The Value of Bayesian Methods for Evidence-Based Medicine

Steven Goodman, MD, MHS, PhD
Professor of Medicine and Epidemiology
Co-director, **METRICS**
(**Met**a-**r**esearch **I**nnovation **C**enter at **S**tanford)

FDA Workshop on Substantial Evidence
April 21, 2016


PITCHER'S
LIVURA
PORTERS'
NEW LIFE
TONIC
Soothing
Calming
Quieting
DR. KIRBY'S
GOLDEN
FORMULA
PRICE $1.00
NECTARINE
SYRUP
FIGS
PEPSIN
HADDAD'S
ELIXIR
RHEU-
MATIC
COMPOUND
WHITE
LINIMENT
THE FAMOUS
WHITE LILY
FACE WASH
FOR BEAUTIFYING
THE COMPLEXION
SWITTS
FRUITY SYRUP
SYRUP
FIGS
PEPSIN
SENNA
FOR CONSTIPATION
JOHNSTONS
Sarsaparilla
LIVER COMPLAINT
GIN
PILLS
KIDNEY
Shake the Bottle Well
PRICE ONE DOLLAR


PITCHER'S
LIVURA
CURES
Shake the Bottle Well
DOSE
PRICE ONE DOLLAR
GIN PILLS
KIDNEYS
FATHER JOHN'S
HOFFS
LINIMENT
HADDAD'S
ELIXIR
DR. HOBSON'S
RHEU-
MATIC
COMPOUND
Rheumatism
Gout
Neuralgia
and
Lumbago
THE FAMOUS
WHITE LILY
FACE WASH
FOR BEAUTIFYING
THE
COMPLEXION
DR. J. W. KERMOYT'S
WHITE
LINIMENT
DE WITT'S
FRUITY SYRUP
A COMBINATION OF
THE
SYRUP
OF
FIGS
WITH
PEPSIN
SENNA
FOR CONSTIPATION
JOHNSTON
Sarsapa

"Can Bayesian Approaches to Studyinig New Ireatmenits Improve Regulatory Decision Making?" 5/20/2004

Norris Aldersen
Janet Woodcock
Bob Temple
Steve Goodman
Greg Campbell
Don ***Berry***
Don Rubin
Tom Louis
Larry Kessler
Telba Irony
Constantine Gatsonis
Janet Wittes
Andy Grieve
Frank Rockhold
Bill DuMouchel
Ralph D'Agostino
Susan Ellenberg
Stacy Lindborg
Robert O'Neill
Jay Siegel

CLINICAL TRIALS

Impact Factor: 1.925 | **Ranking:** Medicine, Research & Experimental 77 out of 123

Table of Contents

August 2005; 2 (4)

Articles

Evidence-Based Medicine

A New Approach to Teaching the Practice of Medicine

Evidence-Based Medicine Working Group

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. Evidence-based medicine requires new skills of the physician, including efficient literature searching and the application of formal rules of evidence evaluating the clinical literature

dose of phenytoin intravenously and the drug is continued orally. A computed tomographic head scan is completely normal, and an electroencephalogram shows only nonspecific findings. The patient is very concerned about his risk of seizure recurrence. How might the resident proceed?

The Way of the Past

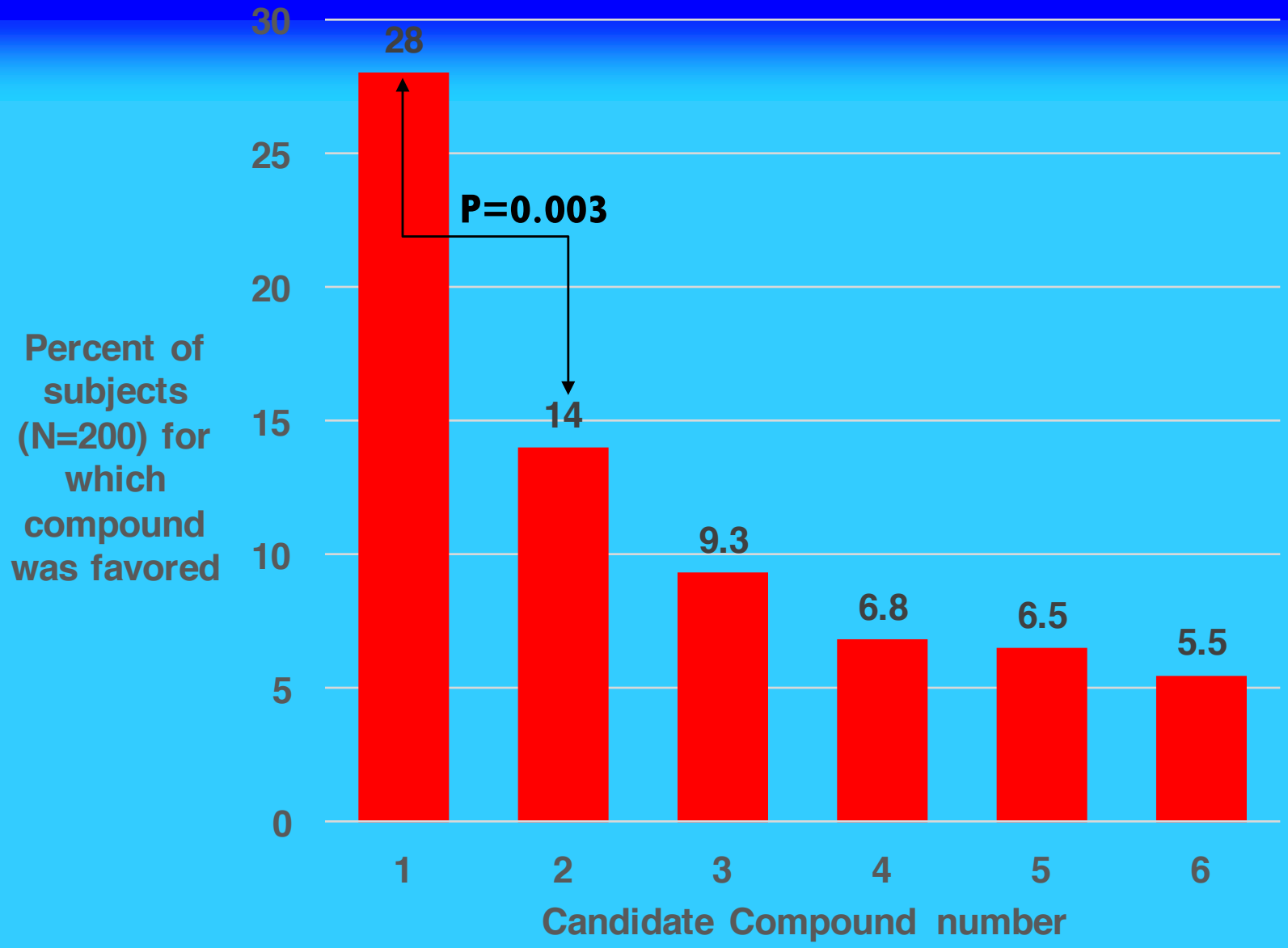
Faced with this situation as a clinical clerk, the resident was told by her se-

year is between 43% and 51%, and at 3 years the risk is between 51% and 60%. After a seizure-free period of 18 months his risk of recurrence would likely be less than 20%. She conveys this information to the patient, along with a recommendation that he take his medication, see his family doctor regularly, and have a review of his need for medication if he remains seizure-free for 18 months. The patient leaves with a clear idea of his likely prognosis

