

Observational Studies and RCTs: A Trialist (?) Perspective

Steven Goodman, MD, MHS, PhD
Johns Hopkins University
Schools of Medicine and Public Health
sgoodman@jhmi.edu



Things identified as cancer risks (Altman and Simon, JNCI, 1992)

- Electric Razors
- Broken Arms (only in women)
- Fluorescent lights
- Allergies
- Breeding Reindeer
- Being a waiter
- Owning a pet bird
- Hot dogs
- Being short
- Being tall

Having a refrigerator

Goodman ACE Talk 9/8/03

The unobservability of causal effects

With Factor

Person A ----->

Person B ----->

Person C ----->

Average (A,B,C)

W/O Factor

?

?

?

D

E

F

Average (D, E, F)

Goodman ACE Talk 9/8/03

Basis for causal inference

- Assumption that the subjects without the risk factor are “like” the subjects with the risk factor in with respect to every other causal risk factor, known or unknown.
- The reliability of a observational results depends on our judgment that we have appropriately identified, measured and controlled all relevant confounders, i.e. that we understand the mechanism of action.

Goodman ACE Talk 9/8/03

Randomization...

- Does not guarantee confounder balance between groups.
- Does provide a mathematical (i.e. “objective”) distribution for confounders - both known and unknown - which provides a basis for our uncertainty statements (CIs) around the estimated effect size.
- A key element is that a risk factor is actively set by the experimenter, rather than just observed.

Goodman ACE Talk 9/8/03

The E=mc² of Epidemiology

$\Pr(\text{Outcome} \mid X=x) = \Pr(\text{Outcome} \mid \text{Set}(X=x))$

In English

The probability of an outcome when we observe a risk factor having a value or is unchanged when we actively set it to have that value.

Goodman ACE Talk 9/8/03

The Essential Tension

The tension between observational results - whose validity ultimately rests on our judgment about how well we understand the causal mechanism - and RCTs - whose validity derives from mathematical theory - is part of a longstanding struggle within science and medicine about the relative primacy of mechanistic/theoretical understanding versus empirical knowledge.

Goodman ACE Talk 9/8/03

A little history...

The French Debates

"In statistical affairs... the first care above all else is to lose sight of the man taken in isolation in order to consider him as only a fraction of a species. It is necessary to strip him of his individuality in order to arrive at the elimination of all accidental effects that individuality would have been able to introduce into the question... it is altogether different in the domain of medicine..."

(Poisson et al. 1835)

Goodman ACE Talk 9/8/03

Claude Bernard on Statistics

"Empiricism precedes science..."

...never have statistics taught anything, and never can they teach anything about the nature of phenomena. ...statistics teach absolutely nothing about the mode of action of medicine nor the mechanism of cure..."
(Bernard, p. 137)

Goodman ACE Talk 9/8/03

The Rise of Evidence-based Medicine (EBM)

Evidence-Based Medicine: A New Approach to Teaching the Practice of Medicine

Evidence-Based Medicine Working Group, JAMA (1992)

"A new paradigm for medical practice is emerging. Evidence-based medicine de-emphasizes intuition, unsystematic clinical experience, and pathophysiologic rationale as sufficient grounds for clinical decision making and stresses the examination of evidence from clinical research. EBM requires new skills of the physician, including efficient literature searching and the application of formal rules of evidence evaluating the clinical literature."

Goodman ACE Talk 9/8/03

The Rise of Evidence-based Medicine (EBM)

The Dark Side of Evidence-Based Medicine:

Horwitz R, 1996

“ The unspoken conviction...is that impersonal knowledge of the probability of an event is the...precondition for effective clinical medicine and superior to ...more traditional methods.”

“....If you were a 38-year-old with anemia and thrombocytopenia...Would you prefer to be treated by a naïve clinician guided by a set of pathways or by an expert physician with experience treating hundreds of patients with anemia?”

