

We thank Marc Mitchell, Megan Murray, and Joshua Salomon for valuable comments and references. The findings, interpretations, and conclusions expressed here are those of the authors and do not necessarily reflect the views of the board of executive directors of the World Bank or the governments they represent.

Contributors and sources: EB and ME conduct research on and implement health interventions in developing countries, with focus on cross sectoral programmes and policies. This paper originated from their field experiences in poor rural areas.

Funding: ME was supported by a grant from National Institute on Aging (PO1-AG17625).

Competing interests: None declared.

- 1 World Health Organization. *World health report 2002: reducing risks, promoting healthy life*. Geneva: WHO, 2002.
- 2 Ezzati M, Lopez AD, Rodgers A, Vander Hoorn S, Murray CJL, Comparative Risk Assessment Collaborative Group. Selected major risk factors and global and regional burden of disease. *Lancet* 2002;360:1347-60.
- 3 Frieden TR, Sterling TR, Munsiff SS, Watt CJ, Dye C. Tuberculosis. *Lancet* 2003;362:887-99.
- 4 Gajalakshmi V, Peto R, Kanaka TS, Jha P. Smoking and mortality from tuberculosis and other diseases in India: retrospective study of 43 000 adult male deaths and 35 000 controls. *Lancet* 2003;362:507-15.
- 5 Kolappan C, Gopi PG. Tobacco smoking and pulmonary tuberculosis. *Thorax* 2002;57:964-6.
- 6 Lam TH, Ho SY, Hedley AJ, Mak KH, Peto R. Mortality and smoking in Hong Kong: case-control study of all adult deaths in 1998. *BMJ* 2001;323:361-3.
- 7 Leung CC, Yew WW, Chan CK, Tam CM, Lam CW, Chang KC, et al. Smoking and tuberculosis in Hong Kong. *Int J Tuberc Lung Dis* 2003;7:980-6.

- 8 Liu BQ, Peto R, Chen ZM, Boreham J, Wu YP, Li JY, et al. Emerging tobacco hazards in China: 1. Retrospective proportional mortality study of one million deaths. *BMJ* 1998;317:1411-22.
- 9 Mishra VK, Retherford RD, Smith KR. Biomass cooking fuels and prevalence of tuberculosis in India. *Int J Infect Dis* 1999;3:119-29.
- 10 Perez-Padilla R, Perez-Guzman C, Baez-Saldana R, Torres-Cruz A. Cooking with biomass stoves and tuberculosis: a case control study. *Int J Tuberc Lung Dis* 2001;5:1-7.
- 11 Thomas P, Zelkoff J. Air pollutants: moderators of pulmonary host resistance against infection. In: Holgate ST, Samet JM, Koren HS, Maynard RL, eds. *Air pollution and health*. San Diego, CA: Academic Press, 1999.
- 12 Dye C, Watt CJ, Bleed DM. Low access to a highly effective therapy: a challenge for international tuberculosis control. *Bull World Health Organ* 2002;80:437-44.
- 13 Raviglione MC, Pio A. Evolution of WHO policies for tuberculosis control, 1948-2001. *Lancet* 2002;359:775-80.
- 14 Dye C, Watt CJ, Bleed DM, Williams BG. What is the limit to case detection under the DOTS strategy for tuberculosis control? *Tuberculosis* 2003;83:35-43.
- 15 Squire SB, Wilkinson D. Strengthening "DOTS" through community care for tuberculosis. *BMJ* 1997;315:1395-6.
- 16 China Tuberculosis Control Collaboration. The effect of tuberculosis control in China. *Lancet* 2004;364:417-22.
- 17 Squire SB, Tang S. How much of China's success in tuberculosis control is really due to DOTS? *Lancet* 2004;364:391-2.
- 18 Florig HK. China's air pollution risks. *Environ Sci Tech* 1997;31:274-9A.
- 19 Smith KR, Shuhua G, Kun H, Daxiong Q. One hundred million improved cookstoves in China: how was it done? *World Development* 1993;21:941-61.
- 20 Raviglione MC, Dye C, Schmidt S, Kochi A. Assessment of worldwide tuberculosis control. WHO global surveillance and monitoring project. *Lancet* 1997;350:624-9.

(Accepted 6 September 2004)

Users' guide to detecting misleading claims in clinical research reports

Victor M Montori, Roman Jaeschke, Holger J Schünemann, Mohit Bhandari, Jan L Brozek, P J Devereaux, Gordon H Guyatt

Plenty of advice is available to help readers identify studies with weak methods, but would you be able to identify misleading claims in a report of a well conducted study?

