Contradicted and Initially Stronger Effects in Highly Cited Clinical Research

John P. A. Ioannidis, MD

Context Controversy and uncertainty ensue when the results of clinical research on the effectiveness of interventions are subsequently contradicted. Controversies are most prominent when high-impact research is involved.

Objectives To understand how frequently highly cited studies are contradicted or find effects that are stronger than in other similar studies and to discern whether specific characteristics are associated with such refutation over time.

Design All original clinical research studies published in 3 major general clinical journals or high-impact-factor specialty journals in 1990-2003 and cited more than 1000 times in the literature were examined.

Main Outcome Measure The results of highly cited articles were compared against subsequent studies of comparable or larger sample size and similar or better controlled designs. The same analysis was also performed comparatively for matched studies that were not so highly cited.

Results Of 49 highly cited original clinical research studies, 45 claimed that the intervention was effective. Of these, 7 (16%) were contradicted by subsequent studies, 7 others (16%) had found effects that were stronger than those of subsequent studies, 20 (44%) were replicated, and 11 (24%) remained largely unchallenged. Five of 6 highly-cited nonrandomized studies had been contradicted or had found stronger effects vs 9 of 39 randomized controlled trials ($P=.008$). Among randomized trials, studies with contradicted or stronger effects were smaller ($P=.009$) than replicated or unchallenged studies although there was no statistically significant difference in their early or overall citation impact. Matched control studies did not have a significantly different share of refuted results than highly cited studies, but they included more studies with "negative" results.

Conclusions Contradiction and initially stronger effects are not unusual in highly cited research of clinical interventions and their outcomes. The extent to which high citations may provoke contradictions and vice versa needs more study. Controversies are most common with highly cited nonrandomized studies, but even the most highly cited randomized trials may be challenged and refuted over time, especially small ones.

JAMA. 2005;294:218-228 www.jama.com

CLINICAL RESEARCH ON IMPORTANT questions about the efficacy of medical interventions is sometimes followed by subsequent studies that either reach opposite conclusions or suggest that the original claims were too strong. Such disagreements may upset clinical practice and acquire publicity in both scientific circles and in the lay press. Several empirical investigations have tried to address whether specific types of studies are more likely to be contradicted and to explain observed controversies. For example, evidence exists that small studies may sometimes be refuted by larger ones.[1,2]

Similarly, there is some evidence on disagreements between epidemiological studies and randomized trials.[3-5] Prior investigations have focused on a variety of studies without any particular attention to their relative importance and scientific impact. Yet, most research publications have little impact while a small minority receives most attention and dominates scientific thinking and clinical practice. Impact is difficult to measure in all its dimensions. However, the number of citations received by a publication is a surrogate of the attention it has received in the scientific literature and its influence on scientific debate and progress. Citations are readily and objectively counted in established databases.[6] High citation count does not necessarily mean that these studies are accepted; citations may sometimes be critical of an article. Nevertheless, citation count is a measure of how much a study has occupied the thinking of other scientists and has drawn attention—for good or bad.

It is important to evaluate the replication of clinical research studies that have the highest citation impact. How frequently are such studies eventually contradicted by other research or are found to have too strong results compared with subsequent evidence? Is this more common for specific types of studies? Answering these questions would be useful for interpreting the results of influential clinical research.

METHODS

Eligible Original Studies

Eligible original studies for this analysis included all publications that had received more than 1000 Institute for Scientific Information (ISI)–indexed[6]

Author Affiliations: Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece, and the Institute for Clinical Research and Health Policy Studies, Department of Medicine, Tufts-New England Medical Center, Boston, Mass.
Corresponding Author: John P. A. Ioannidis, MD, Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina 45110, Greece (jioannid@cc.uoi.gr).

citations; had been published between 1990 and 2003 in the 3 general medical journals with the current highest impact factor (*New England Journal of Medicine, JAMA, Lancet*) or in medical specialty journals with impact factor exceeding 7.0 (according to the Journal Citation Reports 2003) that are likely to publish clinical research (including in decreasing impact factor, the *Journal of the National Cancer Institute*, *Gastroenterology*, *Annals of Internal Medicine*, *Circulation*, *Journal of Clinical Oncology*, *Archives of General Psychiatry*, *Blood, Hepatology*, *American Journal of Respiratory and Critical Care Medicine*, *Diabetes*, *Brain*, *Annals of Neurology*, *Journal of the American College of Cardiology*, *Diabetes Care*, *Journal of the American Society of Nephrology*, *Arthritis and Rheumatism*, and the *American Journal of Psychiatry*); addressed the efficacy of therapeutic or preventive interventions; and pertained to primary data (excluding reviews and meta-analyses).

Citation counts for articles published between January 1, 1990, and December 31, 2003, in these journals were downloaded from ISI. Citation counts are censored on August 20, 2004. All articles with more than 1000 citations were screened further. Studies with group authorship may be cited in various ways; therefore, I summed up citations cataloged under different entries for the same article (using the first author name, group abbreviations, and anonymous entries).[7] The total citation count does not capture the few citations for which wrong name, journal, volume, or page might have been cited. Since citations depend on the time interval since publication, a separate citation count was limited to the first 3 years after the publication year.

Other Clinical Research on the Same Questions

For each eligible original study, a search was performed to identify whether there had been any other concurrently or subsequently published clinical research addressing the same question. Other research was considered eligible, only if the sample size was close to or larger than that of the highly cited original study or if it used a theoretically better controlled design. Thus, for highly cited randomized trials, I perused all randomized trials having at least 30% of the sample size of the eligible highly cited original study. Whenever available, quantitative meta-analyses of trials were used as summaries of trial results. Whenever several pertinent meta-analyses were available, the one including the largest number of studies was preferred. For highly cited nonrandomized studies, subsequently published pertinent randomized trials and meta-analyses thereof were eligible regardless of sample size; nonrandomized evidence was also considered, if randomized trials were not available.

Concurrently or subsequently published evidence was identified in PubMed using searches that combined terms pertaining to the tested interventions, disease and outcome, and terms pertinent to the search of randomized trials and meta-analyses. Searches followed the Cochrane algorithms for finding meta-analyses and randomized trials.[8]

Data Extraction and Classification of Studies

For each eligible original study, I recorded the study name, intervention, disease and outcomes of interest, study design, sample size, main conclusions, and citation counts. For the articles presenting or summarizing other relevant research, I recorded the study design, total sample size, and the findings as compared with those of the original highly cited study.

Highly cited studies were classified as negative (when they claimed the tested experimental intervention was ineffective, harmful, or no better from the control intervention), unchallenged (when no other clinical research of eligible design and sample size was available to validate the claimed efficacy), contradicted, initially stronger effects, or replicated effects. The classification of studies in these categories was based on the final interpretation of the results by the authors in the "Abstract" and "Discussion" sections of their original publications. Highly cited articles were classified according to whether their authors suggested that an intervention was overall effective or ineffective. When both benefits and harms or caveats were presented, I focused on the net conclusion of whether the experimental intervention merits consideration for use in clinical practice. Subsequent research was classified in the same manner. Contradiction was declared when the original highly cited study claimed the intervention to be effective, while subsequent research showed it to be ineffective. When both original and subsequent research claimed the intervention was effective, studies were compared further regarding the effect size for the major clinical outcome, the durability of the treatment effect, and the generalizability and applicability to various settings. Initially stronger effects were defined when the relative risk reduction for the main outcome in the subsequent research was half or less compared with what had been proposed by the original highly cited study (regardless of whether confidence intervals might overlap or not), or when the subsequent research showed that the originally proposed benefit was of short duration or its applicability and generalizability was limited. Classification of the studies independently by another investigator yielded a highly similar profile (weighted Cohen $\kappa = 0.92$).

Correlates of Contradicted or Initially Stronger Effects

Among original highly cited studies with efficacy claims, analyses examined whether those with contradicted or initially stronger effects differed from the replicated and unchallenged ones in study design, publication year, sample size, type of disease (heart disease vs other), journal of publication, citation count, early citation count, and average citations per year after publication. Comparisons used the Mann-Whitney *U* test for continuous variables and Fisher exact test for binary variables.

Comparison of Highly Cited Articles Against Less Cited Articles

To evaluate whether highly cited studies differ from other studies that are not so highly cited in their findings and potential for contradiction, a control group of articles pertaining to the assessment of interventions was also assembled. Control-group articles were 1:1 matched for journal, year of publication, and design (randomized vs nonrandomized) against each of the highly cited articles. Control articles were selected by screening chronologically the contents of the pertinent journals for each pertinent year starting July 1 (to ensure approximately similar follow-up for citations with the highly cited articles against which they were matched). Other research was searched and the control articles were categorized in a similar fashion as described for the highly cited articles above. Differences between highly cited and control articles were examined with conditional logistic regression to account for matching.

Analyses

Analyses were performed in SPSS version 12.0 (SPSS Inc, Chicago, Ill) and StatXact (Cytel Corp, Boston, Mass). *P* values are 2-tailed, and $P<.05$ was considered statistically significant.

RESULTS

Eligible Studies

One hundred fifteeen articles published between 1990 and 2003 had received more than 1000 citations (major general clinical journals, n=91; specialty journals, n=24). Of those, 66 were excluded (nonsystematic reviews or editorials, n=20; meta-analyses, n=7; case-control studies of risk factors, n=12; prevalence or incidence studies, n=8; cohort studies of risk factors, n=3; recommendations, n=3; prognostic models, n=4; time-trend analysis, n=1; case series, n=1; presentations of interviews, instruments, or assays n=3, classification criteria n=4). The remaining 49 articles were eligible (TABLE 1)[9-57] of which 47 had appeared in major general medical journals. They included 43 randomized trials, 4 prospective cohorts, and 2 case series. In recent years (1998 through 2003), the 3 general journals have published an almost equal number of highly cited articles (*New England Journal of Medicine*, n=4; *JAMA*, n=3; *Lancet*, n=3). A smaller proportion of highly cited articles published in specialty journals than those published in general journals were eligible for the analysis (2/24 vs 47/91, $P<.001$), because highly cited articles in specialized journals were mostly nonsystematic reviews or editorials (10/24); classification criteria (4/24); or descriptions of standardized interviews, instruments, and assays (3/24). Many diverse disciplines were represented, but the most common topic was heart disease (n=27).

Four eligible highly cited studies showed no efficacy for the tested interventions. They contradicted prior claims for potential efficacy of vitamin E, beta carotene, and retinol for lung cancer and/or coronary artery disease; and showed an increased risk of coronary artery disease with hormone therapy in postmenopausal women (TABLE 2).

Of the 45 eligible highly cited studies with efficacy claims (Table 2), 7 (16%) were contradicted by subsequent research, and another 7 (16%) were found to have initially stronger effects. In all these 14 cases (BOX 1), subsequent studies were either larger or better controlled (randomized vs a nonrandomized original study). The findings of 20 highly cited articles (44%) were replicated (also with a larger sample size in subsequent research compared with the original highly cited study) and 11 (24%) had remained largely unchallenged.[58-78]

Comparison of Contradicted or Initially Stronger vs Replicated or Unchallenged Findings

Five of 6 highly cited nonrandomized studies had been contradicted or had initially stronger effects while this was seen in only 9 of 39 highly cited randomized trials ($P=.008$). TABLE 3 shows that trials with contradicted or initially stronger effects had significantly smaller sample sizes and tended to be older than those with replicated or unchallenged findings. There were no significant differences on the type of disease. The proportion of contradicted or initially stronger effects did not differ significantly across journals ($P=.60$). There was also no significant difference in the number of citations received in the first 3 years between these 2 groups or in the overall number of citations over time although the citations per year tended to be nonsignificantly fewer in trials with contradicted or initially stronger effects.

Comparison of Highly Cited Articles Against Less-Cited Control Articles

Of the 49 articles in the control group[79-127] (with median of 157 citations, range 38-815, until 2004), the findings of 2 articles[91,119] were contradicted[128,129] and 8 studies* had initially stronger effects[130-137] (BOX 2); 20 articles† contained "positive" findings that were replicated,[68,138-155] 8 studies‡ remained unchallenged, and 11 studies§ did not have any "positive" results; in 7 articles with some "positive" finding,[79,87,91,98,108,112,120] there were also other interventions evaluated that had "negative" results although this mixture of "positive" and "negative" results had not been observed in any of the highly cited articles. The control articles had a larger number of "negative" findings compared with the highly cited articles (matched odds ratio [OR], 8; 95% confidence interval [CI], 1.8-34; $P=.006$ for any "negative" finding; and matched OR, 3.3; 95% CI, 0.92-12.0, $P=.07$ for exclusively "negative" findings). The highly cited articles did not have a smaller proportion of contradicted or initially stronger effects than the control articles if anything

*References 82, 90, 92, 95, 96, 109, 110, 117.
†References 79-81, 83, 86-89, 101, 103, 104, 106, 108, 111, 112, 118, 123, 125-127.
‡References 93, 97, 98, 102, 107, 114, 115, 120.
§References 84, 85, 94, 99, 100, 105, 113, 114, 119, 120, 122.