Goodman ACE Talk 9/8/03

COMMENTARIES

In Defense of Black Box Epidemiology

David A. Savitz

Assume that an epidemiologic study has been conducted reporting a positive association of an exposure and a disease. The investigators selected subjects in a haphazard manner. The association is excellent.

haphazardly most useful when extended into novel territory rather than in addressing well-understood biological phenomena.

The Emptiness of the Black Box

Petr Skrabanek

The "black box" strategy is a current paradigm of epidemiologic research, better described by the term "risk factor epidemiology." In the hope of unraveling causes of disease, associations are sought between disease and various "exposures." "Black box" is an untested procedure linking the exposure and the disease in a causal sequence. An association, by itself a fortuitous finding, is thus converted, by logical sleight-of-hand, into a causal sequence. The "black box" strategy is ever discovered by "stabs in the dark." In science, at least, one proceeds from an interesting problem, embedded within a larger body of systematic knowledge, toward its solution or rejection. Reasoning, such as "the existence of cars is associated with car accidents; ergo, let us ban cars and there will be no more car accidents," may be relevant for public health, but it is not science.

Black box epidemiologic disease understanding.

Modern-day revolt

“The evaluation of quality of RCTs is not an easy task. Consequently, interpretive decisions by old pre-EBM experts may be replaced by interpretive decisions from a new group of experts with EBM credentials....”

“A new form of dogmatic authoritarianism may...be revived in modern medicine, but the pronouncements will come from Cochranian Oxford rather than Galenic Rome.”

(Feinstein, AJM, 1997)

Goodman ACE Talk 9/8/03

Judgment in RCTs

- RCT combination: Heterogeneity and quality
- RCT Generalizability
- RCT Subgroup analysis
- RCT plausibility

Goodman ACE Talk 9/8/03

Judgment in Combining RCTs

Mammography!

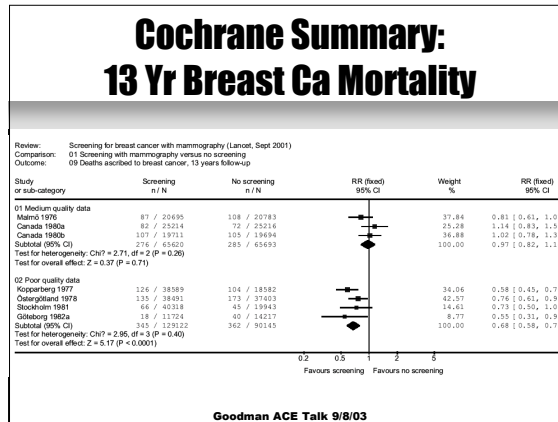
Goodman ACE Talk 9/8/03

Goodman ACE Talk 9/8/03

Cochrane Summary

Study	Rand.	Exclus.	Cause	Verdict
HIP	Adequate w/ imbalance	Flawed	Bias susp. Flawed?	Flawed
Malmö	Adequate w/ variable N's	Adequate -	Adequate	Medium
Two-County	Bias suspected	Bias suspected	Bias possible(?)	Poor / "very likely" flawed
Edinburgh	Bias strongly suspected	Bias suspected	No data	Flawed
Canadian	Adequate	Adequate	Blinded, low autopsy rate	Medium
Stockholm	Inadequate - imbalances, variable Ns	Bias suspected	Only study det. cancer confirmed	Poor
Göteborg	? variable ratio ? analysis	Adequate, but dates uncertain	31% autopsy	Poor

Goodman ACE Talk 9/8/03



USPSTF R&R Statement

"Recently, a 2001 Cochrane Collaboration review of the same trials concluded that 6 of the 8 trials were "flawed" or of "poor quality" and that the pooled results from the remaining 2 better trials did not support a benefit from mammography. Although the USPSTF was concerned about many (but not all) of the flaws identified in this review, it did not consider the presence of flaws sufficient reason in itself for rejecting trial results. Instead, it examined whether observed mortality reductions in the trials were likely to be explained by the biases potentially introduced by such flaws. Studies rated to be of "fair" quality by the USPSTF contained flaws that were considered unlikely to account for observed benefits (or lack of benefits)."

