



Published in final edited form as:

*Ann Emerg Med.* 2012 May ; 59(5): 425–431.e11. doi:10.1016/j.annemergmed.2011.05.024.

## Understanding Commonly Encountered Limitations in Clinical Research: An Emergency Medicine Resident's Perspective

**Christopher W. Jones, MD and Timothy F. Platts-Mills, MD**

From the Department of Emergency Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC

### INTRODUCTION

The breadth of emergency medicine and the rapid growth in relevant research makes an ability to assess new research findings particularly important for emergency physicians. Improvements in the treatment of acute myocardial infarction, evolution of thrombolytic use in acute stroke, and the demise of military antishock trousers for traumatic shock provide examples of the dynamic relationship between emergency medicine research and clinical practice. The purpose of this article is to provide an overview of common research limitations and flaws relevant to emergency medicine. We explain and provide published examples of problems related to external validity, experimenter bias, publication bias, straw man comparisons, incorporation bias, randomization, composite outcomes, clinical importance versus statistical significance, and disease-oriented versus patient-oriented outcomes.

For residents, familiarity with these concepts will allow for better interpretation of evidence-based lectures and improved understanding and participation during journal clubs. For those who learn best from review articles, textbooks, or other summaries of primary literature, awareness of these issues is needed to understand critiques raised by reviewers. For all readers, knowledge of these commonly encountered methodological problems will improve the emergency provider's ability to determine whether and how new scientific developments serve to inform clinical practice. By necessity, this article is only a starting point for learning about these subjects, many of which are complex. Readers are encouraged to refer to questions and answers from the *Annals of Emergency Medicine* Journal Club series. Available topics from this series are described in Appendix E1, available online at <http://www.annemergmed.com>.

### EXTERNAL VALIDITY

External validity is the generalizability of a study's conclusions beyond the specific sample examined. In some studies, the study patients will be very similar to relevant clinical populations; in other studies, the lack of generalizability will render the results useless to the reader. The applicability of study findings may differ among geographic regions or specialties, or even among providers with different skills and experience practicing in the same emergency department (ED). Common threats to external validity include studies examining restricted or atypical demographic groups, studies performed at different sites of care (eg, specialty clinic versus community ED versus academic ED), and studies conducted in countries or settings in which available resources differ from the setting in which the reader practices. Issues related to external validity are particularly important in emergency

medicine because the patient population is inherently heterogeneous in terms of demographics, acuity, comorbidities, and referral source.

A recent investigation into the need for lumbar puncture after head computed tomography (CT) for evaluation of subarachnoid hemorrhage illustrates the importance of considering external validity.<sup>1</sup> In this study, head CT within 5 days of the onset of headache was found to be 100% sensitive for subarachnoid hemorrhage. The authors used this finding to suggest that lumbar puncture is no longer necessary to conclusively rule out subarachnoid hemorrhage in patients presenting with thunderclap headache. However, the study sample consisted of patients referred to a specialty neurosurgical center for evaluation of known or possible subarachnoid hemorrhage. Patients with suspected subarachnoid hemorrhage referred to a neurosurgical center for evaluation are arguably a very different population than undifferentiated patients initially presenting to an ED. If, for example, referred patients with subarachnoid hemorrhage tend to have larger bleeding episodes than ED patients, referred patients would also be more likely to have abnormal CT scan results. In a neurosurgery referral center population, perhaps it is true that 100% of patients with subarachnoid hemorrhage will have a positive CT result and there is no additional value to performing a lumbar puncture. However, we should be cautious about applying this conclusion to ED patients with a sudden-onset headache.

Whereas external validity depends on whether study conclusions are generalizable to other populations, internal validity depends on whether study conclusions are valid within the study sample. There are many potential threats to internal validity; several are described below. When a study is designed, there is often a tradeoff between internal and external validity. A highly controlled study of a carefully specified patient group will tend to have good internal validity, but this can limit its external validity.