Table 1. Eligible Highly Cited Studies

Study	Type of Intervention and Disease	Design	Sample Size	No. of Citations	
				All	3-Year
ACTG019,[9] 1990	Zidovudine in asymptomatic HIV-1 infection	RCT	1338	1179	549
Brown et al,[10] 1990	Lipid lowering to decrease coronary lesions and CAD	RCT	146	1312	394
Moertel et al,[11] 1990	Levamisole and fluorouracil for colon cancer	RCT	246	1050	259
V-HeFT II,[12] 1991	Enalapril vs hydralazine + isosorbide for CHF	RCT	804	1469	386
Nurses' Health Study,[13] 1991	Postmenopausal hormonal therapy for CAD prevention	Cohort	48 470	1355	230
NASCET,[14] 1991	Carotid endarterectomy in high-grade stenosis	RCT	659	2434	347
HA-1A Sepsis,[15] 1991	Monoclonal antibody to endotoxin for gram-negative sepsis	RCT	200	1028	435
SOLVD,[16] 1991	Enalapril in patients with LV dysfunction	RCT	2569	2798	1113
SAVE,[17] 1992	Captopril for patients after MI	RCT	2231	2803	632
PAMI,[18] 1993	Angioplasty vs tPA thrombolysis in acute MI	RCT	395	1642	868
Captopril Collaborative,[19] 1993	Captopril for slowing disease progression in diabetic nephropathy	RCT	409	2090	388
Health Professionals,[20] 1993	Vitamin E for CAD prevention in men	Cohort	39 910	1281	409
Nurses' Health Study,[21] 1993	Vitamin E for CAD prevention in women	Cohort	87 245	1131	409
Rossaint et al,[22] 1993	Nitric oxide inhalation for acute respiratory distress syndrome	Case series	9	1025	399
DCCT,[23] 1993	Intensive management to reduce type 1 diabetes complications	RCT	1441	6005	1260
EPIC,[24] 1994	7E3 in high-risk angioplasty	RCT	2099	1467	203
ACTG076,[25] 1994	Zidovudine to reduce perinatal HIV-1 transmission	RCT	477	1449	461
STRESS,[26] 1994	Stent vs balloon angioplasty in CAD	RCT	410	2153	543
BENESTENT,[27] 1994	Stent vs balloon angioplasty in single-vessel CAD	RCT	520	2295	633
ABC,[28] 1994	Vitamin E and beta carotene for lung cancer	RCT	29 133	1872	542
NINDS rt-PA,[29] 1995	rt-PA in acute stroke	RCT	624	1939	485
WOSCOPS,[30] 1995	Pravastatin in hypercholesterolemia	RCT	6595	3163	901
CARE,[31] 1996	Pravastatin after MI with average cholesterol	RCT	4195	2795	908
US Carvedilol,[32] 1996	Carvedilol for CHF	RCT	1094	1544	543
BERET,[33] 1996	Beta carotene/retinol for preventing lung cancer/CAD	RCT	18 314	1044	439
Physicians' Health,[34] 1997	Aspirin to prevent MI in men with various C-reactive protein levels	RCT	1086	1597	539
ACTG320,[35] 1997	Triple therapy with indinavir vs 2 nucleosides in HIV-1 infection	RCT	1156	1293	728
EPILOG,[36] 1997	Abciximab glycoprotein IIb/IIIa blockade in PCI	RCT	2792	1066	596
HIT,[37] 1998	Interferon alfa-2b + ribavirin vs interferon alone for chronic hepatitis C	RCT	912	1319	612
LIPID,[38] 1998	Pravastatin for secondary CAD prevention	RCT	9014	1641	750
RALES,[39] 1999	Spironolactone in severe CHF	RCT	1663	1085	635
HOPE,[40] 2000	Ramipril to prevent CAD in high-risk patients without LV dysfunction/CHF	RCT	9297	1777	1323
SHEP,[41] 1991	Treatment of systolic hypertension in elderly adults	RCT	4736	1872	397
PEPI,[42] 1995	Postmenopausal estrogen/progestin for CAD risk factors	RCT	875	1300	320
ACAS,[43] 1995	Endarterectomy in asymptomatic stenosis >60%	RCT	1662	1427	416
HERS,[44] 1998	Estrogen/progestin for secondary CAD prevention	RCT	2763	2050	987
AFCAPS/TexCAPS,[45] 1998	Lovastatin for primary CAD prevention with average cholesterol	RCT	6605	1559	731
WHI,[46] 2002	Estrogen/progestin for CAD prevention	RCT	16 608	1468	2000*
MRC Vitamin,[47] 1991	Folate to prevent neural tube defects	RCT	1817	1096	378
Zutphen Elderly,[48] 1993	Flavonoids for CAD prevention	Cohort	805	1233	151
4S,[49] 1994	Simvastatin in hypercholesterolemia with previous CAD	RCT	4444	4614	990
CAPRIE,[50] 1996	Clopidogrel vs aspirin in patients at risk of ischemic events	RCT	19 185	1139	280
CHAOS,[51] 1996	Vitamin E to prevent MI and death in patients with CAD	RCT	2002	1004	425
HOT,[52] 1998	Intensive blood-pressure lowering/low-dose aspirin in hypertension	RCT	18 790	1539	799
IHIT,[53] 1998	Interferon alfa-2b + ribavirin vs interferon alone for chronic hepatitis C	RCT	832	1004	486
UKPDS 34,[54] 1998	Intensive management of type 2 diabetes with insulin or sulphonylureas	RCT	3867	2748	1238
CIBIS-II,[55] 1999	Bisoprolol for CHF	RCT	2647	1064	653
Castaigne et al,[56] 1990	All-trans retinoic acid for acute promyelocytic leukemia	Case series	22	1030	270
NSABP P-1,[57] 1998	Tamoxifen for breast cancer prevention	RCT	13 388	1470	745

Abbreviations: ABC, Alpha-Tocopherol, Beta Carotene Cancer Prevention; ACAS, Asymptomatic Carotid Atherosclerosis Study; ACTG, AIDS Clinical Trials Group; AFCAPS/TexCAPS, Air Force/Texas Coronary Atherosclerosis Prevention Study; BENESTENT, Belgian Netherlands Stent; BERET, Beta Carotene and Retinol Efficacy Trial; CAD, coronary artery disease; CAPRIE, Clopidogrel vs Aspirin in Patients at Risk of Ischemic Events; CARE, Cholesterol and Recurrent Events; CHAOS, Cambridge Heart Antioxidant Study; CHF, congestive heart failure; CIBIS-II, Cardiac Insufficiency Bisoprolol Study II; DCCT, Diabetes Control and Complications Trial; EPIC, Evaluation of 7E3 for the Prevention of Ischemic Complications; EPILOG, Evaluation in PTCA to Improve Long-Term Outcome with Abciximab Glycoprotein IIb/IIIa Blockade; HA-1H, human IgM monoclonal antibody; HERS, Heart and Estrogen/progestin Replacement Study; HIT, Hepatitis Interventional Therapy; HIV-1, human immunodeficiency virus type 1; HOPE, Heart Outcomes Prevention Evaluation; HOT, Hypertension Optimal Treatment; IHIT, International Hepatitis Interventional Therapy; LIPID, Long-Term Intervention with Pravastatin in Ischaemic Disease; LV, left ventricular; MI, myocardial infarction; MRC, Medical Research Council; NASCET, North American Symptomatic Carotid Endarterectomy Trial; NINDS rt-PA, National Institute of Neurological Disorders and Stroke recombinant tissue-Plasminogen Activator; NSABP P-1, National Surgical Adjuvant Breast and Bowel Project P-1; PAMI, Primary Angioplasty in Myocardial Infarction; PCI, percutaneous coronary intervention; PEPI, Postmenopausal Estrogen/Progestin Interventions; RALES, Randomized Aldactone Evaluation Study; RCT, randomized controlled trial; SAVE, Survival And Ventricular Enlargement; SHEP, Systolic Hypertension in the Elderly Program; SOLVD, Studies of Left Ventricular dysfunction; STRESS, Stent Restenosis Study; UKPDS 34, UK Prospective Diabetes Study 34; V-HeFT II, Vasodilator-Heart Failure Trial II; WHI, Women's Health Initiative; WOSCOPS, West of Scotland Coronary Prevention Study; 4S, Scandinavian Simvastatin Survival Study.

*Projected.

Table 2. Other Research and Current State of Knowledge

Highly Cited Study	Other Research	No. of Participants*	Comment on Current State of Knowledge
Contradicted studies			
Nurses' Health Study[13]	RCT[46]	16 608	Estrogen/progestin do not protect from but increase CAD risk in postmenopausal women
HA-1A Sepsis[15]	RCT[62]	2199	Contrary to initial findings, HA-1A did not improve survival in gram-negative sepsis
Health Professionals[20]	RCT[66]	6996	Contrary to initial findings, vitamin E supplementation does not reduce CAD in men
Nurses' Health[21]	RCT[66]	2545	Contrary to initial findings, vitamin E supplementation does not reduce CAD in women
Rossaint et al,[22] (nitric oxide)	MA RCT[67]	535	Despite initial claims of better oxygenation, nitric oxide does not improve survival in respiratory distress syndrome
PEPI[42]	RCT[46]	16 608	Estrogen/progestin do not protect from but increase CAD risk in postmenopausal women
CHAOS[51]	RCT[66]	9541	Contrary to initial findings, vitamin E does not prevent coronary events
Initially stronger effects			
ACTG019[9]	MA RCT[58]	5566	The early benefit of zidovudine against HIV-1 disease progression decreases over time
PAMI[18]	MA RCT[65]	2593	Superiority of angioplasty over tPA thrombolysis may be less prominent than originally proposed and pertinent mostly to specialized centers
STRESS[26]	MA RCT[69]	9918	Stents reduce restenosis and need for revascularization compared with simple angioplasty, but the effect may be inflated by lack of blinding and is probably modest
BENESTENT[27]	MA RCT[69]	9918	Stents reduce restenosis and need for revascularization compared with simple angioplasty, but the effect may be inflated by lack of blinding and is probably modest
NINDS rt-PA[29]	MA RCT[70]	2775	rt-PA may improve outcomes in acute ischemic stroke, but benefit is limited and seen only when treatment is given very early
ACAS[43]	MA RCT[75]	2440	Carotid endarterectomy has a small absolute benefit in asymptomatic stenosis >60%
Zutphen Elderly[48]	MA cohorts[76]	105 000	Flavonoids reduce the risk of CAD modestly
Replicated studies			
Brown et al,[10] (lipid lowering)	MA RCT[59]	148 321	Cholesterol and LDL lowering achieves significant risk reductions in CAD
Moertel et al,[11] (levamisole/5-FU)	MA RCT[60]	3302	Fluorouracil adjuvant therapy improves survival in colon cancer
NASCET[14]	MA RCT[61]	6092	Carotid endarterectomy is effective in symptomatic patients with 70%-99% stenosis
SOLVD[16]	MA RCT[63]	7105	ACE inhibition reduces mortality and hospitalizations in patients with CHF
SAVE[17]	MA RCT[64]	105 337	ACE inhibition reduces mortality after MI
EPIC[24]	MA RCT[68]	20 137	Glycoprotein IIB/IIIA antagonists reduce cardiovascular events in percutaneous revascularization
WOSCOPS[30]	MA RCT[59]	148 321	Statins achieve significant risk reductions in CAD
CARE[31]	MA RCT[59]	148 321	Statins achieve significant risk reductions in CAD
US Carvedilol[32]	MA RCT[71]	10 135	β-Blockers decrease mortality in patients with CHF
ACTG320[35]	MA RCT[72]	4686	Protease-inhibitor–based triple therapy improves survival compared with double nucleosides in HIV-1 infection
EPILOG[36]	MA RCT[68]	20 137	Glycoprotein IIB/IIIA antagonists reduce cardiovascular events in percutaneous revascularization
HIT[37]	MA RCT[73]	6585	Interferon alfa-2b + ribavirin has better outcomes than interferon alone in chronic hepatitis C
LIPID[38]	MA RCT[59]	148 321	Statins achieve significant risk reductions in CAD
SHEP[41]	MA RCT[74]	15 693	Treatment of isolated hypertension in elderly patients reduces the risk of stroke
AFCAPS/TexCAPS[45]	MA RCT[59]	148 321	Cholesterol and LDL lowering achieves significant risk reductions in CAD
4S[49]	MA RCT[59]	148 321	Statins achieve significant risk reductions in CAD
IHIT[53]	MA RCT[73]	6585	Interferon alfa-2b plus ribavirin has better outcomes than interferon alone in chronic hepatitis C
CIBIS-II[55]	MA RCT[71]	10 135	β-Blockers decrease mortality in patients with CHF
All-trans-retinoid acid[56]	RCT[77]	346	All-trans retinoic acid is effective for acute promyelocytic leukemia
NSABP P-1[57]	MA RCT[78]	28 406	Tamoxifen is effective for the prevention of breast cancer
Unchallenged studies			
V-HeFT II,[12]			ACE inhibition is superior to vasodilators for CHF
Captopril Collaborative[19]			ACE inhibition slows renal disease progression in diabetes with macroproteinuria (benefit subsequently extended to microproteinuria and patients without diabetes)
DCCT[23]			Intensive insulin management of type 1 diabetes reduces microvascular complications (subsequent research has addressed increasingly intensive management)
ACTG076[25]			Zidovudine reduces the risk of perinatal HIV-1 transmission (subsequent research has addressed shorter and more convenient regimens)
Physicians' Health[34]			Aspirin prevents MI especially in men with high levels of C-reactive protein
MRC Vitamin[47]			Folate supplementation significantly reduces the risk of neural tube defects (subsequent research has addressed various doses and modes of administration of folate)
RALES[39]			Spironolactone reduces morbidity and mortality in CHF (no other similar trial)
HOPE[40]			Ramipril prevents CAD events in high-risk patients without left ventricular dysfunction (no other similar trial)
CAPRIE[50]			Clopidogrel seems superior to aspirin in preventing stroke and MI in patients at risk of ischemic stroke (subsequent research has addressed the combination of clopidogrel and aspirin)
HOT[52]			Intensive blood pressure lowering decreases the risk of cardiovascular events (2 much smaller trials have shown similar effects of intensive blood pressure lowering in patients with diabetes)
UKPDS 34[54]			Intensive management of type 2 diabetes reduces the risk of microvascular complications
Negative studies			
ABC[28]			Neither α-tocopherol nor beta carotene prevents lung cancer
BERET[33]			Neither beta carotene nor retinol prevent lung cancer or CAD
HERS[44]			Estrogen/progestin are ineffective for secondary CAD prevention in postmenopausal women
WHI[46]			Estrogen/progestin do not protect from but increase CAD risk in postmenopausal women

Abbreviations: The abbreviations of the highly cited studies correspond to the popular names listed in Table 1. ACE, angiotensin-converting enzyme; CAD, coronary artery disease; CHF, congestive heart failure; HA-1A, human IgM monoclonal antibody; HIV-1, human immunodeficiency virus type 1; LDL, low-density lipoprotein; MA, meta-analysis; RCT, randomized controlled trial; rt-PA, recombinant tissue-type plasminogen activator; tPA, tissue plasminogen activator.
*For meta-analyses, the number of participants refers to the total sample size of all studies (large and small ones) and includes the sample size of the original highly cited study.