Goodman ACE Talk 9/8/03

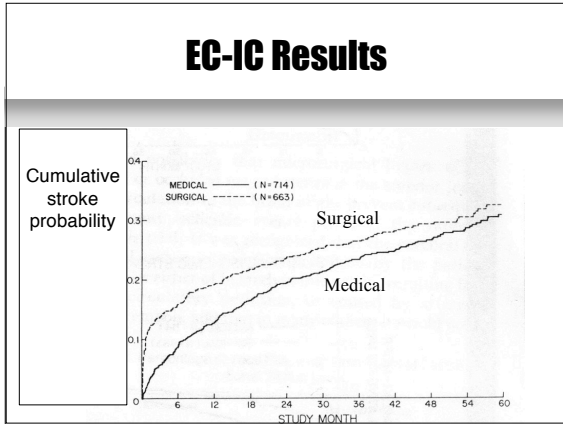
- ### Areas for Judgment in the Mammography Debate
- Developing quality criteria.
 - Applying quality criteria.
 - Role of non-RCT data in forming conclusions.
 - Natural history, biology, prognosis
 - Definition of primary endpoint.
 - Definitions of harms and benefits.
 - Balancing of harms and benefits.
 - How to balance qualitative uncertainty (about the effect of RCT flaws) against quantitative estimates.
- Goodman ACE Talk 9/8/03

Judgment in Generalizing RCTs

The EC-IC Trial

Goodman ACE Talk 9/8/03

- ### The EC-IC Trial
- Surgery involved connecting temporal artery (extracranial) to middle cerebral artery (intracranial) in attempt to bypass stenotic areas and decrease stroke risk.
 - Surgery was long established.
 - Patients randomized to surgery or medical care.
 - 1377 patients randomized, 1977-82, trials results reported in 1985.
- Goodman ACE Talk 9/8/03



EC-IC Reaction

- Shock and disbelief
- From a participating MD, "horrified" by results; "We know the procedure benefits some patients; if we did not we would not have performed it as many times as we have."

Bannister, NEJM, 1986.

Goodman ACE Talk 9/8/03

Medical News & Perspectives

Neurosurgeons Address EC/IC Study; Question Controlled Surgical Trials

LAST YEAR, Sydney J. Peerless, MD, was among the first to learn the results of an eight-year international surgical study. "My jaw dropped. I couldn't believe what I found," he said. "What the data were telling me was that with 95% confidence we can rule out a 5% benefit from extracranial/intracranial (EC/IC) bypass surgery."

Peerless, who is professor of neurosurgery at Canada's University of Western Ontario, London, was principal neurosurgical investigator in the Bannister study. His incredulity was

overall postoperative stroke rate was 25%—lower than that demonstrated in larger uncontrolled EC/IC bypass studies.

Of the 1377 patients enrolled, 714 were randomly assigned to medical management (325 mg aspirin qid and antihypertensive drugs prn). The remaining 663 received the operation followed by the same medical care.

Among patients who were originally selected and then excluded were 115 patients who refused to enter the trial; 52 whose clinicians insisted they undergo bypass surgery; and 11 for

on."

Other neurosurgeons think they may have better luck identifying such subgroups. Thoralf M. Sundt, Jr, MD, a Mayo Clinic neurosurgeon who offered a rebuttal to Peerless' presentation at the Denver meeting, is one.

Sundt maintained that the investigators did not note the condition of the patient's vessels, and thus might have overlooked factors that could have adversely influenced the outcome. He said that these include a small fibrotic, atrophic, temporal artery; inadequate perfusion gradient across the anastomosis; and the needle at the

JAMA 256: 165-167, 1986

EC-IC Reaction

- A large proportion (between 30%-70%) of potentially eligible patients were alleged to have not been entered into trial.
- Surgeons claimed that trial results were not generalizable; that enrolled patients were those selectively identified by surgeons as least likely to benefit.
- Committee of neurosurgeons appointed to review trial.

