Canadian Cancer Trials Group

Groupe canadien des essais sur le cancer

Randomized Clinical Trials

How did we get here?

Where are we going?

Joe Pater

Focus of this presentation

- How did we get here?
  - Long ago until 1948 – the refinement of control selection
  - 1948 until now – the growth of randomized trials
    - In medicine
    - Elsewhere
- Where are we going?
  - Big Data/Real World Data and Evidence
    - The causal inference “revolution”
      - Will there continue to be a role for randomized trials?
    - Can observational data be part of a solution?
  - Surrogate endpoints
  - Right-sizing trials

How Did We Get Here?

Long Ago Until 1948

Topics

- Recognition of the need to compare
- Understanding that the comparison had to be “fair”
- Methods of ensuring fairness
- When did the idea of randomization creep in – and why?

Sources

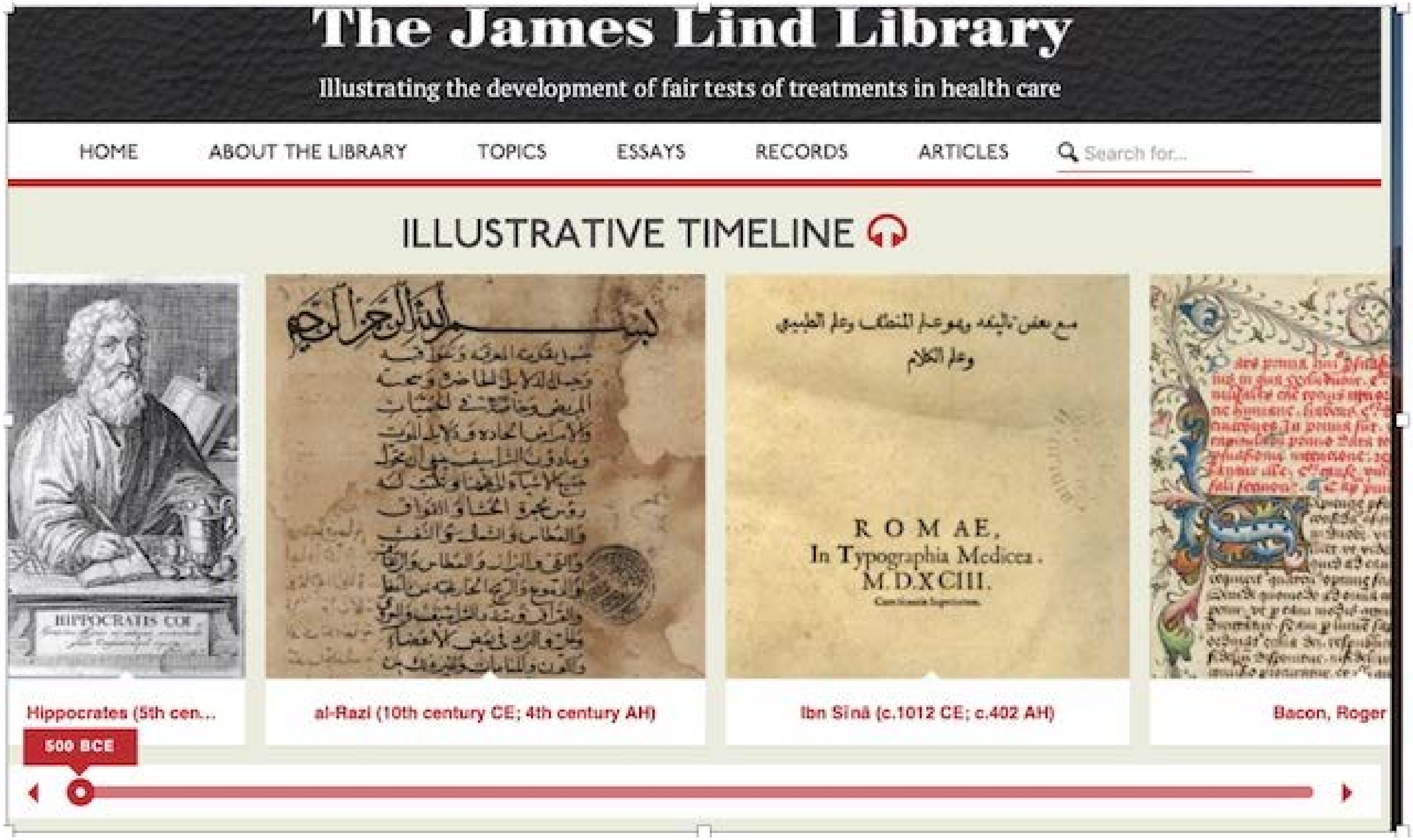
http://www.jameslindlibrary.org

The Entry of Randomized Assignment into the Social Sciences
Julian C. Jamison WPS8062

Assessing the Gold Standard — Lessons from the History of RCTs
Laura E. Bothwell, Ph.D., Jeremy A. Greene, M.D., Ph.D., Scott H. Podolsky, M.D., and David S. Jones, M.D., Ph.D. N Engl J Med 374;22 nejm.org June 2, 2016

The Emergence of the Randomized, Controlled Trial
Laura E. Bothwell, Ph.D., and Scott H. Podolsky, M.D. N Engl J Med 375;6 nejm.org August 11, 2016

The advent of fair treatment allocation schedules in clinical trials during the 19th and early 20th centuries.
Iain Chalmers, Estela Dukan, Scott Podolsky, George Davey Smith.
J R Soc Med 2012: 105: 221–227. DOI 10.1258/jrsm.2012.12k029



The James Lind Library

Illustrating the development of fair tests of treatments in health care

ILLUSTRATIVE TIMELINE

Hippocrates (5th cen...

al-Razi (10th century CE; 4th century AH)

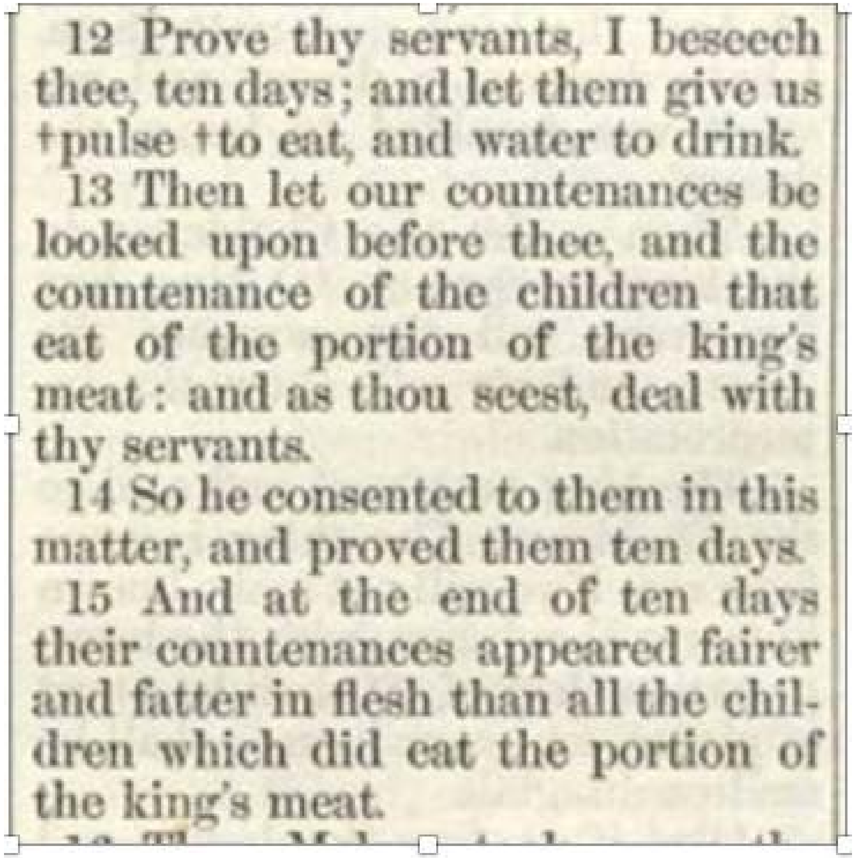
Ibn Sīnā (c.1012 CE; c.402 AH)

Bacon, Roger

http://www.jameslindlibrary.org

Recognition of the need to compare

- Book of Daniel



12 Prove thy servants, I beseech thee, ten days; and let them give us †pulse †to eat, and water to drink.

13 Then let our countenances be looked upon before thee, and the countenance of the children that eat of the portion of the king's meat: and as thou seest, deal with thy servants.

14 So he consented to them in this matter, and proved them ten days.

15 And at the end of ten days their countenances appeared fairer and fatter in flesh than all the children which did eat the portion of the king's meat.

http://www.jameslindlibrary.org

Recognition of the need to compare

Although this episode nicely captures the idea of a comparison group, there is an obvious problem with endogeneity and selection bias. Hence not only is randomization in any form missing, but there is no sense of a controlled or fair experiment.

Jamison

Understanding that the comparison had to be "fair"

- Petrarch letter to Boccaccio (1364)

  "I solemnly affirm and believe, if a hundred or a thousand men of the same age, same temperament and habits, together with the same surroundings, were attacked at the same time by the same disease, that if one half followed the prescriptions of the doctors of the variety of those practicing at the present day, and that the other half took no medicine but relied on Nature's instincts, I have no doubt as to which half would escape".

  Jamison

Methods of ensuring fairness

- In the assignment of therapy
  - Planned selection
    - James Lind

Their cases were as similar as I could have them. They all in general had putrid gums, the spots and lassitude, with weakness of their knees. They lay together in one place, being a proper apartment of the sick in the fore-hold; and had one diet common to all.

Jamison

Methods of ensuring fairness

- In the assignment of therapy
  - Alternation
    - Most cited study is that by Fibinger who administered diptheria antitoxin to 484 patients admitted on alternate days
    - However, alternation was used as a method of “fair treatment allocation” before that and continued to be used throughout the first half of the 20th century
  - Randomization (see later)

When did the idea of randomization creep in?

History of Clinical Trials
The Emergence of the Randomized, Controlled Trial
Laura E. Bothwell, Ph.D., and Scott H. Podolsky, M.D.

The birth of the randomized, controlled trial (RCT) is typically dated to a 1948 evaluation by the British Medical Research Council (MRC) of streptomycin for the treatment of tuberculosis.

When did the idea of randomization creep in – and why?

- Two narratives:
  - The statisticians did it
  - It was done solely to control selection bias

The statisticians did It

Chalmers: “Harry Marks judges the randomized clinical trial to have been "an extension of the statistician R.A. Fisher's ideas about experimental design" and that "the statisticians' randomized controlled trial came to represent the symbol and substance of the statistical method in medicine." Jean-Paul Gaudilliere observes: "The history of randomized clinical trials may be traced back to the biometricians' work and it seems to be a good example of 'applied statistics'. On the one hand there was a direct lineage from Pearson to Bradford Hill via Fisher and Major Greenwood ... On the other hand, it is not too difficult to argue for conceptual legacy, since the basic concepts grounding the choice of randomisation can be traced back to R.A. Fisher's work."

The goal was to control selection bias

Although one of the reasons that the streptomycin trial has become iconic is that the treatment allocation schedule was based on random number tables, this was not for any esoteric statistical reason. It was because successful concealment of allocation schedules and prevention of foreknowledge of upcoming allocations among clinicians entering patients in trials is more likely to be achieved with allocation schedules based on random numbers than with schedules using alternation.

Chalmers

Two components to randomization

Subverting Randomization in Controlled Trials
Kenneth F. Schulz, PhD, MBA

(JAMA. 1995;274:1456-1458)

WHAT IS RANDOMIZATION'?
Randomization, if successfully accomplished, prevents bias in allocation of participants to comparison groups. Its success depends on two interrelated processes. First, an unpredictable allocation sequence must be generated based on a random procedure. Second, strict implementation of that schedule must be secured through an assignment mechanism (allocation concealment process) that prevents foreknowledge of treatment assignment. Crucially, allocation concealment shields those who admit patients to a trial from knowing the upcoming assignments. The decision to accept or reject a participant must be made and informed consent obtained without knowledge of the treatment to be assigned.

1948 to Now

The Rise of RCTs

The rise of RCTs

- RCTs became the dominant method for assessing the role of medical intervention in the decades after 1948
- The major drivers were:
  - Academic proselytizers: Sackett, Chalmers X2, etc.
  - Government funding bodies: UK MRC, then NIH
  - FDA

Academic proselytizers

Clinical epidemiologists, meanwhile, promoted RCTs as the best means to make medicine more rational. By the early 1980s, they had labeled RCTs the gold standard of medical knowledge. As evidence-based medicine rose to prominence in ensuing decades, methodologic hierarchies emerged, with case reports at the bottom and RCTs at the top.

Bothwell June 2016

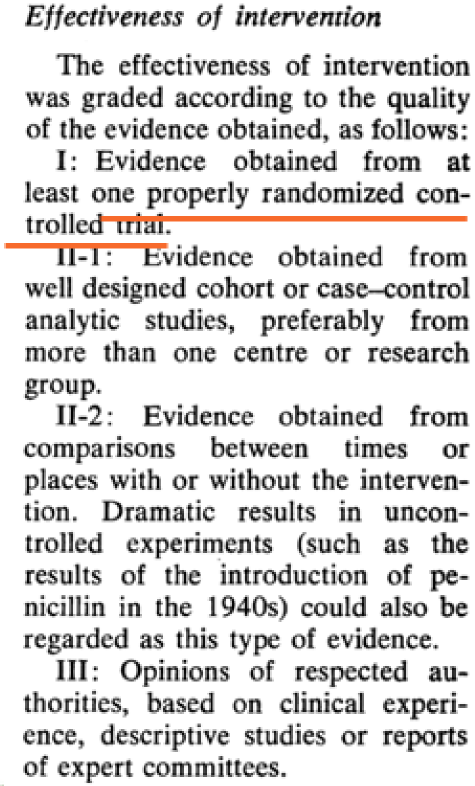
Levels of Evidence

The periodic health examination

Canadian Task Force on the Periodic Health Examination*

CMAJ November 1979

Walter Spitzer … Suzanne Fletcher … David Sackett, et al



Effectiveness of intervention

The effectiveness of intervention was graded according to the quality of the evidence obtained, as follows:

I: Evidence obtained from at least one properly randomized controlled trial.