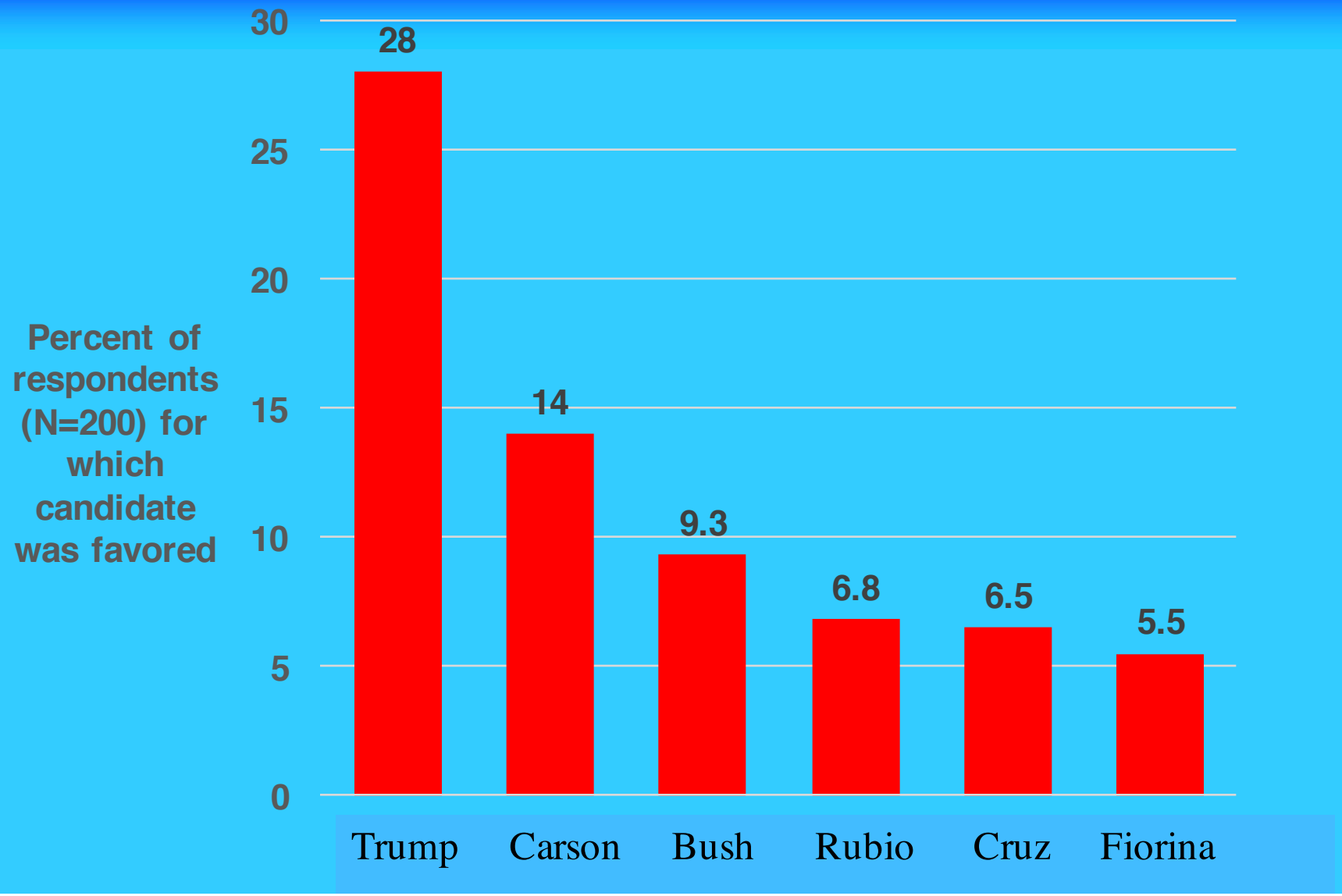
Definition of EBM

- **1992**: 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. Evidence-based medicine requires new skills of the physician, including efficient literature searching and the application of formal rules of evidence evaluating the clinical literature.
- **2000**: A systematic approach to clinical problem solving which allows the **integration of the best available research evidence with clinical expertise** and patient values.

Which candidate compound would be best in a future population?



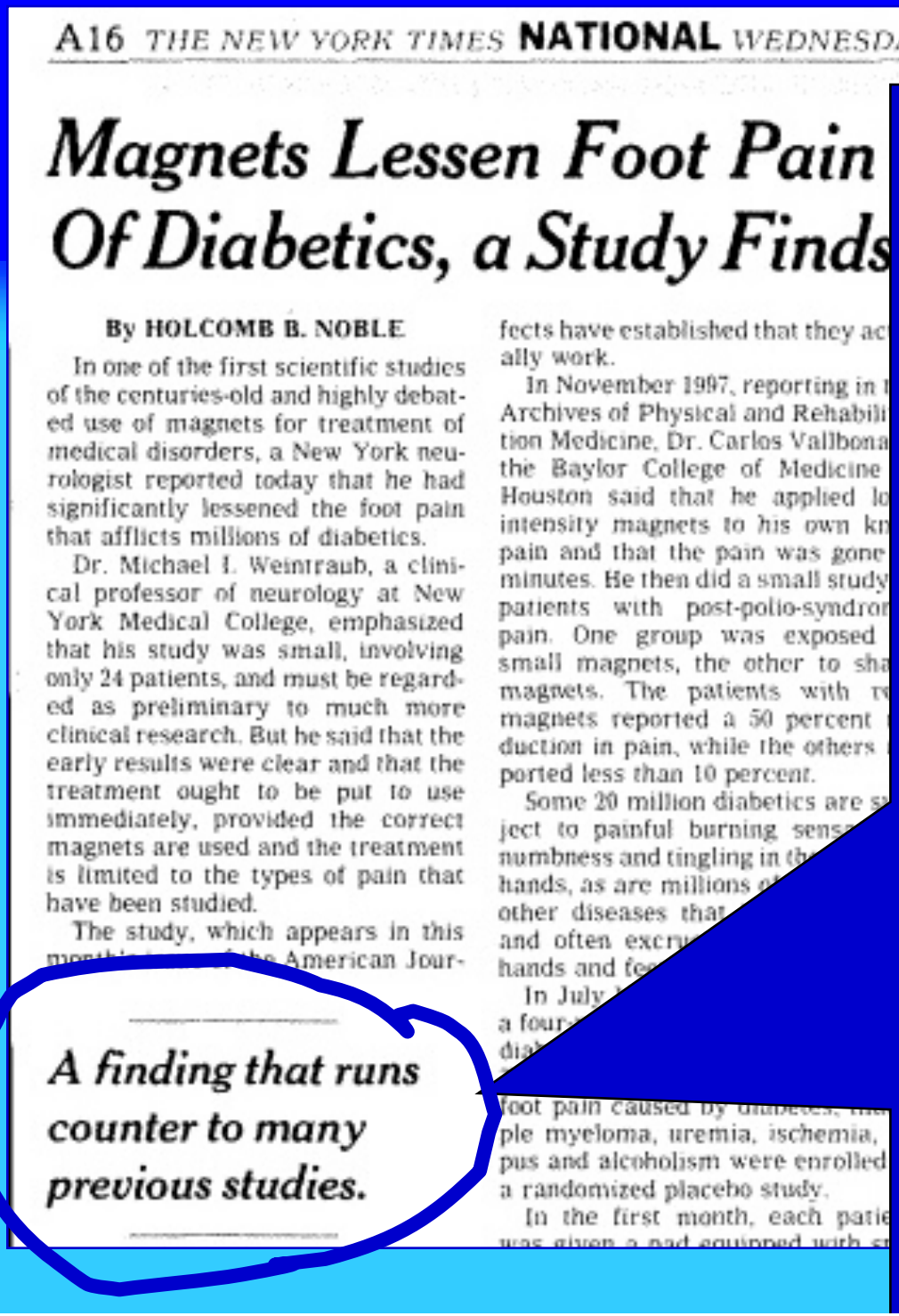
Which candidate would be best in a future population?



Things identified as cancer risks
(JNCI news, 1992)

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

Having a refrigerator



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

Magnets Lessen Foot Pain Of Diabetics, a Study Finds

By HOLCOMB 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. Weintraub, 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 use immediately, provided the correct magnets are used and the treatment is limited to the types of pain that have been studied.

The study, which appears in this

fects have established that they ac
ally work.

In November 1997, reporting in t
Archives of Physical and Rehabili
tion Medicine, Dr. Carlos Vallbona
the Baylor College of Medicine
Houston said that he applied lo
intensity magnets to his own kn
pain and that the pain was gone
minutes. He then did a small study
patients with post-polio-syndro
pain. One group was exposed
small magnets, the other to sha
magnets. The patients with
magnets reported a 50 percent
duction in pain, while the others
ported less than 10 percent.

Some 20 million diabetics are s
ject to painful burning sens
numbness and tingling in th
hands, as are millions o
other diseases that
and often excru
hands and fe

In July
a four
dia

foot pain caused by diabetes, tha
ple myeloma, uremia, ischemia,
pus and alcoholism were enrolled
a randomized placebo study.

In the first month, each patie
was given a pad equipped with s

A finding that runs counter to many previous studies.

"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."

Is everything we eat associated with cancer? A systematic cookbook review[1–3]

Jonathan D Schoenfeld and John PA Ioannidis

ABSTRACT

Background: Nutritional epidemiology is a highly prolific field. Debates on associations of nutrients with disease risk are common in the literature and attract attention in public media.

Objective: We aimed to examine the conclusions, statistical significance, and reproducibility in the literature on associations between specific foods and cancer risk.

Design: We selected 50 common ingredients from random recipes in a cookbook. PubMed queries identified recent studies that evaluated the relation of each ingredient to cancer risk. Information regarding author conclusions and relevant effect estimates were extracted. When >10 articles were found, we focused on the 10 most recent articles.

Results: Forty ingredients (80%) had articles reporting on their cancer risk. Of 264 single-study assessments, 191 (72%) concluded that the tested food was associated with an increased (n = 103) or a decreased (n = 88) risk; 75% of the risk estimates had weak ($0.05 > P \geq 0.001$) or no statistical ($P > 0.05$) significance. Statistically significant results were more likely than nonsignificant findings to be published in the study abstract than in only the full text ($P < 0.0001$). Meta-analyses (n = 36) presented more conservative results; only 13 (26%) reported an increased (n = 4) or a decreased (n = 9) risk (6 had more than weak statistical support). The median RRs (IQRs) for studies that concluded an increased or a decreased risk were 2.20 (1.60, 3.44) and 0.52 (0.39, 0.66), respectively. The RRs from the meta-analyses were on average null (median: 0.96; IQR: 0.85, 1.10).