Goodman ACE Talk 9/8/03

EC-IC Reaction

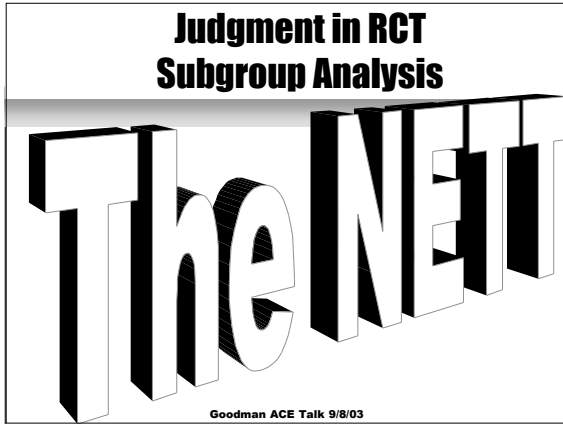
- **The Extracranial-Intracranial Bypass Study. A report of the committee appointed by the American Association of Neurological Surgeons to examine the study.**
Goldring S, Zervas N, and Langfitt T. NEJM 316: 817-820, 1987.
- **Was the international randomized trial of extracranial-intracranial arterial bypass representative of the population at risk?**
Sundt TM, Jr. NEJM 316: 814-816, 1987.
- **Are the results of the extracranial-intracranial bypass trial generalizable?**
Barnett HJ, Sackett D, Taylor DW, et al. NEJM 316: 820-824, 1987.

Goodman ACE Talk 9/8/03

EC-IC Dénouement

- No empirical support for claim that non-enrolled patients benefited from surgery.
- Reimbursement for operation halted; surgery decreased 75%.
- Later studies supported result.

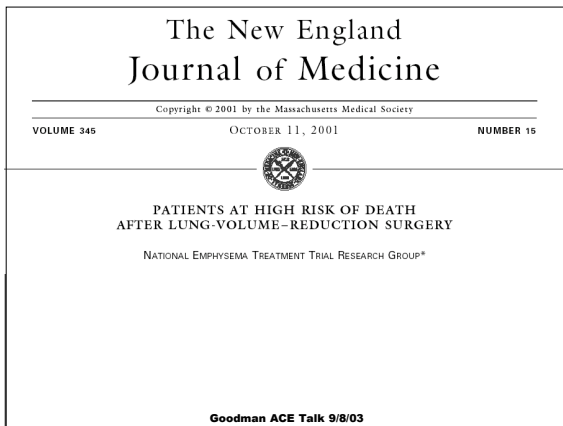
Goodman ACE Talk 9/8/03



Lung Volume Reduction Surgery

- Last ditch effort to ameliorate symptoms of terminal emphysema.
- Involves removing lobes of the lung to allow space in the chest for the remainder to “reinflate”.
- Many strong surgical proponents, w/claims based on short term observational data.
- CMS (then HCFA) agreed to fund a trial - NETT - National Emphysema Treatment Trial for \$100 million, and would only pay for surgery for patients enrolled in RCT.

Goodman ACE Talk 9/8/03



Patients at High Risk...

- Surgery raised mortality risk (0.43/py vs. 0.11/py) in the 69/1033 patients who had :
 - FEV1 < 20%
 - AND
 - Homogeneous emphysema OR Carbon monoxide diffusion capacity < 20% normal

Goodman ACE Talk 9/8/03

NETT Lesson

The NEW ENGLAND
JOURNAL of MEDICINE

ESTABLISHED IN 1812 MAY 22, 2003 VOL. 348 NO. 21

A Randomized Trial Comparing Lung-Volume-Reduction Surgery
with Medical Therapy for Severe Emphysema

National Emphysema Treatment Trial Research Group*

“Overall mortality was 0.11 deaths/py in both treatment groups, RR=1.01, p=0.90.”