Science is often not objective.¹ Emotional investment in particular ideas and personal interest in academic success may lead investigators to overemphasise the importance of their findings and the quality of their work. Even more serious conflicts arise when for-profit organisations, including pharmaceutical companies, provide funds for research and consulting, conduct data management and analyses, and write reports on behalf of the investigators.

Although guides to help recognise methodological weaknesses that may introduce bias are now widely available,²⁻³ these criteria do not protect readers against misleading interpretations of methodologically sound

studies. In this article, we present a guide that provides clinicians with tools to defend against biased inferences from research studies (box).

Read methods and results only

The discussion section of research reports often offers inferences that differ from those a dispassionate reader would draw from the methods and results.⁴ The table gives details of two systematic reviews summarising a similar set of randomised trials assessing the effect of albumin for fluid resuscitation. The trials included in both reviews were small and methodologically weak, and their results are heterogeneous. Both the reviews provide point estimates suggesting that albumin may increase mortality and confidence intervals that include the possibility of a considerable increase in mortality. Nevertheless, one set of authors took a strong position that albumin is dangerous, the other that it is not. Their positions were consistent with the interests of funders of their reviews.⁵

This is not an idiosyncratic example. Systematic examinations of the relation between funding and

Department of Medicine, Mayo Clinic College of Medicine, Rochester MN, USA

Victor M Montori
assistant professor

Department of Medicine, McMaster University, Hamilton ON, Canada

Roman Jaeschke
clinical professor of medicine

continued over

BMJ 2004;329:1093-6

Guide to avoid being misled by biased presentation and interpretation of data

1. Read only the Methods and Results sections; bypass the Discussion section
2. Read the abstract reported in evidence based secondary publications
3. Beware faulty comparators
4. Beware composite endpoints
5. Beware small treatment effects
6. Beware subgroup analyses



Illustrative examples and references w1-w15 are on bmj.com

Department of Clinical Epidemiology and Biostatistics, McMaster University
 Holger J Schünemann *associate professor of medicine*
 Mohit Bhandari *assistant professor of medicine*
 P J Devereaux *assistant professor of medicine*
 Gordon H Guyatt *professor of medicine*
 Polish Institute for Evidence Based Medicine, Jagiellonian University Medical School, Krakow, Poland
 Jan L Brozek *lecturer in medicine*
 Correspondence to: G H Guyatt guyatt@mcmaster.ca



conclusions have found that the odds of recommending an experimental drug as treatment of choice increases fivefold with for-profit organisation funding (odds ratio 5.3, 95% confidence interval 2.0 to 14.4) compared with not-for-profit funding.⁶

If editors insisted that discussion sections of original articles included a systematic review of the relevant literature, this first pointer would no longer be relevant. However, few original trial reports include systematic reviews,⁷ and this may not change in the foreseeable future.

To follow this advice, readers must be able to make sense of the methods and results. Fortunately, clinicians can access many educational materials to acquire skills in interpreting studies' designs and their findings.²⁻³

Read the abstract reported in pre-appraised resources

Secondary journals, such as the *ACP Journal Club*, *Evidence-Based Medicine*, and *Evidence-Based Mental Health*, publish structured abstracts produced by a team of clinicians and methodologists. These abstracts often include critical information about the research methods (allocation concealment, blinding, completeness of follow up) omitted from the original papers.⁸ The conclusions of this "secondary" abstract are the product of critical appraisal by people without competing financial or personal interests.

The objectivity and methodological sophistication of those preparing the independent structured abstracts will often provide additional value for clinicians. For example, substantial discrepancies occur between the full report of the PROGRESS trial and the abstract in *ACP Journal Club*.^{w3 w4} The title of the

original publication describes the study as testing "A perindopril-based blood pressure lowering regimen" and reports that the perindopril regimen resulted in a 28% relative risk reduction in the risk of recurrent stroke (95% confidence interval 17-38%).^{w3} The *ACP Journal Club* abstract and its commentary identified the publication as describing two parallel but separate randomised placebo controlled trials including about 6100 patients with a history of stroke or transient ischaemic attack:

- In one trial, patients were randomised to receive perindopril or placebo, and active treatment had no appreciable effect on stroke (relative risk reduction = 5%, 95% confidence interval - 19% to 23%).
- In the second trial, patients were allocated to receive perindopril plus indapamide or double placebo. Combined treatment resulted in a 43% relative risk reduction (30% to 54%) in recurrent stroke.