## EXPERIMENTER BIAS

Experimenter bias is introduced by study investigators because of the inability of investigators to be completely objective. Experimenter bias may be conscious or unconscious and may influence the study through choices about study design,<sup>2</sup> implementation, reporting and interpretation of results,<sup>3</sup> or even a decision to publish results at all.<sup>4</sup> Few investigations are initiated without preconceptions on the part of the investigators about possible outcomes. For clinical scientists, the desire for academic success and the associated desire to publish is a source of bias. Scientists may also be influenced by the desire to provide further evidence to support their previous work and to obtain or maintain grant funding. Investigators conducting industry-sponsored studies may be influenced by financial incentives. Financial conflicts of interest are rampant in both original research articles<sup>5</sup> and clinical guidelines.<sup>6</sup> Most journals require authors to disclose financial conflicts of interest. These disclosures do not eliminate the possibility of experimenter bias but can at least inform readers about the presence of external influences on decisions made by investigators. Unfortunately, the presence of experimenter bias is difficult to determine with certainty.

Nesiritide, recombinant brain-type natriuretic peptide, provides an example of possible experimenter bias. The Vasodilation in the Management of Acute Congestive Heart Failure (VMAC) study was a large randomized controlled trial comparing nesiritide with nitroglycerin for acutely decompensated congestive heart failure.<sup>7</sup> The study was funded by Scios Inc, the manufacturers of nesiritide. Additionally, the study was designed, implemented, and analyzed by a steering committee appointed by Scios. Multiple investigators, including the principal investigator, received consulting fees from Scios. In 2000, after the completion of the VMAC study but before peer-reviewed publication of

study results, Scios estimated that peak sales of nesiritide would reach \$200 to \$300 million per year. This was a strong incentive for Scios to ensure the success of its product, particularly given that total company revenues that year were just \$12.7 million.<sup>8</sup> The initial article detailing VMAC results endorsed wide use of the drug, stating “nesiritide ... is a useful addition to initial therapy of patients hospitalized with acutely decompensated CHF [congestive heart failure],”<sup>7</sup> despite data showing a trend toward increased mortality in the nesiritide group.<sup>9</sup>

## PUBLICATION BIAS

Interventional studies that show a positive effect from the intervention are more likely to be published than those that do not, resulting in a pro-intervention bias within the literature base.<sup>4</sup> Publication bias is especially important to keep in mind when meta-analyses are evaluated because it is more difficult to obtain data from unpublished trials than published trials when constructing a meta-analysis. Additionally, meta-analyses frequently consist of numerous small or methodologically challenged studies, types of studies that are likely to be published only if the results are positive. A funnel plot can show evidence of publication bias in a meta-analysis. The funnel is constructed by plotting risk ratio or another measure of treatment effect on the  $x$  axis and standard error or another measure of precision on the  $y$  axis. By convention, the  $y$  axis is arranged so that larger, more precise studies are at the top, with smaller imprecise studies at the bottom. The more precise studies should cluster tightly around the true effect value on the  $x$  axis because they will tend to provide more accurate estimates of the effect. Less precise studies will have a wide variety of effect estimates and will not cluster as tightly, creating the funnel shape. When a subset of these smaller, imprecise studies is not published, the result is a skewed distribution of effect estimates among the smaller studies, creating asymmetry in the funnel plot. Funnel plots lack both sensitivity and specificity for diagnosing publication bias because they are often limited by small numbers of available studies, systematic differences in study methodology among studies of varying size, and subjectivity in interpretation.<sup>10</sup>

Evidence of publication bias is found in studies of intravenous magnesium to decrease mortality in the setting of acute myocardial infarction.<sup>11</sup> Two meta-analyses, based on relatively small studies, suggested that magnesium use conferred a survival benefit in this setting.<sup>12,13</sup> A large randomized controlled trial involving more than 58,000 patients contradicted these findings by showing no benefit from magnesium.<sup>14</sup> A funnel plot shows evidence of publication bias present in the studies used for the initial meta-analyses (Figure).<sup>11</sup>

## STRAW MAN COMPARISON

A comparison between an experimental group and a control group that receives substandard therapy is a straw man comparison. Unfortunately, this technique is frequently used to provide support for novel therapeutic agents or new applications of existing agents. It is easier to find evidence of equivalence or superiority of a novel intervention when the control group receives inadequate therapy. Relevant considerations include the choice of control-group therapy, as well as dose and frequency.