Goodman ACE Talk 9/8/03

But....

Goodman ACE Talk 9/8/03

NETT Results

After exclusion of previous high-risk subgroup in which surgery was worse, mortality in surgery group was:

Disease location	Exercise Capacity	
	Lo	Hi
Predominantly non-Upper lobe	S: 0.15 M: 0.18	SAME S: 0.1 M: 0.05
Predominantly Upper lobe	S: 0.07 M: 0.15	BETTER S: 0.07 M: 0.07

Goodman ACE Talk 9/8/03

Relevant NETT Considerations

- 16 subgroup-defining variables were specified a priori.
- The threshold for low exercise capacity was determined post-hoc.

Goodman ACE Talk 9/8/03

Discussion

“We recognize the pitfalls of subgroup analyses, but we believe that the heterogeneity of the patients and of the outcomes and the considered approach we used make our findings clinically and statistically valid. The subgroup-specific findings were not the result of data mining or the optimization of P values. The candidate prognostic factors we used to identify subgroups were in large part specified in advance on the basis of biologic rationale.”

NETT, NEJM 2003

Goodman ACE Talk 9/8/03

The NEW ENGLAND JOURNAL of MEDICINE

PERSPECTIVE

The National Emphysema Treatment Trial — How Strong Is the Evidence?

James H. Ware, Ph.D.

- “In summary, the findings of NETT provide some evidence. [of subgroup differences] but the evidence does not meet the highest standard of proof. The finding does not correspond to a primary hypothesis, and the statistical significance of the finding is marginal when evaluated from the perspective of the number of hypotheses considered in the exploratory analysis.”
- “... Findings from such explorations are rarely definitive, but they do offer clues for future research and guidance to clinicians.”

Goodman ACE Talk 9/8/03

The Decision...

Centers for Medicare & Medicaid Services

Medicare News

For Immediate Release: Wednesday, August 20, 2003

Contact: CMS Office of Public Affairs 202-696-6345

For questions about Medicare please call 1-800-MEDICARE or visit www.medicare.gov.

MEDICARE ANNOUNCES INTENTION TO COVER LUNG VOLUME REDUCTION SURGERY

The Centers for Medicare & Medicaid Services (CMS) today announced it intends to make lung volume reduction surgery (LVRS) available to certain Medicare beneficiaries who are expected to benefit from the surgery based on results of the National Emphysema Treatment Trial (NETT).

“This decision follows years of research and an exhaustive evaluation of the available scientific evidence,” said CMS Administrator Tom Scully, “and reflects CMS’ efforts to bring the latest and best medical care to its 41 million beneficiaries.”

The new coverage CMS intends to provide will be available to certain Medicare patients with severe emphysema who meet criteria outlined by the NETT.

Emphysema is a progressive disabling disease affecting nearly 2 million Americans and is most common among people over 65. The disease is characterized by the irreversible destruction of air sacs in the lungs, which makes

Goodman ACE Talk 9/8/03

The Consequences...

New Therapies Pose Quandary for Medicare

By GINA KOLATA
August 17, 2003

The federal Medicare program is expected to decide this week whether to pay for an aggressive and expensive lung operation that could offer a lifeline to tens of thousands of elderly patients.

But health economists and medical experts say the treatment, however alluring, is part of an unsettling trend: new and ever pricier treatments for common medical conditions that are part and parcel of aging - procedures that could potentially benefit tens of thousands of patients, at a total cost that would far exceed the kind of prescription drug benefit now being considered by Congress.

The New York Times, p. 1

Goodman ACE Talk 9/8/03

Judgment of RCT Plausibility

Goodman ACE Talk 9/8/03

ORIGINAL INVESTIGATION

A Randomized, Controlled Trial of the Effects of Remote, Intercessory Prayer on Outcomes in Patients Admitted to the Coronary Care Unit

William S. Harris, PhD; Manohar Gowda, MD; Jerry W. Kolb, MDiv; Christopher P. Strychacz, PhD; James L. Vacek, MD; Philip G. Jones, MS; Alan Foraker, MD; James H. O'Keefe, MD; Ben D. McCallister, MD

Context: Intercessory prayer (praying for others) has been a common response to sickness for millennia, but it has received little scientific attention. The positive findings of a previous controlled trial of intercessory prayer have yet to be replicated.

