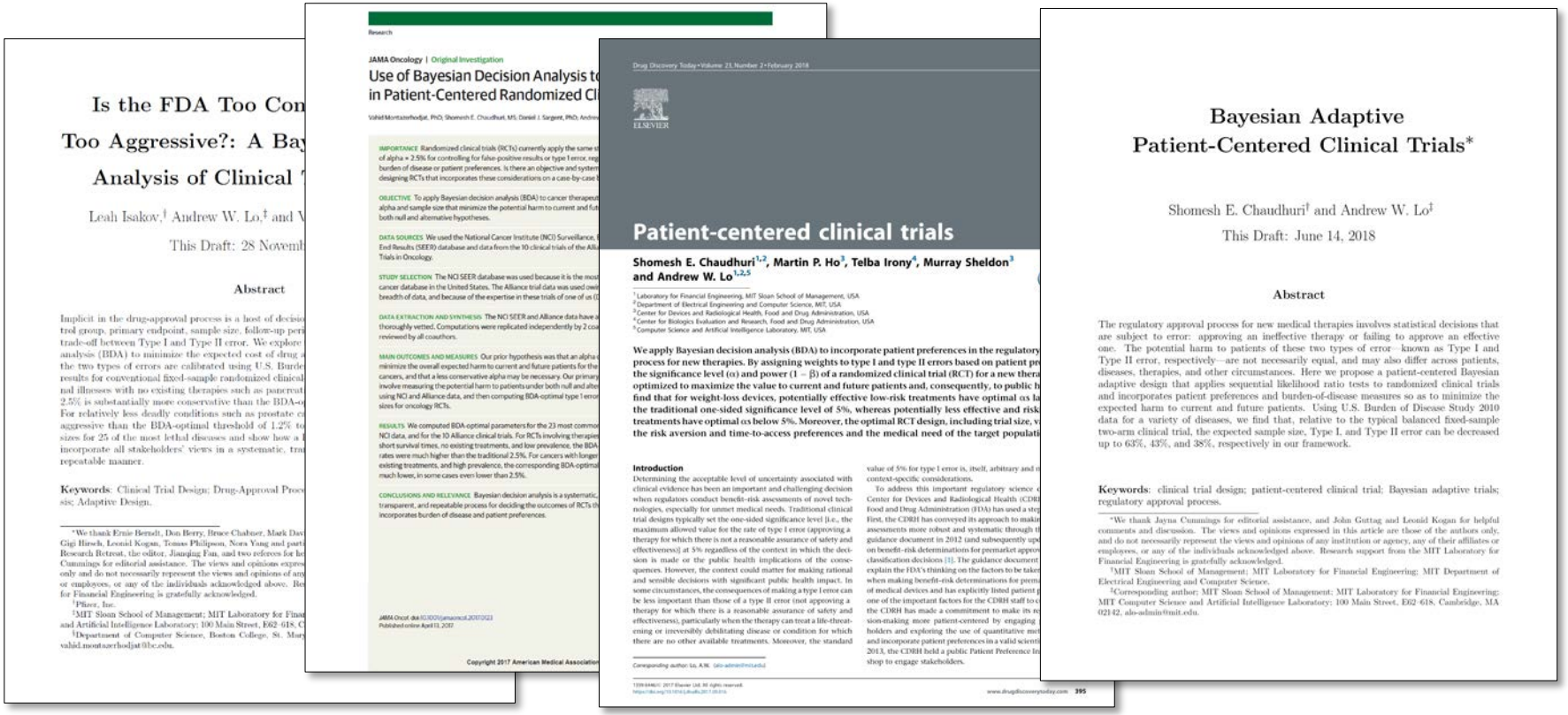
A New Approach

What If We Try To Reduce The Average Loss?

$$\begin{aligned} & \text{Loss(False Positive)} \times \text{Prob(False Positive;}p\text{)} \\ + \; & \text{Loss(False Negative)} \times \text{Prob(False Negative;}p\text{)} \\ \hline & \text{Average Loss(}p\text{)} \end{aligned}$$

- Given the losses from false positives and negatives, choose p to minimize the average loss $\Rightarrow$ BDA

