



Online article and related content
current as of February 21, 2010.

The Need for Large-Scale Randomized Evidence Without Undue Emphasis on Small Trials, Meta-analyses, or Subgroup Analyses

Charles H. Hennekens; David DeMets

JAMA. 2009;302(21):2361-2362 (doi:10.1001/jama.2009.1756)

<http://jama.ama-assn.org/cgi/content/full/302/21/2361>

Correction

[Contact me if this article is corrected.](#)

Citations

This article has been cited 2 times.
[Contact me when this article is cited.](#)

Topic collections

Public Health; Public Health, Other; Quality of Care; Evidence-Based Medicine;
Statistics and Research Methods; Drug Therapy; Drug Therapy, Other
[Contact me when new articles are published in these topic areas.](#)

Subscribe

<http://jama.com/subscribe>

Email Alerts

<http://jamaarchives.com/alerts>

Permissions

permissions@ama-assn.org
<http://pubs.ama-assn.org/misc/permissions.dtl>

Reprints/E-prints

reprints@ama-assn.org

The Need for Large-Scale Randomized Evidence Without Undue Emphasis on Small Trials, Meta-analyses, or Subgroup Analyses

Charles H. Hennekens, MD, DrPH

David DeMets, PhD

RANDOMIZED TRIALS OF ADEQUATE SIZE AND DURATION designed to test a priori hypotheses represent the most reliable design strategy to detect the most realistically small to moderate therapeutic effects of drugs. Such trials should achieve high adherence to an adequate dose of the drug and a sufficient number of clinical end points to distinguish reliably between the null hypothesis and the most plausible alternative hypothesis of small to moderate benefit or harm.¹ With regard to the development of drugs to treat diabetes mellitus, the US Food and Drug Administration (FDA) has developed guidance for industry that somewhat overemphasizes results from meta-analyses of phase 2 trials that were not large enough to test realistic hypotheses about clinical cardiovascular (CV) events.² Even in aggregate, such results should be considered more as hypothesis formulating than as hypothesis testing. The main need now is for trials that are large enough to have adequate statistical power.

The quality and usefulness of any meta-analysis are dependent on the quality and comparability of data from the component trials. In particular, the trials combined should have high adherence and follow-up rates and should have reasonably comparable drugs, doses, and outcomes. The characteristics of the participants and the magnitude of effect from each trial must be sufficiently similar so that their combination will not produce a distorted estimate. Thus, meta-analyses can reduce the role of chance in the interpretation but may introduce bias and confounding.

For example, a meta-analysis of rosiglitazone³ involved 42 randomized trials with a total of 26 011 patients who experienced 158 myocardial infarctions (MIs) and 61 deaths due to CV causes (CV death). The investigators concluded that rosiglitazone was associated with a significant increase in risk of MI (relative risk [RR], 1.43; 95% confidence interval [CI], 1.03-1.98) as well as a nonsignificant increase in CV death (RR, 1.64; 95% CI, 0.98-2.74).³ The chief value of this report should have been to formulate a hypothesis about one possible hazard of rosiglitazone that may offset any potential benefits.⁴ Furthermore, the widths of the CIs suggest that this meta-analysis was unable to distinguish reliably whether rosiglitazone conferred no hazard or a substantial hazard of CV events. Hypothesis formulation should lead to adequate hypothesis testing in

another randomized trial of adequate size and duration designed a priori to address the question.⁴

The Rosiglitazone Evaluated for Cardiovascular Outcomes and Regulation of Glycaemia in Diabetes (RECORD) trial was designed a priori to assess the noninferiority of rosiglitazone added to metformin or sulfonylurea compared with dual therapy metformin and sulfonylurea on reducing CV events among 4447 patients with type 2 diabetes.⁵ The primary prespecified outcome was time to first CV hospitalization or CV death with a hazard ratio (HR) noninferiority margin of 1.20, which is the upper bound of the 95% CI as recommended by the FDA.² During a mean 5.5-year follow up, there were 321 incident primary clinical events in the rosiglitazone group and 323 in the active comparator group (HR, 0.99; 95% CI, 0.85-1.16), meeting the criterion for noninferiority. Thus, the results of the large-scale trial did not support the hypothesis formulated from the meta-analysis of smaller trials.