Straw man comparators were used in 2 manufacturer-sponsored studies comparing lumiracoxib, a selective Cox-2 inhibitor, and ibuprofen.<sup>15,16</sup> The first of these claimed to test the analgesic efficacy of lumiracoxib. It compared the maximum daily dose of lumiracoxib with a single dose of 400 mg of ibuprofen for pain control after tooth extraction.<sup>15</sup> The study concluded that lumiracoxib was superior to ibuprofen. However, one sixth of the standard 24-hour dose of ibuprofen was used as the comparator. A second study involving these 2 drugs, also sponsored by Novartis, compared complications related to

gastrointestinal ulcers between treatment groups.<sup>16</sup> Patients older than 50 years and with osteoarthritis received the maximum daily dose of lumiracoxib or 800 mg ibuprofen 3 times per day for 1 year. They concluded that the risk of complications from upper gastrointestinal tract ulcers was lower in the lumiracoxib group. This ibuprofen dose is much higher than the dose used when the primary outcome was pain control. In both of these cases, dosing in the active control (ibuprofen) group was designed to magnify any advantage of lumiracoxib compared with ibuprofen.

The use of straw man comparators is always misleading, at times unethical, and in some instances illegal. An extreme example of this occurred in Nigeria during a 1996 meningitis epidemic. Pfizer compared trovafloxacin, a novel fluoroquinolone, with ceftriaxone in children with bacterial meningitis. Although the results of this study have not been published in a peer-reviewed journal, details were documented by the *Washington Post*.<sup>17</sup> Their investigation revealed that most of the children in the ceftriaxone group received intramuscular doses of 33 mg/kg per day, far below the recommended intravenous dosing of 100 mg/kg per day for bacterial meningitis. A statement from Pfizer justified this discrepancy by stating that the change in dosing was done “to diminish the significant pain resulting from the intramuscular injection.”<sup>18</sup> Conveniently for Pfizer, this decision also increased the likelihood that trovafloxacin would outperform ceftriaxone. In 2009, Pfizer reached a financial settlement to compensate families involved in the study.<sup>19</sup>

## INCORPORATION BIAS

Incorporation bias occurs in studies of diagnostic tests in which one of the tests being examined and the criterion standard (aka reference standard) are not independent. This commonly occurs when 2 diagnostic tests are compared and one of these tests is also used to define the “correct” answer. The test that is used to define the reference standard will always outperform the other test.

Incorporation bias is present in a retrospective chart review of blunt trauma victims who had both cervical CT and radiography to evaluate for cervical spine fracture.<sup>20</sup> The authors reported that CT was 100% sensitive for the detection of these injuries in the included cohort, whereas radiography was 65% sensitive. However, CT results were used as the criterion standard. This approach ignores the possibility that some of the injuries observed on CT were artifacts or even that the CT missed fractures observed on radiographs. The study used circular reasoning: the authors used CT as the criterion standard because they believed it was superior to radiography, and they then used their results to “prove” this claim. As designed, it was inevitable that the authors would reach the conclusion that “[c]omputed tomography of the cervical spine should replace cervical spine radiographs ....”<sup>20</sup>

## UNSUCCESSFUL RANDOMIZATION

Randomization is the process by which study participants are allocated to particular treatment groups by a random process such as a coin flip or computerized random-number generator. The purpose of randomization is to ensure that aside from the experimental intervention, study groups are as similar as possible. If randomization is unsuccessful or is performed poorly, study results may be due to the presence of differences between the 2 groups other than the intervention, which is more likely to happen in small studies than in large studies. Stratified randomization helps prevent this by balancing rates of key characteristics among study groups. In addition to comparing the characteristics of the treatment and control groups provided by study authors, one should also consider whether the groups might differ in ways not described in the article.

The European Cooperative Acute Stroke Study III is a recently published randomized trial of thrombolytic therapy (alteplase) versus placebo administered between 3.0 and 4.5 hours after the onset of stroke symptoms.<sup>21</sup> The mean pretreatment National Institutes of Health Stroke Scale (NIHSS) score in the alteplase group was 10.7 compared with 11.6 in the placebo group, suggesting that baseline stroke severity was modestly worse in the placebo group. Additionally, 7.7% of patients in the alteplase group and 14.1% in the placebo group had a history of stroke, a factor that has been associated with poor functional outcomes. More significant is that only 5% of patients in the alteplase group had initial NIHSS scores greater than 20 compared with 10% of patients in the placebo group.<sup>22</sup> These figures were not reported in the initial European Cooperative Acute Stroke Study III article; however, they are of critical importance. Compared with other stroke patients, those with NIHSS scores above 20 have very poor outcomes regardless of any intervention, and patients with scores below 10 usually do well regardless of any intervention.<sup>23</sup> Together, these baseline imbalances in stroke severity between study groups potentially account for the entire 7.2% absolute difference in favorable functional outcome reported in this study. Randomization can be performed in ways that limit this problem. For example, a stratified randomization scheme using NIHSS scores to account for stroke severity could have been used to balance the allocation of patients with mild, moderate, and severe strokes among the alteplase and placebo groups.