Bayesian Decision Analysis



Is the FDA Too Con
Too Aggressive?: A Bay
Analysis of Clinical
Leah Isakov,† Andrew W. Lo,‡ and
This Draft: 28 Novemb
Abstract
Implicit in the drug-approval process is a host of decisio
trol group, primary endpoint, sample size, follow-up peri
trade-off between Type I and Type II error. We explore
analysis (BDA) to minimize the expected cost of drug
the two types of errors are calibrated using U.S. Burde
results for conventional fixed-sample randomized clinical
nal illnesses with no existing therapies such as pancreat
2.5% is substantially more conservative than the BDA-o
For relatively less deadly conditions such as prostate c
aggressive than the BDA-optimal threshold of 1.2% to
sizes for 25 of the most lethal diseases and show how a
incorporate all stakeholders' views in a systematic, tra
repeatable manner.
Keywords: Clinical Trial Design; Drug-Approval Proc
sis; Adaptive Design.
*We thank Ernie Berndt, Don Berry, Bruce Chabner, Mark Dav
Gigi Hirsch, Leonid Kogan, Tomas Philipson, Nora Yang and part
Research Retreat, the editor, Jianqing Fan, and two referees for he
Cummings for editorial assistance. The views and opinions expres
only and do not necessarily represent the views and opinions of an
or employees, or any of the individuals acknowledged above. Re
for Financial Engineering is gratefully acknowledged.
†Pfizer, Inc.
‡MIT Sloan School of Management; MIT Laboratory for Finan
and Artificial Intelligence Laboratory; 100 Main Street, E62-618, C
§Department of Computer Science, Boston College, St. Mary
vahid.montazerhodjat@bc.edu.

Research
JAMA Oncology | Original Investigation
Use of Bayesian Decision Analysis t
in Patient-Centered Randomized Cl
Vahid Montazerhodjat, PhD; Shomesh E. Chaudhuri, MS; Daniel J. Sargent, PhD; Andrew
IMPORTANCE Randomized clinical trials (RCTs) currently apply the same
of alpha = 2.5% for controlling for false-positive results or type I error, reg
burden of disease or patient preferences. Is there an objective and system
designing RCTs that incorporates these considerations on a case-by-case
OBJECTIVE To apply Bayesian decision analysis (BDA) to cancer therapeut
alpha and sample size that minimize the potential harm to current and fut
both null and alternative hypotheses.
DATA SOURCES We used the National Cancer Institute (NCI) Surveillance,
End Results (SEER) database and data from the 10 clinical trials of the Alli
Trials in Oncology.
STUDY SELECTION The NCI SEER database was used because it is the mos
cancer database in the United States. The Alliance trial data was used owi
breadth of data, and because of the expertise in these trials of one of us (
DATA EXTRACTION AND SYNTHESIS The NCI SEER and Alliance data have a
thoroughly vetted. Computations were replicated independently by 2 coa
reviewed by all coauthors.
MAIN OUTCOMES AND MEASURES Our prior hypothesis was that an alpha
minimize the overall expected harm to current and future patients for the
cancers, and that a less conservative alpha may be necessary. Our primary
involve measuring the potential harm to patients under both null and alte
using NCI and Alliance data, and then computing BDA-optimal type I erro
sizes for oncology RCTs.
RESULTS We computed BDA-optimal parameters for the 23 most comma
NCI data, and for the 10 Alliance clinical trials. For RCTs involving therapies
short survival times, no existing treatments, and low prevalence, the BDA
rates were much higher than the traditional 2.5%. For cancers with longer
existing treatments, and high prevalence, the corresponding BDA-optima
much lower, in some cases even lower than 2.5%.
CONCLUSIONS AND RELEVANCE Bayesian decision analysis is a systematic,
transparent, and repeatable process for deciding the outcomes of RCTs th
incorporates burden of disease and patient preferences.
JAMA Oncol. doi:10.1001/jamaoncol.2017.0123
Published online April 13, 2017
Copyright 2017 American Medical Association