Likewise, in a meta-analysis of 7 small trials evaluating use of intravenous magnesium during suspected acute MI, there were 25 deaths among 657 patients in the magnesium group vs 53 deaths among 644 patients in the placebo group (HR, 0.45; 95% CI, 0.15-0.74).⁶ Although the existing totality of evidence was compatible with the possibility that intravenous magnesium was an effective, safe, and inexpensive intervention, a prudent approach was to await the results of the large fourth International Study of Infarct Survival (ISIS-4) trial before routinely using this therapy. In ISIS-4, 58 050 patients with suspected MI were randomized to receive either intravenous magnesium or usual care.⁷ Patients treated with magnesium had a nonsignificant 6% increase in mortality, as well as significant increases in heart failure, death attributable to cardiogenic shock, and bradycardia. In subgroup analyses (which are useful to formulate but not test hypotheses) no significant differences were found among patients treated less than 6 hours after the onset of symptoms, those who received intravenous magnesium within 2 hours after thrombolytic therapy, and those who received neither thrombolytic therapy

Author Affiliations: Department of Clinical Science and Medical Education, Charles E. Schmidt College of Biomedical Science and Center of Excellence, Florida Atlantic University, Boca Raton (Dr Hennekens); and Department of Biostatistics & Medical Informatics, University of Wisconsin School of Medicine and Public Health, Madison (Dr DeMets).

Corresponding Author: Charles H. Hennekens, MD, DrPH, Sir Richard Doll Research Professor, Florida Atlantic University, 777 Glades Rd, Research Park Ste 310, Boca Raton, FL 33431 (chenneke@fau.edu).

nor aspirin. As a result, magnesium was no longer recommended as standard therapy of acute evolving MI.

As another example, in a meta-analysis of 9 relatively small randomized trials of angiotensin-receptor blockers (ARBs) for prevention of recurrent atrial fibrillation (AF), there was a statistically significant protective effect associated with use of ARBs (HR, 0.82; 95% CI, 0.70-0.97).⁸ Based on these findings, which formulated the hypothesis, the multicenter Gruppo Italiano per lo Studio della Streptochinasi nell'Infarto Miocardico Atrial Fibrillation (GISSI-AF) trial⁹ was designed to test the hypothesis. GISSI-AF enrolled 1442 patients who were in sinus rhythm at baseline but had experienced either multiple AF episodes in the prior 6 months or had successful cardioversion for AF in the prior 2 weeks. All patients had underlying CV disease, diabetes, or left atrial enlargement. Patients were randomized to receive either the ARB, valsartan, with dose escalation to 320 mg, for 1 year or placebo in addition to their other treatments. For the primary end points at 1 year, the results indicated no significant difference between valsartan and placebo in the rate of first AF recurrence (HR, 0.97; 95% CI, 0.83-1.14) after adjustment for baseline variables.

More generally, for several different therapeutic questions, the results of meta-analyses that were not large enough on their own to be reliable were compared with those of subsequent large randomized trials of the same question that involved at least 1000 patients. The meta-analyses did not predict accurately outcomes of the large randomized trials 35% of the time, but this may have been because in many cases neither was large enough to be reliable.¹⁰

With respect to the interpretation of subgroup analyses of randomized trials as well as their meta-analyses, the caveats needed to compare those defined a priori by baseline characteristics are far less than those required when comparisons are made on the basis of variables derived after randomization. With regard to the former, there is loss of statistical power because only subgroups of the total randomized trial population are being compared. A greater concern, however, is that confounding variables may no longer be distributed at random among the subgroups. Analyses of subgroups defined a posteriori from information accumulated after randomization can only formulate data-derived hypotheses and cannot provide serious evidence for hypothesis testing.¹

In summary, the guiding principle about benefits and risks of interventions should be that rational clinical decisions for individual patients as well as policy decision for the health of the general public should be based on a sufficient totality of evidence. Furthermore, phase 2 trials of drugs should be performed mainly for proof of concept and dose ranging. In addition, meta-analyses and subgroup analyses are useful to formulate but not test hypotheses. If the totality of evidence is incomplete, it is appropriate to remain uncertain. Finally, to detect reliably the most plausible small to moderate effects of interventions a sufficient totality of evidence must include large-scale randomized phase 3 trials of sufficient size and duration with high adherence and a large enough number of clinical

end points to distinguish reliably between the null hypothesis of no effect and the most plausible alternative hypotheses of small to moderate benefit or harm.