Conclusions: Associations with cancer risk or benefits have been claimed for most food ingredients. Many single studies highlight implausibly large effects, even though evidence is weak. Effect sizes shrink in meta-analyses. *Am J Clin Nutr* 2013;97:127–34.

The American Journal of Clinical Nutrition

and such discrepancies in the evidence have fueled hot debates (9–12) rife with emotional and sensational rhetoric that can subject the general public to increased anxiety and contradictory advice (13, 14). One wonders whether this highly charged atmosphere and intensive testing of food-related associations may create a plethora of false-positive findings (15) and questionable research practices, especially when the research is highly exploratory, the analyses and protocols are not preregistered, and the findings are selectively reported. It was previously shown in a variety of other fields that "negative" results are either less likely to be published (16–21) or misleadingly interpreted (19, 22). Studies may spuriously highlight results that barely achieve statistical significance (15, 23) or report effect estimates that either are overblown (24, 25) or cannot be replicated in other studies (24, 26, 27).

To better evaluate the extent to which these factors may affect studies investigating dietary risk factors for malignancy, we surveyed recently published studies and meta-analyses that addressed the potential association between a large random sample of food ingredients and cancer risk of any type of malignancy.

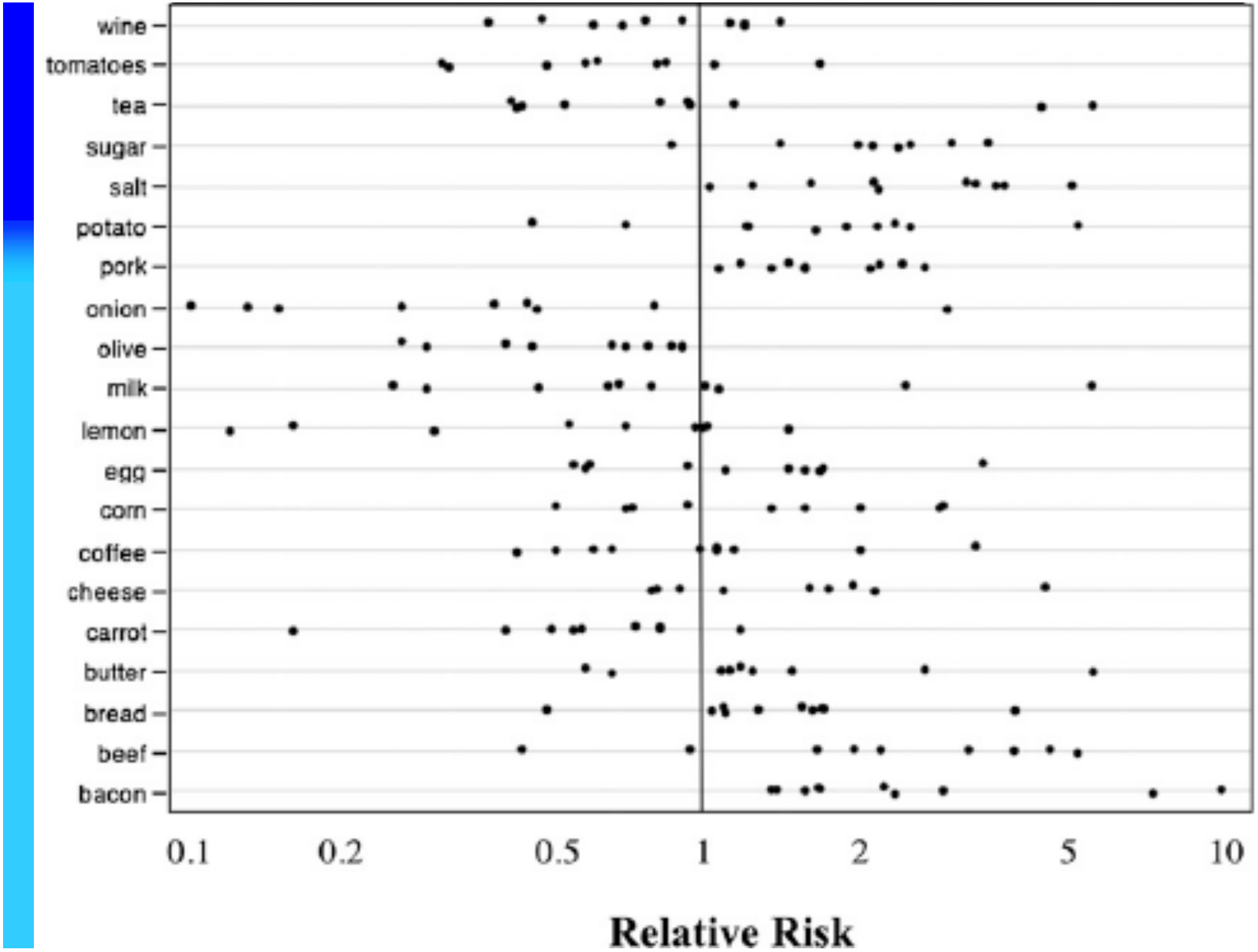
SUBJECTS AND METHODS

Random ingredient selection

We selected ingredients from random recipes included in *The Boston Cooking-School Cook Book* (28), available online at http://archive.org/details/bostoncookingsch00farmrich. A copy of the book was obtained in portable document format and viewed by using *Skim* version 1.3.17 (http://skim-app.sourceforge.net). The recipes (*see* Supplementary Table 1 under "Supplemental data" in the online issue) were selected at ran-

Design: We selected 50 common ingredients from random recipes in a cookbook. PubMed queries identified recent studies that evaluated the relation of each ingredient to cancer risk. Information regarding author conclusions and relevant effect estimates were extracted. When >10 articles were found, we focused on the 10 most recent articles.

Results: Forty ingredients (80%) had articles reporting on their cancer risk. Of 264 single-study assessments, 191 (72%) concluded that the tested food was associated with an increased (n = 103) or a decreased (n = 88) risk; 75% of the risk estimates had weak (0.05


wine
tomatoes
tea
sugar
salt
potato
pork
onion
olive
milk
lemon
egg
corn
coffee
cheese
carrot
butter
bread
beef
bacon
0.1
0.2
0.5
1
2
5
10
Relative Risk

A short research quiz

A well done study is reported on a new electrical stimulator for sickle cell pain control, and the authors state that it has turned out, somewhat surprisingly (e.g. < 20% chance of being true before the experiment), to be effective in reducing migraine frequency by 15%, 95% CI: 1% to 29%, *P*=0.03. The probability that this association is real is:

a.) < 75%

b.) 75% to 94.99...%

c.) 95% or higher

A short research quiz

A well done study is reported on a new electrical stimulator for sickle cell pain control, and the authors state that it has turned out, somewhat surprisingly (e.g. < 20% chance of being true before the experiment), to be effective in reducing migraine frequency by 15%, 95% CI: 1% to 29%, *P*=0.03. The probability that this association is real is:

a.) < 75%

b.) 75% to 94.99...%

c.) 95% or higher

Medicine Residents' Understanding of the Biostatistics and Results in the Medical Literature

Donna M. Windish, MD, MPH
Stephen J. Huot, MD, PhD
Michael L. Green, MD, MSc

Context Physicians depend on the medical literature to keep current with clinical information. Little is known about residents' ability to understand statistical methods or how to appropriately interpret research outcomes.

Objective To evaluate residents' understanding of biostatistics and interpretation of research results.

Design, Setting, and Participants Multiprogram cross-sectional survey of internal medicine residents.

Main Outcome Measure Percentage of questions correct on a biostatistics/study design multiple-choice knowledge test.

Results The survey was completed by 277 of 367 residents (75.5%) in 11 residency programs. The overall mean percentage correct on statistical knowledge and interpretation of results was 41.4% (95% confidence interval [CI], 39.7%-43.3%) vs 71.5% (95% CI, 57.5%-85.5%) for fellows and general medicine faculty with research training ($P<.001$). Higher scores in residents were associated with additional advanced degrees (50.0% [95% CI, 44.5%-55.5%] vs 40.1% [95% CI, 38.3%-42.0%]; $P<.001$); prior biostatistics training (45.2% [95% CI, 42.7%-47.8%] vs 37.9% [95% CI, 35.4%-40.3%]; $P=.001$); enrollment in a university-based training program (43.0% [95% CI, 41.0%-45.1%] vs 36.3% [95% CI, 32.6%-40.0%]; $P=.002$); and male sex (44.0% [95% CI, 41.4%-46.7%] vs 38.8% [95% CI, 36.4%-41.1%]; $P=.004$). On individual knowledge questions, 81.6% correctly interpreted a relative risk. Residents were less likely to know how to interpret an adjusted odds ratio from a multivariate regression analysis (37.4%) or the results of a Kaplan-Meier analysis (10.5%). Seventy-five percent indicated they did not understand all of the statistics they encountered in journal articles, but 95% felt it was important to understand these concepts to be an intelligent reader of the literature.