Drug Discovery Today • Volume 23, Number 2 • February 2018
ELSEVIER
Patient-centered clinical trials
Shomesh E. Chaudhuri[1,2], Martin P. Ho[3], Telba Irony[4], Murray Sheldon[3] and Andrew W. Lo[1,2,5]
[1] Laboratory for Financial Engineering, MIT Sloan School of Management, USA
[2] Department of Electrical Engineering and Computer Science, MIT, USA
[3] Center for Devices and Radiological Health, Food and Drug Administration, USA
[4] Center for Biologics Evaluation and Research, Food and Drug Administration, USA
[5] Computer Science and Artificial Intelligence Laboratory, MIT, USA
We apply Bayesian decision analysis (BDA) to incorporate patient preferences in the regulatory process for new therapies. By assigning weights to type I and type II errors based on patient pre the significance level (α) and power (1 − β) of a randomized clinical trial (RCT) for a new thera optimized to maximize the value to current and future patients and, consequently, to public h find that for weight-loss devices, potentially effective low-risk treatments have optimal αs la the traditional one-sided significance level of 5%, whereas potentially less effective and risk treatments have optimal αs below 5%. Moreover, the optimal RCT design, including trial size, v the risk aversion and time-to-access preferences and the medical need of the target populati
Introduction
Determining the acceptable level of uncertainty associated with clinical evidence has been an important and challenging decision when regulators conduct benefit-risk assessments of novel technologies, especially for unmet medical needs. Traditional clinical trial designs typically set the one-sided significance level [i.e., the maximum allowed value for the rate of type I error (approving a therapy for which there is not a reasonable assurance of safety and effectiveness)] at 5% regardless of the context in which the decision is made or the public health implications of the consequences. However, the context could matter for making rational and sensible decisions with significant public health impact. In some circumstances, the consequences of making a type I error can be less important than those of a type II error (not approving a therapy for which there is a reasonable assurance of safety and effectiveness), particularly when the therapy can treat a life-threatening or irreversibly debilitating disease or condition for which there are no other available treatments. Moreover, the standard value of 5% for type I error is, itself, arbitrary and
context-specific considerations.
To address this important regulatory science
Center for Devices and Radiological Health (CDR
Food and Drug Administration (FDA) has used a ste
First, the CDRH has conveyed its approach to makin
assessments more robust and systematic through t
guidance document in 2012 (and subsequently up
on benefit-risk determinations for premarket approv
classification decisions [1]. The guidance document
explain the FDA's thinking on the factors to be take
when making benefit-risk determinations for prem
of medical devices and has explicitly listed patient p
one of the important factors for the CDRH staff to c
the CDRH has made a commitment to make its re
sion-making more patient-centered by engaging
holders and exploring the use of quantitative met
and incorporate patient preferences in a valid scienti
2013, the CDRH held a public Patient Preference In
shop to engage stakeholders.
Corresponding author: Lo, A.W. (alo-admin@mit.edu)
1359-6446/© 2017 Elsevier Ltd. All rights reserved.
https://doi.org/10.1016/j.drudis.2017.09.016
www.drugdiscoverytoday.com 395

Bayesian Adaptive
Patient-Centered Clinical Trials*
Shomesh E. Chaudhuri† and Andrew W. Lo‡
This Draft: June 14, 2018
Abstract
The regulatory approval process for new medical therapies involves statistical decisions that are subject to error: approving an ineffective therapy or failing to approve an effective one. The potential harm to patients of these two types of error—known as Type I and Type II error, respectively—are not necessarily equal, and may also differ across patients, diseases, therapies, and other circumstances. Here we propose a patient-centered Bayesian adaptive design that applies sequential likelihood ratio tests to randomized clinical trials and incorporates patient preferences and burden-of-disease measures so as to minimize the expected harm to current and future patients. Using U.S. Burden of Disease Study 2010 data for a variety of diseases, we find that, relative to the typical balanced fixed-sample two-arm clinical trial, the expected sample size, Type I, and Type II error can be decreased up to 63%, 43%, and 38%, respectively in our framework.
Keywords: clinical trial design; patient-centered clinical trial; Bayesian adaptive trials; regulatory approval process.
*We thank Jayna Cummings for editorial assistance, and John Guttag and Leonid Kogan for helpful comments and discussion. The views and opinions expressed in this article are those of the authors only, and do not necessarily represent the views and opinions of any institution or agency, any of their affiliates or employees, or any of the individuals acknowledged above. Research support from the MIT Laboratory for Financial Engineering is gratefully acknowledged.
†MIT Sloan School of Management; MIT Laboratory for Financial Engineering; MIT Department of Electrical Engineering and Computer Science.
‡Corresponding author; MIT Sloan School of Management; MIT Laboratory for Financial Engineering; MIT Computer Science and Artificial Intelligence Laboratory; 100 Main Street, E62-618, Cambridge, MA 02142, alo-admin@mit.edu.