Financial Disclosures: Dr Hennekens reported that he is funded by the Charles E. Schmidt College of Biomedical Science, Department of Clinical Science, and Medical Education & Center of Excellence at Florida Atlantic University (FAU) as principal investigator on 2 investigator-initiated research grants funded to FAU by Bayer; serves in an advisory role to investigators and sponsors as chair of the data and safety monitoring boards for Actelion, Amgen, Bristol-Myers Squibb, Daiippon Sumitomo, and Sanofi-Aventis, and as a member of the data and safety monitoring boards for Bayer and the Canadian Institutes of Health Research; serves in an advisory role to the US Food and Drug Administration, US National Institutes of Health, and UpToDate; and serves as an independent scientist in an advisory role to legal counsel for General Electric and GlaxoSmithKline; serves as a speaker for the Association of Research in Vision and Ophthalmology, National Association for Continuing Education, PriMed, the International Atherosclerosis Society, AstraZeneca, and Pfizer; receives royalties for authorship or editorship of 3 textbooks and as coinventor on patents for inflammatory markers and CV disease that are held by Brigham and Women's Hospital; and has an investment management relationship with the West-Bacon Group within SunTrust Investment Services, which has discretionary investment authority. Dr DeMets reported that he is partially supported by a National Institutes of Health grant to the University of Wisconsin for the Clinical Translational Science Award for statistical consultation and collaboration and administrative leadership, as a leader of the Data Management and Biostatistics Core (Core C) of the Wisconsin Alzheimer's Disease Research Center grant, by serving as a principal investigator on University of Wisconsin contracts with industry for statistical analysis center activity for multicenter trials, which are currently sponsored by Amgen, AstraZeneca, and Bristol-Myers Squibb; serves or has recently served as an independent biostatistician in an advisory role to investigators and sponsors as a member of the data and safety monitoring boards for Actelion, Amgen, Astellas, AstraZeneca, Biotronik, Boehringer-Ingelheim, CVRx, Genentech, GlaxoSmithKline, Novartis, Merck, Pfizer, Roche, Sanofi Aventis, Takeda, the Duke Clinical Research Institute, and the Population Health Research Institute of McMaster University, Canadian Institutes of Health Research, Harvard Clinical Research Institute, and Hamilton Clinical Research Institute; receives royalties from publishers of the 3 textbooks that he has coauthored and edited; has tax sheltered retirement accounts in mutual funds with Fidelity and UBS; and has 2 small accounts of stock with Sun and Intel.

Additional Contributions: We are indebted to Wendy Schneider, MSN, RN, affiliated clinical instructor, Charles E. Schmidt College of Biomedical Science, FAU, for her expert advice and help as part of her regular duties and to Sir Richard Peto, professor of medical statistics, Oxford University, Oxford, England for his technical support in the preparation of the manuscript for which he received no compensation.

REFERENCES

1. Hennekens CH, Buring JE. *Epidemiology in Medicine*. Boston, MA: Little Brown & Co; 1987.
2. US Food and Drug Administration. *Guidance for Industry: Diabetes Mellitus—Evaluating Cardiovascular Risk in New Antidiabetic Therapies to Treat Type 2 Diabetes*. Food and Drug Administration, Center for Drug Evaluation and Research, December 2008. <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm071627.pdf>. Accessed August 11, 2009.
3. Nissen SE, Wolski K. Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. *N Engl J Med*. 2007;356(24):2457-2471.
4. Hennekens CH, DeMets D, Bairey-Merz CN, Borzak S, Borer J. Doing more good than harm: the need for a cease fire. *Am J Med*. 2009;122(4):315-316.
5. Home PD, Pocock SJ, Beck-Nielsen H, et al; RECORD Study Team. Rosiglitazone Evaluated for Cardiovascular Outcomes in Oral Agent Combination Therapy for Type 2 Diabetes (RECORD). *Lancet*. 2009;373(9681):2125-2135.
6. Teo KK, Yusuf S, Collins R, Held PH, Peto R. Effects of intravenous magnesium in suspected acute myocardial infarction: overview of randomised trials. *BMJ*. 1991;303(6816):1499-1503.
7. ISIS-4 (Fourth International Study of Infarct Survival) Collaborative Group. ISIS-4: a randomised factorial trial assessing early oral captopril, oral mononitrate, and intravenous magnesium sulphate in 58,050 patients with suspected acute myocardial infarction. *Lancet*. 1995;345(8951):669-685.
8. Healey JS, Baranchuk A, Crystal E, et al. Prevention of atrial fibrillation with angiotensin-converting enzyme inhibitors and angiotensin receptor blockers: a meta-analysis. *J Am Coll Cardiol*. 2005;45(11):1832-1839.
9. Disertori M, Latini R, Barlera S, et al; GISSI-AF Investigators. Valsartan for the prevention of recurrent atrial fibrillation. *N Engl J Med*. 2009;360(16):1606-1617.
10. LeLorier J, Gregoire G, Benhaddad A, Lapierre J, D'Erderian F. Discrepancies between meta-analyses and subsequent large randomized, controlled trials. *N Engl J Med*. 1997;337(8):536-542.