Conclusions Most residents in this study lacked the knowledge in biostatistics needed to interpret many of the results in published clinical research. Residency programs should include more effective biostatistics training in their curricula to successfully prepare residents for this important lifelong learning skill.

JAMA. 2007;298(9):1010-1022 www.jama.com

PHYSICIANS MUST KEEP CURRENT with clinical information to practice evidence-based medicine (EBM). In doing so, most prefer to seek evidence-based summaries, which give the clinical bottom line,[1] or evidence-based practice guidelines.[1-3] Resources that maintain these information summaries, however, currently include a limited number of common conditions.[4] Thus, to answer many of their clinical questions, physicians need to access reports of original research. This requires the reader to critically appraise the design, conduct, and analysis of each study and subsequently interpret the results.

Several surveys in the 1980s demonstrated that practicing physicians, particularly those with no formal education in epidemiology and biostatistics, had a poor understanding of common statistical tests and limited ability to interpret study results.[5-7] Many physicians likely have increased difficulty today because more complicated sta-

From: **Medicine Residents' Understanding of the Biostatistics and Results in the Medical Literature**

JAMA. 2007;298(9):1010-1022. doi:10.1001/jama.298.9.1010

The JAMA Network

Table 3. Percentages of Correct Answers for the Knowledge-Based Questions

Question No.[a]	Objective	Correct (95% CI), %
1a	Identify continuous variable	43.7 (37.8-49.5)
1b	Identify ordinal variable	41.5 (35.7-47.3)
1c	Identify nominal variable	32.9 (27.3-38.4)
2	Recognize a case-control study	39.4 (33.6-45.1)
3	Recognize purpose of double-blind studies	87.4 (83.5-91.3)
4a	Identify ANOVA	47.3 (41.4-53.2)
4b	Identify χ^2 analysis	25.6 (20.5-30.8)
4c	Identify *t* test	58.1 (52.3-63.9)
5	Recognize definition of bias	46.6 (40.7-52.4)
6	Interpret the meaning of *P* value >.05	58.8 (53.0-64.6)
7	Identify Cox proportional hazard regression	13.0 (9.0-17.0)
8	Interpret standard deviation	50.2 (42.3-56.1)
9	Interpret 95% CI and statistical significance	11.9 (8.0-15.7)
10	Recognize power, sample size, and significance-level relationship	30.3 (24.9-35.7)
11	Determine which test has more specificity	56.7 (50.8-62.5)
12	Interpret an unadjusted odds ratio	39.0 (33.3-44.7)
13	Interpret odds ratio in multivariate regression analysis	37.4 (31.9-43.3)
14	Interpret relative risk	81.6 (77.0-86.2)
15	Determine strength of evidence for risk factors	17.0 (12.6-21.4)
16	Interpret Kaplan-Meier analysis results	10.5 (6.9-14.1)

Abbreviations: ANOVA, analysis of variance; CI, confidence interval.
[a] See Appendix.

277 Yale Residents

95% felt it was important to understand these concepts to be an intelligent reader of the literature.

Average score: 41%

Irony…

- 59% reportedly chose the correct “P-value” definition, *but all of the multiple choice answers were incorrect!*

nature International weekly journal of science

NATURE | NEWS

Statisticians issue warning over misuse of *P* values

Policy statement aims to halt missteps in the quest for certainty.

Monya Baker

07 March 2016

Misuse of the *P* value — a common test for judging the strength of scientific evidence — is contributing to the number of research findings that cannot be reproduced, the American Statistical Association (ASA) warns in a statement released today[1]. The group has taken the unusual step of issuing principles to guide use of the *P* value, which it says cannot determine whether a hypothesis is true or whether results are important.

This is the first time that the 177-year-old ASA has made explicit recommendations on such a foundational matter in statistics, says executive director Ron Wasserstein. The society's members had become increasingly concerned that the *P* value was being misapplied in ways that cast doubt on statistics generally, he adds.

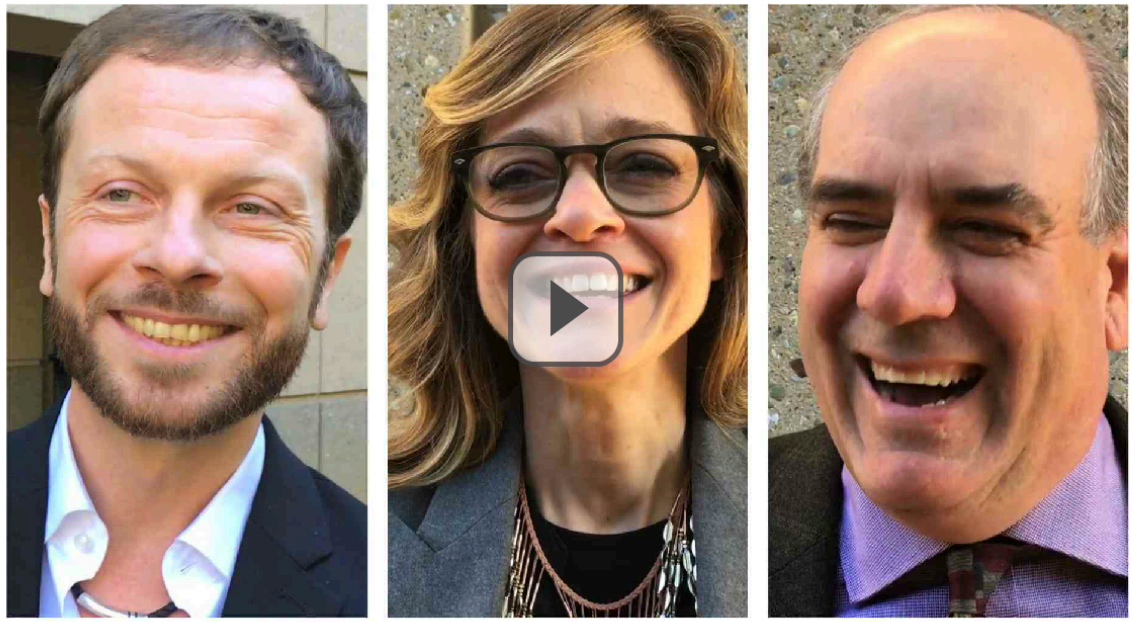
In its statement, the ASA advises researchers to avoid drawing scientific conclusions or making policy decisions based on *P* values alone. Researchers should describe not only the data analyses that produced statistically significant results, the society says, but all statistical tests and choices made in calculations. Otherwise, results may seem falsely

How scientists fool

SCIENTIFIC METHOD | 10:23 AM | MAR 7, 2016

Statisticians Found One Thing They Can Agree On: It's Time To Stop Misusing P-Values

By CHRISTIE ASCHWANDEN



Little p-value

What are you trying to say

Of significance?

Preliminary Communication

Maintenance Therapy With Tumor-Treating Fields Plus Temozolomide vs Temozolomide Alone for Glioblastoma
A Randomized Clinical Trial