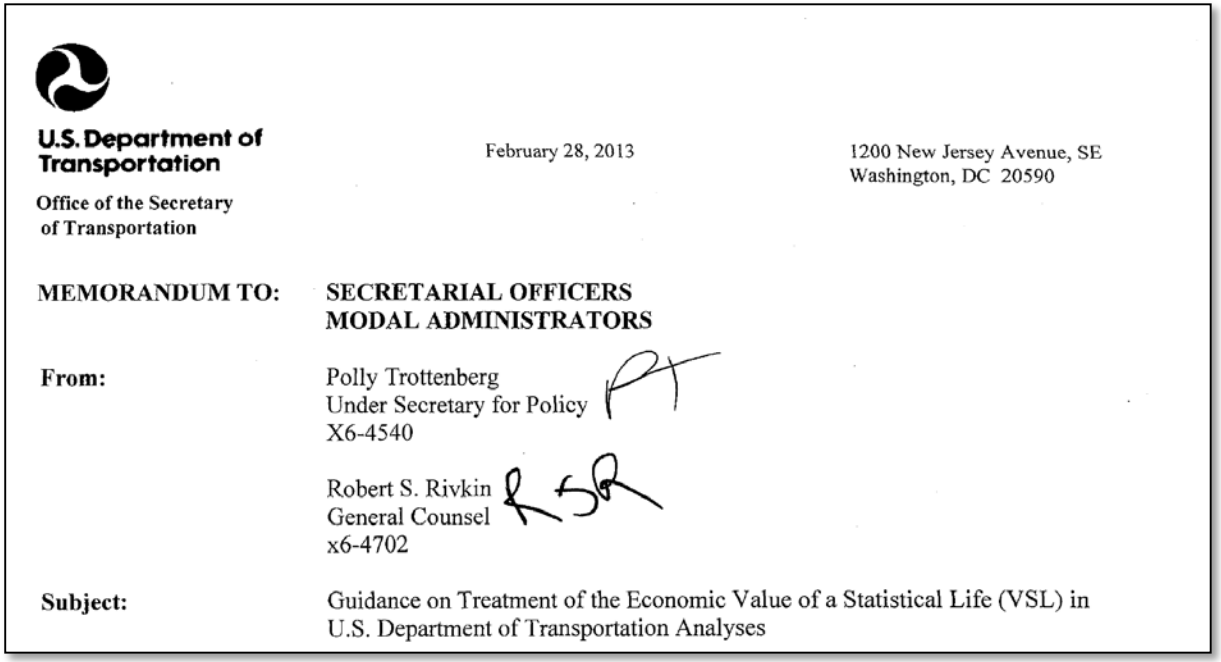
Bayesian Decision Analysis

Collaborators:

- Dawn Bardot, Heather Benz, Brittany Caldwell, Shomesh Chaudhuri, Stephanie Christopher, Katrina Gwin, Brett Hauber, Martin Ho, Telba Irony, Leah Isakov, Brennan Mange, Lauren McLaughlin, Vahid Montazerhodjat, Kyle Myers, John Ruiz, Annie Saha, Dan Sargent, Murray Sheldon, Mo Zhou
- FDA (CDRH), MDIC, MJF Foundation, RTI Health Solutions, MIT LFE

What About the Value of Life?

What About the Value of Life?



U.S. Department of Transportation

Office of the Secretary of Transportation

February 28, 2013

1200 New Jersey Avenue, SE
Washington, DC 20590

MEMORANDUM TO: SECRETARIAL OFFICERS
MODAL ADMINISTRATORS

From: Polly Trottenberg
Under Secretary for Policy
X6-4540

Robert S. Rivkin
General Counsel
x6-4702

Subject: Guidance on Treatment of the Economic Value of a Statistical Life (VSL) in U.S. Department of Transportation Analyses

$9,100,000.00

What About the Value of Life?

PERSPECTIVE

NEJM Aug 2014

UPDATING COST-EFFECTIVENESS

Updating Cost-Effectiveness — The Curious Resilience of the $50,000-per-QALY Threshold

Peter J. Neumann, Sc.D., Joshua T. Cohen, Ph.D., and Milton C. Weinstein, Ph.D.

If one had to select a single threshold outside the context of an explicit resource constraint or opportunity cost, we suggest using either $100,000 or $150,000.

- Compare new drug's price to QALYs
- How much **incremental value** is it providing?