II-1: Evidence obtained from well designed cohort or case–control analytic studies, preferably from more than one centre or research group.

II-2: Evidence obtained from comparisons between times or places with or without the intervention. Dramatic results in uncontrolled experiments (such as the results of the introduction of penicillin in the 1940s) could also be regarded as this type of evidence.

III: Opinions of respected authorities, based on clinical experience, descriptive studies or reports of expert committees.

Government funding

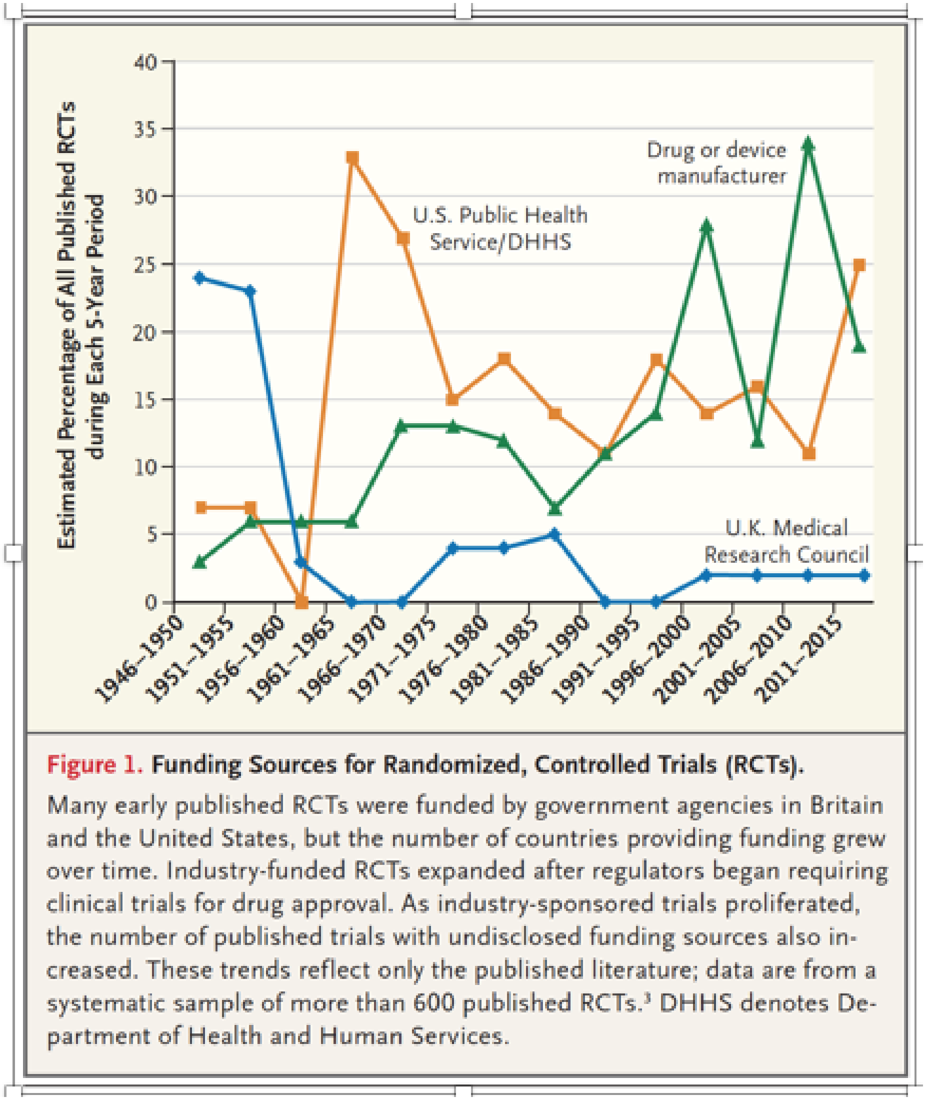
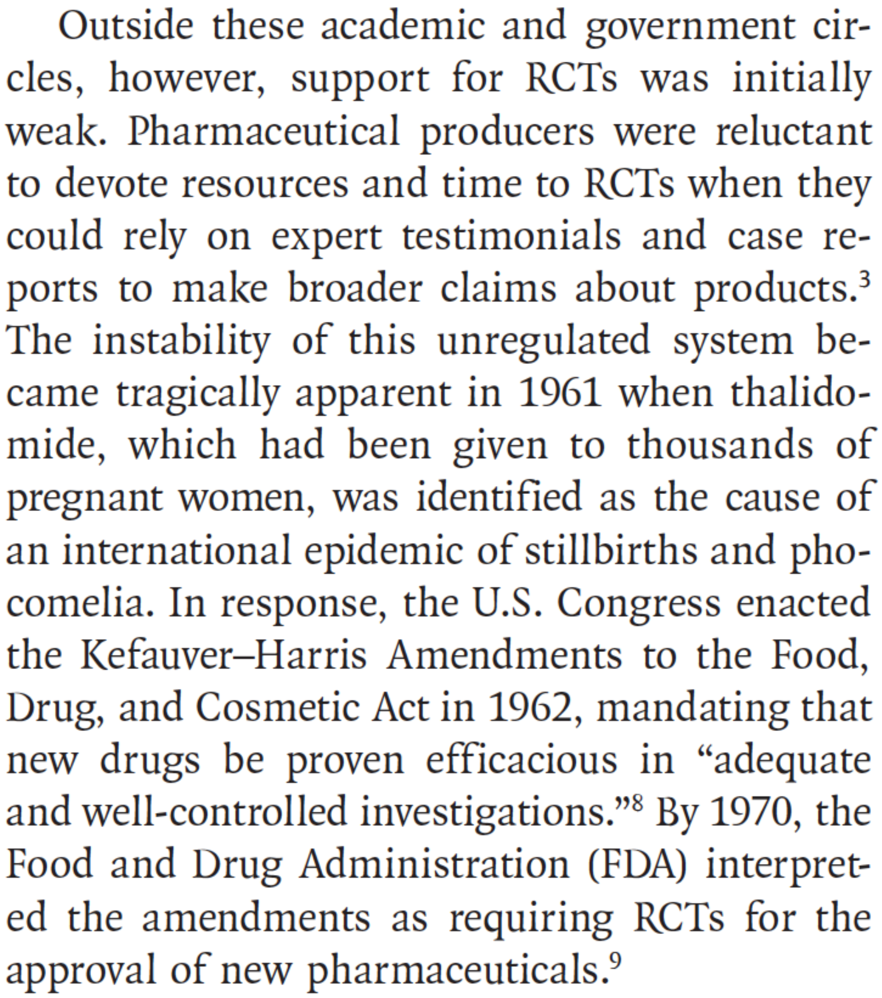


Figure 1. Funding Sources for Randomized, Controlled Trials (RCTs).

Many early published RCTs were funded by government agencies in Britain and the United States, but the number of countries providing funding grew over time. Industry-funded RCTs expanded after regulators began requiring clinical trials for drug approval. As industry-sponsored trials proliferated, the number of published trials with undisclosed funding sources also increased. These trends reflect only the published literature; data are from a systematic sample of more than 600 published RCTs.[3] DHHS denotes Department of Health and Human Services.

Bothwell August 2016

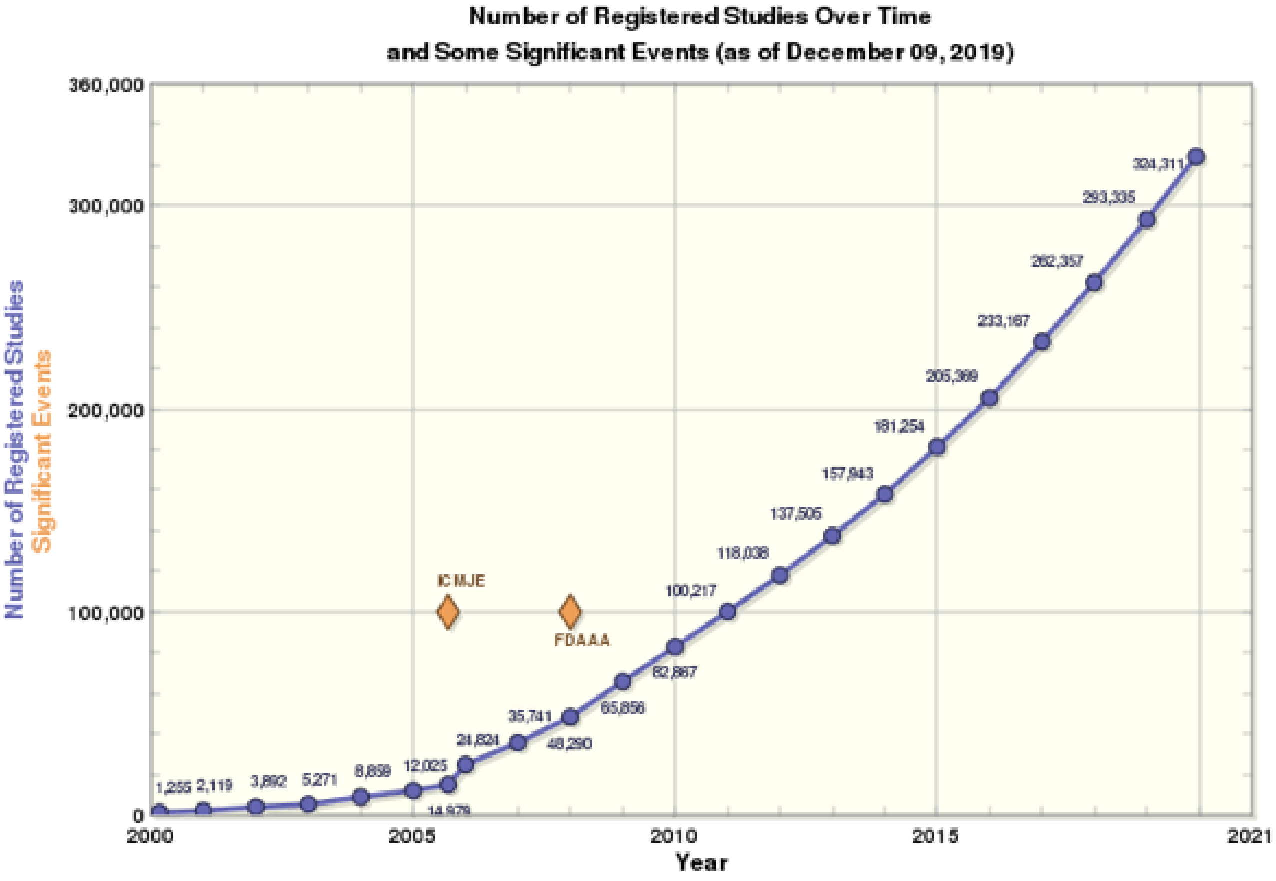
FDA



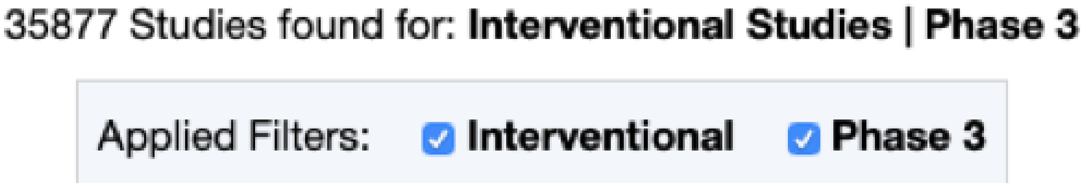
Outside these academic and government circles, however, support for RCTs was initially weak. Pharmaceutical producers were reluctant to devote resources and time to RCTs when they could rely on expert testimonials and case reports to make broader claims about products.[3] The instability of this unregulated system became tragically apparent in 1961 when thalidomide, which had been given to thousands of pregnant women, was identified as the cause of an international epidemic of stillbirths and phocomelia. In response, the U.S. Congress enacted the Kefauver–Harris Amendments to the Food, Drug, and Cosmetic Act in 1962, mandating that new drugs be proven efficacious in "adequate and well-controlled investigations."[8] By 1970, the Food and Drug Administration (FDA) interpreted the amendments as requiring RCTs for the approval of new pharmaceuticals.[9]

Bothwell June 2016

Growth of RCTs in health



Source: https://ClinicalTrials.gov



35877 Studies found for: **Interventional Studies | Phase 3**

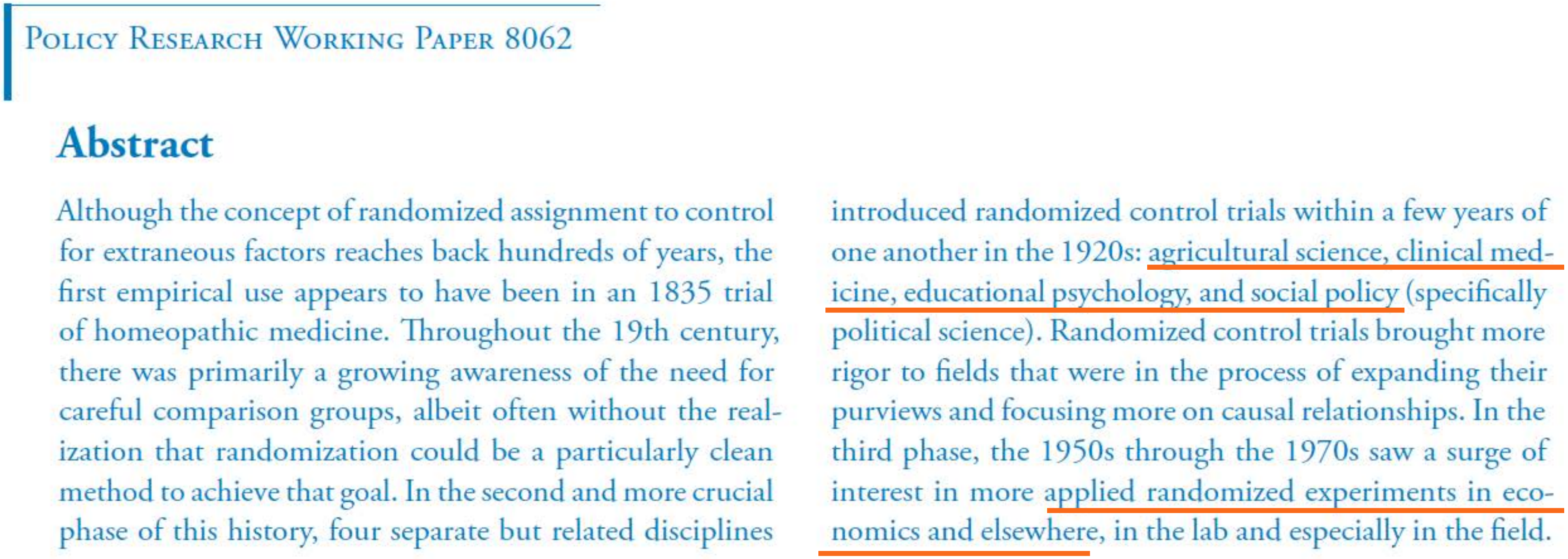
Applied Filters: ☑ **Interventional** ☑ **Phase 3**

RCTs in Other Fields

Policy Research Working Paper 8062

The Entry of Randomized Assignment into the Social Sciences

Julian C. Jamison



Policy Research Working Paper 8062

Abstract

Although the concept of randomized assignment to control for extraneous factors reaches back hundreds of years, the first empirical use appears to have been in an 1835 trial of homeopathic medicine. Throughout the 19th century, there was primarily a growing awareness of the need for careful comparison groups, albeit often without the realization that randomization could be a particularly clean method to achieve that goal. In the second and more crucial phase of this history, four separate but related disciplines introduced randomized control trials within a few years of one another in the 1920s: agricultural science, clinical medicine, educational psychology, and social policy (specifically political science). Randomized control trials brought more rigor to fields that were in the process of expanding their purviews and focusing more on causal relationships. In the third phase, the 1950s through the 1970s saw a surge of interest in more applied randomized experiments in economics and elsewhere, in the lab and especially in the field.