Objectives: To determine whether remote, intercessory prayer for hospitalized, cardiac patients will reduce overall adverse events and length of stay.

Design: Randomized, controlled, double-blind, prospective, parallel-group trial.

Settings: Private, university-associated hospital.

Patients: Nine hundred ninety consecutive patients who were newly admitted to the coronary care unit (CCU).

Intervention: At the time of admission, patients were randomized to receive remote, intercessory prayer (prayer group) or not (usual care group). The first names of patients in the prayer group were given to a team of outside

intercessors who prayed for them daily for 4 weeks. Patients were unaware that they were being prayed for, and the intercessors did not know and never met the patients.

Main Outcome Measures: The medical course from CCU admission to hospital discharge was summarized in a CCU course score derived from blinded, retrospective chart review.

Results: Compared with the usual care group (n = 524), the prayer group (n = 466) had lower mean a SEM weighted (0.35 ± 0.20 vs 7.13 ± 0.27; P = .04) and unweighted (2.7 ± 0.1 vs 3.0 ± 0.1; P = .04) CCU course scores. Lengths of CCU and hospital stays were not different.

Conclusions: Remote, intercessory prayer was associated with lower CCU course scores. This result suggests that prayer may be an effective adjunct to standard medical care.

Arch Intern Med. 1999;159:2273-2278

A16 THE NEW YORK TIMES NATIONAL WEDNESDAY, JANUARY 6, 1999

Magnets Lessen Foot Pain Of Diabetics, a Study Finds

By HOLCOMBE B. NOBLE

In one of the first scientific studies of the centuries-old and highly debated use of magnets for treatment of medical disorders, a New York neurologist reported today that he had significantly lessened the foot pain that afflicts millions of diabetics.

Dr. Michael I. Weisensaul, a clinical professor of neurology at New York Medical College, emphasized that his study was small, involving only 24 patients, and must be regarded as preliminary to much more clinical research. But he said that the early results were clear and that the treatment ought to be put to a wider test.

The study, which appears in this issue of the journal, involved 12 patients with diabetes and peripheral neuropathy, a condition that causes numbness and pain in the hands and feet.

In July 1997, Dr. Weisensaul conducted a four-month study of diabetic and nondiabetic foot pain caused by diabetes, multiple myeloma, syringomyelia, and alcoholism were enrolled in a randomized placebo study. In the first month, each patient

received either a small magnet or a placebo. The patients with magnets reported a 50 percent decrease in pain, while the others reported less than 10 percent.

In November 1997, reporting in Archives of Physical and Rehabilitation Medicine, Dr. Carlos Vallbo, the Baylor College of Medicine Houston said that he applied intercessory magnets to his own leg pain and that the pain was gone minutes later. He then did a small study of patients with post-polio-syndrome. The patients with magnets reported a 50 percent decrease in pain, while the others reported less than 10 percent.

“We have no idea how or why the magnets work.”

“A real breakthrough...”

“...the [study] must be regarded as preliminary.”

“But...the early results were clear and... the treatment ought to be put to use immediately.”

A finding that runs counter to many previous studies.

Summary

- Large, well done, RCTs provide more reliable evidence than do observational studies addressing the same question.
- The need for RCTs in the face of observational evidence is related to the longstanding debate about the appropriate balance between empirical results over groups and mechanistic understanding in individual cases.
- This debate also affects our interpretation of RCT results.
- It will be informed by, but will not end, with the WHI episode.

Goodman ACE Talk 9/8/03

LET'S HOPE LIFE ON MARS IS MORE INTELLIGENT THAN LIFE ON EARTH.