Box 1. Contradicted and Initially Stronger Effects in Highly Cited Studies

Contradicted Findings

The Nurses' Health Study,[13] a prospective cohort, found a 44% relative risk reduction in coronary artery disease events in women receiving hormone therapy. A small randomized trial[42] found major beneficial effects of this intervention on surrogate markers of coronary artery disease (lipoprotein and fibrinogen levels) claiming that this should translate to a major clinical benefit. Although the latter trial was not refuted at the level of surrogate outcomes, inferences for the anticipated effects on clinical outcomes were contradicted. The Women's Health Initiative,[46] a large randomized trial, found that estrogen and progestin significantly increased the relative risk of coronary events by 29% among postmenopausal women, and refuting results were also seen in another large randomized trial, the Heart and Estrogen/progestin Replacement Study (HERS).[44]

Two large prospective cohort studies, the Health Professionals Follow-Up study[20] and the Nurses' Health Study,[21] found that vitamin E was significantly associated with a decreased risk of coronary artery disease and a trial of 2002 patients also suggested a 47% relative risk reduction for cardiovascular deaths or nonfatal myocardial infarction with vitamin E.[51] However, an even larger randomized trial[66] subsequently showed absolutely no beneficial effect for vitamin E on coronary artery disease (relative risk 1.05 for cardiovascular deaths and 1.02 for myocardial infarction).

A small randomized trial (n=200) suggested that the human IgM monoclonal antibody to endotoxin could almost halve mortality due to gram-negative sepsis.[15] A subsequent randomized trial of more than 10-fold larger sample size[62] found a nonsignificant 11% relative risk increase for mortality.

Finally, a small series of 9 patients[22] proposed that nitric oxide inhalation is very effective in patients with respiratory distress syndrome by improving oxygenation. However 5 randomized trials involving 535 patients[67] failed to show any clinical benefit.

Initially Stronger Effects

The early results of a trial on zidovudine monotherapy in asymptomatic patients with human immunodeficiency virus infection[9] showed a significant 60% relative risk reduction against disease progression in the first year. The short-term benefit was not exaggerated. Yet this effect was short-lived and the benefit was lost after 18 months both in the same trial and also as shown in a subsequent meta-analysis.[58]

A randomized trial of 395 patients[18] showed that immediate angioplasty was superior to thrombolysis with tissue plasminogen activator in acute myocardial infarction, achieving a 58% relative risk reduction for death or reinfarction. However, a subsequent meta-analysis with more than 2500 patients[65] suggested that the benefit is probably much smaller (relative risk reduction 30%) and the largest and most recent trial that involved both specialized and nonspecialized centers had not shown any sizeable benefit of angioplasty (nonsignificant 20% risk reduction for death and nonsignificant 33% risk reduction for reinfarction).

Two randomized trials of 410 and 520 patients, respectively,[26,27] showed that stents were superior to balloon angioplasty for management of coronary artery disease with 31% and 42% relative risk reductions, respectively, in the need for revasularization. Current evidence, as summarized by a meta-analysis of almost 10 000 patients, suggests that the benefit is probably much smaller that originally thought (approximately 10% relative risk reduction), and unblinding may have led to an increased effect on repeat angioplasty in these trials.[69]

Another trial suggested a prime role for tissue plasminogen activator in acute ischemic stroke.[29] However, subsequent evidence has narrowed indications and the intervention is considered effective mostly when given very early after symptom onset.[70]

Carotid endarterectomy was initially reported to achieve a 5.9% absolute risk reduction for stroke or death, projected at 5 years,[43] in patients with asymptomatic stenosis of the carotid artery exceeding 60%. A meta-analysis of several trials suggested a more modest benefit with 2% absolute risk reduction at 3.1 years.[75]

Finally, a cohort study of 805 people found a 68% adjusted relative risk reduction for coronary artery disease with flavonoids[48] while a meta-analysis of prospective cohorts with total sample size exceeding 100 000 suggests only a 20% relative risk reduction in the top vs bottom third of flavonoid uptake.[76]

there was a trend for more contradicted or initially stronger effects in the highly cited articles (matched OR, 1.6; 95% CI, 0.6-4.0; $P=.35$; matched OR, 6.0; 95% CI, 0.7-50; $P=.10$ when limited to contradicted findings).

COMMENT

Original highly cited articles about medical interventions are published almost exclusively in 3 general medical journals. Actually, there has been an approximate equal share of very highly cited articles among these 3 journals since 1998 as impact factor differences have diminished among these 3 journals. Articles in specialty journals that reach such high numbers of citations are usually review articles or articles describing tools useful to specific diseases rather than original data. Contradicted and potentially exaggerated findings are not uncommon in the most visible and most influential original clinical research: 16% of the top-cited clinical research articles on postulated effective medical interventions that have been published within the last 15 years have been contradicted by subsequent clinical studies and another 16% have been found to have initially stronger effects than subsequent research. Contradiction or initially stronger effects have been encountered in 5 of 6 cases for which nonrandomized designs were used, but even randomized trials have not escaped controversy. More

Table 3. Comparison of Characteristics and Citation Counts of Randomized Trials With Contradicted or Initially Stronger Effects vs Those With Replicated or Unchallenged Findings

Characteristic	Contradicted or Initially Stronger Effects (n = 9)	Replicated or Unchallenged (n = 30)	*P* Value
Published in 1990-1995	8	15	.06
Heart disease topic	4	13	1.00
Sample size, median (IQR)	624 (403-1500)	2165 (892-5201)	.009
All citations received, median (IQR)	1427 (1104-2046)	1542 (1255-2513)	.43
Citations in 3 y, median (IQR)	485 (421-591)	622 (393-825)	.32
Citations per year, median (IQR)	149 (105-215)	214 (146-263)	.07

Abbreviations: IQR, interquartile range.

than a third of the top-cited randomized trials published from 1990 through 1995 have already been affected, while for more recent trials, the time frame is still early and more may be contradicted in the future. Sample size seems to be important, with smaller sample sizes in trials that have met controversy vs those that have not.

The classification of studies in this analysis involves many judgments pertaining to the complexity of studying a given research question with somewhat different populations, interventions, durations, and outcomes. However, these studies are widely known for their inferences and this is also proven by the high interrater agreement. Nevertheless, it should also be acknowledged that although the classification was performed in duplicate, the searches were performed by only 1 investigator. It is unavoidable that some other investigators may feel differently about the categorization of specific studies, especially for topics that may also have heavy debates surrounding them. However, this is unlikely to change the aggregate picture about refutation rates.

The examination of contradictions and refutations offers a fascinating look at the process of science. Four of the highly cited articles examined herein were refuting investigations with "negative" results. However, in a sense, even the other highly cited articles with "positive" results refuted prior knowledge and practice by introducing new concepts and proposing new interventions. We should acknowledge that there is no proof that the subsequent studies and meta-analyses were necessarily correct. A perfect gold standard is not possible in clinical research, so we can only interpret results of studies relative to other studies. Whenever new research fails to replicate early claims for efficacy or suggests that efficacy is more limited than previously thought, it is not necessary that the original studies were totally wrong and the newer ones are correct simply because they are larger or better controlled. Alternative explanations for these discrepancies may include differences in disease spectrum, eligibility criteria, or the use of concomitant interventions.[156] Different studies on the same question are typically not replicas of each other. In fact discrepancies may be interesting on their own because they require careful scrutiny of the data and reappraisal of our beliefs. Thus, it is probably not surprising that the citation rate of these refuted studies did not seem to be much affected. Nevertheless, the controversy generates considerable uncertainty for clinical practice and none of the contradicted interventions is currently recommended by practice guidelines.

The mere fact that a study is highly cited suggests that there is a strong active interest in the questions addressed from a clinical or research perspective. This may increase the chances that other, larger trials may eventually be conducted. However, for most clinical questions of interest, no large trials are ever conducted and evidence is based only on small trials or nonrandomized studies.[157] Small trials or meta-analyses thereof may often be refuted subsequently by large trials[1,2] when such large trials are performed. Small studies using surrogate markers may also sometimes lead to erroneous clinical inferences.[158] There were only 2 studies with typical surrogate markers among the highly cited studies examined herein, but both were subsequently contradicted in their clinical extrapolations about the efficacy of nitric oxide[22] and hormone therapy.[42] In the case of initially stronger effects, the differences in the effect sizes could often be within the range of what would be expected based on chance variability. This reinforces the notion that results from clinical studies, especially early ones, should be interpreted using not only the point estimates but also the uncertainty surrounding them. However, besides differences in effect sizes, most initially stronger effects pertained also to issues of durability, generalizability, or applicability of the proposed effects, as discussed above. Thus, clinicians should be aware that these important aspects may not be fully settled when an important treatment breakthrough is announced.

A third of the most-cited clinical research seems to have replication problems, and this seems to be as large, if not larger, than the vast majority of other, less-cited clinical research. The current analysis found that matched studies that were not so highly cited had a greater proportion of "negative" findings and similar or smaller proportions of contradicted results as the highly cited ones. Publication bias[159,160] and time-lag bias[161,162] favoring the rapid and prominent publication of "positive" findings may underlie some of the observed phenomena. Highly cited articles are already a selected sample with underrepresentation of "negative" findings compared with the average article on interventions published in major journals. It is possible that high-profile journals may tend to publish occasionally very striking findings and that this may lead to some difficulty in replicating some of these findings.[163] Poynard et al[164] evaluated the conclusions of hepatology-related articles published between 1945 and 1999 and found that, overall, 60% of these conclusions were

Box 2. Contradicted and Initially Stronger Effects in Control Studies

Contradicted Findings

In a prospective cohort,[91] vitamin A was inversely related to breast cancer (relative risk in the highest quintile, 0.84; 95% confidence interval [CI], 0.71-0.98) and vitamin A supplementation was associated with a reduced risk ($P=.03$) in women at the lowest quintile group; in a randomized trial[128] exploring further the retinoid-breast cancer hypothesis, fenretinide treatment of women with breast cancer for 5 years had no effect on the incidence of second breast malignancies.

A trial (n=51) showed that cladribine significantly improved the clinical scores of patients with chronic progressive multiple sclerosis.[119] In a larger trial of 159 patients, no significant treatment effects were found for cladribine in terms of changes in clinical scores.[129]

Initially Stronger Effects

A trial (n=28) of aerosolized ribavirin in infants receiving mechanical ventilation for severe respiratory syncytial virus infection[82] showed significant decreases in mechanical ventilation (4.9 vs 9.9 days) and hospital stay (13.3 vs 15.0 days). A meta-analysis of 3 trials (n=104) showed a decrease of only 1.8 days in the duration of mechanical ventilation and a nonsignificant decrease of 1.9 days in duration of hospitalization.[130]

A trial (n=406) of intermittent diazepam administered during fever to prevent recurrence of febrile seizures[90] showed a significant 44% relative risk reduction in seizures. The effect was smaller in other trials and the overall risk reduction was no longer formally significant[131]; moreover, the safety profile of diazepam was deemed unfavorable to recommend routine preventive use.

A case-control and cohort study evaluation[92] showed that the increased risk of sudden infant death syndrome among infants who sleep prone is increased by use of natural-fiber mattresses, swaddling, and heating in bedrooms. Several observational studies have been done since, and they have provided inconsistent results on these interventions, in particular, they disagree on the possible role of overheating.[132]

A trial of 54 children[95] showed that the steroid budenoside significantly reduced the croup score by 2 points at 4 hours, and significantly decreased readmissions by 86%. A meta-analysis (n=3736)[133] showed a significant improvement in the Westley score at 6 hours (1.2 points), and 12 hours (1.9 points), but not at 24 hours. Fewer return visits and/or (re)admissions occurred in patients treated with glucocorticoids, but the relative risk reduction was only 50% (95% CI, 24%-64%).

A trial (n=55) showed that misprostol was as effective as dinoprostone for termination of second-trimester pregnancy and was associated with fewer adverse effects than dinoprostone.[96] A subsequent trial[134] showed equal efficacy, but a higher rate of adverse effects with misoprostol (74%) than with dinoprostone (47%).

A trial (n=50) comparing botulinum toxin vs glyceryl trinitrate for chronic anal fissure concluded that both are effective alternatives to surgery but botulinum toxin is the more effective nonsurgical treatment (1 failure vs 9 failures with nitroglycerin).[109] In a meta-analysis[135] of 31 trials, botulinum toxin compared with placebo showed no significant efficacy (relative risk of failure, 0.75; 95% CI, 0.32-1.77), and was also no better than glyceryl trinitrate (relative risk of failure, 0.48; 95% CI, 0.21-1.10); surgery was more effective than medical therapy in curing fissure (relative risk of failure, 0.12; 95% CI, 0.07-0.22).

A trial of acetylcysteine (n=83) showed that it was highly effective in preventing contrast nephropathy (90% relative risk reduction).[110] There have been many more trials and many meta-analyses on this topic. The latest meta-analysis[136] shows a nonsignificant 27% relative risk reduction with acetylcysteine.

A trial of 129 stunted Jamaican children found that both nutritional supplementation and psychosocial stimulation improved the mental development of stunted children; children who got both interventions had additive benefits and achieved scores close to those of nonstunted children.[117] With long-term follow-up, however, it was found that the benefits were small and the 2 interventions no longer had additive effects.[137]

considered to be true in 2000 and that there was no difference between randomized and nonrandomized studies or high- vs low-quality studies. Allowing for somewhat different definitions, the higher rates of refutation and the generally worse performance of nonrandomized studies in the present analysis may stem from the fact that I focused on a selected sample of the most noticed and influential clinical research. For such highly cited studies, the turnaround of "truth" may be faster; in particular nonrandomized studies may be more likely to be probed and challenged than nonrandomized studies published in the general literature.