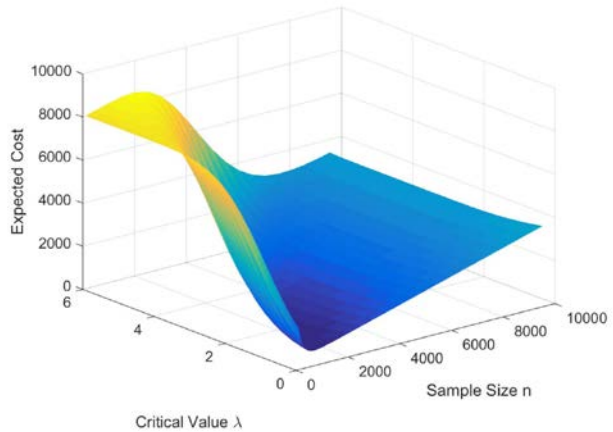
Bayesian Decision Analysis

$$\underset{p,n}{\text{Min}}\ \text{E}[\text{Cost}] \;=\; \underset{p,n}{\text{Min}}\Big(\text{E}[\text{Cost}|\text{H}_0]p_0 \;+\; \text{E}[\text{Cost}|\text{H}_1](1-p_0)\Big)$$

$$\text{E}[\text{Cost}] \;=\; c_1 p_o\left[N\Phi(-\lambda_n) \;+\; N\xi\Phi\left(\lambda_n - \delta_o\sqrt{\mathcal{I}_n}\right) \;+\; n(1+\gamma N\xi)\right]$$

$$\xi \equiv \frac{c_2}{c_1}\frac{1-p_0}{p_0}$$

- BDA-optimal decision **minimizes** expected cost



Measuring "Size" and "Importance"

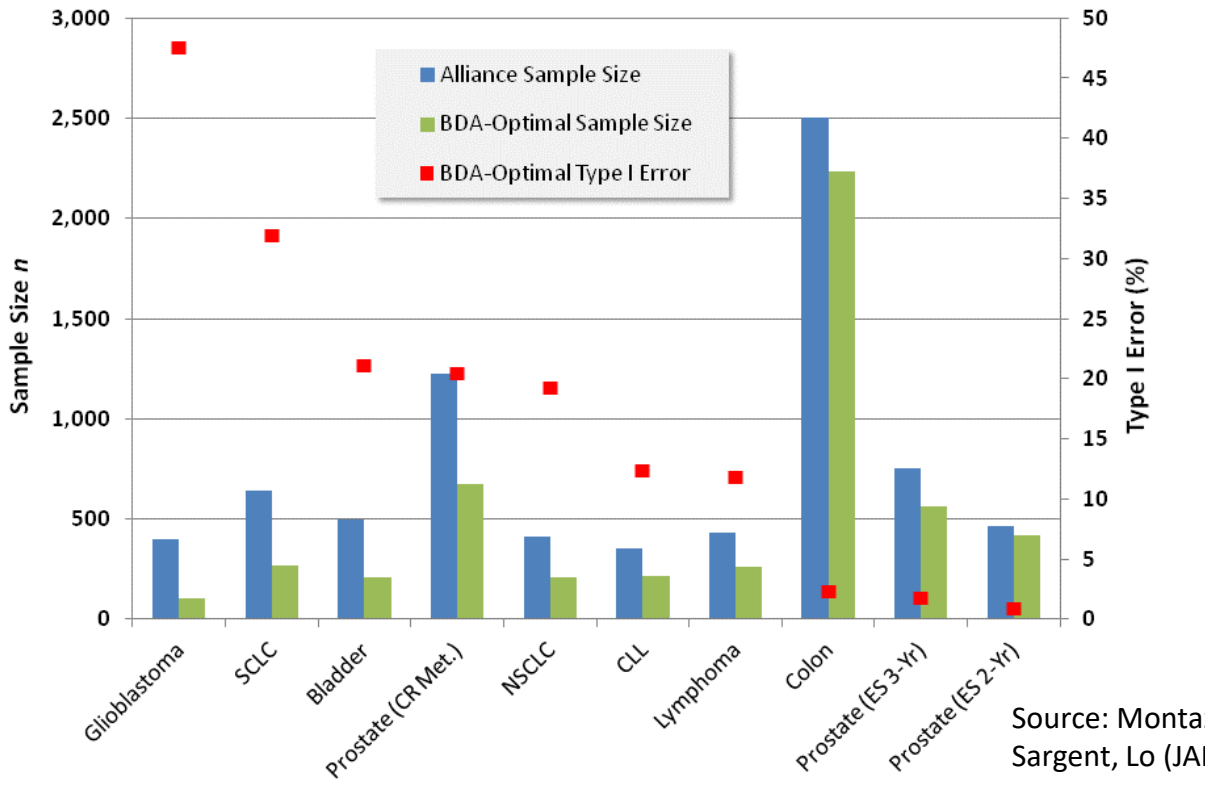
Disease Name	Prevalence (Thousands)	Severity
Ischemic heart disease	8,895.61	0.12
Lung cancer	289.87	0.45
Ischemic stroke	3,932.33	0.15
Hemorrhagic/other non-ischemic stroke	949.33	0.16
Chronic obstructive pulmonary disease	32,372.11	0.06
Diabetes	23,694.90	0.05
Cirrhosis of the liver	78.37	0.49
Alzheimer's disease	5,145.03	0.18
Colorectal cancer	798.90	0.15
Pneumococcal pneumonia	84.14	0.30