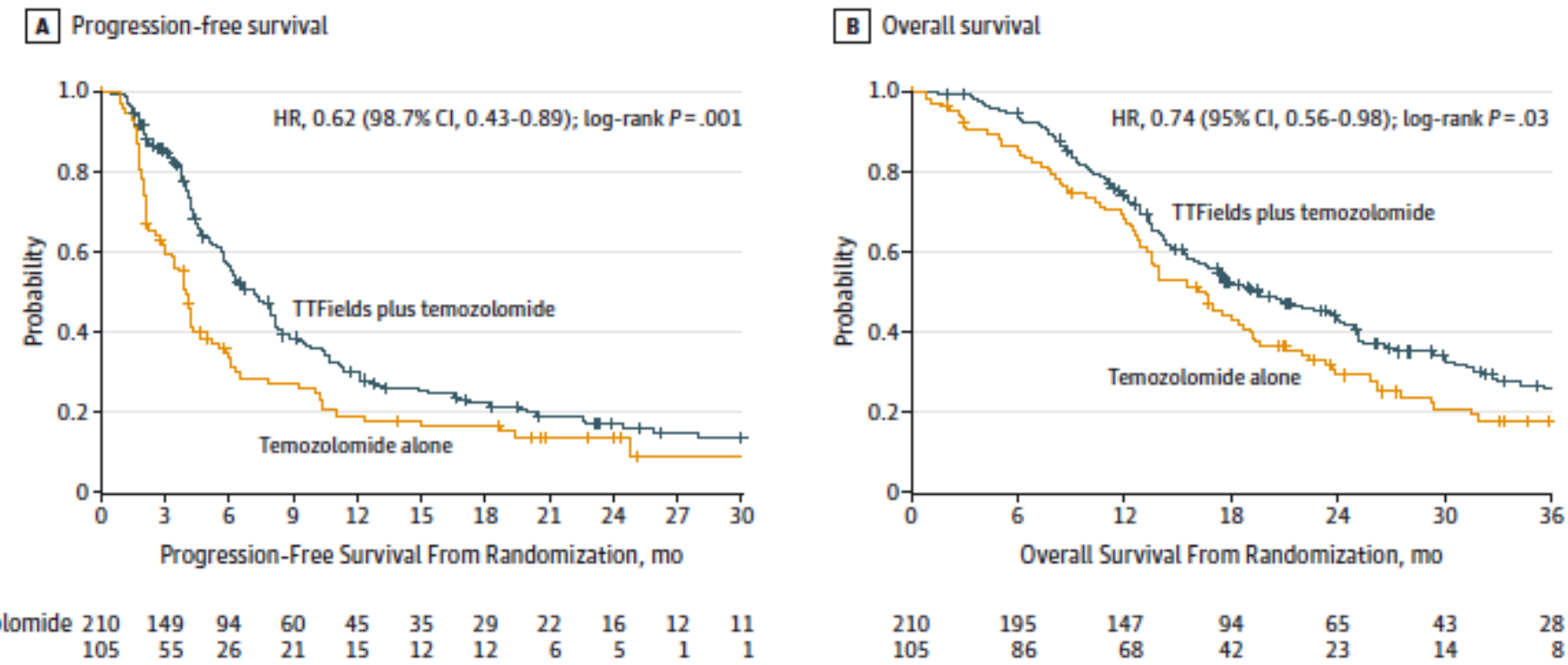
Roger Stupp, MD; Sophie Taillibert, MD; Andrew A. Kanner, MD; Santosh Kesari, MD, PhD; David M. Steinberg, PhD; Steven A. Toms, MD, FACS, MPH; Lynne P. Taylor, MD, FAAN; Frank Lieberman, MD; Antonio Silvani, MD; Karen L. Fink, MD, PhD; Gene H. Barnett, MD, MBA; Jay-Jiguang Zhu, MD, PhD; John W. Henson, MD, MBA, FAAN; Herbert H. Engelhard, MD, PhD; Thomas C. Chen, [illegible] Joseph Landolfi, D[illegible] PhD; Eilon D. Kirson, M[illegible]

MAIN OUTCOMES AND MEASURES The primary end point was progression-free survival in the intent-to-treat population (significance threshold of .01) with overall survival in the per-protocol population (n = 280) as a powered secondary end point (significance threshold of .006). This prespecified interim analysis was to be conducted on the first 315 patients after at least 18 months of follow-up.

RESULTS The interim analysis included 210 patients randomized to TTFields plus temozolomide and 105 randomized to temozolomide alone, and was conducted at a median follow-up of 38 months (range, 18-60 months). Median progression-free survival in the intent-to-treat population was 7.1 months (95% CI, 5.9-8.2 months) in the TTFields plus temozolomide group and 4.0 months (95% CI, 3.3-5.2 months) in the temozolomide alone group (hazard ratio [HR], 0.62 [98.7% CI, 0.43-0.89]; *P* = .001). Median overall survival in the per-protocol population was 20.5 months (95% CI, 16.7-25.0 months) in the TTFields plus temozolomide group (n = 196) and 15.6 months (95% CI, 13.3-19.1 months) in the temozolomide alone group (n = 84) (HR, 0.64 [99.4% CI, 0.42-0.98]; *P* = .004).

CONCLUSIONS AND RELEVANCE In this interim analysis of 315 patients with glioblastoma who had completed standard chemoradiation therapy, adding TTFields to maintenance temozolomide chemotherapy significantly prolonged progression-free and overall survival.

Figure 2. Survival Curves for Patients Included in the Interim Analysis in the Intent-to-Treat Population



No. at risk

TTFields plus temozolomide	210	149	94	60	45	35	29	22	16	12	11
Temozolomide alone	105	55	26	21	15	12	12	6	5	1	1

210	195	147	94	65	43	28
105	86	68	42	23	14	8

Survival analyses on time from date of randomization until tumor progression, death, or last follow-up (censored patients) according to the Kaplan-Meier method. The small vertical ticks on the curves indicate censored patients. HR indicates hazard ratio; TTFields, tumor-treating fields.

Rationale: Prior evidence

Tumor-treating fields (TTFields) are an antimitotic treatment that selectively disrupts the division of cells by delivering low-intensity, intermediate-frequency (200 kHz) alternating electric fields via transducer arrays applied to the shaved scalp.[8-10] In preclinical models, TTFields have been shown to cause mitotic arrest and apoptosis by disrupting mitotic spindle formation during metaphase and causing dielectrophoretic movement of polar molecules during cytokinesis.[8,10-12] In a randomized phase 3 trial in which TTFields were compared with chemotherapy in 237 patients with recurrent glioblastoma, the use of TTFields did not prolong progression-free survival or overall survival, but the therapy was associated with better quality of life without the typical chemotherapy-associated toxic effects.[13]

Research White Paper

Mechanistic Evidence in Evidence-Based Medicine: A Conceptual Framework

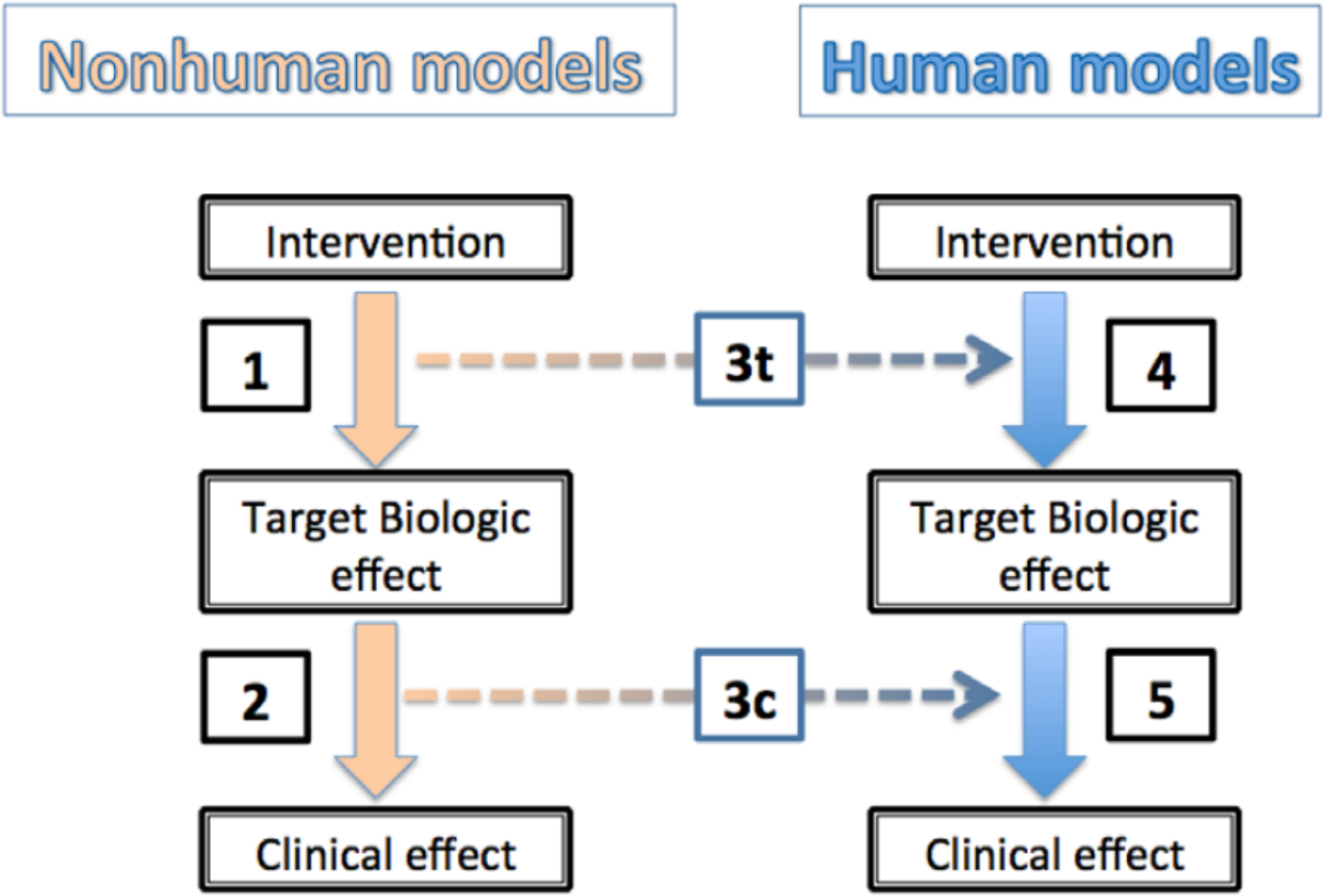
Prepared for:
Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
540 Gaither Road
Rockville, MD 20850
www.ahrq.gov