Finally, a certain proportion of highly cited trials may remain unchallenged. Sometimes the evidence from the original study may seem so overwhelming that further similar studies are deemed unethical to perform. The original study may be widely considered as a milestone for clinical practice and may provide the gold standard for testing new interventions. However, sometimes other, validating research may be in the works. Clinical research is time-consuming and challenging results may take several years to generate and publish. Therefore evidence from recent trials, no mater how impressive, should be interpreted with caution, when only one trial is available. It is important to know whether other similar or larger trials are still ongoing or being planned. Therefore, transparent and thorough trial registration is of paramount im-

portance[165] in order to limit premature claims for efficacy.

Author Contributions: Dr Ioannidis had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.
Financial Disclosures: None reported.
Acknowledgment: I thank Dr Tom Trikalinos for classifying independently the status of the highly cited articles.

REFERENCES

1. LeLorier J, Gregoire G, Benhaddad A, Lapierre J, Derderian F. Discrepancies between meta-analyses and subsequent large randomized, controlled trials. *N Engl J Med*. 1997;337:536-542.
2. Cappelleri JC, Ioannidis JP, Schmid CH, et al. Large trials vs meta-analysis of smaller trials: how do their results compare? *JAMA*. 1996;276:1332-1338.
3. Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. *N Engl J Med*. 2000;342:1878-1886.
4. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *N Engl J Med*. 2000;342:1887-1892.
5. Ioannidis JP, Haidich AB, Pappa M, et al. Comparison of evidence of treatment effects in randomized and nonrandomized studies. *JAMA*. 2001;286:821-830.
6. Institute for Scientific Information. Web of Knowledge. Available at http://www.thompson.isi.com. Accessed August 20, 2004.
7. Dickersin K, Scherer R, Suci ES, Gil-Montero M. Problems with indexing and citation of articles with group authorship. *JAMA*. 2002;287:2772-2774.
8. *Cochrane Reviewer's Handbook*. Available at: http://www.cochrane.org/resources/handbook.
9. Volberding PA, Lagakos SW, Koch MA, et al; the AIDS Clinical Trials Group of the National Institute of Allergy and Infectious Diseases. Zidovudine in asymptomatic human immunodeficiency virus infection: a controlled trial in persons with fewer than 500 CD4-positive cells per cubic millimeter. *N Engl J Med*. 1990;322:941-949.
10. Brown G, Albers JJ, Fisher LD, et al. Regression of coronary artery disease as a result of intensive lipid-lowering therapy in men with high levels of apolipoprotein B. *N Engl J Med*. 1990;323:1289-1298.
11. Moertel CG, Fleming TR, Macdonald JS, et al. Levamisole and fluorouracil for adjuvant therapy of resected colon carcinoma. *N Engl J Med*. 1990;322:352-358.
12. Cohn JN, Johnson G, Ziesche S, et al. A comparison of enalapril with hydralazine-isosorbide dinitrate in the treatment of chronic congestive heart failure. *N Engl J Med*. 1991;325:303-310.
13. Stampfer MJ, Colditz GA, Willett WC, et al. Postmenopausal estrogen therapy and cardiovascular disease: ten-year follow-up from the Nurses' Health Study. *N Engl J Med*. 1991;325:756-762.
14. North American Symptomatic Carotid Endarterectomy Trial Collaborators. Beneficial effect of carotid endarterectomy in symptomatic patients with high-grade carotid stenosis. *N Engl J Med*. 1991;325:445-453.
15. Ziegler EJ, Fisher CJ Jr, Sprung CL, et al; the HA-1A Sepsis Study Group. Treatment of gram-negative bacteremia and septic shock with HA-1A human monoclonal antibody against endotoxin: a randomized, double-blind, placebo-controlled trial. *N Engl J Med*. 1991;324:429-436.
16. The SOLVD Investigators. Effect of enalapril on survival in patients with reduced left ventricular ejection fractions and congestive heart failure. *N Engl J Med*. 1991;325:293-302.
17. Pfeffer MA, Braunwald E, Moye LA, et al; the SAVE Investigators. Effect of captopril on mortality and morbidity in patients with left ventricular dysfunction after myocardial infarction: results of the survival and ventricular enlargement trial. *N Engl J Med*. 1992;327:669-677.
18. Grines CL, Browne KF, Marco J, et al; the Primary Angioplasty in Myocardial Infarction Study Group. A comparison of immediate angioplasty with thrombolytic therapy for acute myocardial infarction. *N Engl J Med*. 1993;328:673-679.
19. Lewis EJ, Hunsicker LG, Bain RP, Rohde RD; the Collaborative Study Group. The effect of angiotensin-converting-enzyme inhibition on diabetic nephropathy. *N Engl J Med*. 1993;329:1456-1462.
20. Rimm EB, Stampfer MJ, Ascherio A, Giovannucci E, Colditz GA, Willett WC. Vitamin E consumption and the risk of coronary heart disease in men. *N Engl J Med*. 1993;328:1450-1456.
21. Stampfer MJ, Hennekens CH, Manson JE, Colditz GA, Rosner B, Willett WC. Vitamin E consumption and the risk of coronary disease in women. *N Engl J Med*. 1993;328:1444-1449.
22. Rossaint R, Falke KJ, Lopez F, Slama K, Pison U, Zapol WM. Inhaled nitric oxide for the adult respiratory distress syndrome. *N Engl J Med*. 1993;328:399-405.
23. The Diabetes Control and Complications Trial Research Group. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. *N Engl J Med*. 1993;329:977-986.
24. The EPIC Investigation. Use of a monoclonal antibody directed against the platelet glycoprotein IIb/IIIa receptor in high-risk coronary angioplasty. *N Engl J Med*. 1994;330:956-961.
25. Connor EM, Sperling RS, Gelber R, et al; Pediatric AIDS Clinical Trials Group Protocol 076 Study Group. Reduction of maternal-infant transmission of human immunodeficiency virus type 1 with zidovudine treatment. *N Engl J Med*. 1994;331:1173-1180.
26. Fischman DL, Leon MB, Baim DS, et al; Stent Restenosis Study Investigators. A randomized comparison of coronary-stent placement and balloon angioplasty in the treatment of coronary artery disease. *N Engl J Med*. 1994;331:496-501.
27. Serruys PW, de Jaegere P, Kiemeneij F, et al; BENESTENT Study Group. A comparison of balloon-expandable-stent implantation with balloon angioplasty in patients with coronary artery disease. *N Engl J Med*. 1994;331:489-495.
28. The Alpha-Tocopherol, Beta Carotene Cancer Prevention Study Group. The effect of vitamin E and beta carotene on the incidence of lung cancer and other cancers in male smokers. *N Engl J Med*. 1994;330:1029-1035.
29. The National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group. Tissue plasminogen activator for acute ischemic stroke. *N Engl J Med*. 1995;333:1581-1587.
30. Shepherd J, Cobbe SM, Ford I, et al; West of Scotland Coronary Prevention Study Group. Prevention of coronary heart disease with pravastatin in men with hypercholesterolemia. *N Engl J Med*. 1995;333:1301-1307.
31. Sacks FM, Pfeffer MA, Moye LA, et al; Cholesterol and Recurrent Events Trial Investigators. The effect of pravastatin on coronary events after myocardial infarction in patients with average cholesterol levels. *N Engl J Med*. 1996;335:1001-1009.
32. Packer M, Bristow MR, Cohn JN, et al; US Carvedilol Heart Failure Study Group. The effect of carvedilol on morbidity and mortality in patients with chronic heart failure. *N Engl J Med*. 1996;334:1349-1355.
33. Omenn GS, Goodman GE, Thornquist MD, et al. Effects of a combination of beta carotene and vitamin A on lung cancer and cardiovascular disease. *N Engl J Med*. 1996;334:1150-1155.
34. Ridker PM, Cushman M, Stampfer MJ, Tracy RP, Hennekens CH. Inflammation, aspirin, and the risk of cardiovascular disease in apparently healthy men. *N Engl J Med*. 1997;336:973-979.
35. Hammer SM, Squires KE, Hughes MD, et al; AIDS Clinical Trials Group 320 Study Team. A controlled trial of two nucleoside analogues plus indinavir in persons with human immunodeficiency virus infection and CD4 cell counts of 200 per cubic millimeter or less. *N Engl J Med*. 1997;337:725-733.
36. The EPILOG Investigators. Platelet glycoprotein IIb/IIIa receptor blockade and low-dose heparin during percutaneous coronary revascularization. *N Engl J Med*. 1997;336:1689-1696.
37. McHutchison JG, Gordon SC, Schiff ER, et al; Hepatitis Interventional Therapy Group. Interferon alfa-2b alone or in combination with ribavirin as initial treatment for chronic hepatitis C. *N Engl J Med*. 1998;339:1485-1492.
38. The Long-Term Intervention with Pravastatin in Ischaemic Disease (LIPID) Study Group. Prevention of cardiovascular events and death with pravastatin in patients with coronary heart disease and a broad range of initial cholesterol levels. *N Engl J Med*. 1998;339:1349-1357.
39. Pitt B, Zannad F, Remme WJ, et al; Randomized Aldactone Evaluation Study Investigators. The effect of spironolactone on morbidity and mortality in patients with severe heart failure. *N Engl J Med*. 1999;341:709-717.
40. Yusuf S, Sleight P, Pogue J, Bosch J, Davies R, Dagenais G; the Heart Outcomes Prevention Evaluation Study Investigators. Effects of an angiotensin-converting-enzyme inhibitor, ramipril, on cardiovascular events in high-risk patients. *N Engl J Med*. 2000;342:145-153.
41. SHEP Cooperative Research Group. Prevention of stroke by antihypertensive drug treatment in older persons with isolated systolic hypertension: final results of the Systolic Hypertension in the Elderly Program (SHEP). *JAMA*. 1991;265:3255-3264.
42. The Writing Group for the PEPI Trial. Effects of estrogen or estrogen/progestin regimens on heart disease risk factors in postmenopausal women. *JAMA*. 1995;273:199-208.
43. Executive Committee for the Asymptomatic Carotid Atherosclerosis Study. Endarterectomy for asymptomatic carotid artery stenosis. *JAMA*. 1995;273:1421-1428.
44. Hulley S, Grady D, Bush T, et al; Heart and Estrogen/progestin Replacement Study (HERS) Research Group. Randomized trial of estrogen plus progestin for secondary prevention of coronary heart disease in postmenopausal women. *JAMA*. 1998;280:605-613.
45. Downs JR, Clearfield M, Weis S, et al. Primary prevention of acute coronary events with lovastatin in men and women with average cholesterol levels: results of AFCAPS/TexCAPS. *JAMA*. 1998;279:1615-1622.
46. Rossouw JE, Anderson GL, Prentice RL, et al. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomized controlled trial. *JAMA*. 2002;288:321-333.
47. MRC Vitamin Study Research Group. Prevention of neural tube defects: results of the Medical Research Council Vitamin Study. *Lancet*. 1991;338:131-137.
48. Hertog MG, Feskens EJ, Hollman PC, Katan MB, Kromhout D. Dietary antioxidant flavonoids and risk of coronary heart disease: the Zutphen Elderly Study. *Lancet*. 1993;342:1007-1011.
49. Randomised trial of cholesterol lowering in 4444 patients with coronary heart disease: the Scandinavian Simvastatin Survival Study (4S). *Lancet*. 1994;344:1383-1389.
50. CAPRIE Steering Committee. A randomised, blinded, trial of clopidogrel vs aspirin in patients at risk of ischaemic events (CAPRIE). *Lancet*. 1996;348:1329-1339.
51. Stephens NG, Parsons A, Schofield PM, Kelly F, Cheeseman K, Mitchinson MJ. Randomised controlled trial of vitamin E in patients with coronary disease: Cambridge Heart Antioxidant Study (CHAOS). *Lancet*. 1996;347:781-786.