## COMPOSITE OUTCOMES

Composite outcomes are created by combining multiple endpoints into a single common outcome measure. Investigators sometimes use this technique because finding a statistically significant difference between outcomes is easier when outcomes are common. Composite outcomes can make the interpretation of results difficult because the individual factors making up a given composite outcome are frequently not equivalent, clouding the clinical significance of the composite measure. The problem of composite endpoints can be at least partially addressed by comparing the frequency of the individual components if these results are provided by the authors. Others have recommended specific questions to ask when assessing the usefulness of composite outcomes (Table).<sup>24</sup>

The Clopidogrel in Unstable Angina to Prevent Recurrent Events Trial (CURE) was an industry-sponsored study comparing clopidogrel with placebo in patients with non-ST-elevation myocardial infarction.<sup>25</sup> The primary outcome was a composite of death from a cardiovascular cause, stroke, or recurrent nonfatal myocardial infarction. Recurrent myocardial infarction was defined as an increase in cardiac biomarker levels to at least double the upper limit of normal, whether or not this increase was considered clinically significant. The difference in composite outcome was largely driven by increased recurrent myocardial infarction in the placebo group (6.7% versus 5.2%) rather than by a nonsignificant absolute reduction in death of 0.4% in the treatment group. In fact, recurrent myocardial infarction was the only component of the composite outcome that reached individual statistical significance. Reporting that “clopidogrel significantly reduces the risk of the composite outcome of death from cardiovascular causes, nonfatal myocardial infarction, or stroke”<sup>25</sup> is more likely to promote clopidogrel use than stating “clopidogrel significantly reduces the risk of an asymptomatic rise in cardiac biomarkers, without a significant reduction in mortality or recurrent stroke.”

## CLINICAL IMPORTANCE VERSUS STATISTICAL SIGNIFICANCE

Statistical significance is reached when study results justify the rejection of the study's null hypothesis with a predetermined level of probability. Traditionally, but somewhat arbitrarily, if the probability of obtaining an observed result or one more extreme by chance

alone is less than 5% (ie,  $P < .05$ ), then a result is judged statistically significant. The clinical importance of a finding is influenced by the importance of the outcome measured, the magnitude of differences observed, and the existence of alternative therapies. Findings that are clinically important but not statistically significant ought to be repeated with a larger sample; findings that are statistically significant but not clinically important are irrelevant. A growing viewpoint in the research community, termed Bayesian inference, represents an alternative to frequentist statistics. Rather than focusing on whether a particular finding is mathematically significant, a Bayesian approach seeks to determine the truth of a particular finding by incorporating experimentally observed evidence together with the prior probability of a given hypothesis based on previous research.<sup>26,27</sup>

The use of dexamethasone in patients with migraine headaches provides an example of possible discordance between clinical importance and statistical significance. A randomized controlled trial was performed in which 130 patients with migraine headache received 15 mg dexamethasone or placebo.<sup>28</sup> Within the 48 hours after ED discharge, 32% of patients in the placebo arm developed a severe recurrent headache compared with 22% of patients in the dexamethasone group. The odds ratio in favor of dexamethasone was 0.6, with a 95% confidence interval of 0.3 to 1.3. Although this result is not statistically significant, a 10% difference in headache relapse is clinically important and justifies further investigation. A subsequent, adequately powered meta-analysis demonstrated a statistically significant improvement in relapse rate of similar magnitude.<sup>29</sup>

## DISEASE-ORIENTED (INTERMEDIATE) VERSUS PATIENT-ORIENTED OUTCOMES

Clinical decisions should be based on outcomes that matter to patients. Disease-oriented outcomes describe the presence or severity of a particular disease. Examples of disease-oriented outcomes include laboratory values, ECG findings, and abnormalities on imaging studies. Studies using patient-oriented outcomes, on the other hand, measure survival or some aspect of quality of life such as disability, pain, out-of-pocket costs, time in a hospital, or time away from work. Disease-oriented outcomes are often favored in clinical research because they are easier to measure. Examining disease-oriented outcomes may be a necessary step in understanding pharmacology or pathophysiology, but ideally, clinical decisions should be supported by studies demonstrating an effect on patient-oriented outcomes.