Contract No: 290-2007-10061-I

Prepared by:
The Johns Hopkins University Evidence-based Practice Center
Baltimore, MD

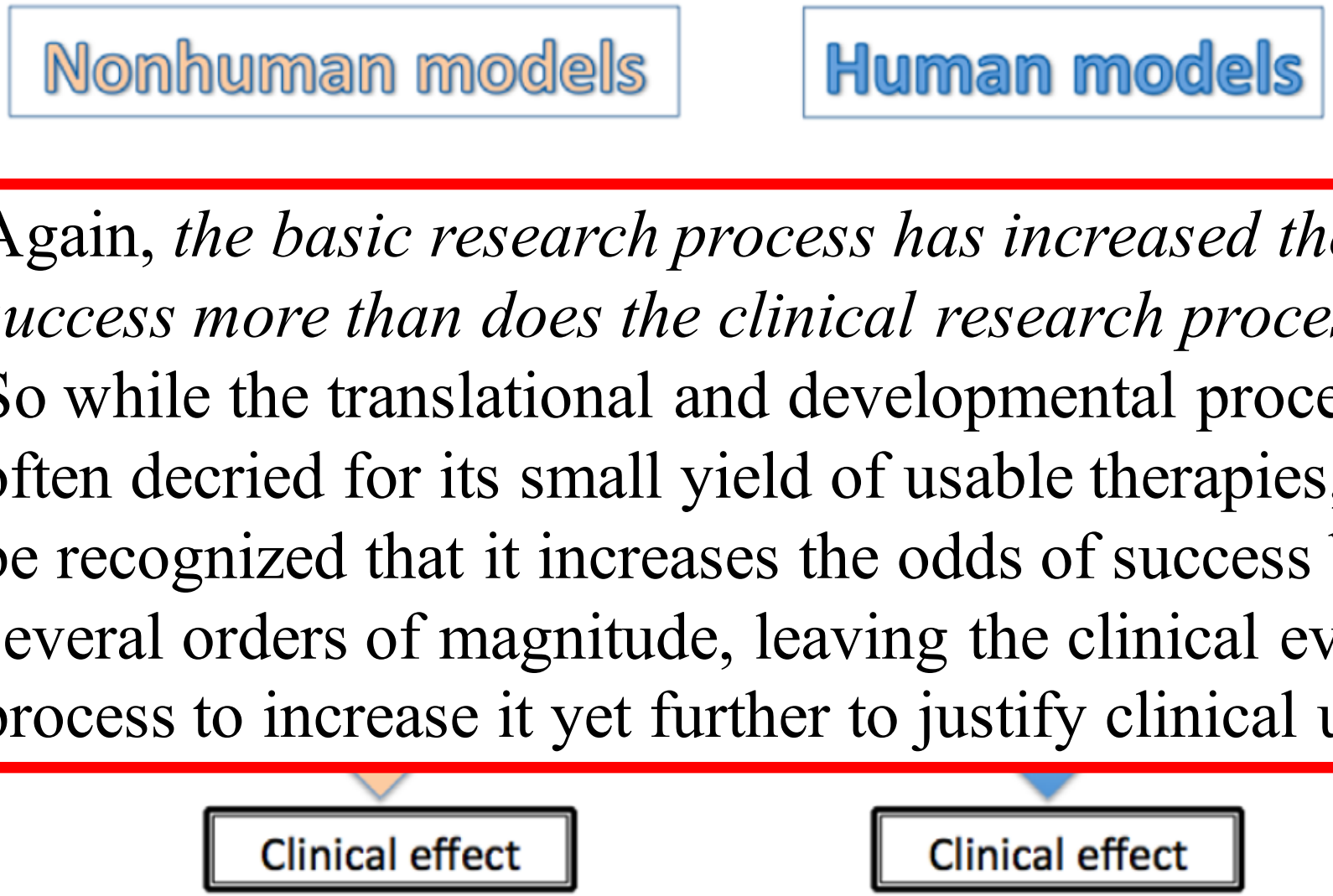
Investigators:
Steven N. Goodman, M.D., Ph.D.
Jason Gerson, Ph.D.

Figure A. Schematic of mechanistic framework model



The propagation of the strength of evidence is through a Bayesian algorithm, with the strength of evidence represented by the degree to which the probability of a clinical effect is modified by evidence from the component steps. This modeling makes clear how strong mechanistic evidence can be necessary for proper inferences, yet still, by itself, yield very low probabilities of success for a given intervention.

Figure A. Schematic of mechanistic framework model

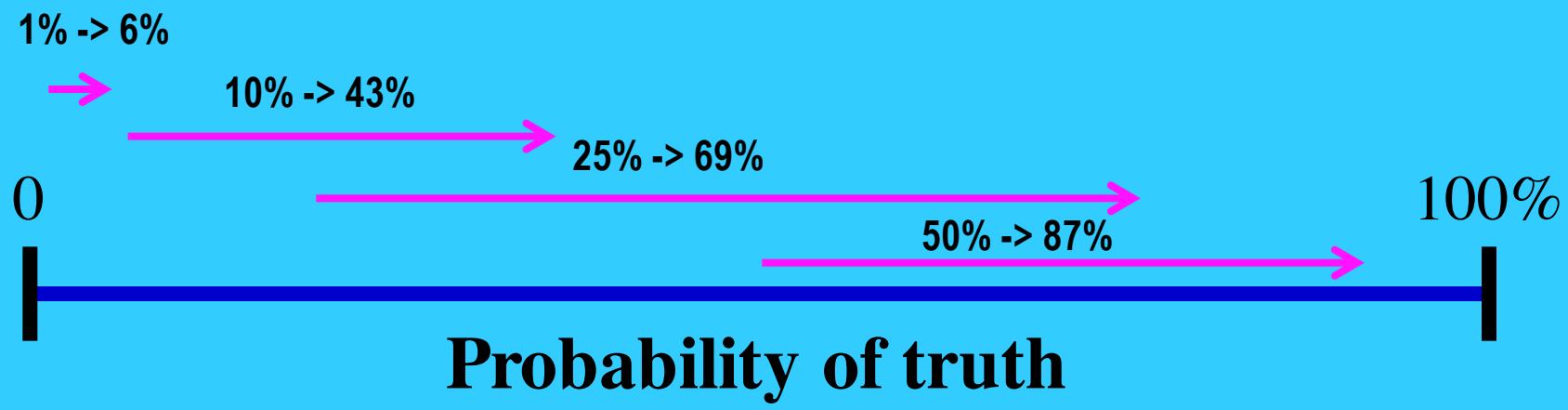


The propagation of the strength of evidence is through a Bayesian algorithm, with the strength of evidence represented by the degree to which the probability of a clinical effect is modified by evidence from the component steps. This modeling makes clear how strong mechanistic evidence can be necessary for proper inferences, yet still, by itself, yield very low probabilities of success for a given intervention.

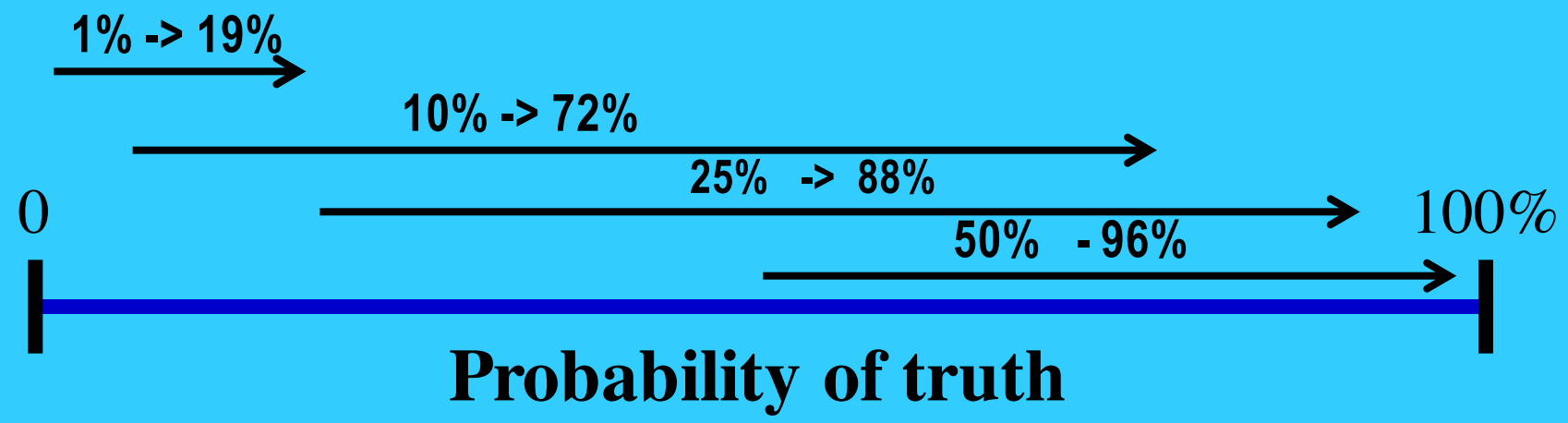
P-values: Bayesian Translations

	$\text{Exp}(-Z^2/2)$			Decrease in probability of the null hypothesis, %	
P-value *(Z-score)*	Minimum Bayes factor	-e p ln(p)	Strength of evidence	From	To no less than
0.10 *(1.64)*	.26	.6	Weak	75 50 17	44 21 5
0.05 *(1.96)*	.15	.4	Weak to Moderate	75 50 26	31 13 5
0.03 *(2.17)*	.1	.3	Moderate	75 50 33	22 - 47 9 - 23 5 - 13
0.01 *(2.58)*	.04	.13	Moderate to strong	75 50 60	10 3.5 5
0.001 *(3.28)*	.005	.02	Strong to very strong	75 50 92	1 0.5 5

P=0.05
(strongest BF =0.14)



P=0.01
(strongest BF =0.04)



Scientific claims are...