The *ACP Journal Club* commentary notes that the authors disagree with the interpretation of the publication as reporting two separate trials (which explains why it is difficult for even the knowledgeable reader to get a clear picture of the design from the original publication).

Beware faulty comparators

Several systematic reviews have shown that industry funded studies typically yield larger treatment effects than not-for-profit funded studies.⁹⁻¹⁰ One likely explanation is choice of comparators.¹¹ Researchers with an interest in a positive result may choose a placebo comparator rather than an alternative drug with proved effectiveness. For instance, in a study of 136 trials of new treatments for multiple myeloma, 60% of studies funded by for-profit organisations, but only 21% of trials funded by not-for-profit organisations, compared their new interventions against placebo or no treatment.¹² Box A on bmj.com gives other examples.

When reading reports of randomised trials, clinicians should ask themselves: "Should the comparator have been another active agent rather than placebo; if investigators chose an active comparator, was the dose, formulation, and administration regimen optimal?"

Beware composite end points

Investigators often use composite end points to enhance the statistical efficiency of clinical trials. Problems in the interpretation of these trials arise when composite end points include component outcomes to which patients attribute very different importance—for example, the primary outcome in a trial of irbesartan v

Comparison of two systematic reviews of albumin for fluid resuscitation

| | Wilkes, 2001w1 | Alderson, 2002w2 |
|---|--|--|
| Funding body | Plasma Proteins Therapeutic Association | NHS |
| Trials included | 42 short term trials reporting mortality | 31 short term trials reporting mortality |
| Overall difference in mortality between albumin and crystalloid (relative risk, 95% CI) | 1.11 (0.95 to 1.28) | 1.52 (1.17 to 1.99) |
| Difference in burns patients | 1.76 (0.97 to 3.17) | 2.4 (1.11 to 5.19) |
| Interests of funding body | Promotes access to and reimbursement for the use of albumin | Pays for use of albumin |
| Authors' conclusions | Results "should serve to allay concerns regarding the safety of albumin" | Recommend banning use of albumin outside rigorously conducted randomised controlled trials |

placebo for diabetic nephropathy was a composite end point including death, end stage renal disease, and doubling of serum creatinine concentration.¹³ Problems may also arise when the most important end point occurs infrequently or when the apparent effect on component end points differs. Thus, the reported effect on “important cardiovascular end points” may reflect mostly the effect of treatment on angina rather than on mortality and non-fatal myocardial infarction (see bmj.com for another example).

When the more important outcomes occur infrequently, clinicians should focus on individual outcomes rather than on composite end points. Under these circumstances, inferences about the effect of treatment on the more important end points (which, because they occur infrequently will have very wide confidence intervals) will be weak. In focusing on the more reliable estimates of effects on the less important outcomes, readers will see that, for instance, a putative reduction in death, myocardial infarction, and revascularisation is really just an effect on the frequency of revascularisation.¹⁴

Beware small treatment effects

Increasingly, investigators are conducting very large trials to detect small treatment effects. Results suggest small treatment effects when either the point estimate is close to no effect (a relative or absolute risk reduction close to 0; a relative risk or odds ratio close to 1) or the confidence interval includes values close to no effect. In one large trial, investigators randomly allocated just over 6000 participants to receive angiotensin converting enzyme inhibitors or diuretics for hypertension and concluded “initiation of antihypertensive treatment involving ACE inhibitors in older subjects . . . seems to lead to better outcomes than treatment with diuretic agents.”¹⁵ In absolute terms, the difference between the regimens was very small: there were 4.2 events per 100 patient years in the angiotensin converting enzyme group and 4.6 events per 100 patient years in the diuretic group. The relative risk reduction corresponding to this absolute difference (11%) had an associated 95% confidence interval of -1% to 21%.

Here, we have two reasons to doubt the importance of the apparent difference between the two types of antihypertensive drug. Firstly, the point estimate suggests a very small absolute difference (0.4 events per 100 patient years) and, secondly, the confidence interval suggests it may have been even smaller—indeed, there may have been no true difference at all.

When the absolute risk of adverse events in untreated patients is low, the presentation may focus on relative risk reduction and de-emphasise or ignore absolute risk reduction. Other techniques for making treatment effects seem large include misleading graphical representations,¹⁶ and using different time frames to present harms and benefits (see box C on bmj.com).