52. Hansson L, Zanchetti A, Carruthers SG, et al; HOT Study Group. Effects of intensive blood-pressure lowering and low-dose aspirin in patients with hypertension: principal results of the Hypertension Optimal Treatment (HOT) randomised trial. *Lancet*. 1998;351:1755-1762.
53. Poynard T, Marcellin P, Lee SS, et al; International Hepatitis Interventional Therapy Group (IHIT). Randomised trial of interferon alpha-2b plus ribavirin for 48 weeks or for 24 weeks vs interferon alpha-2b plus placebo for 48 weeks for treatment of chronic infection with hepatitis C virus. *Lancet*. 1998;352:1426-1432.
54. UK Prospective Diabetes Study (UKPDS) Group. Effect of intensive blood-glucose control with metformin on complications in overweight patients with type 2 diabetes (UKPDS 34). *Lancet*. 1998;352:854-865.
55. The Cardiac Insufficiency Bisoprolol Study II (CIBIS-II): a randomised trial. *Lancet*. 1999;353:9-13.
56. Castaigne S, Cjomienne C, Daniel MT, et al. All-trans-retinoic acid as a differentiation therapy for acute promyelocytic leukemia, 1: clinical results. *Blood*. 1990;76:1704-1709.
57. Fisher B, Costantino JP, Wickerham DL, et al. Tamoxifen for prevention of breast cancer: report of the National Surgical Adjuvant Breast and Bowel Project P-1 Study. *J Natl Cancer Inst*. 1998;90:1371-1388.
58. Ioannidis JP, Cappelleri JC, Lau J, et al. Early or deferred zidovudine therapy in HIV-infected patients without an AIDS-defining illness. *Ann Intern Med*. 1995;122:856-866.
59. Law MR, Wald NJ, Rudnicka AR. Quantifying effect of statins on low density lipoprotein cholesterol, ischaemic heart disease, and stroke: systematic review and meta-analysis. *BMJ*. 2003;326:1423.
60. Gill S, Loprinzi CL, Sargent DJ, et al. Pooled analysis of fluorouracil-based adjuvant therapy for stage II and III colon cancer: who benefits and by how much? *J Clin Oncol*. 2004;22:1797-1806.
61. Rothwell PM, Eliasziw M, Gutnikov SA, et al. Analysis of pooled data from the randomised controlled trials of endarterectomy for symptomatic carotid stenosis. *Lancet*. 2003;361:107-116.
62. McCloskey RV, Straube RC, Sanders C, Smith SM, Smith CR; CHESS Trial Study Group. Treatment of septic shock with human monoclonal antibody HA-1A: a randomized, double-blind, placebo-controlled trial. *Ann Intern Med*. 1994;121:1-5.
63. Garg R, Yusuf S; Collaborative Group on ACE Inhibitor Trials. Overview of randomized trials of angiotensin-converting enzyme inhibitors on mortality and morbidity in patients with heart failure. *JAMA*. 1995;273:1450-1456.
64. Rodrigues EJ, Eisenberg MJ, Pilote L. Effects of early and late administration of angiotensin-converting enzyme inhibitors on mortality after myocardial infarction. *Am J Med*. 2003;115:473-479.
65. Cucherat M, Bonnefoy E, Tremeau G. Primary angioplasty vs intravenous thrombolysis for acute myocardial infarction. *Cochrane Database Syst Rev*. 2003;3:CD001560.
66. Yusuf S, Dagenais G, Pogue J, Bosch J, Sleight P; The Heart Outcomes Prevention Evaluation Study Investigators. Vitamin E supplementation and cardiovascular events in high-risk patients. *N Engl J Med*. 2000;342:154-160.
67. Sokol J, Jacobs SE, Bohn D. Inhaled nitric oxide for acute hypoxemic respiratory failure in children and adults. *Cochrane Database Syst Rev*. 2003;1:CD002787.
68. Karvouni E, Katritsis DG, Ioannidis JP. Intravenous glycoprotein IIb/IIIa receptor antagonists reduce mortality after percutaneous coronary interventions. *J Am Coll Cardiol*. 2003;41:26-32.
69. Brophy JM, Belisle P, Joseph L. Evidence for use of coronary stents: a hierarchical bayesian meta-analysis. *Ann Intern Med*. 2003;138:777-786.
70. Hacke W, Donnan G, Fieschi C, et al. Association of outcome with early stroke treatment: pooled analysis of ATLANTIS, ECASS, and NINDS rt-PA stroke trials. *Lancet*. 2004;363:768-774.
71. Brophy JM, Joseph L, Rouleau JL. Beta-blockers in congestive heart failure: a Bayesian meta-analysis. *Ann Intern Med*. 2001;134:550-560.
72. Yazdanpanah Y, Sissoko D, Egger M, Mouton Y, Zwahlen M, Chene G. Clinical efficacy of antiretroviral combination therapy based on protease inhibitors or non-nucleoside analogue reverse transcriptase inhibitors: indirect comparison of controlled trials. *BMJ*. 2004;328:249.
73. Kjaergard LL, Krogsgaard K, Gluud C. Interferon alfa with or without ribavirin for chronic hepatitis C: systematic review of randomised trials. *BMJ*. 2001;323:1151-1155.
74. Staessen JA, Gasowski J, Wang JG, et al. Risks of untreated and treated isolated systolic hypertension in the elderly: meta-analysis of outcome trials. *Lancet*. 2000;355:865-872.
75. Benavente O, Moher D, Pham B. Carotid endarterectomy for asymptomatic carotid stenosis: a meta-analysis. *BMJ*. 1998;317:1477-1480.
76. Huxley RR, Neil HA. The relation between dietary flavonol intake and coronary heart disease mortality: a meta-analysis of prospective cohort studies. *Eur J Clin Nutr*. 2003;57:904-908.
77. Tallman MS, Andersen JW, Schiffer CA, et al. All-trans-retinoic acid in acute promyelocytic leukemia. *N Engl J Med*. 1997;337:1021-1028.
78. Cuzick J, Powles T, Veronesi U, et al. Overview of the main outcomes in breast-cancer prevention trials. *Lancet*. 2003;361:296-300.
79. Watts NB, Harris ST, Genant HK, et al. Intermittent cyclical etidronate treatment of postmenopausal osteoporosis. *N Engl J Med*. 1990;323:73-79.
80. Smith P, Arnesen H, Holme I. The effect of warfarin on mortality and reinfarction after myocardial infarction. *N Engl J Med*. 1990;323:147-152.
81. Hussey GD, Klein M. A randomized, controlled trial of vitamin A in children with severe measles. *N Engl J Med*. 1990;323:160-164.
82. Smith DW, Frankel LR, Mathers LH, Tang AT, Ariagno RL, Prober CG. A controlled trial of aerosolized ribavirin in infants receiving mechanical ventilation for severe respiratory syncytial virus infection. *N Engl J Med*. 1991;325:24-29.
83. The National Institute of Child Health and Human Developments Intravenous Immunoglobulin Study Group. Intravenous immune globulin for the prevention of bacterial infections in children with symptomatic human immunodeficiency virus infection. *N Engl J Med*. 1991;325:73-80.
84. Green DM, Zevon MA, Lowrie G, Seigelstein N, Hall B. Congenital anomalies in children of patients who received chemotherapy for cancer in childhood and adolescence. *N Engl J Med*. 1991;325:141-146.
85. Bradley JD, Brandt KD, Katz BP, Kalasinski LA, Ryan SI. Comparison of an antiinflammatory dose of ibuprofen, an analgesic dose of ibuprofen, and acetaminophen in the treatment of patients with osteoarthritis of the knee. *N Engl J Med*. 1991;325:87-91.
86. Crawford J, Ozer H, Stoller R, et al. Reduction by granulocyte colony-stimulating factor of fever and neutropenia induced by chemotherapy in patients with small-cell lung cancer. *N Engl J Med*. 1991;325:164-170.
87. Maggioni AP, Franzosi MG, Santoro E, White H, Van de Werf F, Tognoni G; Gruppo Italiano per lo Studio della Sopravvivenza nell'Infarto Miocardico II (GISSI-2) and The International Study Group. The risk of stroke in patients with acute myocardial infarction after thrombolytic and antithrombotic treatment. *N Engl J Med*. 1992;327:1-6.
88. Packer M, Gheorghiade M, Young JB, et al; RADIANCE Study. Withdrawal of digoxin from patients with chronic heart failure treated with angiotensin-converting enzyme inhibitors. *N Engl J Med*. 1993;329:1-7.
89. Jenkins DJ, Wolever TM, Rao AV, et al. Effect on blood lipids of very high intakes of fiber in diets low in saturated fat and cholesterol. *N Engl J Med*. 1993;329:21-26.
90. Rosman NP, Colton T, Labazzo J, et al. A controlled trial of diazepam administered during febrile illnesses to prevent recurrence of febrile seizures. *N Engl J Med*. 1993;329:79-84.
91. Hunter DJ, Manson JE, Colditz GA, et al. A prospective study of the intake of vitamins C, E, and A and the risk of breast cancer. *N Engl J Med*. 1993;329:234-240.
92. Ponsonby AL, Dwyer T, Gibbons LE, Cochrane JA, Wang YG. Factors potentiating the risk of sudden infant death syndrome associated with the prone position. *N Engl J Med*. 1993;329:377-382.
93. Winterkorn JM, Kupersmith MJ, Wirtschafter JD, Forman S. Treatment of vasospastic amaurosis fugax with calcium-channel blockers. *N Engl J Med*. 1993;329:396-398.
94. Greenberg ER, Baron JA, Tosteson TD, et al; Polyp Prevention Study Group. A clinical trial of antioxidant vitamins to prevent colorectal adenoma. *N Engl J Med*. 1994;331:141-147.
95. Klassen TP, Feldman ME, Watters LK, Sutcliffe T, Rowe PC. Nebulized budesonide for children with mild-to-moderate croup. *N Engl J Med*. 1994;331:285-289.
96. Jain JK, Mishell DR Jr. A comparison of intravaginal misoprostol with prostaglandin E2 for termination of second-trimester pregnancy. *N Engl J Med*. 1994;331:290-293.
97. Loprinzi CL, Michalak JC, Quella SK, et al. Megestrol acetate for the prevention of hot flashes. *N Engl J Med*. 1994;331:347-352.
98. O'Connell MJ, Martenson JA, Wieand HS, et al. Improving adjuvant therapy for rectal cancer by combining protracted-infusion fluorouracil with radiation therapy after curative surgery. *N Engl J Med*. 1994;331:502-507.
99. Suarez FL, Savaiano DA, Levitt MD. A comparison of symptoms after the consumption of milk or lactose-hydrolyzed milk by people with self-reported severe lactose intolerance. *N Engl J Med*. 1995;333:1-4.
100. Singh SN, Fletcher RD, Fisher SG, et al; Survival Trial of Antiarrhythmic Therapy in Congestive Heart Failure. Amiodarone in patients with congestive heart failure and asymptomatic ventricular arrhythmia. *N Engl J Med*. 1995;333:77-82.
101. Bhasin S, Storer TW, Berman N, et al. The effects of supraphysiologic doses of testosterone on muscle size and strength in normal men. *N Engl J Med*. 1996;335:1-7.
102. Stegeman CA, Tervaert JW, de Jong PE, Kallenberg CG; Dutch Co-Trimoxazole Wegener Study Group. Trimethoprim-sulfamethoxazole (co-trimoxazole) for the prevention of relapses of Wegener's granulomatosis. *N Engl J Med*. 1996;335:16-20.
103. van Hensbroek MB, Onyiorah E, Jaffar S, et al. A trial of artemether or quinine in children with cerebral malaria. *N Engl J Med*. 1996;335:69-75.
104. van der Horst CM, Saag MS, Cloud GA, et al; National Institute of Allergy and Infectious Diseases Mycoses Study Group and AIDS Clinical Trials Group. Treatment of cryptococcal meningitis associated with the acquired immunodeficiency syndrome. *N Engl J Med*. 1997;337:15-21.
105. Levine RJ, Hauth JC, Curet LB, et al. Trial of calcium to prevent preeclampsia. *N Engl J Med*. 1997;337:69-76.
106. Musch DC, Martin DF, Gordon JF, Davis MD, Kuppermann BD; the Ganciclovir Implant Study Group. Treatment of cytomegalovirus retinitis with a sustained-release ganciclovir implant. *N Engl J Med*. 1997;337:83-90.
107. Adams RJ, McKie VC, Hsu L, et al. Prevention of a first stroke by transfusions in children with sickle cell anemia and abnormal results on transcranial Doppler ultrasonography. *N Engl J Med*. 1998;339:5-11.
108. Stefanick ML, Mackey S, Sheehan M, Ellsworth N, Haskell WL, Wood PD. Effects of diet and exercise in men and postmenopausal women with low levels of