NOT

Rather...

Borrowing information from adults to children:

A Case Study of Guillan-Barré Syndrome

Adult vs. Pediatric GBS

- No known difference in pathophysiology or clinical course, except that children recover more quickly, almost never die, and have fewer serious sequelae.

What do we know in adults?: PE vs. Placebo

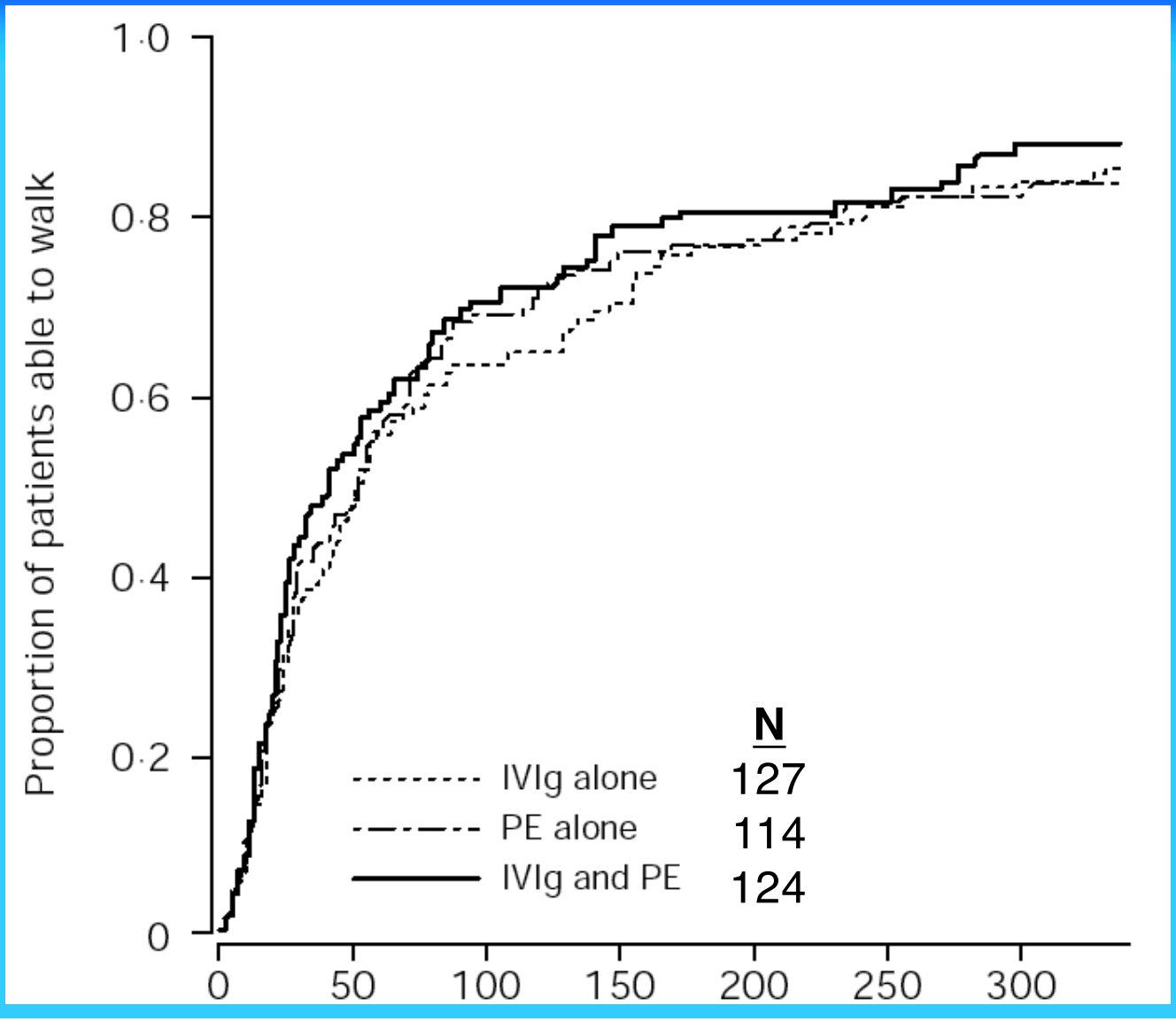
- Two RCTs of plasma exchange versus placebo showed identical effects on median time to unaided walking (Time to grade 2):

Study	N (age)	Placebo	PE	HR	P-value
French GBS Study (1985)	245 (>16)	111 d	70 d	0.63	<0.001
GBS Study Group (1985)	220 (>12)	85 d	53 d	0.62	<0.001

What do we know in children? PE vs. Historical Control Studies

Author	Mean treated (N)	Mean untreated (N)	HR	S.D.
Epstein 1990	24 (9)	60 (14)	0.4	0.17-0.93
Lamont 1991	17 (6)	43 (18)	0.4	0.15 - 0.99
Jansen 1993	16 (8)	29 (11)	0.55	0.23 -1.37
Graf 1999	76 (6)	50 (9)	1.52	0.54 - 4.3
TOTAL		**RE model**	**0.58**	**0.32-1.0**