Controversy has surrounded the association between etomidate use and adrenal insufficiency. A recent randomized controlled trial compared etomidate with ketamine for use in rapid sequence intubation.<sup>30</sup> Etomidate was associated with a significant increase in adrenal insufficiency as defined by failure to respond to a cortisol stimulation test, or a random cortisol level below 276 nmol/L measured within 48 hours after tracheal intubation. Clearly, an individual patient will not know or care whether his or her serum cortisol level is slightly above or below 276 nmol/L. As addressed by the authors, this study did not demonstrate any significant difference in mortality or mechanical ventilation-free days, both of which are patient-oriented outcomes of greater clinical importance than serum cortisol test results.

Military antishock trousers provide a more dramatic example of how disease-oriented outcomes and patient-oriented outcomes can be at odds. This device, which encases the lower half of the body in high-pressure air bladders, became a staple of emergency medical services care for the treatment of hypovolemic shock in the United States during the 1970s.<sup>31</sup> Their use was justified because they increased blood pressure for hypotensive trauma patients.<sup>32</sup> Hypotension, however, is a disease-oriented outcome that is important



only because of its assumed relationship to mortality or other patient-oriented outcomes. In fact, a subsequent meta-analysis suggested that military antishock trousers use was associated with increased mortality.<sup>33</sup> The decision to use the trousers according to their effects on blood pressure cost money and probably cost lives.

## CONCLUSIONS

These examples are intended to illustrate limitations commonly encountered in research relevant to the practice of emergency medicine. No study is perfect. Even studies that include significant biases can be valuable if interpreted with an awareness of relevant limitations. Ideally, the discussion section of an original research article will both identify these limitations and include a Bayesian interpretation of the findings, in which the meaning of the new results is added to existing literature on the subject to obtain a poststudy result. If this is not done by the study authors, this work is left to the reader and to those who write article summaries and reviews. This method of integrating new data with existing information is particularly important in light of the constantly expanding medical knowledge base. A recent analysis of interventional studies published in top medical journals between 1990 and 2003 found that by 2005, a third had been contradicted or faced significant challenges about the magnitude of initially reported effect sizes.<sup>34</sup> These findings highlight the need for clinicians to question both new and established research results (Table). Ultimately, clinicians, not researchers, make decisions that affect patients. Now more than ever, good clinical care should be rooted in the process of judging new research according to its scientific merit, considering the relevance of new data to one's own patient population, and ensuring that the outcomes we care about are the ones that matter for our patients.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

*Funding and support:* By *Annals* policy, all authors are required to disclose any and all commercial, financial, and other relationships in any way related to the subject of this article as per ICMJE conflict of interest guidelines (see [www.icmje.org](http://www.icmje.org)). The authors have stated that no such relationships exist.