Source: Isakov, Montazerhodjat, Lo (2015)

Measuring "Size" and "Importance"

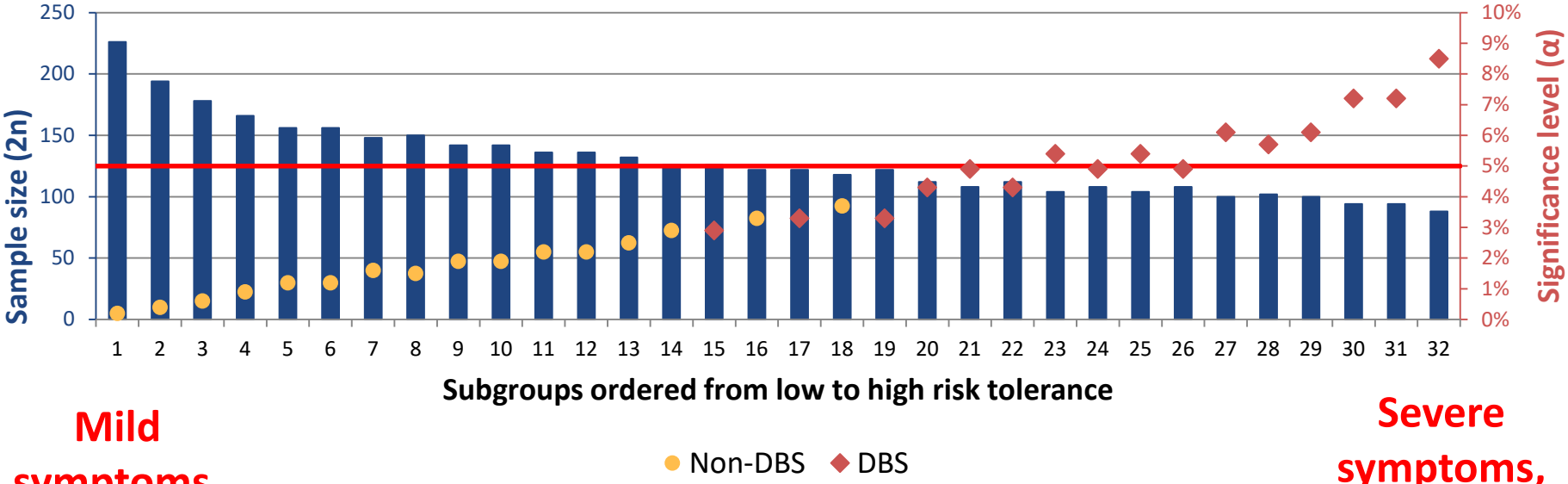
	Post-Trial		In-Trial
	$\widehat{H} = H_0$	$\widehat{H} = H_1$	
$H = H_0$	0	C_1	nc_1
$H = H_1$	C_2	0	$n\gamma C_2$

Application to Alliance Cancer Trials



Source: Montazerhodjat, Chaudhuri, Sargent, Lo (JAMA Onc. 2017)

Results for Parkinson's Device



Mild symptoms, no DBS

Severe symptoms, DBS

Is This Really Practical?

Proposal: "Right-to-Try" License:

- Two-year license, no off-label use, strict monitoring and data collection/sharing requirements (paid by company)
- Can be revoked any time during two-year period
- At the end of two years, either license expires or it converts to regular approval
- Similar to adaptive trials, but much greater economic incentives for bipoharma industry

Qualifications

- How to choose parameters? (ODAC, but with patients)
- Whose preferences should be reflected? (patients!)
- Potential backlash from toxicities and side effects?
- Ethical considerations

But these issues already exist in one form or another for current methods (e.g., eteplirsen); BDA provides a more systematic, transparent, objective, repeatable, rational framework for addressing them

Conclusion

"We should care about patient values as well as *p*-values."

– Donald Berry, MD Anderson

Thank You!