HDL cholesterol and high levels of LDL cholesterol. *N Engl J Med*. 1998;339:12-20.
109. Brisinda G, Maria G, Bentivoglio AR, Cassetta E, Gui D, Albanese A. A comparison of injections of botulinum toxin and topical nitroglycerin ointment for the treatment of chronic anal fissure. *N Engl J Med*. 1999; 341:65-69.
110. Tepel M, van der Giet M, Schwarzfeld C, Laufer U, Liermann D, Zidek W. Prevention of radiographic-contrast-agent-induced reductions in renal function by acetylcysteine. *N Engl J Med*. 2000;343:180-184.
111. Levine AM, Wernz JC, Kaplan L, et al. Low-dose chemotherapy with central nervous system prophylaxis and zidovudine maintenance in AIDS-related lymphoma: a prospective multi-institutional trial. *JAMA*. 1991;266:84-88.
112. Freeman EW, Rickels K, Sondheimer SJ, Polansky M. A double-blind trial of oral progesterone, alprazolam, and placebo in treatment of severe premenstrual syndrome. *JAMA*. 1995;274:51-57.
113. Williams-Russo P, Sharrock NE, Mattis S, Szatrowski TP, Charlson ME. Cognitive effects after epidural vs general anesthesia in older adults: a randomized trial. *JAMA*. 1995;274:44-50.
114. Meduri GU, Headley AS, Golden E, et al. Effect of prolonged methylprednisolone therapy in unresolving acute respiratory distress syndrome: a randomized controlled trial. *JAMA*. 1998;280:159-165.
115. Gulick RM, Mellors JW, Havlir D, et al. Simultaneous vs sequential initiation of therapy with indinavir, zidovudine, and lamivudine for HIV-1 infection: 100-week follow-up. *JAMA*. 1998;280:35-41.
116. Grady D, Herrington D, Bittner V, et al. Cardiovascular disease outcomes during 6.8 years of hormone therapy: Heart and Estrogen/progestin Replacement Study follow-up (HERS II). *JAMA*. 2002;288: 49-57.
117. Grantham-McGregor SM, Powell CA, Walker SP, Himes JH. Nutritional supplementation, psychosocial stimulation, and mental development of stunted children: the Jamaican Study. *Lancet*. 1991;338:1-5.
118. Ghana VAST Study Team. Vitamin A supplementation in northern Ghana: effects on clinic attendances, hospital admissions, and child mortality. *Lancet*. 1993;342:7-12.
119. Sipe JC, Romine JS, Koziol JA, McMillan R, Zyroff J, Beutler E. Cladribine in treatment of chronic progressive multiple sclerosis. *Lancet*. 1994;344:9-13.
120. Karonga Prevention Trial Group. Randomised controlled trial of single BCG, repeated BCG, or combined BCG and killed *Mycobacterium leprae* vaccine for prevention of leprosy and tuberculosis in Malawi. *Lancet*. 1996;348:17-24.
121. Waldo AL, Camm AJ, deRuyter H, et al; the SWORD Investigators. Effect of d-sotalol on mortality in patients with left ventricular dysfunction after recent and remote myocardial infarction. *Lancet*. 1996;348:7-12.
122. Veronesi U, Maisonneuve P, Costa A, et al; Italian Tamoxifen Prevention Study. Prevention of breast cancer with tamoxifen: preliminary findings from the Italian randomised trial among hysterectomised women. *Lancet*. 1998;352:93-97.
123. The EPISTENT Investigators. Randomised placebo-controlled and balloon-angioplasty-controlled trial to assess safety of coronary stenting with use of platelet glycoprotein-IIb/IIIa blockade. *Lancet*. 1998;352:87-92.
124. CLIP Group (Cancer of the Liver Italian Programme). Tamoxifen in treatment of hepatocellular carcinoma: a randomised controlled trial. *Lancet*. 1998;352:17-20.
125. Torp-Pedersen C, Kober L; TRACE Study Group. Effect of ACE inhibitor trandolapril on life expectancy of patients with reduced left-ventricular function after acute myocardial infarction. *Lancet*. 1999;354:9-12.
126. Negrin RS, Haeuber DH, Nagler A, et al. Maintenance treatment of patients with myelodysplastic syndromes using recombinant human granulocyte colony-stimulating factor. *Blood*. 1990;76:36-43.
127. Budman DR, Berry DA, Cirrincione CT, et al; the Cancer and Leukemia Group B. Dose and dose intensity as determinants of outcome in the adjuvant treatment of breast cancer. *J Natl Cancer Inst*. 1998;90:1205-1211.
128. Veronesi U, De Palo G, Marubini E, et al. Randomized trial of fenretinide to prevent second breast malignancy in women with early breast cancer. *J Natl Cancer Inst*. 1999;91:1847-1856.
129. Rice GP, Filippi M, Comi G; Cladribine MRI Study Group. Cladribine and progressive MS: clinical and MRI outcomes of a multicenter controlled trial. *Neurology*. 2000;54:1145-1155.
130. Ventre K, Randolph A. Ribavirin for respiratory syncytial virus infection of the lower respiratory tract in infants and young children. *Cochrane Database Syst Rev*. 2004;4:CD000181.
131. Rantala H, Tarkka R, Uhari M. A meta-analytic review of the preventive treatment of recurrences of febrile seizures. *J Pediatr*. 1997;131:922-925.
132. Creery D, Mikrogianakis A. Sudden infant death syndrome. *Clin Evid*. 2003;10:457-467.
133. Russell K, Wiebe N, Saenz A, et al. Glucocorticoids for croup. *Cochrane Database Syst Rev*. 2004;1: CD001955.
134. Makhlouf AM, Al-Hussaini TK, Habib DM, Makarem MH. Second-trimester pregnancy termination: comparison of three different methods. *J Obstet Gynaecol*. 2003;23:407-411.
135. Nelson R. A systematic review of medical therapy for anal fissure. *Dis Colon Rectum*. 2004;47:422-431.
136. Nallamothu BK, Shojania KG, Saint S, et al. Is acetylcysteine effective in preventing contrast-related nephropathy? a meta-analysis. *Am J Med*. 2004;117:938-947.
137. Grantham-McGregor SM, Walker SP, Chang SM, Powell CA. Effects of early childhood supplementation with and without stimulation on later development in stunted Jamaican children. *Am J Clin Nutr*. 1997; 66:247-253.
138. Adachi JD, Roux C, Pitt PI, et al. A pooled data analysis on the use of intermittent cyclical etidronate therapy for the prevention and treatment of corticosteroid induced bone loss. *J Rheumatol*. 2000;27:2424-2431.
139. Anand SS, Yusuf S. Oral anticoagulant therapy in patients with coronary artery disease: a meta-analysis. *JAMA*. 1999;282:2058-2067.
140. Fawzi WW, Chalmers TC, Herrera MG, Mosteller F. Vitamin A supplementation and child mortality: a meta-analysis. *JAMA*. 1993;269:898-903.
141. Spector SA, Gelber RD, McGrath N, et al; Pediatric AIDS Clinical Trials Group. A controlled trial of intravenous immune globulin for the prevention of serious bacterial infections in children receiving zidovudine for advanced human immunodeficiency virus infection. *N Engl J Med*. 1994;331:1181-1187.
142. Lyman GH, Kuderer NM, Djulbegovic B. Prophylactic granulocyte colony-stimulating factor in patients receiving dose-intensive cancer chemotherapy: a meta-analysis. *Am J Med*. 2002;112:406-411.
143. Dundar Y, Hill R, Dickson R, Walley T. Comparative efficacy of thrombolytics in acute myocardial infarction: a systematic review. *QJM*. 2003;96:103-113.
144. Hood WB Jr, Dans AL, Guyatt GH, Jaeschke R, McMurray JJ. Digitalis for treatment of congestive heart failure in patients in sinus rhythm. *Cochrane Database Syst Rev*. 2004;2:CD002901.
145. Anderson JW, Randles KM, Kendall CW, Jenkins DJ. Carbohydrate and fiber recommendations for individuals with diabetes: a quantitative assessment and meta-analysis of the evidence. *J Am Coll Nutr*. 2004;23: 5-17.
146. Giorgi A, Weatherby RP, Murphy PW. Muscular strength, body composition and health responses to the use of testosterone enanthate: a double blind study. *J Sci Med Sport*. 1999;2:341-355.
147. Artemether-Quinine Meta-analysis Study Group. A meta-analysis using individual patient data of trials comparing artemether with quinine in the treatment of severe falciparum malaria. *Trans R Soc Trop Med Hyg*. 2001;95:637-650.
148. Saag MS, Cloud GA, Graybill JR, et al; National Institute of Allergy and Infectious Diseases Mycoses Study Group. A comparison of itraconazole vs fluconazole as maintenance therapy for AIDS-associated cryptococcal meningitis. *Clin Infect Dis*. 1999;28:291-296.
149. Martin DF, Kuppermann BD, Wolitz RA, Palestine AG, Li H, Robinson CA; Roche Ganciclovir Study Group. Oral ganciclovir for patients with cytomegalovirus retinitis treated with a ganciclovir implant. *N Engl J Med*. 1999;340:1063-1070.
150. Nieman DC, Brock DW, Butterworth D, Utter AC, Nieman CC. Reducing diet and/or exercise training decreases the lipid and lipoprotein risk factors of moderately obese women. *J Am Coll Nutr*. 2002;21:344-350.
151. Kaplan LD, Straus DJ, Testa MA, et al; National Institute of Allergy and Infectious Diseases AIDS Clinical Trials Group. Low-dose compared with standard-dose m-BACOD chemotherapy for non-Hodgkin's lymphoma associated with human immunodeficiency virus infection. *N Engl J Med*. 1997;336:1641-1648.
152. Diegoli MS, da Fonseca AM, Diegoli CA, Pinotti JA. A double-blind trial of four medications to treat severe premenstrual syndrome. *Int J Gynaecol Obstet*. 1998;62:63-67.
153. Flather MD, Yusuf S, Kober L, et al; ACE-Inhibitor Myocardial Infarction Collaborative Group. Long-term ACE-inhibitor therapy in patients with heart failure or left-ventricular dysfunction: a systematic overview of data from individual patients. *Lancet*. 2000;355: 1575-1581.
154. Ossenkoppele GJ, van der Holt B, Verhoef GE, et al; Dutch-Belgian Hemato-Oncology Cooperative Group. A randomized study of granulocyte colony-stimulating factor applied during and after chemotherapy in patients with poor risk myelodysplastic syndromes: a report from the HOVON Cooperative Group. *Leukemia*. 1999;13:1207-1213.
155. Citron ML, Berry DA, Cirrincione C, et al. Randomized trial of dose-dense vs conventionally scheduled and sequential vs concurrent combination chemotherapy as postoperative adjuvant treatment of node-positive primary breast cancer: first report of Intergroup Trial C9741/Cancer and Leukemia Group B Trial 9741. *J Clin Oncol*. 2003;21:1431-1439.
156. Lau J, Ioannidis JP, Schmid CH. Summing up evidence: one answer is not enough. *Lancet*. 1998;351: 123-127.
157. Ioannidis JP, Cappelleri JC, Lau J. Issues in comparisons between meta-analyses and large trials. *JAMA*. 1998;279:1089-1093.
158. Fleming TR, DeMets DL. Surrogate end points in clinical trials: are we being misled? *Ann Intern Med*. 1996;125:605-613.
159. Dickersin K, Min YI. Publication bias: the problem that won't go away. *Ann N Y Acad Sci*. 1993;703: 135-146.
160. Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. *Lancet*. 1991; 337:867-872.
161. Stern JM, Simes RJ. Publication bias: evidence of delayed publication in a cohort study of clinical research projects. *BMJ*. 1997;315:640-645.
162. Ioannidis JP. Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials. *JAMA*. 1998;279:281-286.
163. Wheatley K, Clayton D. Be skeptical about unexpected large treatment effects: the case of an MRC AML12 randomization. *Control Clin Trials*. 2003;24:66-70.
164. Poynard T, Munteanu M, Ratziu V, et al. Truth survival in clinical research: an evidence-based requiem? *Ann Intern Med*. 2002;136:888-895.
165. DeAngelis C, Drazen JM, Frizelle FA, et al. Clinical trial registration: a statement from the International Committee of Medical Journal Editors. *N Engl J Med*. 2004;351:1250-1251.

Essay

Why Most Published Research Findings Are False

John P. A. Ioannidis

Summary

There is increasing concern that most current published research findings are false. The probability that a research claim is true may depend on study power and bias, the number of other studies on the same question, and, importantly, the ratio of true to no relationships among the relationships probed in each scientific field. In this framework, a research finding is less likely to be true when the studies conducted in a field are smaller; when effect sizes are smaller; when there is a greater number and lesser preselection of tested relationships; where there is greater flexibility in designs, definitions, outcomes, and analytical modes; when there is greater financial and other interest and prejudice; and when more teams are involved in a scientific field in chase of statistical significance. Simulations show that for most study designs and settings, it is more likely for a research claim to be false than true. Moreover, for many current scientific fields, claimed research findings may often be simply accurate measures of the prevailing bias. In this essay, I discuss the implications of these problems for the conduct and interpretation of research.

Published research findings are sometimes refuted by subsequent evidence, with ensuing confusion and disappointment. Refutation and controversy is seen across the range of research designs, from clinical trials and traditional epidemiological studies [1–3] to the most modern molecular research [4,5]. There is increasing concern that in modern research, false findings may be the majority or even the vast majority of published research claims [6–8]. However, this should not be surprising. It can be proven that most claimed research findings are false. Here I will examine the key factors that influence this problem and some corollaries thereof.

The Essay section contains opinion pieces on topics of broad interest to a general medical audience.

Modeling the Framework for False Positive Findings

Several methodologists have pointed out [9–11] that the high rate of nonreplication (lack of confirmation) of research discoveries is a consequence of the convenient, yet ill-founded strategy of claiming conclusive research findings solely on the basis of a single study assessed by formal statistical significance, typically for a *p*-value less than 0.05. Research is not most appropriately represented and summarized by *p*-values, but, unfortunately, there is a widespread notion that medical research articles

It can be proven that most claimed research findings are false.

should be interpreted based only on *p*-values. Research findings are defined here as any relationship reaching formal statistical significance, e.g., effective interventions, informative predictors, risk factors, or associations. "Negative" research is also very useful. "Negative" is actually a misnomer, and the misinterpretation is widespread. However, here we will target relationships that investigators claim exist, rather than null findings.

As has been shown previously, the probability that a research finding is indeed true depends on the prior probability of it being true (before doing the study), the statistical power of the study, and the level of statistical significance [10,11]. Consider a 2 × 2 table in which research findings are compared against the gold standard of true relationships in a scientific field. In a research field both true and false hypotheses can be made about the presence of relationships. Let R be the ratio of the number of "true relationships" to "no relationships" among those tested in the field. R is characteristic of the field and can vary a lot depending on whether the field targets highly likely relationships or searches for only one or a few true relationships among thousands and millions of hypotheses that may be postulated. Let us also consider, for computational simplicity, circumscribed fields where either there is only one true relationship (among many that can be hypothesized) or the power is similar to find any of the several existing true relationships. The pre-study probability of a relationship being true is $R/(R + 1)$. The probability of a study finding a true relationship reflects the power $1 - \beta$ (one minus the Type II error rate). The probability of claiming a relationship when none truly exists reflects the Type I error rate, α. Assuming that c relationships are being probed in the field, the expected values of the 2 × 2 table are given in Table 1. After a research finding has been claimed based on achieving formal statistical significance, the post-study probability that it is true is the positive predictive value, PPV. The PPV is also the complementary probability of what Wacholder et al. have called the false positive report probability [10]. According to the 2 × 2 table, one gets $PPV = (1 - \beta)R/(R - \beta R + \alpha)$. A research finding is thus

Citation: Ioannidis JPA (2005) Why most published research findings are false. PLoS Med 2(8): e124.

Abbreviation: PPV, positive predictive value

John P. A. Ioannidis is in the Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece, and Institute for Clinical Research and Health Policy Studies, Department of Medicine, Tufts-New England Medical Center, Tufts University School of Medicine, Boston, Massachusetts, United States of America. E-mail: jioannid@cc.uoi.gr

Competing Interests: The author has declared that no competing interests exist.

DOI: 10.1371/journal.pmed.0020124

Table 1. Research Findings and True Relationships

Research Finding	True Relationship		
	Yes	No	Total
Yes	$c(1-\beta)R/(R+1)$	$c\alpha/(R+1)$	$c(R+\alpha-\beta R)/(R+1)$
No	$c\beta R/(R+1)$	$c(1-\alpha)/(R+1)$	$c(1-\alpha+\beta R)/(R+1)$
Total	$cR/(R+1)$	$c/(R+1)$	c

DOI: 10.1371/journal.pmed.0020124.t001

more likely true than false if $(1-\beta)R > \alpha$. Since usually the vast majority of investigators depend on $\alpha = 0.05$, this means that a research finding is more likely true than false if $(1-\beta)R > 0.05$.

What is less well appreciated is that bias and the extent of repeated independent testing by different teams of investigators around the globe may further distort this picture and may lead to even smaller probabilities of the research findings being indeed true. We will try to model these two factors in the context of similar 2 × 2 tables.