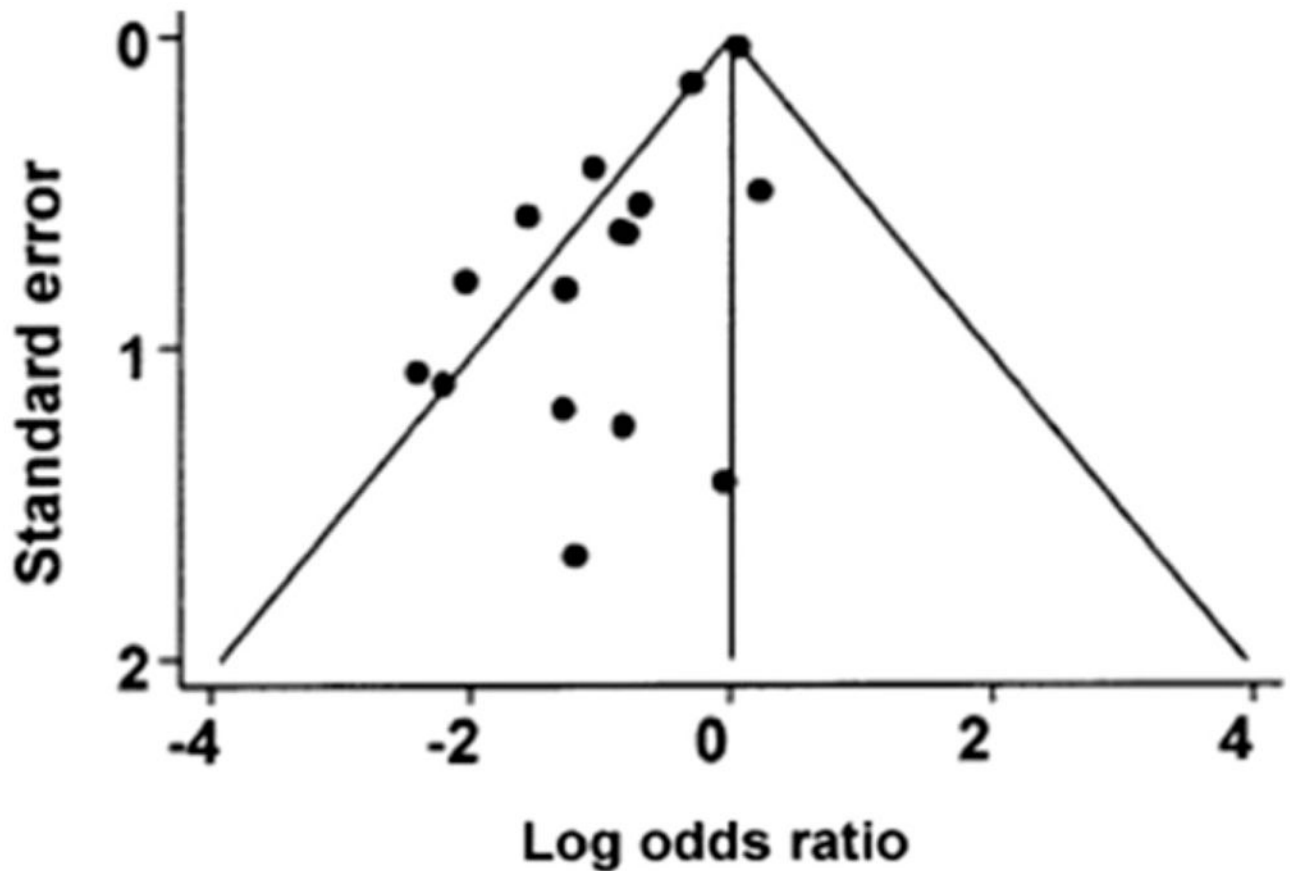
## References

1. Cortnum S, Sorensen P, Jorgensen J. Determining the sensitivity of computed tomography scanning in early detection of subarachnoid hemorrhage. *Neurosurgery*. 2010; 66:900–902. discussion 903. [PubMed: 20404693]
2. Heres S, Davis J, Maino K, et al. Why olanzapine beats risperidone, risperidone beats quetiapine, and quetiapine beats olanzapine: an exploratory analysis of head-to-head comparison studies of second-generation antipsychotics. *Am J Psychiatry*. 2006; 163:185–194. [PubMed: 16449469]
3. Als-Nielsen B, Chen W, Gluud C, et al. Association of funding and conclusions in randomized drug trials: a reflection of treatment effect or adverse events? *JAMA*. 2003; 290:921–928. [PubMed: 12928469]
4. Ramsey S, Scoggins J. Commentary: practicing on the tip of an information iceberg? evidence of underpublication of registered clinical trials in oncology. *Oncologist*. 2008; 13:925–929. [PubMed: 18794216]
5. Patsopoulos NA, Ioannidis JP, Analatos AA. Origin and funding of the most frequently cited papers in medicine: database analysis. *BMJ*. 2006; 332:1061–1064. [PubMed: 16547014]
6. Choudhry NK, Stelfox HT, Detsky AS. Relationships between authors of clinical practice guidelines and the pharmaceutical industry. *JAMA*. 2002; 287:612–617. [PubMed: 11829700]