Scientific Background on the Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2019

UNDERSTANDING DEVELOPMENT AND POVERTY ALLEVIATION

The Committee for the Prize in Economic Sciences in Memory of Alfred Nobel

(From left) Nobel Laureates in Economic Sciences Michael Kremer, Esther Duflo, and Abhijit Banerjee attend a press conference at The Royal Swedish Academy of Sciences in Stockholm, Sweden, on Dec. 7, 2019. JONAS EKSTROMER/TT NEWS AGENCY/AFP/GETTY IMAGES

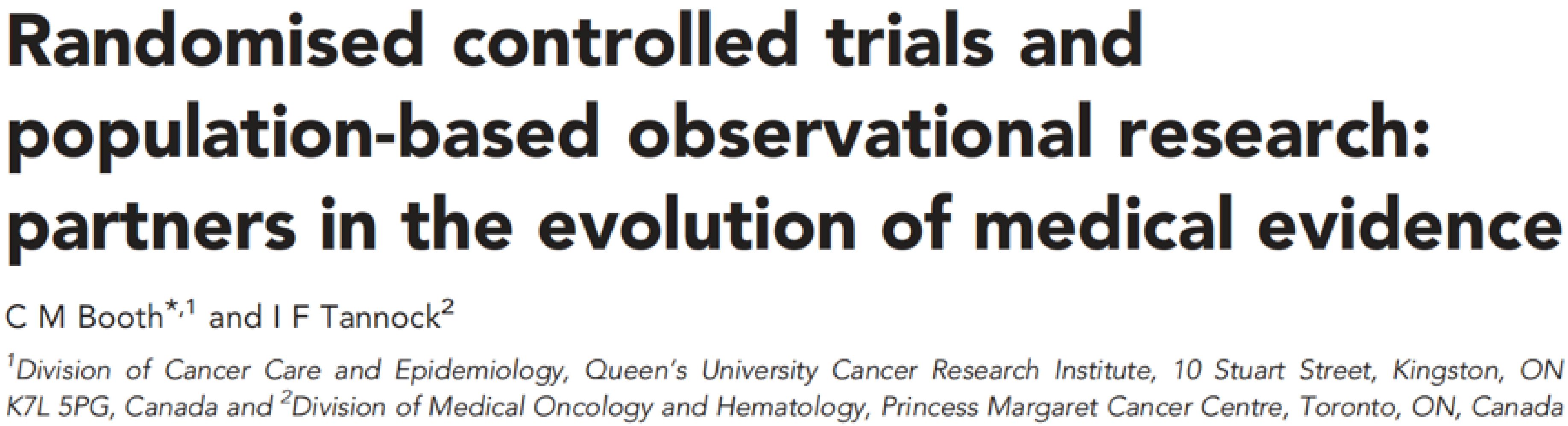
The modern approach to development economics relies on two simple but powerful ideas. One idea is that empirical micro-level studies guided by economic theory can provide crucial insights into the design of policies for effective poverty alleviation. The other is that the best way to draw precise conclusions about the true path from causes to effects is often to conduct a randomized controlled field trial. The systematic application of these ideas over the past 20 years has paved the way for the transformation of development research.

Where Are We Going?

Where are we going?

- Can RCTs be replaced?
  - The “causal revolution”
    - Can we derive reliable information on the effects of intervention from observational data?
- Are we doing the right kind of RCTs?
  - Right endpoints?
  - Right-sized trials?
  - Right populations?
- These questions are inter-related
  - Although it seems illogical, will deal with second set first and then consider whether part of the long-term solution lies in the answer to the first.

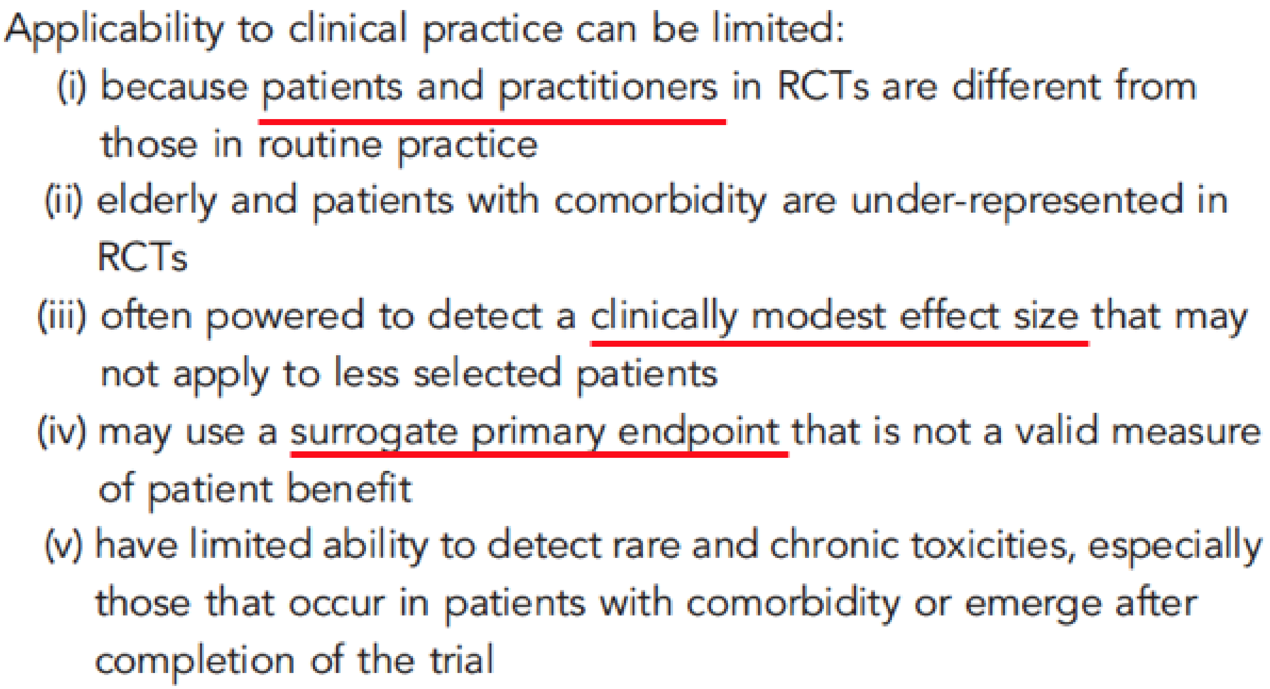
Right Kinds of RCTs?



Randomised controlled trials and population-based observational research: partners in the evolution of medical evidence

C M Booth*,[1] and I F Tannock[2]

[1]Division of Cancer Care and Epidemiology, Queen's University Cancer Research Institute, 10 Stuart Street, Kingston, ON K7L 5PG, Canada and [2]Division of Medical Oncology and Hematology, Princess Margaret Cancer Centre, Toronto, ON, Canada



Applicability to clinical practice can be limited:

(i) because patients and practitioners in RCTs are different from those in routine practice

(ii) elderly and patients with comorbidity are under-represented in RCTs

(iii) often powered to detect a clinically modest effect size that may not apply to less selected patients

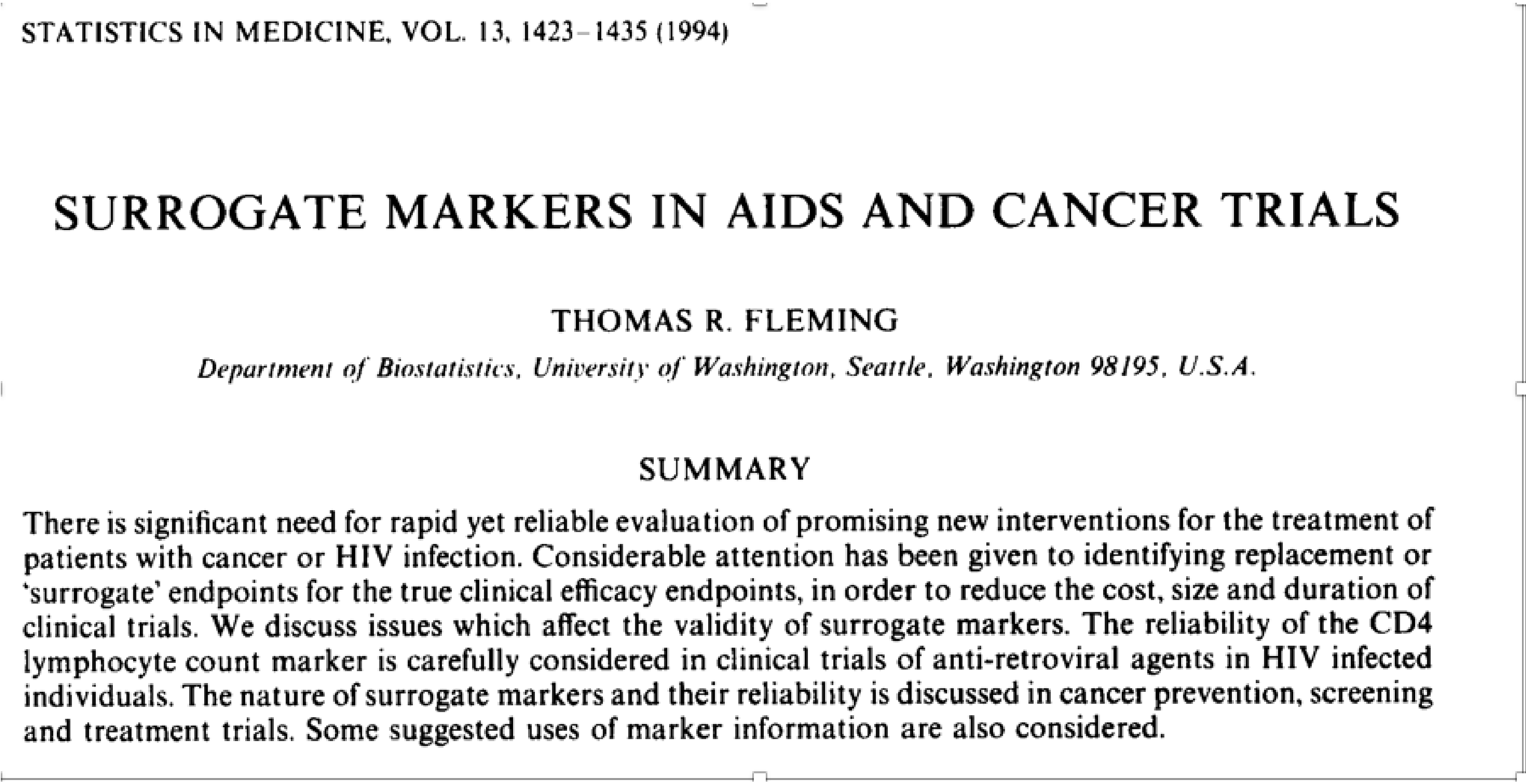
(iv) may use a surrogate primary endpoint that is not a valid measure of patient benefit

(v) have limited ability to detect rare and chronic toxicities, especially those that occur in patients with comorbidity or emerge after completion of the trial

Canadian Cancer Trials Group
Groupe canadien des essais sur le cancer

Right Endpoints

The Surrogate Endpoint Problem



STATISTICS IN MEDICINE, VOL. 13, 1423–1435 (1994)

SURROGATE MARKERS IN AIDS AND CANCER TRIALS

THOMAS R. FLEMING

Department of Biostatistics, University of Washington, Seattle, Washington 98195, U.S.A.