Beware subgroup analyses

Clinicians often wonder if a subgroup of patients will achieve larger benefits when exposed to a particular treatment. Because the play of chance can lead to apparent but spurious differences in effects between

Summary points

Many interests may influence researchers towards favourable interpretation and presentation of their findings

Guides helping clinicians to identify weak studies that are open to bias are available

Clinicians also need to be able to detect when authors of methodologically strong studies make misleading claims

Six pointers to help clinicians avoid being misled are presented

subgroups, clinicians should be cautious about reports of subgroup analyses.

In the trial of angiotensin converting enzyme inhibitor versus diuretic based antihypertensive treatment above, the complete conclusion reads “initiation of antihypertensive treatment involving ACE inhibitors in older subjects, particularly men, appears to lead to better outcomes than treatment with diuretic agents.”¹⁵ The possibility that effect differed by sex was not one of a small number of prior hypotheses, and the size of the difference in effect was small (relative risk reductions of 17% (95% confidence interval 3 to 29%) in men and 0% (-20% to 17%) in women).

The difference between the angiotensin converting enzyme inhibitors and diuretics was significant in men but not women. Investigators might use this fact to argue for differences in the effect between men and women. The real issue, however, is whether chance is a sufficient explanation for the difference in effect between women and men, and it is in this case. The P value associated with the null hypothesis that the underlying relative risk is identical in men and women is 0.15. Thus, if there were no true difference in effect between men and women, we would see apparent differences of 17%, or greater, 15% of the time. Overall, the inference that the effect differs in men satisfies only one of seven criteria for a valid subgroup analysis.¹⁷

Clinicians may sometimes encounter parallel examples when investigators discount effects of beneficial treatments because of apparent lack of effects in subgroups. Here, clinicians must ask themselves whether there is strong evidence that patients in a subgroup should be denied the beneficial treatment. The answer will usually be “no.” In general, we advise clinicians to be sceptical concerning claims of differential treatment effects in subgroups of patients.

Conclusion

We have presented six pointers to help clinicians protect themselves and their patients from potentially misleading presentations and interpretations of research findings. These strategies are unlikely to be foolproof. Decreasing the dependence of the research endeavour on for-profit funding, implementing a requirement for mandatory registration of clinical trials, and instituting more structured approaches to reviewing and reporting research^{18 19} may reduce biased reporting. At the same time, it is likely that potentially misleading reporting will always be with us, and the guide we have presented will help clinicians to stay armed.

Contributors and sources: The authors are clinical epidemiologists who are involved in critical review of medical and surgical studies as part of their educational and editorial activities. In that context, they have collected some of the examples presented here. Other examples were collected by searching titles of records in PubMed using the textwords "misleading," "biased," and "spin" using the *Related Articles* feature for every pertinent hit, and by noting relevant references from retrieved papers. VMM and GHG wrote the first draft of the article. All authors contributed to the ideas represented in the article, made critical contributions and revisions to the first draft, and approved the final version. GHG is guarantor.

Funding: VMM is a Mayo Foundation scholar. PJD is supported by a Canadian Institutes of Health Research senior research fellowship award. MB was supported by a Detweiler fellowship, Royal College of Physicians and Surgeons of Canada and is currently supported by a Canada Research Chair.

Competing interests: VMM, RJ, HJS, PJD, and GHG are associate editors of the *ACP Journal Club* and *Evidence-Based Medicine*. JLB and RJ edit an evidence based medicine journal in Poland. MB is editor of the evidence based orthopaedic trauma section in the *Journal of Orthopaedic Trauma*.

- 1 Horton R. The rhetoric of research. *BMJ* 1995;310:985-7.
- 2 Greenhalgh T. *How to read a paper*. London: BMJ Books, 2001.
- 3 Guyatt G, Rennie D, eds. *Users' guides to the medical literature. A manual for evidence-based clinical practice*. Chicago, IL: AMA Press, 2002.
- 4 Bero LA, Rennie D. Influences on the quality of published drug studies. *Int J Technol Assess Health Care* 1996;12:209-37.
- 5 Cook D, Guyatt G. Colloid use for fluid resuscitation: evidence and spin. *Ann Intern Med* 2001;135:205-8.
- 6 Als-Nielsen B, Chen W, Gluud C, Kjaergard LL. Association of funding and conclusions in randomized drug trials: a reflection of treatment effect or adverse events? *JAMA* 2003;290:921-8.
- 7 Clarke M, Alderson P, Chalmers I. Discussion sections in reports of controlled trials published in general medical journals. *JAMA* 2002;287:2799-801.