Bias

First, let us define bias as the combination of various design, data, analysis, and presentation factors that tend to produce research findings when they should not be produced. Let u be the proportion of probed analyses that would not have been "research findings," but nevertheless end up presented and reported as such, because of bias. Bias should not be confused with chance variability that causes some findings to be false by chance even though the study design, data, analysis, and presentation are perfect. Bias can entail manipulation in the analysis or reporting of findings. Selective or distorted reporting is a typical form of such bias. We may assume that u does not depend on whether a true relationship exists or not. This is not an unreasonable assumption, since typically it is impossible to know which relationships are indeed true. In the presence of bias (Table 2), one gets PPV = $([1-\beta]R + u\beta R)/(R + \alpha - \beta R + u - u\alpha + u\beta R)$, and PPV decreases with increasing u, unless $1-\beta \le \alpha$, i.e., $1-\beta \le 0.05$ for most situations. Thus, with increasing bias, the chances that a research finding is true diminish considerably. This is shown for different levels of power and for different pre-study odds in Figure 1.

Conversely, true research findings may occasionally be annulled because of reverse bias. For example, with large measurement errors relationships are lost in noise [12], or investigators use data inefficiently or fail to notice statistically significant relationships, or there may be conflicts of interest that tend to "bury" significant findings [13]. There is no good large-scale empirical evidence on how frequently such reverse bias may occur across diverse research fields. However, it is probably fair to say that reverse bias is not as common. Moreover measurement errors and inefficient use of data are probably becoming less frequent problems, since measurement error has decreased with technological advances in the molecular era and investigators are becoming increasingly sophisticated about their data. Regardless, reverse bias may be modeled in the same way as bias above. Also reverse bias should not be confused with chance variability that may lead to missing a true relationship because of chance.

Testing by Several Independent Teams

Several independent teams may be addressing the same sets of research questions. As research efforts are globalized, it is practically the rule that several research teams, often dozens of them, may probe the same or similar questions. Unfortunately, in some areas, the prevailing mentality until now has been to focus on isolated discoveries by single teams and interpret research experiments in isolation. An increasing number of questions have at least one study claiming a research finding, and this receives unilateral attention. The probability that at least one study, among several done on the same question, claims a statistically significant research finding is easy to estimate. For n independent studies of equal power, the 2 × 2 table is shown in Table 3: PPV = $R(1-\beta^n)/(R+1-[1-\alpha]^n - R\beta^n)$ (not considering bias). With increasing number of independent studies, PPV tends to decrease, unless $1-\beta < \alpha$, i.e., typically $1-\beta < 0.05$. This is shown for different levels of power and for different pre-study odds in Figure 2. For n studies of different power, the term β^n is replaced by the product of the terms β_i for $i = 1$ to n, but inferences are similar.

Corollaries

A practical example is shown in Box 1. Based on the above considerations, one may deduce several interesting corollaries about the probability that a research finding is indeed true.

Corollary 1: The smaller the studies conducted in a scientific field, the less likely the research findings are to be true. Small sample size means smaller power and, for all functions above, the PPV for a true research finding decreases as power decreases towards $1-\beta = 0.05$. Thus, other factors being equal, research findings are more likely true in scientific fields that undertake large studies, such as randomized controlled trials in cardiology (several thousand subjects randomized) [14] than in scientific fields with small studies, such as most research of molecular predictors (sample sizes 100-fold smaller) [15].

Corollary 2: The smaller the effect sizes in a scientific field, the less likely the research findings are to be true. Power is also related to the effect size. Thus research findings are more likely true in scientific fields with large effects, such as the impact of smoking on cancer or cardiovascular disease (relative risks 3–20), than in scientific fields where postulated effects are small, such as genetic risk factors for multigenetic diseases (relative risks 1.1–1.5) [7]. Modern epidemiology is increasingly obliged to target smaller

Table 2. Research Findings and True Relationships in the Presence of Bias

Research Finding	True Relationship		
	Yes	No	Total
Yes	$(c[1-\beta]R + uc\beta R)/(R+1)$	$c\alpha + uc(1-\alpha)/(R+1)$	$c(R+\alpha-\beta R+u-u\alpha+u\beta R)/(R+1)$
No	$(1-u)c\beta R/(R+1)$	$(1-u)c(1-\alpha)/(R+1)$	$c(1-u)(1-\alpha+\beta R)/(R+1)$
Total	$cR/(R+1)$	$c/(R+1)$	c

DOI: 10.1371/journal.pmed.0020124.t002

effect sizes [16]. Consequently, the proportion of true research findings is expected to decrease. In the same line of thinking, if the true effect sizes are very small in a scientific field, this field is likely to be plagued by almost ubiquitous false positive claims. For example, if the majority of true genetic or nutritional determinants of complex diseases confer relative risks less than 1.05, genetic or nutritional epidemiology would be largely utopian endeavors.

Corollary 3: The greater the number and the lesser the selection of tested relationships in a scientific field, the less likely the research findings are to be true. As shown above, the post-study probability that a finding is true (PPV) depends a lot on the pre-study odds (*R*). Thus, research findings are more likely true in confirmatory designs, such as large phase III randomized controlled trials, or meta-analyses thereof, than in hypothesis-generating experiments. Fields considered highly informative and creative given the wealth of the assembled and tested information, such as microarrays and other high-throughput discovery-oriented research [4,8,17], should have extremely low PPV.

Corollary 4: The greater the flexibility in designs, definitions, outcomes, and analytical modes in a scientific field, the less likely the research findings are to be true. Flexibility increases the potential for transforming what would be "negative" results into "positive" results, i.e., bias, *u*. For several research designs, e.g., randomized controlled trials [18–20] or meta-analyses [21,22], there have been efforts to standardize their conduct and reporting. Adherence to common standards is likely to increase the proportion of true findings. The same applies to outcomes. True findings may be more common when outcomes are unequivocal and universally agreed (e.g., death) rather than when multifarious outcomes are devised (e.g., scales for schizophrenia outcomes) [23]. Similarly, fields that use commonly agreed, stereotyped analytical methods (e.g., Kaplan-Meier plots and the log-rank test) [24] may yield a larger proportion of true findings than fields where analytical methods are still under experimentation (e.g., artificial intelligence methods) and only "best" results are reported. Regardless, even in the most stringent research designs, bias seems to be a major problem. For example, there is strong evidence that selective outcome reporting, with manipulation of the outcomes and analyses reported, is a common problem even for randomized trails [25]. Simply abolishing selective publication would not make this problem go away.

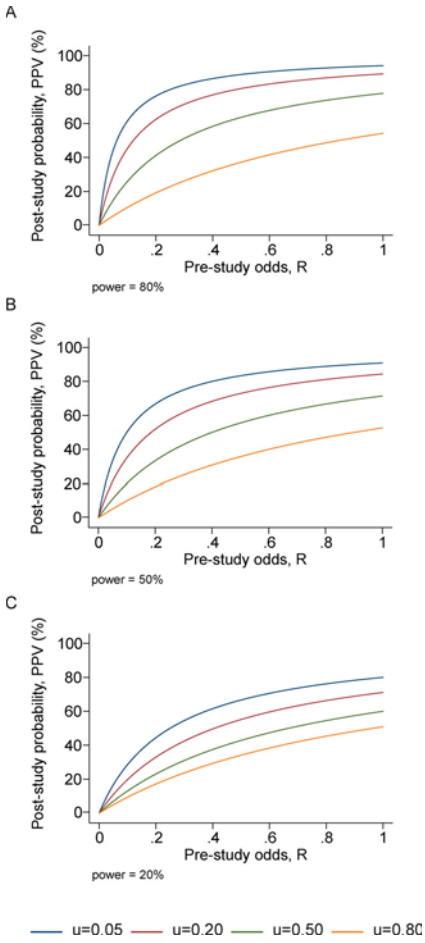
Corollary 5: The greater the financial and other interests and prejudices in a scientific field, the less likely the research findings are to be true. Conflicts of interest and prejudice may increase bias, *u*. Conflicts of interest are very common in biomedical research [26], and typically they are inadequately and sparsely reported [26,27]. Prejudice may not necessarily have financial roots. Scientists in a given field may be prejudiced purely because of their belief in a scientific theory or commitment to their own findings. Many otherwise seemingly independent, university-based studies may be conducted for no other reason than to give physicians and researchers qualifications for promotion or tenure. Such nonfinancial conflicts may also lead to distorted reported results and interpretations. Prestigious investigators may suppress via the peer review process the appearance and dissemination of findings that refute their findings, thus condemning their field to perpetuate false dogma. Empirical evidence on expert opinion shows that it is extremely unreliable [28].

Corollary 6: The hotter a scientific field (with more scientific teams involved), the less likely the research findings are to be true.

Table 3. Research Findings and True Relationships in the Presence of Multiple Studies

Research Finding	True Relationship		
	Yes	No	Total
Yes	$cR(1-\beta^n)/(R+1)$	$c(1-[1-\alpha]^n)/(R+1)$	$c(R+1-[1-\alpha]^n-R\beta^n)/(R+1)$
No	$cR\beta^n/(R+1)$	$c(1-\alpha)^n/(R+1)$	$c([1-\alpha]^n+R\beta^n)/(R+1)$
Total	$cR/(R+1)$	$c/(R+1)$	c

DOI: 10.1371/journal.pmed.0020124.t003

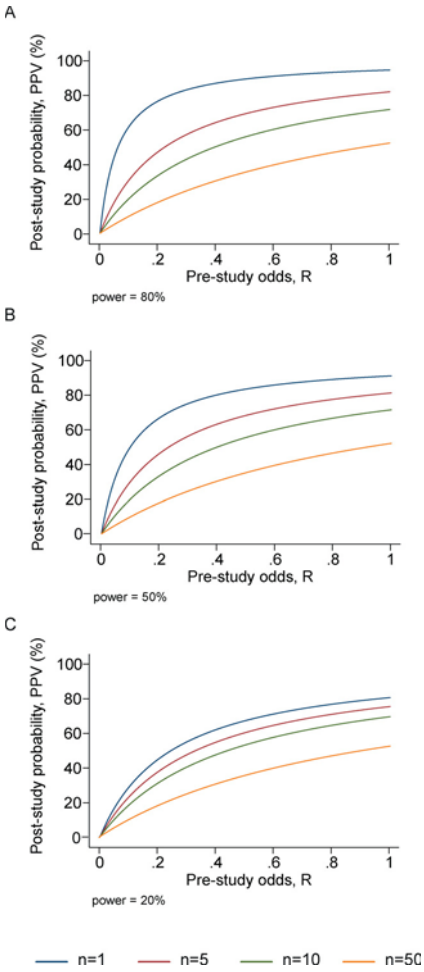


DOI: 10.1371/journal.pmed.0020124.g001

Figure 1. PPV (Probability That a Research Finding Is True) as a Function of the Pre-Study Odds for Various Levels of Bias, *u*

Panels correspond to power of 0.20, 0.50, and 0.80.

This seemingly paradoxical corollary follows because, as stated above, the PPV of isolated findings decreases when many teams of investigators are involved in the same field. This may explain why we occasionally see major excitement followed rapidly by severe disappointments in fields that draw wide attention. With many teams working on the same field and with massive experimental data being produced, timing is of the essence in beating competition. Thus, each team may prioritize on pursuing and disseminating its most impressive "positive" results. "Negative" results may become attractive for dissemination only if some other team has found a "positive" association on the same question. In that case, it may be attractive to refute a claim made in some prestigious journal. The term Proteus phenomenon has been coined to describe this phenomenon of rapidly



DOI: 10.1371/journal.pmed.0020124.g002

Figure 2. PPV (Probability That a Research Finding Is True) as a Function of the Pre-Study Odds for Various Numbers of Conducted Studies, *n*

Panels correspond to power of 0.20, 0.50, and 0.80.

Box 1. An Example: Science at Low Pre-Study Odds

Let us assume that a team of investigators performs a whole genome association study to test whether any of 100,000 gene polymorphisms are associated with susceptibility to schizophrenia. Based on what we know about the extent of heritability of the disease, it is reasonable to expect that probably around ten gene polymorphisms among those tested would be truly associated with schizophrenia, with relatively similar odds ratios around 1.3 for the ten or so polymorphisms and with a fairly similar power to identify any of them. Then $R = 10/100{,}000 = 10^{-4}$, and the pre-study probability for any polymorphism to be associated with schizophrenia is also $R/(R + 1) = 10^{-4}$. Let us also suppose that the study has 60% power to find an association with an odds ratio of 1.3 at $\alpha = 0.05$. Then it can be estimated that if a statistically significant association is found with the *p*-value barely crossing the 0.05 threshold, the post-study probability that this is true increases about 12-fold compared with the pre-study probability, but it is still only 12×10^{-4}.

Now let us suppose that the investigators manipulate their design, analyses, and reporting so as to make more relationships cross the $p = 0.05$ threshold even though this would not have been crossed with a perfectly adhered to design and analysis and with perfect comprehensive reporting of the results, strictly according to the original study plan. Such manipulation could be done, for example, with serendipitous inclusion or exclusion of certain patients or controls, post hoc subgroup analyses, investigation of genetic contrasts that were not originally specified, changes in the disease or control definitions, and various combinations of selective or distorted reporting of the results. Commercially available "data mining" packages actually are proud of their ability to yield statistically significant results through data dredging. In the presence of bias with $u = 0.10$, the post-study probability that a research finding is true is only 4.4×10^{-4}. Furthermore, even in the absence of any bias, when ten independent research teams perform similar experiments around the world, if one of them finds a formally statistically significant association, the probability that the research finding is true is only 1.5×10^{-4}, hardly any higher than the probability we had before any of this extensive research was undertaken!

alternating extreme research claims and extremely opposite refutations [29]. Empirical evidence suggests that this sequence of extreme opposites is very common in molecular genetics [29].