SUMMARY

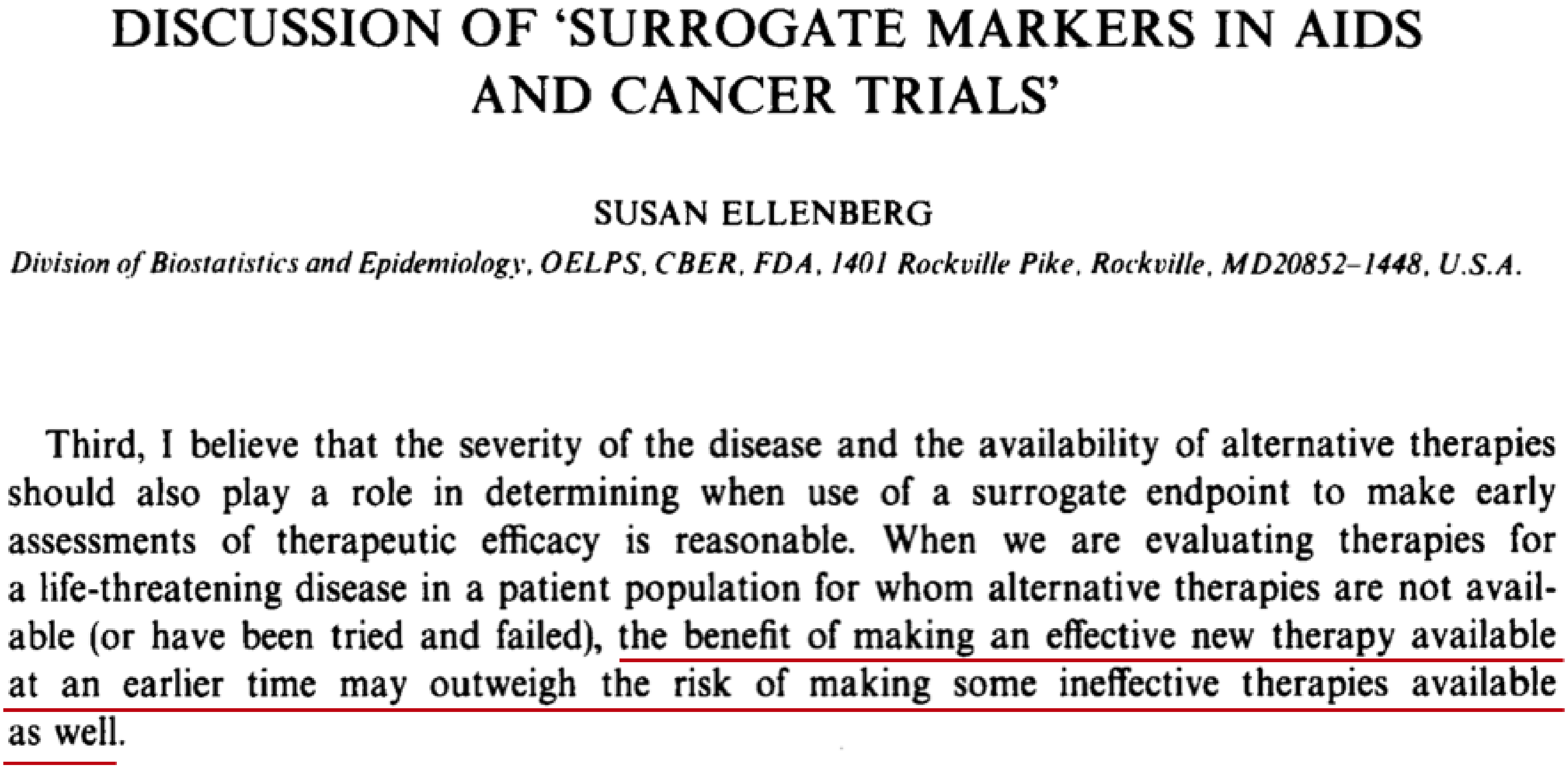
There is significant need for rapid yet reliable evaluation of promising new interventions for the treatment of patients with cancer or HIV infection. Considerable attention has been given to identifying replacement or 'surrogate' endpoints for the true clinical efficacy endpoints, in order to reduce the cost, size and duration of clinical trials. We discuss issues which affect the validity of surrogate markers. The reliability of the CD4 lymphocyte count marker is carefully considered in clinical trials of anti-retroviral agents in HIV infected individuals. The nature of surrogate markers and their reliability is discussed in cancer prevention, screening and treatment trials. Some suggested uses of marker information are also considered.

In life threatening diseases there often is a sense of urgency for rapid yet reliable evaluation of promising new interventions. Trials using patient survival as the primary endpoint frequently require very lengthy follow-up intervals and large numbers of patients. The QOL assessments are made using very subjective outcome measures, and thus provide additional difficulties through the need to identify validated and widely accepted QOL instruments which can be uniformly completed across study centres. To reduce the trial cost, size and duration and to avoid complexities of QOL assessments, considerable attention has been given, in the design of definitive phase I11 trials, to identifying surrogate or replacement endpoints for the true clinical efficacy endpoint. Measures of biological activity, such as tumour shrinkage or CD4 lymphocyte count, have been frequently chosen surrogates because this information usually is readily available, requiring relatively brief follow-up, and because observational databases reveal that these surrogate variables are strong predictors of the true clinical efficacy endpoints.

Unfortunately, as illustrated and discussed in Fleming,’.’ treatment effects on the true clinical efficacy endpoints may not be reliably predicted by the observed effects on replacement endpoints (to be called surrogate or biological markers), even when natural history data reveal these surrogate markers are strongly correlated with the longer term true clinical efficacy outcomes.

Points

- Surrogates can be misleading
- Two ways surrogates could be useful
  - As a predictor of a future clinically relevant event (common usage)
  - As an indirect indicator of a concurrent clinically relevant endpoint
    - E.G., disease progression could in itself be associated with QOL deterioration



DISCUSSION OF 'SURROGATE MARKERS IN AIDS AND CANCER TRIALS'

SUSAN ELLENBERG

Division of Biostatistics and Epidemiology, OELPS, CBER, FDA, 1401 Rockville Pike, Rockville, MD20852–1448, U.S.A.

Third, I believe that the severity of the disease and the availability of alternative therapies should also play a role in determining when use of a surrogate endpoint to make early assessments of therapeutic efficacy is reasonable. When we are evaluating therapies for a life-threatening disease in a patient population for whom alternative therapies are not available (or have been tried and failed), the benefit of making an effective new therapy available at an earlier time may outweigh the risk of making some ineffective therapies available as well.

What happened?

Is the number of cancer drug approvals a surrogate for regulatory success?

Bishal Gyawali[a,b,*], Shubham Sharma[a], Christopher M. Booth[a]

[a] *Division of Cancer Care and Epidemiology, Cancer Research Institute, Queen's University, Canada*
[b] *Program on Regulation, Therapeutics and Law, Brigham and Women's Hospital, Boston, MA, United States*

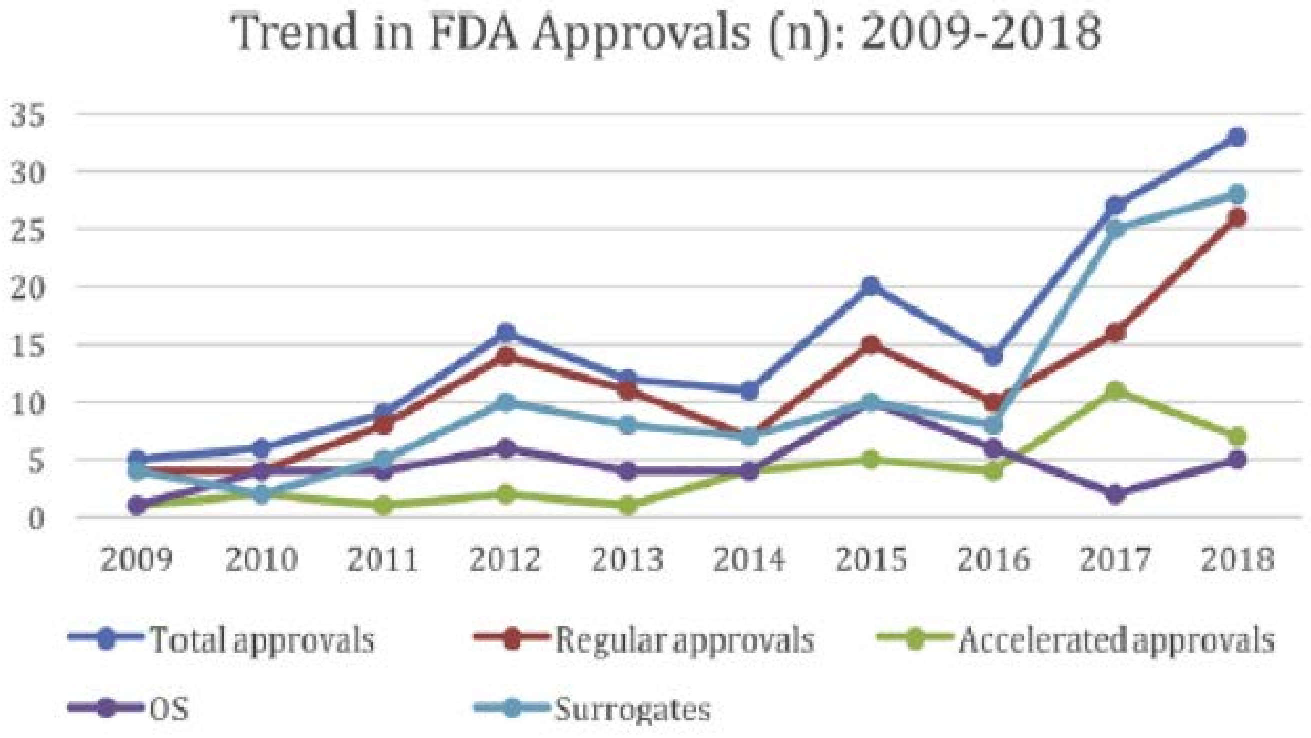
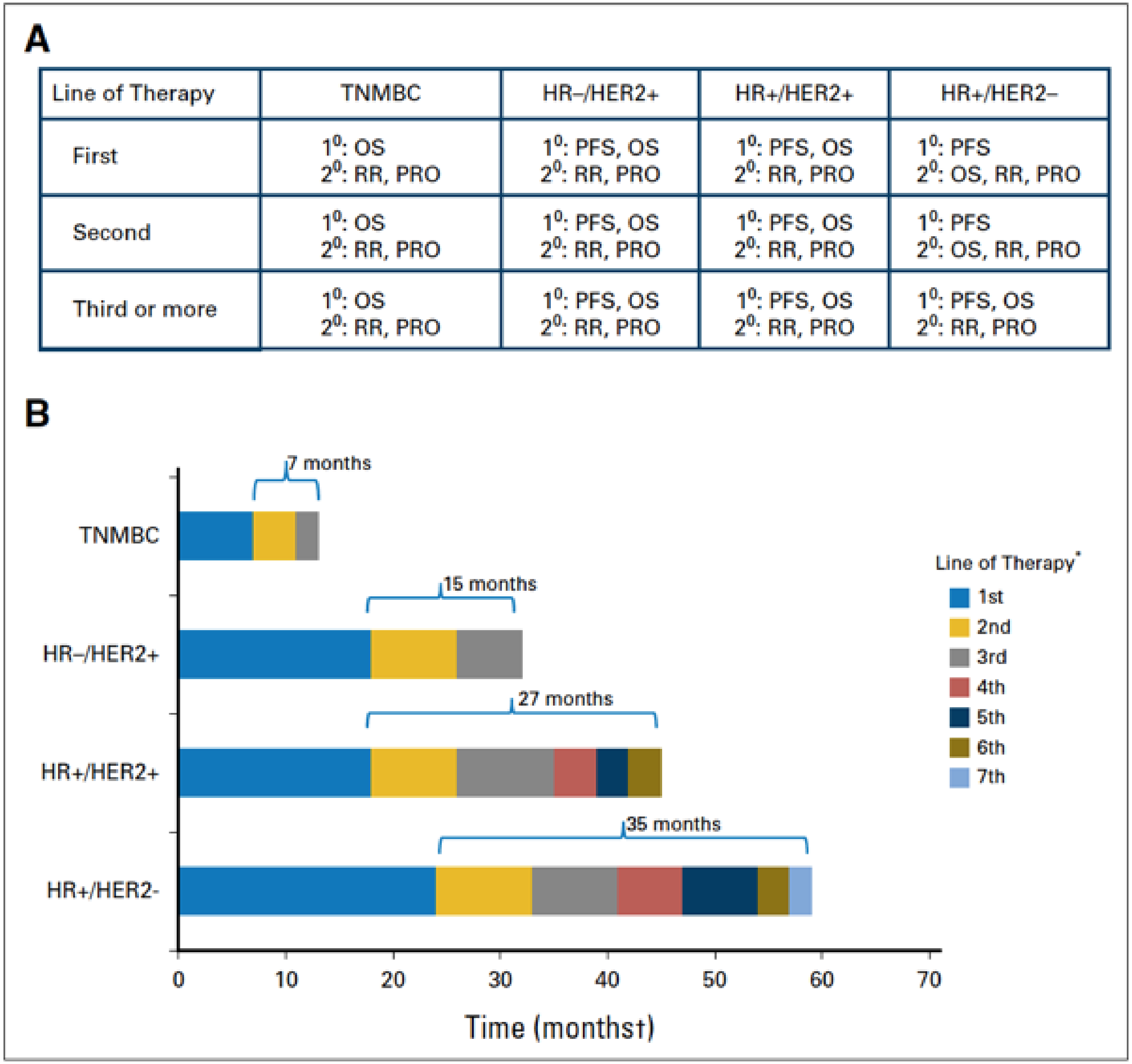


Fig. 1. Trend in FDA approvals of cancer drugs.

National Cancer Institute Breast Cancer Steering Committee Working Group Report on Meaningful and Appropriate End Points for Clinical Trials in Metastatic Breast Cancer

Andrew D. Seidman, Louise Bordeleau, Louis Fehrenbacher, William E. Barlow, Jane Perlmutter, Lawrence Rubinstein, Suparna B. Wedam, Dawn L. Hershman, Jennifer Fallas Hayes, Lynn Pearson Butler, Mary Lou Smith, Meredith M. Regan, Julia A. Beaver, Laleh Amiri-Kordestani, Priya Rastogi, Jo Anne Zujewski, and Larissa A. Korde



A

Line of Therapy	TNMBC	HR–/HER2+	HR+/HER2+	HR+/HER2–
First	1^0: OS 2^0: RR, PRO	1^0: PFS, OS 2^0: RR, PRO	1^0: PFS, OS 2^0: RR, PRO	1^0: PFS 2^0: OS, RR, PRO
Second	1^0: OS 2^0: RR, PRO	1^0: PFS, OS 2^0: RR, PRO	1^0: PFS, OS 2^0: RR, PRO	1^0: PFS 2^0: OS, RR, PRO
Third or more	1^0: OS 2^0: RR, PRO	1^0: PFS, OS 2^0: RR, PRO	1^0: PFS, OS 2^0: RR, PRO	1^0: PFS, OS 2^0: RR, PRO

Fig 2. (A) Working group consensus on preferred end points by biologic subtype and line of therapy. (B) Hypothetical scenarios for expected postprogression survival (PPS) and choice of preferred end point. In settings such as first-line treatment of triple-negative metastatic breast cancer (TNMBC) where expected PPS is < 12 months, overall survival (OS) is the preferred primary end point. In settings such as hormone receptor–negative (HR–)/human epidermal growth factor receptor 2–positive (HER2+) or HR+/HER2+ MBC where PPS is > 12 months, in both the first- and later-line settings, progression-free survival (PFS) is the end point of choice, and OS could be considered as a coprimary end point. In settings such as HR+/HER2– MBC, given the expected long PPS, PFS is the most appropriate end point. When such patients have disease that is refractory to endocrine therapy and have been exposed to several lines of chemotherapy, where PPS is expected to be much shorter, OS may be the most meaningful and appropriate end point. (*) Line of therapy may be endocrine therapy, chemotherapy, HER2-targeted therapy, combinations, and so forth. (†) Months shown are for illustrative purposes only. 1°, primary end point; 2°, secondary end point; PRO, patient-reported outcome; RR, response rate.