- 8 Devereaux PJ, Manns BJ, Ghali WA, Quan H, Guyatt GH. Reviewing the reviewers: the quality of reporting in three secondary journals. *CMAJ* 2001;164:1573-6.
- 9 Bekelman JE, Li Y, Gross CP. Scope and impact of financial conflicts of interest in biomedical research: a systematic review. *JAMA* 2003;289:454-65.
- 10 Lexchin J, Bero LA, Djulbegovic B, Clark O. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. *BMJ* 2003;326:1167-70.
- 11 Mann H, Djulbegovic B. Biases due to differences in the treatments selected for comparison (comparator bias). *James Lind Library*. www.jameslindlibrary.org/essays/bias/comparator_bias.html (accessed 2 Oct 2004).
- 12 Djulbegovic B, Lacey M, Cantor A, Fields KK, Bennett CL, Adams JR, et al. The uncertainty principle and industry-sponsored research. *Lancet* 2000;356:635-8.
- 13 Lewis EJ, Hunsicker LG, Clarke WR, Berl T, Pohl MA, Lewis JB, et al. Renoprotective effect of the angiotensin-receptor antagonist irbesartan in patients with nephropathy due to type 2 diabetes. *N Engl J Med* 2001;345:851-60.
- 14 Waksman R, Ajani AE, White RL, Chan RC, Satler LF, Kent KM, et al. Intravascular gamma radiation for in-stent restenosis in saphenous-vein bypass grafts. *N Engl J Med* 2002;346:1194-9.
- 15 Wing LM, Reid CM, Ryan P, Beilin LJ, Brown MA, Jennings GL, et al. A comparison of outcomes with angiotensin-converting-enzyme inhibitors and diuretics for hypertension in the elderly. *N Engl J Med* 2003;348:583-92.
- 16 Tufte E. *The visual display of quantitative information*. Cheshire, CT: Graphics Press, 1983.
- 17 Oxman A, Guyatt G, Green L, Craig J, Walter S, Cook D. When to believe a subgroup analysis. In: Guyatt G, Rennie D, eds. *Users' guides to the medical literature. A manual for evidence-based clinical practice*. Chicago, IL: AMA Press, 2002:553-65.
- 18 Docherty M, Smith R. The case for structuring the discussion of scientific papers. *BMJ* 1999;318:1224-5.
- 19 Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet* 2001;357:1191-4.

(Accepted 15 September 2004)

Monitoring global health: time for new solutions

Christopher J L Murray, Alan D Lopez, Suwit Wibulprasert

Improved global health monitoring requires new technologies and methods, strengthened national capacity, norms and standards, and gold standard global reporting. The World Health Organization's many functions limit its capacity for global reporting, and a new global health monitoring organisation is needed to provide independent gold standard health information to the world

Sound information on financial and human resources invested in health, health interventions delivered to people in need, and the impact of these efforts on people's health is critical for planning health systems, implementing programmes, epidemic response, allocating budgets for research and development, monitoring progress, and evaluating what works and what does not. Although all countries collect a wide range of health information through registries, surveys, and vital registration systems, huge gaps hinder our ability to respond to global health challenges, which are alarming at a time when global investments in AIDS, tuberculosis, and malaria are increasing and when there is renewed focus on health goals as exemplified by the UN Millennium Declaration.¹

The availability of valid, reliable, and comparable health information to inform local, regional, national, and global decisions can be furthered through four interconnected efforts: improving the technology and methods for population health measurement; strengthening national capacity and motivating governments to collect and analyse useful health data;

establishing global norms and standards for what are the core health related measurements and how to measure them; and reporting to the globe valid, reliable, and comparable assessments of inputs, service delivery, and achievements for health. Although many challenges and initiatives are under way for the first three of these components, the fourth area is currently the weakest and getting worse, not better. We explore the problem and the potential solution to global monitoring and evaluation and briefly review the other three areas as necessary.

Technology and methods of health measurement

For several critical measures of health, delivery of interventions, and financial and human resources including many millennium development goal indicators, the current measurement technology is inadequate. Although validated methods have been developed to measure child mortality through household surveys in settings where vital registration is incomplete or non-existent,²⁻⁴ no adequate method

Harvard Initiative for Global Health, Harvard University, 104 Mount Auburn Street, Cambridge, MA 02138, USA

Christopher J L Murray
director

School of Population Health, University of Queensland, Herston Road, Herston, Queensland 4006, Australia

Alan D Lopez
head of school

Office of Permanent Secretary, Ministry of Public Health, Tiwanond Road, Nonthaburi 11000, Thailand

Suwit Wibulprasert
senior adviser on health economics

Correspondence to: C J L Murray (christopher_murray@harvard.edu)

BMJ 2004;329:1096-100