What do we know in Adults? : PE vs. IV Ig



PSGBS Study Group, Lancet 1997; **349**: 225-230

Median time to walking:
IVIg: 51 d
PE: 49 d
IVIg+PE: 45 d

HR IVIG vs PE:
1.04, CI 0.8 to 1.4

What do we know in Adults? : PE vs. IV Ig

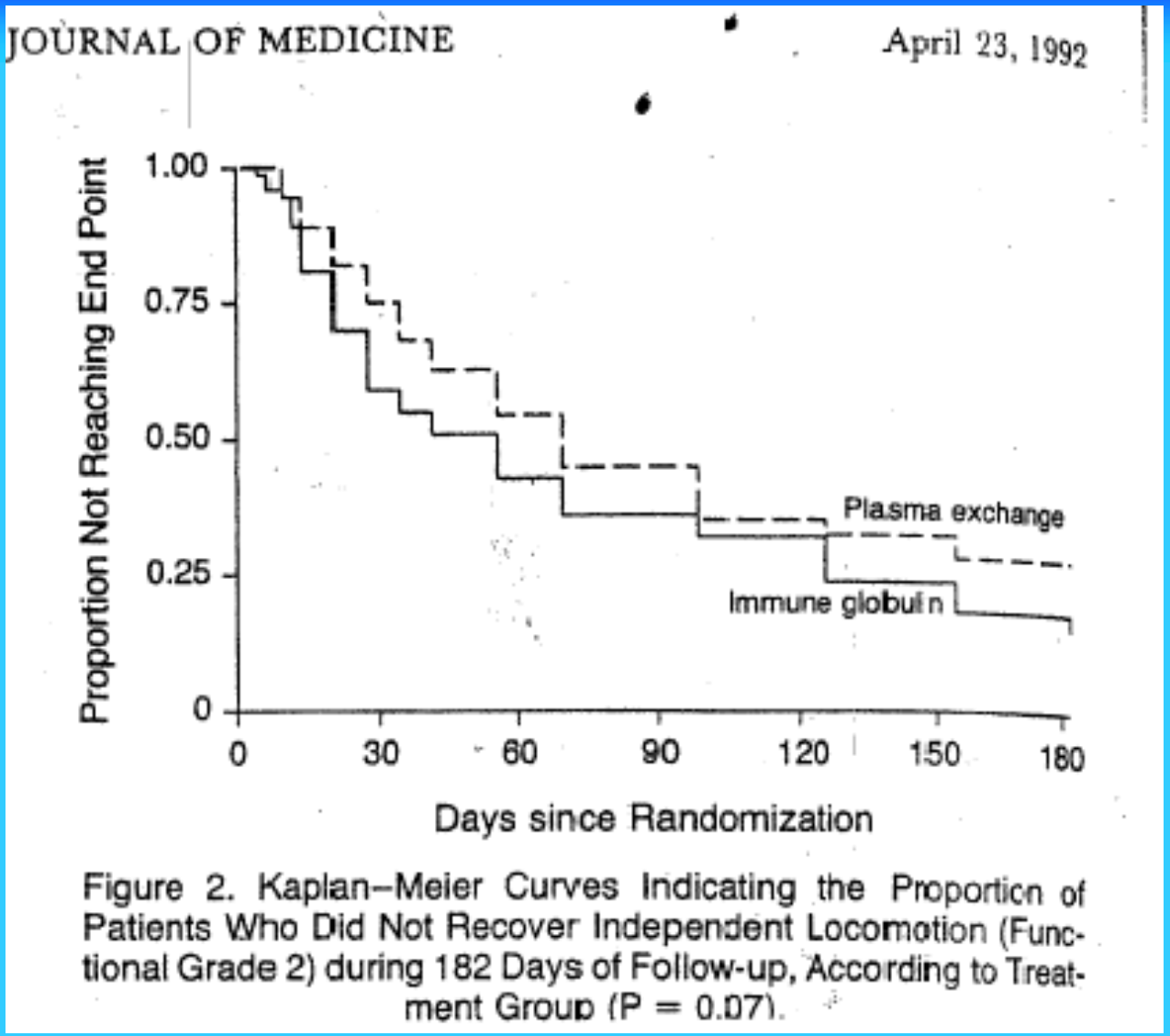


Figure 2. Kaplan–Meier Curves Indicating the Proportion of Patients Who Did Not Recover Independent Locomotion (Functional Grade 2) during 182 Days of Follow-up, According to Treatment Group (P = 0.07).

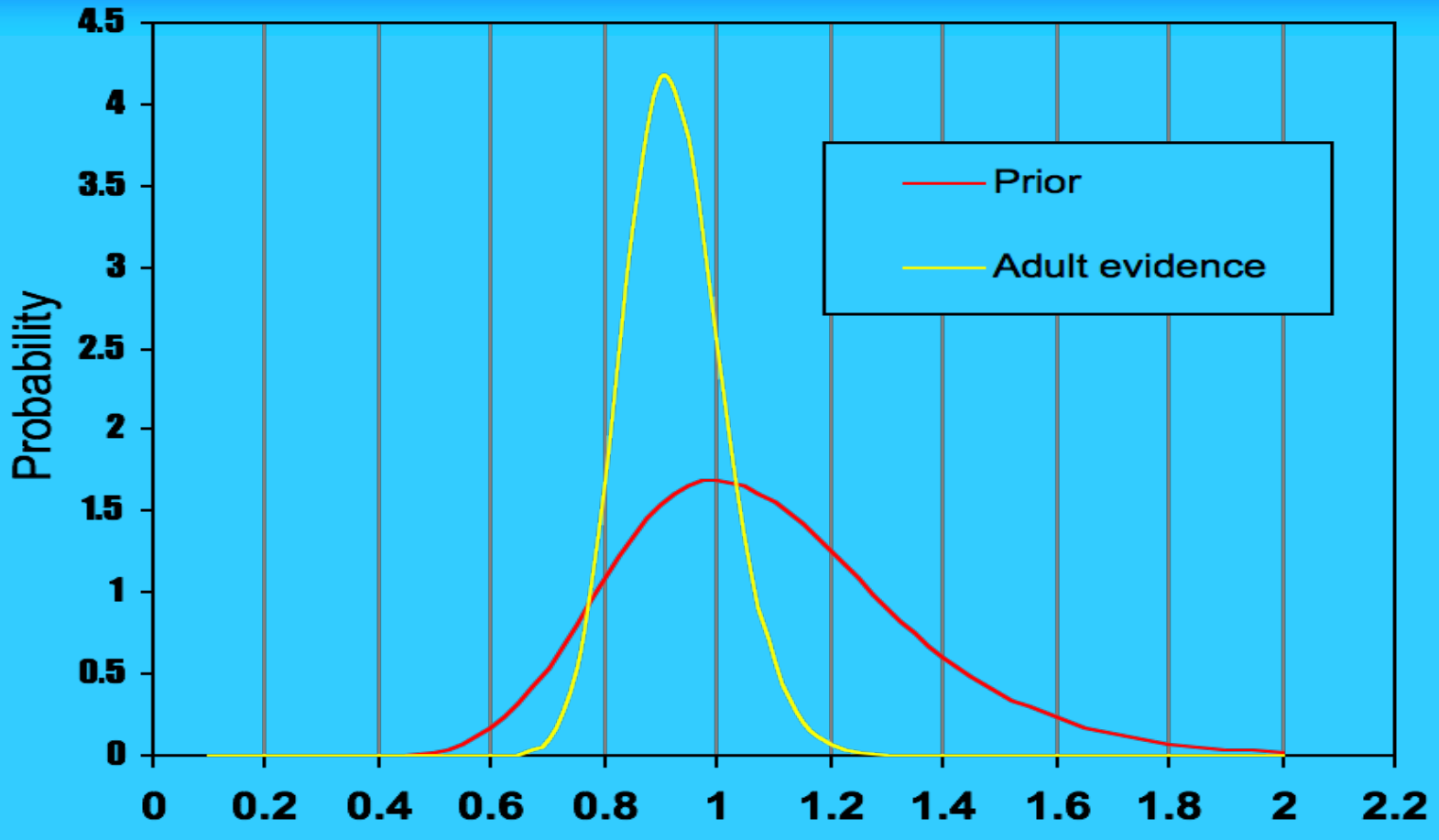
Van der Meché, NEJM 1992

Median time to walking:
IVIg: 55 d (N=74)
PE: 69 d (N=73)

HR:
0.80, CI 0.62 to 1.02, p=0.07

Meta-analytic average of two PE vs. IVIg studies:
HR = 0.91, CI 0.75 to 1.1

Prior on IVIG vs PE Effect



The value of borrowing

- Save 80% of the sample size of a new trial.
- Define the amount of evidence needed to be convincing (BF=3), on top of the adult evidence.
- Promote conversations among scientists about the things they know something about (e.g. similarity of clinical course and pathophysiology)

FDA Discussion

(Fisher, CCT, 20:16-39,1999)

L. Moyé, MD, PhD

"What we have to wrestle with is how to interpret p-values for secondary endpoints in a trial which frankly was negative for the primary. …In a trial with a positive endpoint…you haven't spent all of the alpha on that primary endpoint, and so you have some alpha to spend on secondary endpoints….In a trial with a negative finding for the primary endpoint, you have no more alpha to spend for the secondary endpoints."

FDA Discussion, cont.

(Fisher, CCT, 20:16-39,1999)

Dr. Lipicky: What are the p-values needed for the secondary endpoints? …Certainly we're not talking 0.05 anymore. …You're out of this 0.05 stuff and I would have like to have seen what you thought was significant and at what level…

What p-value tells you that it's there study after study?

Dr. Konstam: …what kind of statistical correction would you have to do that survival data given the fact that it's not a specified endpoint? I have no idea how to do that from a mathematical viewpoint.

Summary

- The sine qua non of EBM is integration of evidence of different types; empirical, mechanistic, clinical.
- Without Bayesian formalism and measures, there is no quantitative language, no conceptual framework, and barely even a qualitative language to do this.
- The social-scientific conventions in current use (e.g. evidence hierarchies and $P \leq 0.05$, Power > 80%) are unmoored from either proper measures of evidential strength or probability of truth. These conventions have also inhibited progress in developing new standards.
- Meta-research can be terrifically informative regarding the reliability of scientific methods.
- The needs and role of FDA are not the same as an EBM practitioner, but it can play a huge role in helping to establish new evidential standards and a regulatory framework for the 21st century that reflects a coherent and consistent approach to inference.

ScienceInsider
Breaking news and analysis from the world of science policy

ALESZU BAJAK

Lectures Aren't Just Boring, They're Ineffective, Too, Study Finds

12 May 2014 3:00 pm | 122 Comments

Wikimedia

Blah? Traditional lecture classes have higher undergraduate failure rates than those using active learning techniques, new research finds.

Are your lectures droning on? Change it up every 10 minutes with more active teaching techniques and more students will succeed, researchers say. A new study finds that

New study says studies are wrong

AFP RELAXNEWS / Friday, August 28, 2015, 8:37 AM

HALFPOINT/SHUTTERSTOCK.COM

Some studies aren't worth stressing over.

Scientific studies about how people act or think can rarely be replicated by outside experts, said a study Thursday that raised new questions about the seriousness of psychology research.

A team of 270 scientists tried reproducing 100 psychology and social science studies that had been published in three top peer-reviewed U.S. journals in

Thank you.