7. Publication Committee for the VMAC Investigators (Vasodilatation in the Management of Acute CHF). Intravenous nesiritide vs nitroglycerin for treatment of decompensated congestive heart failure: a randomized controlled trial. *JAMA*. 2002; 287:1531–1540. [PubMed: 11911755]
8. Newswire. [Accessed March 22, 2011.] PR Scios reports fourth quarter and year end 2000 financial results. 2001. Available at: <http://www.highbeam.com/doc/1G1-71068400.html>
9. Noviasky JA, Kelberman M, Whalen KM, et al. Science or fiction: use of nesiritide as a first-line agent? *Pharmacotherapy*. 2003; 23:1081–1083. [PubMed: 12921256]
10. Lau J, Ioannidis JP, Terrin N, et al. The case of the misleading funnel plot. *BMJ*. 2006; 333:597–600. [PubMed: 16974018]
11. Sterne JA, Egger M. Funnel plots for detecting bias in meta-analysis: guidelines on choice of axis. *J Clin Epidemiol*. 2001; 54:1046–1055. [PubMed: 11576817]
12. Horner SM. Efficacy of intravenous magnesium in acute myocardial infarction in reducing arrhythmias and mortality. Meta-analysis of magnesium in acute myocardial infarction. *Circulation*. 1992; 86:774–779. [PubMed: 1387591]
13. Teo KK, Yusuf S, Collins R, et al. Effects of intravenous magnesium in suspected acute myocardial infarction: overview of randomised trials. *BMJ*. 1991; 303:1499–1503. [PubMed: 1838289]
14. 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; 345:669–685. [PubMed: 7661937]
15. Zelenakas K, Fricke JR Jr, Jayawardene S, et al. Analgesic efficacy of single oral doses of lumiracoxib and ibuprofen in patients with postoperative dental pain. *Int J Clin Pract*. 2004; 58:251–256. [PubMed: 15117091]
16. Schnitzer TJ, Burmester GR, Mysler E, et al. Comparison of lumiracoxib with naproxen and ibuprofen in the Therapeutic Arthritis Research and Gastrointestinal Event Trial (TARGET), reduction in ulcer complications: randomised controlled trial. *Lancet*. 2004; 364:665–674. [PubMed: 15325831]
17. Stephens J. The body hunters: where profits and lives hang in balance. *Washington Post*. Dec 17.2000 :A01.
18. Pfizer, Inc. [Accessed March 22, 2011.] Trovan fact sheet. Available at: [http://www.pfizer.com/files/news/trovan\\_fact\\_sheet\\_final.pdf](http://www.pfizer.com/files/news/trovan_fact_sheet_final.pdf)
19. Reuters. [Accessed March 22, 2011.] Pfizer to pay \$75 mln in Trovan settlement-source. Available at: <http://www.reuters.com/article/2009/07/27/pfizer-trovan-idUSWEN137320090727>
20. Griffen MM, Frykberg ER, Kerwin AJ, et al. Radiographic clearance of blunt cervical spine injury: plain radiograph or computed tomography scan? *J Trauma*. 2003; 55:222–226. discussion 226-227. [PubMed: 12913629]
21. Hacke W, Kaste M, Bluhmki E, et al. Thrombolysis with alteplase 3 to 4.5 hours after acute ischemic stroke. *N Engl J Med*. 2008; 359:1317–1329. [PubMed: 18815396]
22. Bluhmki E, Chamorro A, Davalos A, et al. Stroke treatment with alteplase given 3.0–4.5 h after onset of acute ischaemic stroke (ECASS III): additional outcomes and subgroup analysis of a randomised controlled trial. *Lancet Neurol*. 2009; 8:1095–1102. [PubMed: 19850525]
23. Kwiatkowski TG, Libman RB, Frankel M, et al. Effects of tissue plasminogen activator for acute ischemic stroke at one year. National Institute of Neurological Disorders and Stroke Recombinant Tissue Plasminogen Activator Stroke Study Group. *N Engl J Med*. 1999; 340:1781–1787. [PubMed: 10362821]
24. Montori VM, Permyer-Miralda G, Ferreira-Gonzalez I, et al. Validity of composite end points in clinical trials. *BMJ*. 2005; 330:594–596. [PubMed: 15761002]
25. Yusuf S, Zhao F, Mehta SR, et al. Effects of clopidogrel in addition to aspirin in patients with acute coronary syndromes without ST-segment elevation. *N Engl J Med*. 2001; 345:494–502. [PubMed: 11519503]
26. Brown AM, Schriger DL, Barrett TW. *Annals of Emergency Medicine Journal Club*. Outcome measures, interim analyses, and bayesian approaches to randomized trials: answers to the