These corollaries consider each factor separately, but these factors often influence each other. For example, investigators working in fields where true effect sizes are perceived to be small may be more likely to perform large studies than investigators working in fields where true effect sizes are perceived to be large. Or prejudice may prevail in a hot scientific field, further undermining the predictive value of its research findings. Highly prejudiced stakeholders may even create a barrier that aborts efforts at obtaining and disseminating opposing results. Conversely, the fact that a field is hot or has strong invested interests may sometimes promote larger studies and improved standards of research, enhancing the predictive value of its research findings. Or massive discovery-oriented testing may result in such a large yield of significant relationships that investigators have enough to report and search further and thus refrain from data dredging and manipulation.

Most Research Findings Are False for Most Research Designs and for Most Fields

In the described framework, a PPV exceeding 50% is quite difficult to get. Table 4 provides the results of simulations using the formulas developed for the influence of power, ratio of true to non-true relationships, and bias, for various types of situations that may be characteristic of specific study designs and settings. A finding from a well-conducted, adequately powered randomized controlled trial starting with a 50% pre-study chance that the intervention is effective is eventually true about 85% of the time. A fairly similar performance is expected of a confirmatory meta-analysis of good-quality randomized trials: potential bias probably increases, but power and pre-test chances are higher compared to a single randomized trial. Conversely, a meta-analytic finding from inconclusive studies where pooling is used to "correct" the low power of single studies, is probably false if $R \leq 1{:}3$. Research findings from underpowered, early-phase clinical trials would be true about one in four times, or even less frequently if bias is present. Epidemiological studies of an exploratory nature perform even worse, especially when underpowered, but even well-powered epidemiological studies may have only a one in five chance being true, if $R = 1{:}10$. Finally, in discovery-oriented research with massive testing, where tested relationships exceed true ones 1,000-fold (e.g., 30,000 genes tested, of which 30 may be the true culprits) [30,31], PPV for each claimed relationship is extremely low, even with considerable

standardization of laboratory and statistical methods, outcomes, and reporting thereof to minimize bias.

Claimed Research Findings May Often Be Simply Accurate Measures of the Prevailing Bias

As shown, the majority of modern biomedical research is operating in areas with very low pre- and post-study probability for true findings. Let us suppose that in a research field there are no true findings at all to be discovered. History of science teaches us that scientific endeavor has often in the past wasted effort in fields with absolutely no yield of true scientific information, at least based on our current understanding. In such a "null field," one would ideally expect all observed effect sizes to vary by chance around the null in the absence of bias. The extent that observed findings deviate from what is expected by chance alone would be simply a pure measure of the prevailing bias.

For example, let us suppose that no nutrients or dietary patterns are actually important determinants for the risk of developing a specific tumor. Let us also suppose that the scientific literature has examined 60 nutrients and claims all of them to be related to the risk of developing this tumor with relative risks in the range of 1.2 to 1.4 for the comparison of the upper to lower intake tertiles. Then the claimed effect sizes are simply measuring nothing else but the net bias that has been involved in the generation of this scientific literature. Claimed effect sizes are in fact the most accurate estimates of the net bias. It even follows that between "null fields," the fields that claim stronger effects (often with accompanying claims of medical or public health importance) are simply those that have sustained the worst biases.

For fields with very low PPV, the few true relationships would not distort this overall picture much. Even if a few relationships are true, the shape of the distribution of the observed effects would still yield a clear measure of the biases involved in the field. This concept totally reverses the way we view scientific results. Traditionally, investigators have viewed large and highly significant effects with excitement, as signs of important discoveries. Too large and too highly significant effects may actually be more likely to be signs of large bias in most fields of modern research. They should lead investigators to careful critical thinking about what might have gone wrong with their data, analyses, and results.

Of course, investigators working in any field are likely to resist accepting that the whole field in which they have spent their careers is a "null field." However, other lines of evidence, or advances in technology and experimentation, may lead eventually to the dismantling of a scientific field. Obtaining measures of the net bias in one field may also be useful for obtaining insight into what might be the range of bias operating in other fields where similar analytical methods, technologies, and conflicts may be operating.

How Can We Improve the Situation?

Is it unavoidable that most research findings are false, or can we improve the situation? A major problem is that it is impossible to know with 100% certainty what the truth is in any research question. In this regard, the pure "gold" standard is unattainable. However, there are several approaches to improve the post-study probability.

Better powered evidence, e.g., large studies or low-bias meta-analyses, may help, as it comes closer to the unknown "gold" standard. However, large studies may still have biases and these should be acknowledged and avoided. Moreover, large-scale evidence is impossible to obtain for all of the millions and trillions of research questions posed in current research. Large-scale evidence should be targeted for research questions where the pre-study probability is already considerably high, so that a significant research finding will lead to a post-test probability that would be considered quite definitive. Large-scale evidence is also particularly indicated when it can test major concepts rather than narrow, specific questions. A negative finding can then refute not only a specific proposed claim, but a whole field or considerable portion thereof. Selecting the performance of large-scale studies based on narrow-minded criteria, such as the marketing promotion of a specific drug, is largely wasted research. Moreover, one should be cautious that extremely large studies may be more likely to find a formally statistical significant difference for a trivial effect that is not really meaningfully different from the null [32–34].

Second, most research questions are addressed by many teams, and it is misleading to emphasize the statistically significant findings of any single team. What matters is the

Table 4. PPV of Research Findings for Various Combinations of Power (1 − β), Ratio of True to Not-True Relationships *(R)*, and Bias *(u)*

1 − β	*R*	*u*	Practical Example	PPV
0.80	1:1	0.10	Adequately powered RCT with little bias and 1:1 pre-study odds	0.85
0.95	2:1	0.30	Confirmatory meta-analysis of good-quality RCTs	0.85
0.80	1:3	0.40	Meta-analysis of small inconclusive studies	0.41
0.20	1:5	0.20	Underpowered, but well-performed phase I/II RCT	0.23
0.20	1:5	0.80	Underpowered, poorly performed phase I/II RCT	0.17
0.80	1:10	0.30	Adequately powered exploratory epidemiological study	0.20
0.20	1:10	0.30	Underpowered exploratory epidemiological study	0.12
0.20	1:1,000	0.80	Discovery-oriented exploratory research with massive testing	0.0010
0.20	1:1,000	0.20	As in previous example, but with more limited bias (more standardized)	0.0015

The estimated PPVs (positive predictive values) are derived assuming $\alpha = 0.05$ for a single study.
RCT, randomized controlled trial.
DOI: 10.1371/journal.pmed.0020124.t004

totality of the evidence. Diminishing bias through enhanced research standards and curtailing of prejudices may also help. However, this may require a change in scientific mentality that might be difficult to achieve. In some research designs, efforts may also be more successful with upfront registration of studies, e.g., randomized trials [35]. Registration would pose a challenge for hypothesis-generating research. Some kind of registration or networking of data collections or investigators within fields may be more feasible than registration of each and every hypothesis-generating experiment. Regardless, even if we do not see a great deal of progress with registration of studies in other fields, the principles of developing and adhering to a protocol could be more widely borrowed from randomized controlled trials.

Finally, instead of chasing statistical significance, we should improve our understanding of the range of R values—the pre-study odds—where research efforts operate [10]. Before running an experiment, investigators should consider what they believe the chances are that they are testing a true rather than a non-true relationship. Speculated high R values may sometimes then be ascertained. As described above, whenever ethically acceptable, large studies with minimal bias should be performed on research findings that are considered relatively established, to see how often they are indeed confirmed. I suspect several established "classics" will fail the test [36].

Nevertheless, most new discoveries will continue to stem from hypothesis-generating research with low or very low pre-study odds. We should then acknowledge that statistical significance testing in the report of a single study gives only a partial picture, without knowing how much testing has been done outside the report and in the relevant field at large. Despite a large statistical literature for multiple testing corrections [37], usually it is impossible to decipher how much data dredging by the reporting authors or other research teams has preceded a reported research finding. Even if determining this were feasible, this would not inform us about the pre-study odds. Thus, it is unavoidable that one should make approximate assumptions on how many relationships are expected to be true among those probed across the relevant research fields and research designs. The wider field may yield some guidance for estimating this probability for the isolated research project. Experiences from biases detected in other neighboring fields would also be useful to draw upon. Even though these assumptions would be considerably subjective, they would still be very useful in interpreting research claims and putting them in context. ■

References

1. Ioannidis JP, Haidich AB, Lau J (2001) Any casualties in the clash of randomised and observational evidence? BMJ 322: 879–880.
2. Lawlor DA, Davey Smith G, Kundu D, Bruckdorfer KR, Ebrahim S (2004) Those confounded vitamins: What can we learn from the differences between observational versus randomised trial evidence? Lancet 363: 1724–1727.
3. Vandenbroucke JP (2004) When are observational studies as credible as randomised trials? Lancet 363: 1728–1731.
4. Michiels S, Koscielny S, Hill C (2005) Prediction of cancer outcome with microarrays: A multiple random validation strategy. Lancet 365: 488–492.
5. Ioannidis JPA, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG (2001) Replication validity of genetic association studies. Nat Genet 29: 306–309.
6. Colhoun HM, McKeigue PM, Davey Smith G (2003) Problems of reporting genetic associations with complex outcomes. Lancet 361: 865–872.
7. Ioannidis JP (2003) Genetic associations: False or true? Trends Mol Med 9: 135–138.
8. Ioannidis JPA (2005) Microarrays and molecular research: Noise discovery? Lancet 365: 454–455.
9. Sterne JA, Davey Smith G (2001) Sifting the evidence—What's wrong with significance tests. BMJ 322: 226–231.
10. Wacholder S, Chanock S, Garcia-Closas M, El ghormli L, Rothman N (2004) Assessing the probability that a positive report is false: An approach for molecular epidemiology studies. J Natl Cancer Inst 96: 434–442.
11. Risch NJ (2000) Searching for genetic determinants in the new millennium. Nature 405: 847–856.
12. Kelsey JL, Whittemore AS, Evans AS, Thompson WD (1996) Methods in observational epidemiology, 2nd ed. New York: Oxford U Press. 432 p.
13. Topol EJ (2004) Failing the public health—Rofecoxib, Merck, and the FDA. N Engl J Med 351: 1707–1709.
14. Yusuf S, Collins R, Peto R (1984) Why do we need some large, simple randomized trials? Stat Med 3: 409–422.
15. Altman DG, Royston P (2000) What do we mean by validating a prognostic model? Stat Med 19: 453–473.
16. Taubes G (1995) Epidemiology faces its limits. Science 269: 164–169.
17. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, et al. (1999) Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. Science 286: 531–537.
18. Moher D, Schulz KF, Altman DG (2001) The CONSORT statement: Revised recommendations for improving the quality of reports of parallel-group randomised trials. Lancet 357: 1191–1194.
19. Ioannidis JP, Evans SJ, Gotzsche PC, O'Neill RT, Altman DG, et al. (2004) Better reporting of harms in randomized trials: An extension of the CONSORT statement. Ann Intern Med 141: 781–788.
20. International Conference on Harmonisation E9 Expert Working Group (1999) ICH Harmonised Tripartite Guideline. Statistical principles for clinical trials. Stat Med 18: 1905–1942.
21. Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, et al. (1999) Improving the quality of reports of meta-analyses of randomised controlled trials: The QUOROM statement. Quality of Reporting of Meta-analyses. Lancet 354: 1896–1900.
22. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, et al. (2000) Meta-analysis of observational studies in epidemiology: A proposal for reporting. Meta-analysis of Observational Studies in Epidemiology (MOOSE) group. JAMA 283: 2008–2012.
23. Marshall M, Lockwood A, Bradley C, Adams C, Joy C, et al. (2000) Unpublished rating scales: A major source of bias in randomised controlled trials of treatments for schizophrenia. Br J Psychiatry 176: 249–252.
24. Altman DG, Goodman SN (1994) Transfer of technology from statistical journals to the biomedical literature. Past trends and future predictions. JAMA 272: 129–132.
25. Chan AW, Hrobjartsson A, Haahr MT, Gotzsche PC, Altman DG (2004) Empirical evidence for selective reporting of outcomes in randomized trials: Comparison of protocols to published articles. JAMA 291: 2457–2465.
26. Krimsky S, Rothenberg LS, Stott P, Kyle G (1998) Scientific journals and their authors' financial interests: A pilot study. Psychother Psychosom 67: 194–201.
27. Papanikolaou GN, Baltogianni MS, Contopoulos-Ioannidis DG, Haidich AB, Giannakakis IA, et al. (2001) Reporting of conflicts of interest in guidelines of preventive and therapeutic interventions. BMC Med Res Methodol 1: 3.
28. Antman EM, Lau J, Kupelnick B, Mosteller F, Chalmers TC (1992) A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts. Treatments for myocardial infarction. JAMA 268: 240–248.
29. Ioannidis JP, Trikalinos TA (2005) Early extreme contradictory estimates may appear in published research: The Proteus phenomenon in molecular genetics research and randomized trials. J Clin Epidemiol 58: 543–549.
30. Ntzani EE, Ioannidis JP (2003) Predictive ability of DNA microarrays for cancer outcomes and correlates: An empirical assessment. Lancet 362: 1439–1444.
31. Ransohoff DF (2004) Rules of evidence for cancer molecular-marker discovery and validation. Nat Rev Cancer 4: 309–314.
32. Lindley DV (1957) A statistical paradox. Biometrika 44: 187–192.
33. Bartlett MS (1957) A comment on D.V. Lindley's statistical paradox. Biometrika 44: 533–534.
34. Senn SJ (2001) Two cheers for P-values. J Epidemiol Biostat 6: 193–204.
35. De Angelis C, Drazen JM, Frizelle FA, Haug C, Hoey J, et al. (2004) Clinical trial registration: A statement from the International Committee of Medical Journal Editors. N Engl J Med 351: 1250–1251.
36. Ioannidis JPA (2005) Contradicted and initially stronger effects in highly cited clinical research. JAMA 294: 218–228.
37. Hsueh HM, Chen JJ, Kodell RL (2003) Comparison of methods for estimating the number of true null hypotheses in multiplicity testing. J Biopharm Stat 13: 675–689.