Where do go from here?

- Drug approvals based on surrogate endpoints like PFS are not going to go away in the foreseeable future
- Similarly, funding decisions are being made on the basis of data from trials where PFS was the primary endpoint
- Options (not mutually exclusive)
  - Continue to try to persuade regulatory agencies and funders to use robust endpoints in decision-making
  - Try to define circumstances where PFS prolongation might be clinically meaningful, not as a surrogate for OS, but as an indicator of patient benefit or as something that patients value in itself

Right Sizing Trials

Have cancer trials gotten too large?

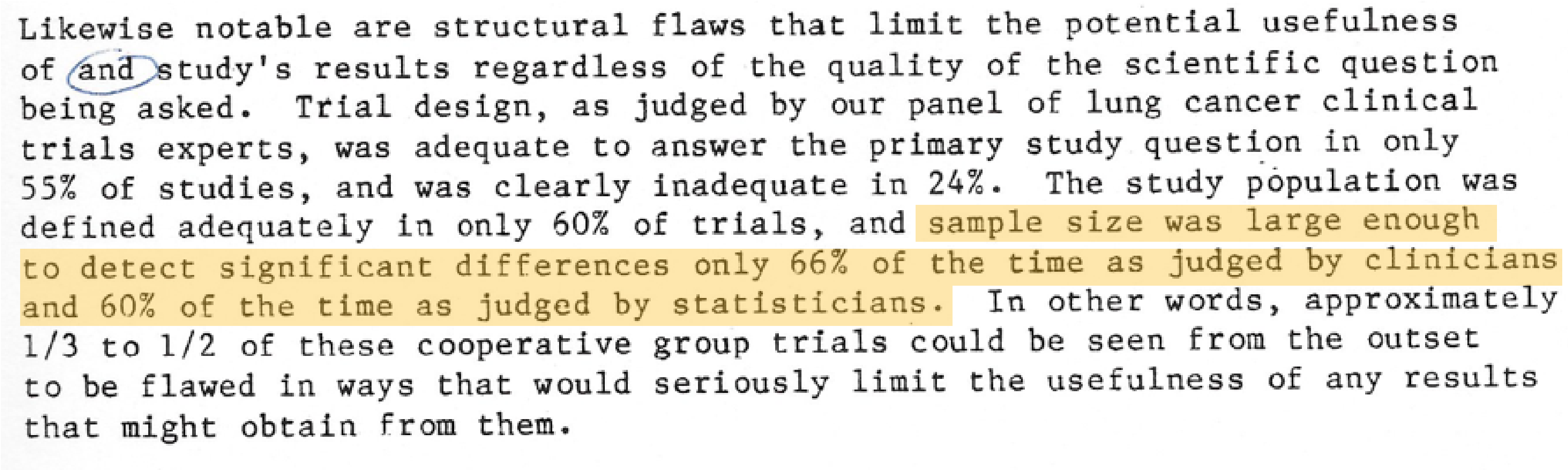
And, if so, what should we do about it?

INTRODUCTION

The performance of the CTEP and the cooperative clinical trials groups over the past eight years has recently been assessed by peer review panels for selected disease sites. The purpose of these reviews has been to identify deficiencies in the mechanisms of conducting clinical trials, as well as possible flaws specific to the individual diseases. The intent specifically has not been to evaluate or criticize the performance of any single group. Following is a summary of the clinical trials review in non-small cell lung cancer, conducted June 6-7, 1985.

METHODS

Reviewers. The panel consisted of three medical oncologists (Drs. Robert Livingston, Ronald Natale, John Ruckdeschel), two statisticians (Drs. Judith O'Fallon and Joseph Pater), two radiation oncologists (Drs. Zvi Fuks and John Earle), and two surgical oncologists (Drs. Martin McKneally and Barry Kahan). All are actively involved in clinical trials in lung cancer.

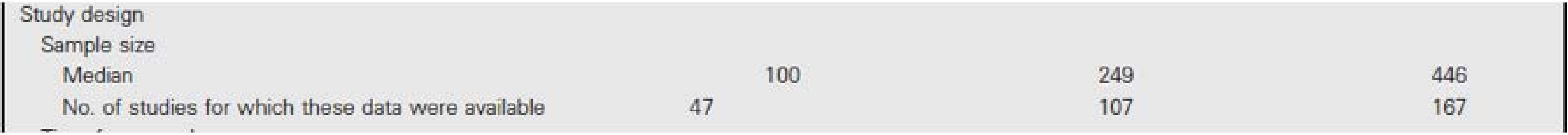


Likewise notable are structural flaws that limit the potential usefulness of and study's results regardless of the quality of the scientific question being asked. Trial design, as judged by our panel of lung cancer clinical trials experts, was adequate to answer the primary study question in only 55% of studies, and was clearly inadequate in 24%. The study population was defined adequately in only 60% of trials, and sample size was large enough to detect significant differences only 66% of the time as judged by clinicians and 60% of the time as judged by statisticians. In other words, approximately 1/3 to 1/2 of these cooperative group trials could be seen from the outset to be flawed in ways that would seriously limit the usefulness of any results that might obtain from them.

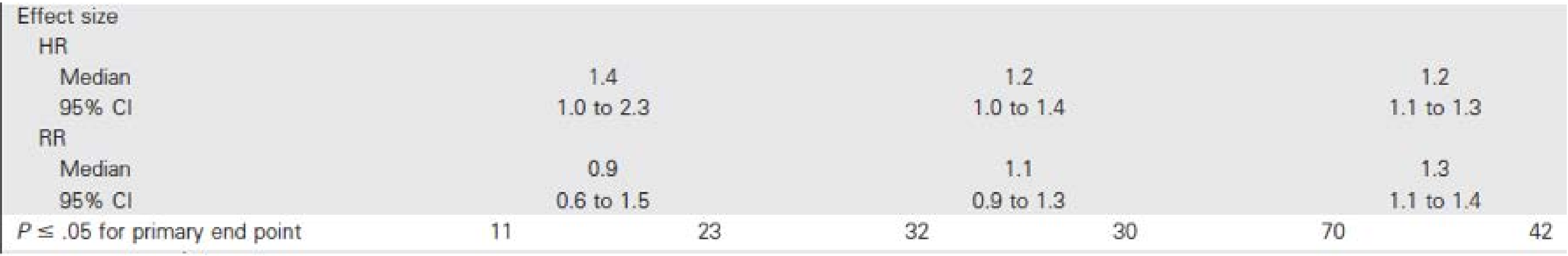
Evolution of the Randomized Controlled Trial in Oncology Over Three Decades

Christopher M. Booth, David W. Cescon, Lisa Wang, Ian F. Tannock, and Monika K. Krzyzanowska

Canadian Cancer Trials Group
Groupe canadien des essais sur le cancer

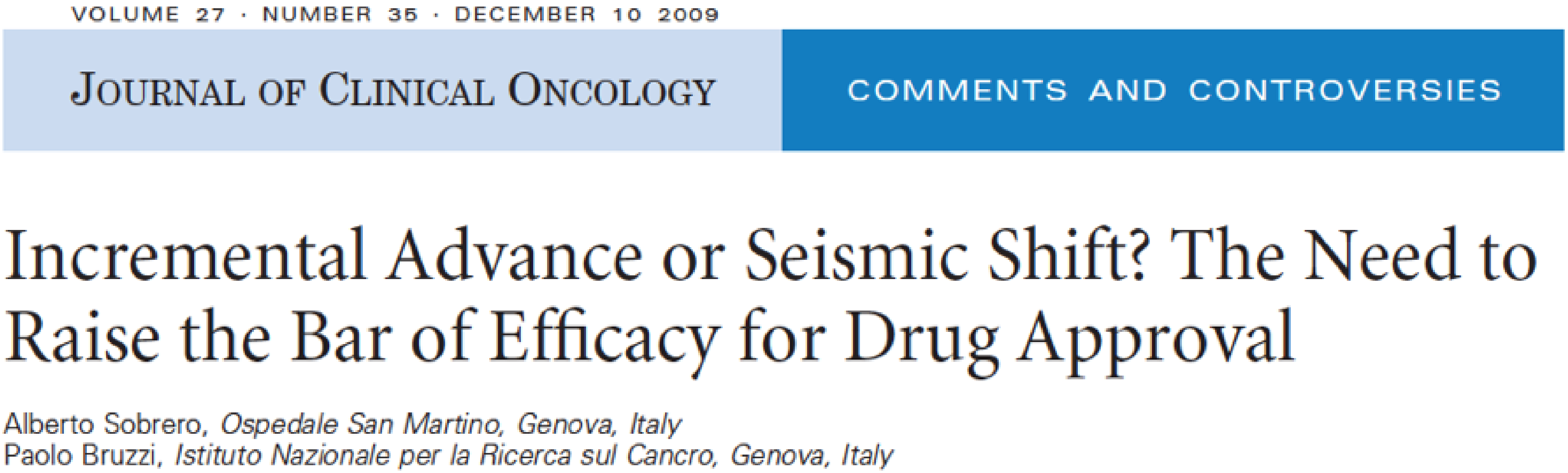


Study design			
Sample size			
Median	100	249	446
No. of studies for which these data were available	47	107	167



Effect size						
HR						
Median	1.4		1.2		1.2	
95% CI	1.0 to 2.3		1.0 to 1.4		1.1 to 1.3	
RR						
Median	0.9		1.1		1.3	
95% CI	0.6 to 1.5		0.9 to 1.3		1.1 to 1.4	
P ≤ .05 for primary end point	11	23	32	30	70	42

Finally, clinicians, investigators, and policy makers should maintain and refine perspective on what constitutes a meaningful benefit to patients beyond the P value associated with the result. Further research is needed to determine whether newly adopted therapies are truly worthwhile to patients.



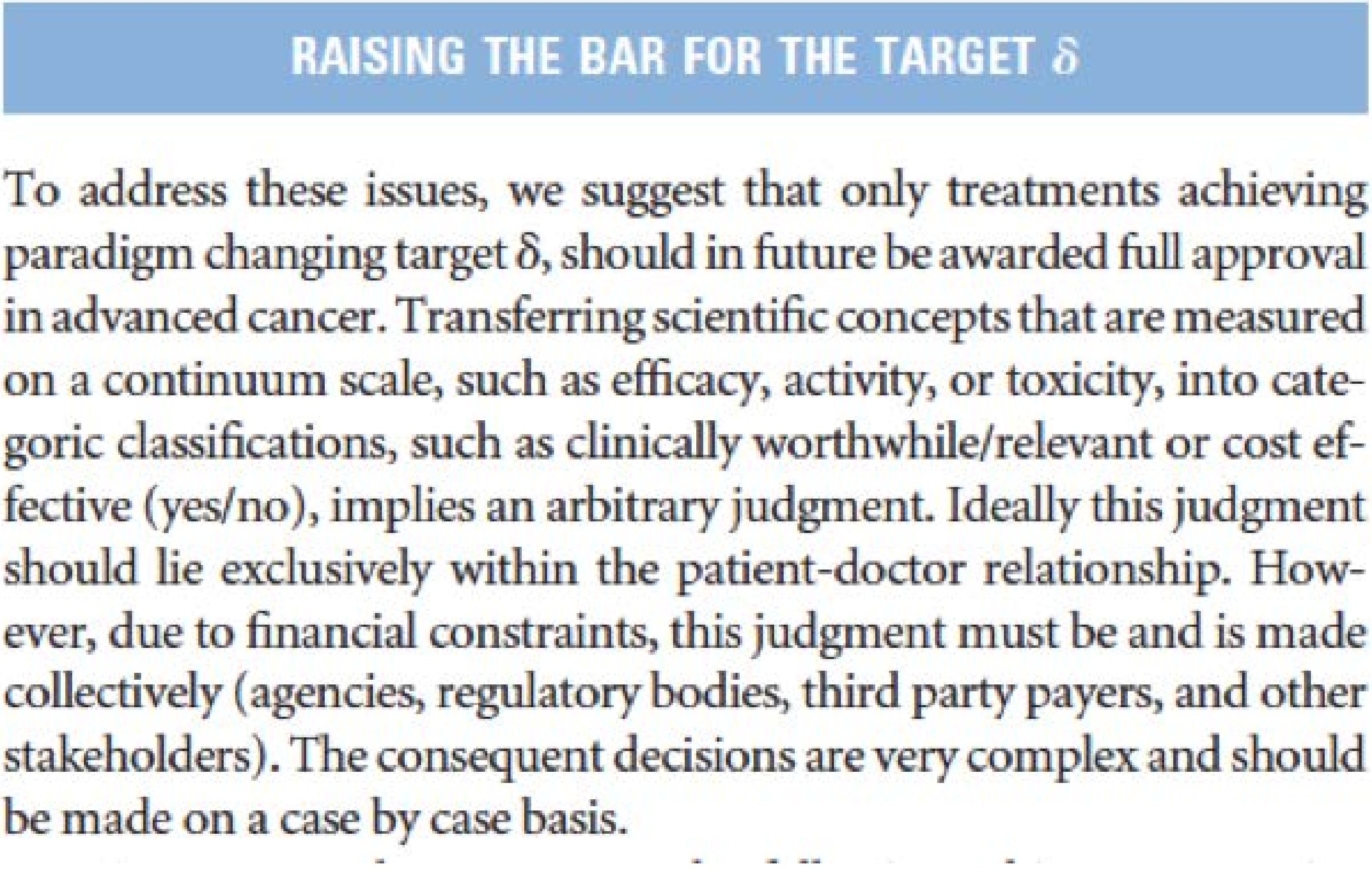
VOLUME 27 · NUMBER 35 · DECEMBER 10 2009