- September 2009 Journal Club questions. *Ann Emerg Med.* 2010; 55:216–224. e211. [PubMed: 20116032]
27. Goodman SN. Toward evidence-based medical statistics. 1: The P value fallacy. *Ann Intern Med.* 1999; 130:995–1004. [PubMed: 10383371]
  28. Rowe BH, Colman I, Edmonds ML, et al. Randomized controlled trial of intravenous dexamethasone to prevent relapse in acute migraine headache. *Headache.* 2008; 48:333–340. [PubMed: 18047499]
  29. Singh A, Alter HJ, Zaia B. Does the addition of dexamethasone to standard therapy for acute migraine headache decrease the incidence of recurrent headache for patients treated in the emergency department? a meta-analysis and systematic review of the literature. *Acad Emerg Med.* 2008; 15:1223–1233. [PubMed: 18976336]
  30. Jabre P, Combes X, Lapostolle F, et al. Etomidate versus ketamine for rapid sequence intubation in acutely ill patients: a multicentre randomised controlled trial. *Lancet.* 2009; 374:293–300. [PubMed: 19573904]
  31. Kaplan BC, Civetta JM, Nagel EL, et al. The military anti-shock trouser in civilian pre-hospital emergency care. *J Trauma.* 1973; 13:843–848. [PubMed: 4743141]
  32. McSwain NE Jr. Mast pneumatic trousers: a mechanical device to support blood pressure. *Med Instrum.* 1977; 11:334–336. [PubMed: 600127]
  33. Dickinson K, Roberts I. Medical anti-shock trousers (pneumatic anti-shock garments) for circulatory support in patients with trauma. *Cochrane Database Syst Rev.* 2000; (2):CD001856. [PubMed: 10796828]
  34. Ioannidis JP. Contradicted and initially stronger effects in highly cited clinical research. *JAMA.* 2005; 294:218–228. [PubMed: 16014596]



**Figure.**

Funnel plot for magnesium trials. Dots represent treatment effects from individual trials. The diagonal lines represent the expected 95% confidence intervals surrounding the pooled treatment effect. The absence of trials in the lower right corner of the figure suggests publication bias. Adapted from Sterne JA, Egger M. Funnel plots for detecting bias in meta-analysis: guidelines on choice of axis. *J Clin Epidemiol.* 2001;54: 1048. Figure 1 used with permission from Elsevier.

**Table**

Questions used to assess for limitations commonly encountered in clinical research.\*

Topic	Study Type	Important Questions
External validity	All types	Are there likely to be systematic differences between the study setting and my own clinical setting about factors related to the patient population, disease process, and available clinical resources? Are these differences likely to be clinically important?
Experimenter bias	All types, especially industry-sponsored trials	What motivated the investigators and study sponsor to perform the study? Who was the study funded by? Did they have control over study data?
Publication bias	Meta-analysis, pooled data analysis	Was an effort made to include data from unpublished trials? Was it successful? Are the articles included of sufficient quality and sufficiently representative of existing knowledge to adequately capture our state of knowledge on this topic?
Straw man comparison	Randomized controlled trials with active comparator	Did the control group receive appropriate treatment or an appropriate diagnostic test? Was the dosing amount and frequency equivalent between groups? Does the standard treatment have advantages over the experimental treatment that are not accounted for by the study (cost, availability, ease of use, etc)?
Incorporation bias	Evaluation of diagnostic tests	Are the reference standard and the modalities being studied interdependent? Were the clinicians or investigators responsible for judging the reference standard blinded to results from the methods being studied?
Randomization	Randomized controlled trials	How was randomization performed? Is the sample size sufficient to make successful randomization likely? Do groups seem sufficiently similar with respect to important characteristics to make confounding unlikely?
Composite outcome	Randomized controlled trials, cohort studies	Are the component endpoints of similar importance to patients? If not, what was the difference in frequency of the most important endpoint? <sup>24</sup>
Clinical importance versus statistical significance	All types	Is each statistically significant result also clinically important or useful? Are there clinically important results that were not statistically significant? How should the results (statistically significant or not) change previously held beliefs on this topic?
Disease-oriented versus patient-oriented outcomes	All types	What patient-oriented outcomes would be relevant to this study? Were each of these measured, and did these differ between study groups? How predictive of a patient-oriented outcome are the disease-oriented outcomes that the authors chose to measure?

\* More detailed explanations of these and related topics can be found in the *Annals of Emergency Medicine* Journal Club archives, available at <http://www.annemerg-med.com/content/journalclub>.