JOURNAL OF CLINICAL ONCOLOGY

COMMENTS AND CONTROVERSIES

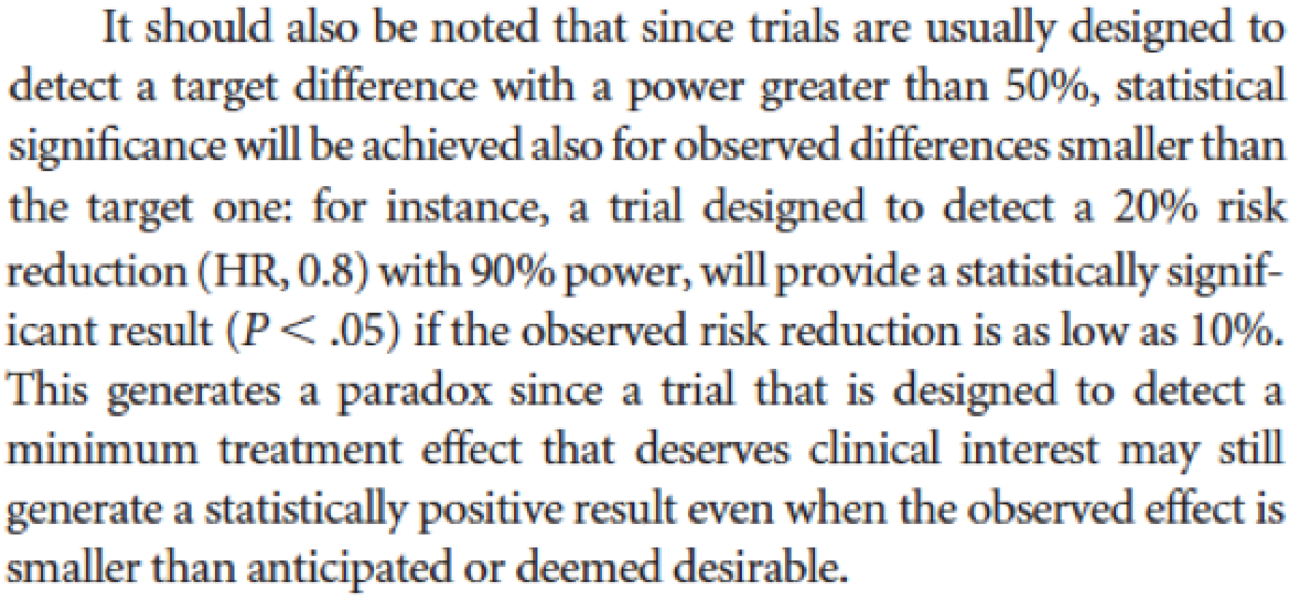
Incremental Advance or Seismic Shift? The Need to Raise the Bar of Efficacy for Drug Approval

Alberto Sobrero, *Ospedale San Martino, Genova, Italy*
Paolo Bruzzi, *Istituto Nazionale per la Ricerca sul Cancro, Genova, Italy*

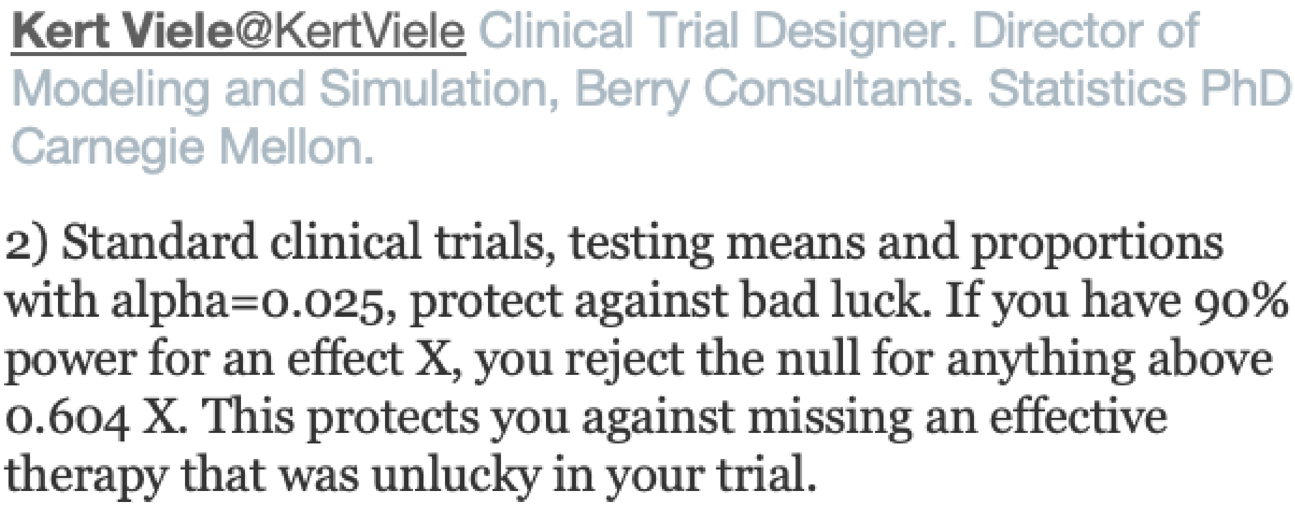


RAISING THE BAR FOR THE TARGET δ

To address these issues, we suggest that only treatments achieving paradigm changing target δ, should in future be awarded full approval in advanced cancer. Transferring scientific concepts that are measured on a continuum scale, such as efficacy, activity, or toxicity, into categoric classifications, such as clinically worthwhile/relevant or cost effective (yes/no), implies an arbitrary judgment. Ideally this judgment should lie exclusively within the patient-doctor relationship. However, due to financial constraints, this judgment must be and is made collectively (agencies, regulatory bodies, third party payers, and other stakeholders). The consequent decisions are very complex and should be made on a case by case basis.

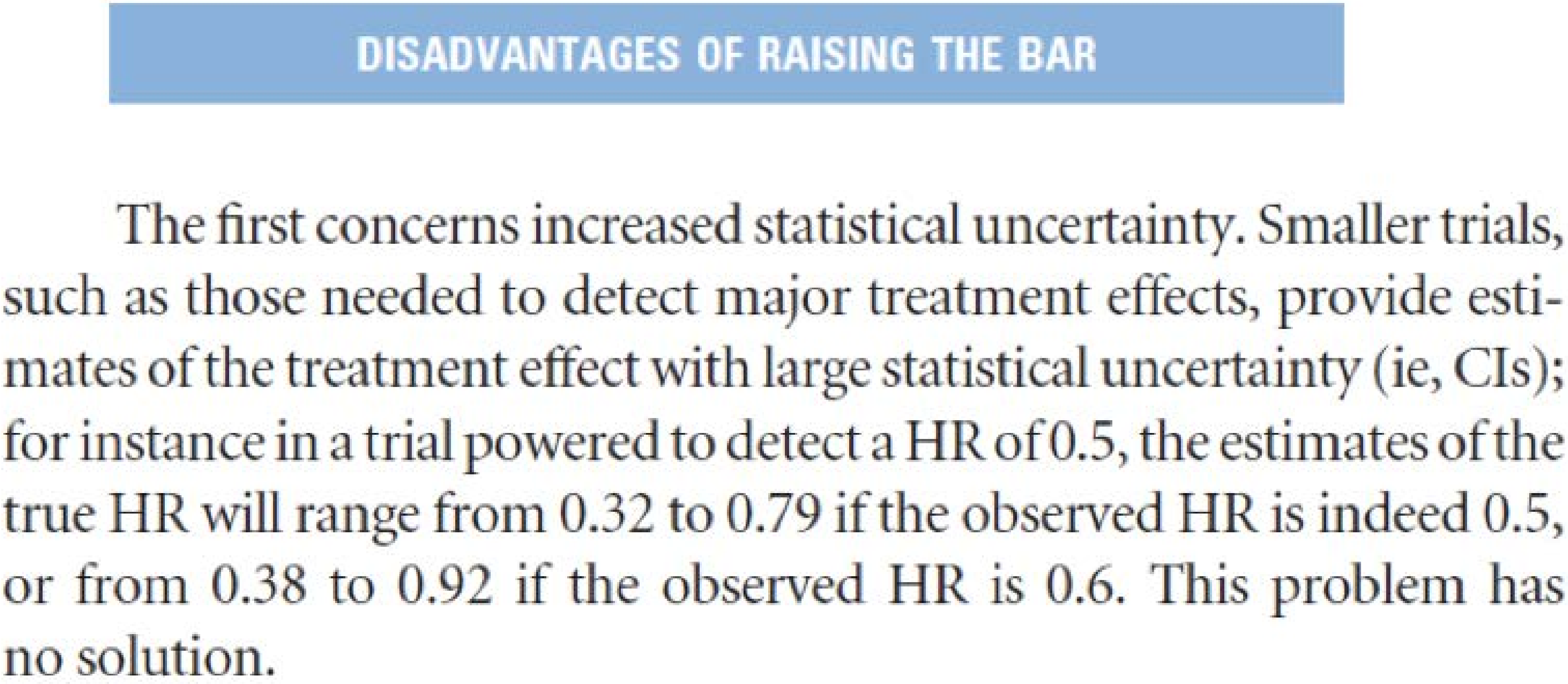


It should also be noted that since trials are usually designed to detect a target difference with a power greater than 50%, statistical significance will be achieved also for observed differences smaller than the target one: for instance, a trial designed to detect a 20% risk reduction (HR, 0.8) with 90% power, will provide a statistically significant result ($P < .05$) if the observed risk reduction is as low as 10%. This generates a paradox since a trial that is designed to detect a minimum treatment effect that deserves clinical interest may still generate a statistically positive result even when the observed effect is smaller than anticipated or deemed desirable.



Kert Viele@KertViele Clinical Trial Designer. Director of Modeling and Simulation, Berry Consultants. Statistics PhD Carnegie Mellon.

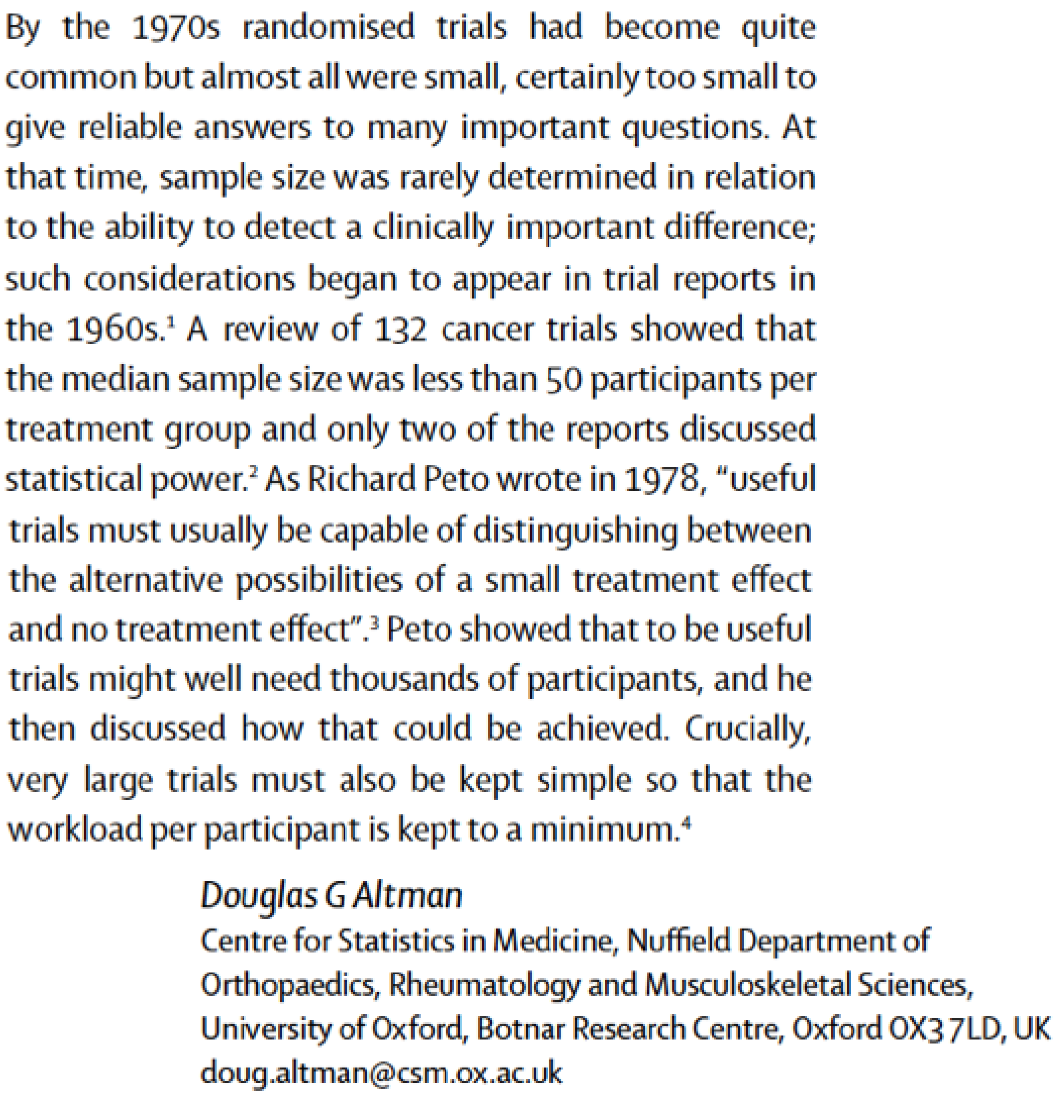
2) Standard clinical trials, testing means and proportions with alpha=0.025, protect against bad luck. If you have 90% power for an effect X, you reject the null for anything above 0.604 X. This protects you against missing an effective therapy that was unlucky in your trial.



DISADVANTAGES OF RAISING THE BAR

The first concerns increased statistical uncertainty. Smaller trials, such as those needed to detect major treatment effects, provide estimates of the treatment effect with large statistical uncertainty (ie, CIs); for instance in a trial powered to detect a HR of 0.5, the estimates of the true HR will range from 0.32 to 0.79 if the observed HR is indeed 0.5, or from 0.38 to 0.92 if the observed HR is 0.6. This problem has no solution.

Another (British) perspective



By the 1970s randomised trials had become quite common but almost all were small, certainly too small to give reliable answers to many important questions. At that time, sample size was rarely determined in relation to the ability to detect a clinically important difference; such considerations began to appear in trial reports in the 1960s.[1] A review of 132 cancer trials showed that the median sample size was less than 50 participants per treatment group and only two of the reports discussed statistical power.[2] As Richard Peto wrote in 1978, "useful trials must usually be capable of distinguishing between the alternative possibilities of a small treatment effect and no treatment effect".[3] Peto showed that to be useful trials might well need thousands of participants, and he then discussed how that could be achieved. Crucially, very large trials must also be kept simple so that the workload per participant is kept to a minimum.[4]

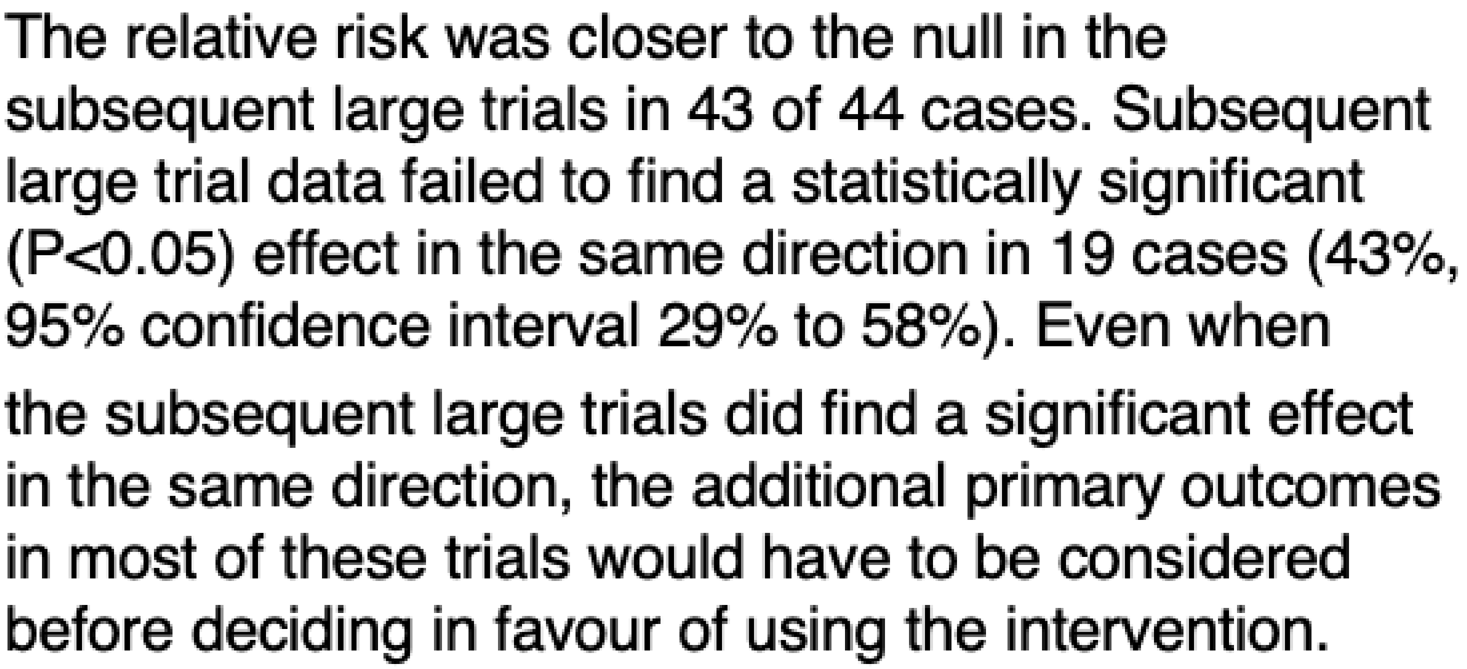
Douglas G Altman
Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Botnar Research Centre, Oxford OX3 7LD, UK
doug.altman@csm.ox.ac.uk

Very large treatment effects in randomised trials as an empirical marker to indicate whether subsequent trials are necessary: meta-epidemiological assessment

Myura Nagendran,[1] Tiago V Pereira,[2] Grace Kiew,[3] Douglas G Altman,[4] Mahiben Maruthappu,[5] John P A Ioannidis,[6] Peter McCulloch[7]

thebmj | *BMJ* 2016;355:i5432 | doi: 10.1136/bmj.i5432

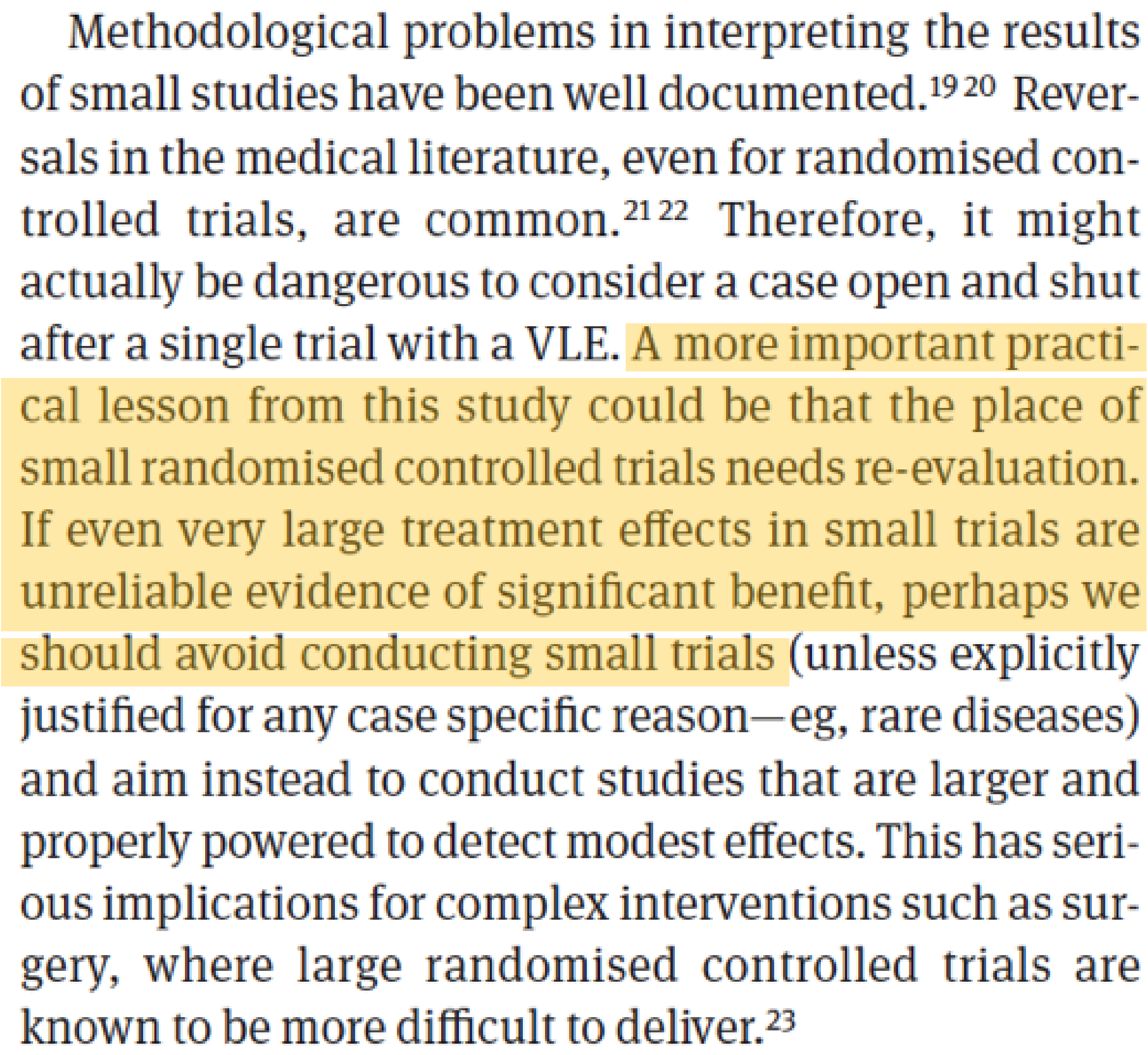
Results



The relative risk was closer to the null in the subsequent large trials in 43 of 44 cases. Subsequent large trial data failed to find a statistically significant ($P<0.05$) effect in the same direction in 19 cases (43%, 95% confidence interval 29% to 58%). Even when

the subsequent large trials did find a significant effect in the same direction, the additional primary outcomes in most of these trials would have to be considered before deciding in favour of using the intervention.

Conclusion



Methodological problems in interpreting the results of small studies have been well documented.[19,20] Reversals in the medical literature, even for randomised controlled trials, are common.[21,22] Therefore, it might actually be dangerous to consider a case open and shut after a single trial with a VLE. A more important practical lesson from this study could be that the place of small randomised controlled trials needs re-evaluation. If even very large treatment effects in small trials are unreliable evidence of significant benefit, perhaps we should avoid conducting small trials (unless explicitly justified for any case specific reason—eg, rare diseases) and aim instead to conduct studies that are larger and properly powered to detect modest effects. This has serious implications for complex interventions such as surgery, where large randomised controlled trials are known to be more difficult to deliver.[23]

Way forward

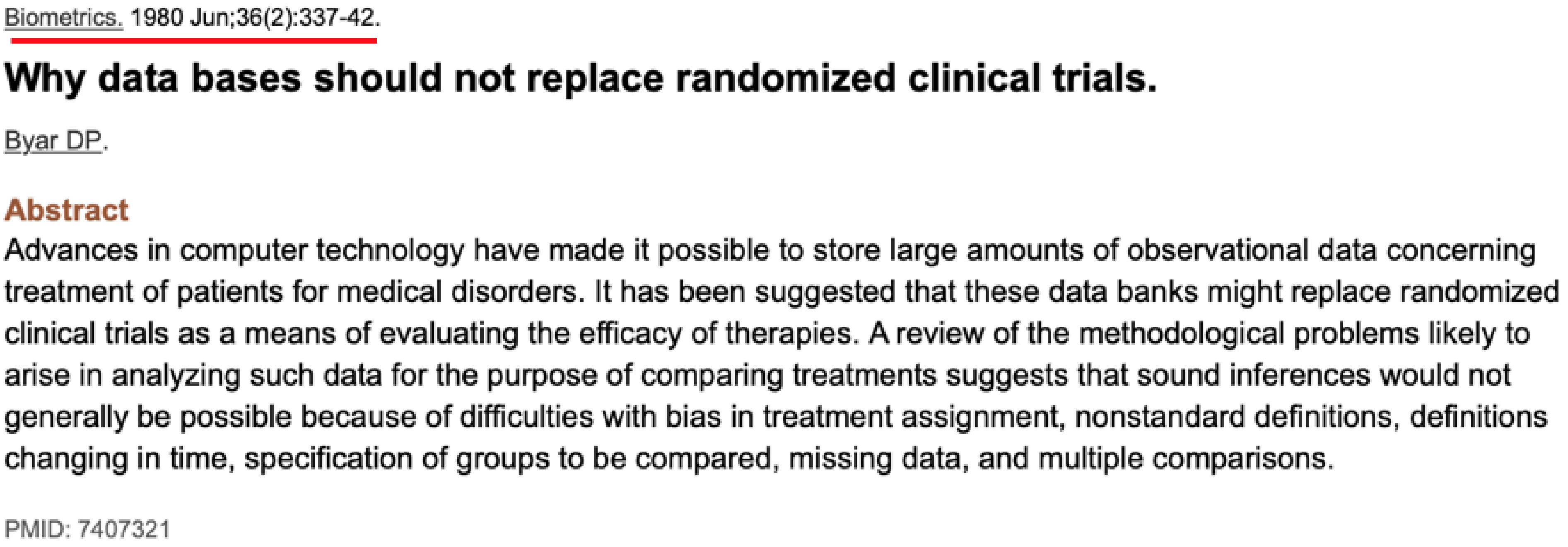
- Dilemma
  - (Too) large trials are resource intensive and may identify as statistically significant clinically insignificant results
  - (Too) small trials may produce imprecise or unreliable results
- Potential solutions
  - In calculating sample sizes, consider whether the smallest difference that will be statistically significant will be clinically significant = do small trials
    - "3) You should always ask your statistician "what is the smallest observed effect where I reject the null?". If this value (0.604 X for standard trials, might be different in others) is clinically meaningless, you should rethink your experiment" (Kert Viele).
  - Do "adequately" sized trials and abandon dichotomania
    - What really matters are the observed results and the confidence limits (or credible intervals) around them, not whether an arbitrary threshold has been crossed
      - In this context, abandon median differences in favour of restricted means
  - Use Real World Data to complement trial results

The Right Populations

The "Big Data/Real World Evidence" Challenge/Opportunity

Observational vs Experimental Data

Not a new issue



Biometrics. 1980 Jun;36(2):337-42.

Why data bases should not replace randomized clinical trials.

Byar DP.

Abstract

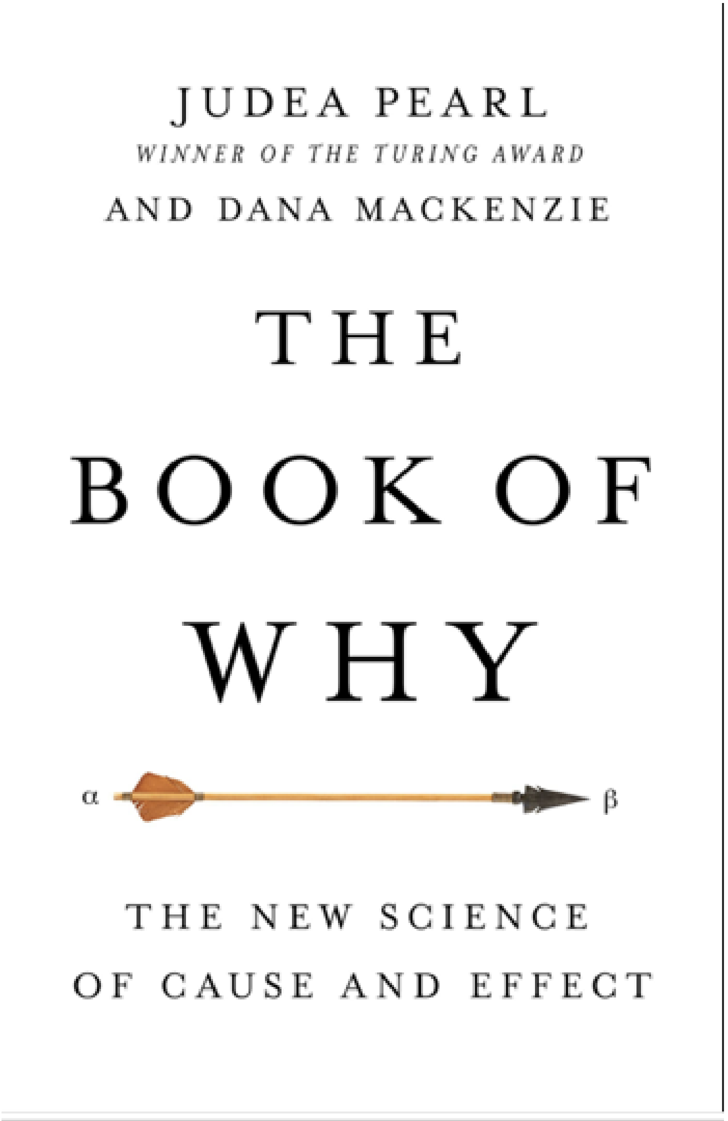
Advances in computer technology have made it possible to store large amounts of observational data concerning treatment of patients for medical disorders. It has been suggested that these data banks might replace randomized clinical trials as a means of evaluating the efficacy of therapies. A review of the methodological problems likely to arise in analyzing such data for the purpose of comparing treatments suggests that sound inferences would not generally be possible because of difficulties with bias in treatment assignment, nonstandard definitions, definitions changing in time, specification of groups to be compared, missing data, and multiple comparisons.

PMID: 7407321

Can RWD/RWE replace RCTs?

- There are certainly people who believe using RWD/RWE as a substitute for RCTs is possible and worth investing in (Roche buying Flatiron Health, for example)
  - The FDA has opened the door slightly to using RWD/RWE as a substitute for randomized controls in some settings
- Others believe what David Byar had to say almost 40 years ago is still fundamentally true
- This is a big topic with lots of local expertise, so I thought I would focus on a line of thought that you may not be as familiar with, i.e., the "causal revolution"

The Causal Revolution



JUDEA PEARL

WINNER OF THE TURING AWARD

AND DANA MACKENZIE

THE BOOK OF WHY

α β

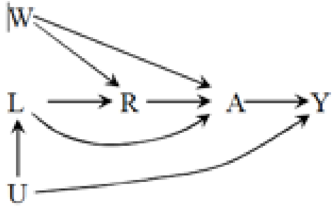
THE NEW SCIENCE OF CAUSE AND EFFECT

The Book of Why

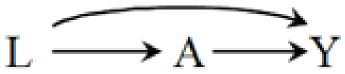
What I described as a "transformation" turned out to be a "revolution" that has changed the thinking in many of the sciences. Many now call it "the Causal Revolution," and the excitement that it has generated in research circles is spilling over to education and applications. I believe the time is ripe to share it with a broader audience.

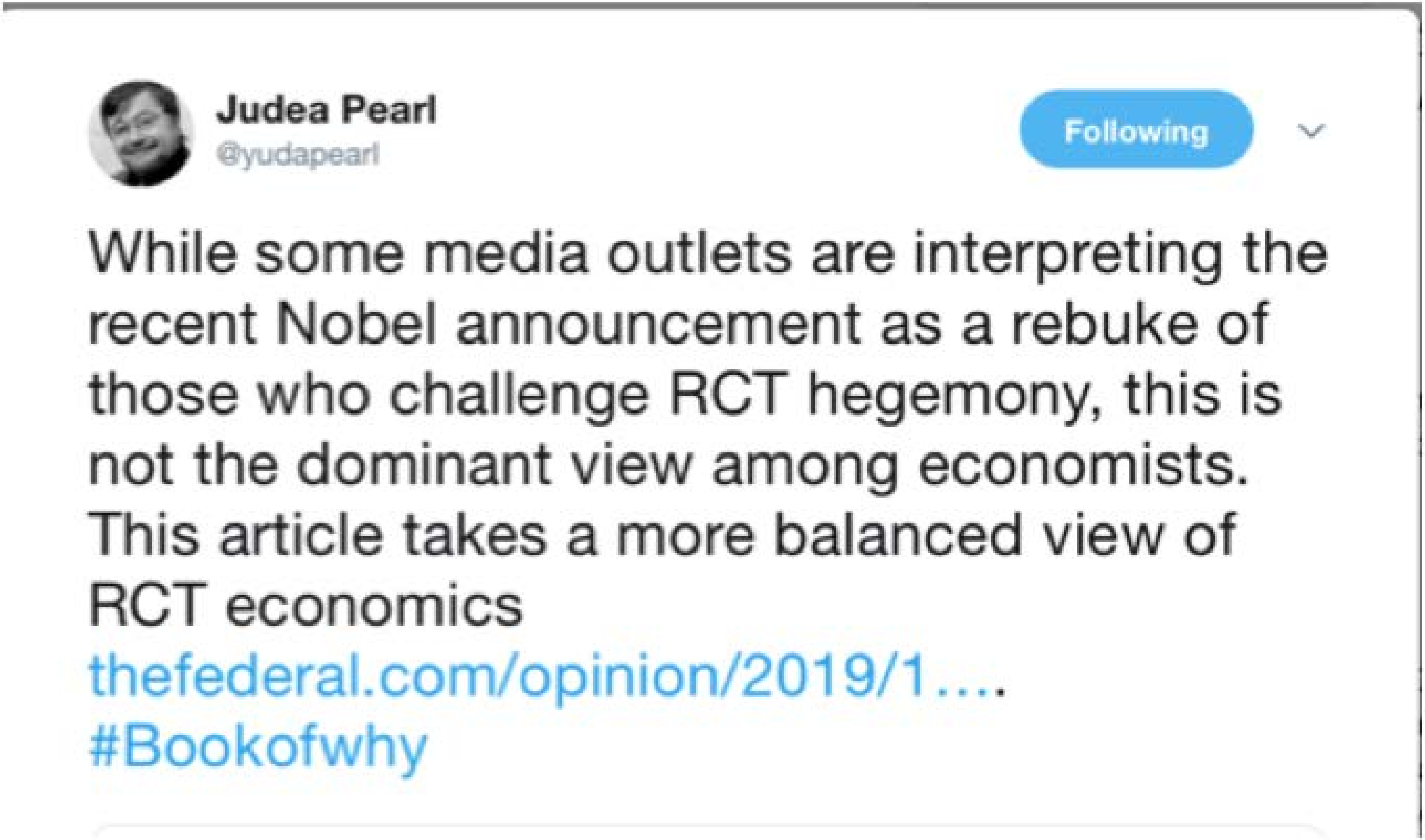
One of my goals in this chapter is to explain, from the point of view of causal diagrams*, precisely why RCTs allow us to estimate the causal effect X Y without falling prey to confounder bias. Once we have understood why RCTs work, there is no need to put them on a pedestal and treat them as the gold standard of causal analysis, which all other methods should emulate. Quite the opposite: we will see that the so-called gold standard in fact derives its legitimacy from more basic principles.

This chapter will also show that causal diagrams make possible a shift of emphasis from confounders to deconfounders. The former cause the problem; the latter cure it. The two sets may overlap, but they don't have to. If we have data on a sufficient set of deconfounders, it does not matter if we ignore some or even all of the confounders.



*DAGs – Directed Acyclic Graphs

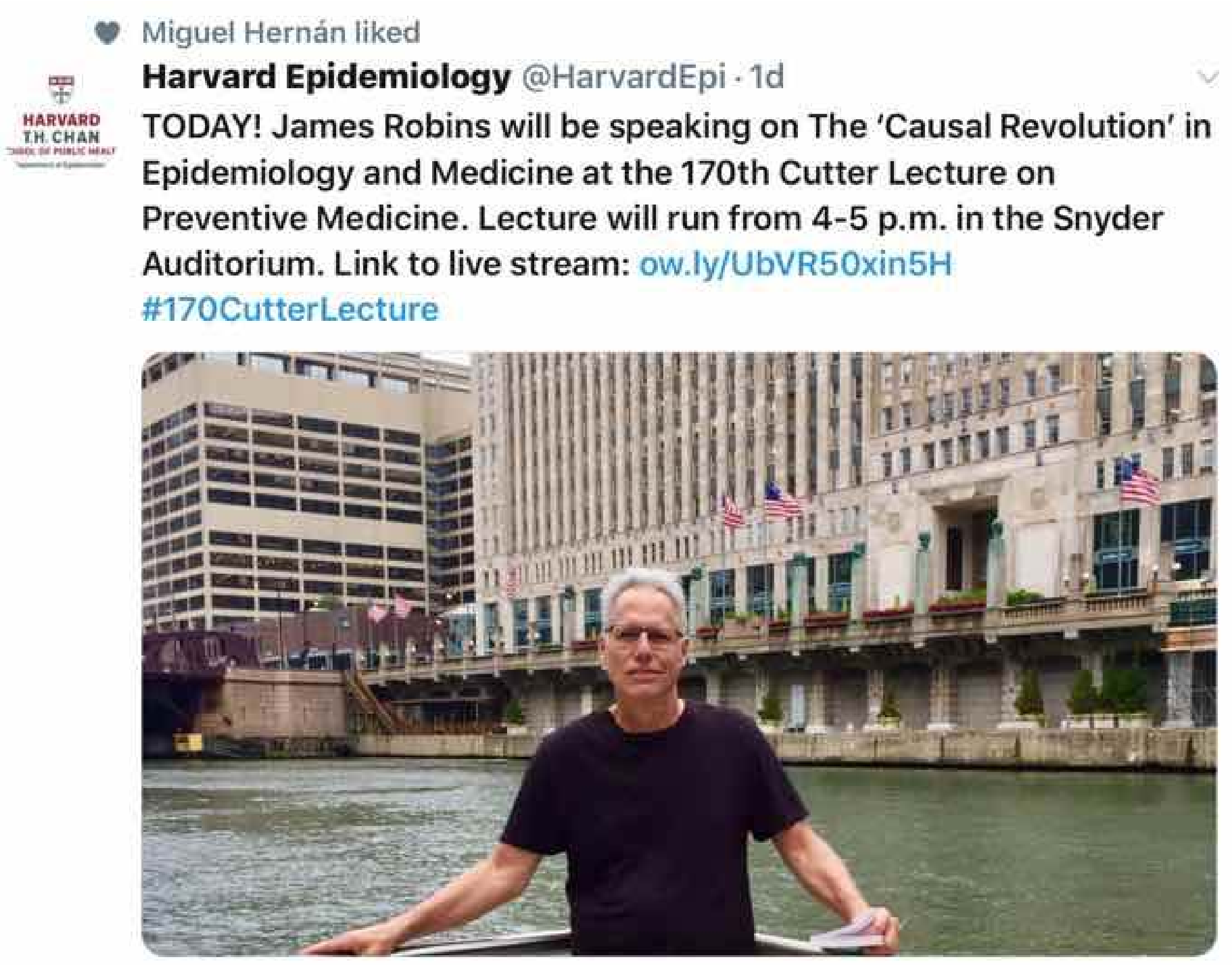




Judea Pearl
@yudapearl

While some media outlets are interpreting the recent Nobel announcement as a rebuke of those who challenge RCT hegemony, this is not the dominant view among economists. This article takes a more balanced view of RCT economics thefederal.com/opinion/2019/1…. #Bookofwhy

Another perspective from the other coast



Miguel Hernán liked

Harvard Epidemiology @HarvardEpi · 1d

TODAY! James Robins will be speaking on The 'Causal Revolution' in Epidemiology and Medicine at the 170th Cutter Lecture on Preventive Medicine. Lecture will run from 4-5 p.m. in the Snyder Auditorium. Link to live stream: ow.ly/UbVR50xin5H #170CutterLecture

Background/Motivation

Data science is science's second chance to get causal inference right: A classification of data science tasks

Miguel A. Hernán,1,2 John Hsu3,4, Brian Healy5,6
1. Departments of Epidemiology and Biostatistics,
Harvard T.H. Chan School of Public Health, Boston, MA

Introduction

For much of science's recent history, learning from data was the academic realm of Statistics. But, in the early 20th century, the founders of modern Statistics made a momentous decision about what could and could not be learned from data. They proclaimed that statistics could be applied to make causal inferences when using data from randomized experiments, but not when using nonexperimental (observational) data. This decision classified an entire class of scientific questions in the health and social sciences as not amenable to formal quantitative inference.

JAMA Oncology Reporting Guidelines

Use of Causal Language

Causal language (including use of terms such as effect and efficacy) should be used only for randomized clinical trials. For all other study designs (including meta-analyses of randomized clinical trials), methods and results should be described in terms of association or correlation and should avoid cause-and-effect wording.

Causal Inference: What If

Miguel A. Hernán, James M. Robins
November 10, 2019

https://cdn1.sph.harvard.edu/wp-content/uploads/sites/1268/2019/11/ci_hernanrobins_10nov19.pdf

https://view6.workcast.net/ControlUsher.aspx?cpak=7893749515199787&pak=7155440261855153

The target trial

The target trial–or its logical equivalents–is central to the causal inference framework. Dorn (1953), Cochran (1972), Rubin (1974), Feinstein (1971), and Dawid (2000) used it. Robins (1986) generalized the concept to time-varying treatments.

While recognizing that randomized experiments have intrinsic advantages for causal inference, sometimes we are stuck with observational studies to answer causal questions. What do we do? We analyze our data as if treatment had been randomly assigned conditional on measured covariates–though we often know this is at best an approximation. Causal inference from observational data then revolves around the hope that the observational study can be viewed as a conditionally randomized experiment.

TUTORIAL

Extending inferences from a randomized trial to a new target population

Issa J. Dahabreh MD ScD[1-4] | Sarah E. Robertson MS[1,2] | Jon A. Steingrimsson PhD[5] | Elizabeth A. Stuart PhD[6] | Miguel A. Hernán MD DrPH[4,7,8]

https://arxiv.org/pdf/1805.00550.pdf

Questions/Comments?