

# What is the best evidence for determining harms of medical treatment?

Jan P. Vandembroucke

∞ See related article page 635

A proper balance of benefits and harms is necessary to assess the overall effect of medical interventions.<sup>1</sup> Most evidence on harms from medical treatments is obtained from observational research. Randomized controlled trials (RCTs) are often not useful in determining rates of adverse effects: the frequency of such events during RCTs may be low owing to restrictive inclusion and exclusion criteria; in addition, follow-up periods are relatively short, and the number of participants included in an RCT is limited. As a result, systematic reviews based on evidence from RCTs often fail to provide accurate data on adverse events. Evidence from nonrandomized studies on adverse effects is often dismissed, simply because the studies were not randomized; however, this philosophy should not be considered the best approach to practising evidence-based medicine.<sup>2</sup>

Observational studies on adverse effects provide data that are as valid as evidence from RCTs.<sup>3</sup> “Unintended effects” of treatment, such as adverse effects, are often unexpected and unpredictable and not linked to indications for treatment.<sup>4</sup> Therefore, there is little need for randomization to quantify them. To rule out “confounding by indication” as much as possible, studies can be limited to idiopathic adverse effects of a drug (i.e., the patient has no risk factor for the adverse effect that could have guided treatment).<sup>5</sup> In considering the risk of venous thrombosis from different oral contraceptives in healthy young women without any known risk factors for thrombosis, it is reasonable to assume that, because the choice of contraceptive cannot be linked to risk factors, any difference in rates of venous thrombosis is due to a difference in the contraceptive. Another classic example of an unexpected and unpredictable adverse effect is a drug-related rash.

Despite such theoretical arguments, and many practical examples, people who conduct systematic reviews, and those who read them, often wonder whether there is “empirical evidence” that data from observational studies are as good as data from RCTs. However, empirical comparisons are possible only in rather exceptional cases. One needs to find instances of adverse effects (to similar drugs used for similar indications) that would be equally good candidates for study in observational research as in randomized trials. To assess rare and late effects, studies should be very large and long-term.

In this issue (page 635), Papanikolaou and colleagues present their comparison of evidence on harms from medical interventions reported in randomized and nonrandomized studies.<sup>6</sup> The authors build on their earlier work that

showed that the reporting of adverse effects in randomized trials is often inadequate and needs strengthening.<sup>7</sup> In their current study, they examined specific harms of various medical interventions for which data were already available from systematic reviews of RCTs representing at least 4000 subjects and compared them with the same harms reported in nonrandomized studies that each had to have at least 4000 subjects. In most instances, the observational studies estimated smaller risks (absolute and relative) than the randomized trials.

## Data from observational studies cannot be dismissed simply because the study was not randomized.

The study by Papanikolaou and colleagues has limitations. They may have excluded some useful case-control studies, since these studies need fewer numbers for the same statistical efficiency. In addition, they included observational studies that had a wider array of indications and applications than those in the RCTs, and they accepted different measures of relative risk. Finally, most of the harms that the authors report were pharmacologically predictable (e.g., bleeding with anticoagulant or antiplatelet therapy, which could have been foreseen when the trial was designed and occurred early). Thus, their analysis has few data on adverse effects that were completely unpredictable (i.e., that were initially unknown for these therapies, that were very rare or that came late).

Whatever the limitations, the findings are reassuring. If anything, observational studies are more conservative than randomized trials. In only 2 instances did the observational studies estimate a clearly greater risk than did the respective randomized trials. Clearly, evidence on adverse effects derived from observational research cannot be dismissed simply because the study was not randomized. In fact, the actual harm may be even greater than that reported in an observational study.

Researchers involved in observational pharmacoepidemiology have always suspected this — it is tacit knowledge in the field. Papanikolaou and colleagues pinpoint the mecha-

nism: randomized trials have strict protocols for data collection and harm detection. In contrast, much observational research on adverse effects uses existing data from routine patient care. Unlike data from randomized trials, data collected from real patient encounters contain diverse inaccuracies, such as those that pertain to exposure to the imputed drug, reporting of the potential adverse effect and potential confounders. The cumulative effect of these inaccuracies dilutes the findings.

The 2 instances Papanikolaou and colleagues point out in which the observational studies estimated a clearly greater harm than did the respective randomized trials merit attention. In one case, a population registry of patients receiving anticoagulant therapy showed more intracranial bleeding than that reported in the respective RCTs. It is likely that these RCTs restricted the type of patients to be enrolled and therefore did not represent the true population in which the drug is applied, which would contain patients at increased risk of bleeding. In the second case, observational studies comparing laparoscopic repair and open surgical repair of inguinal hernias reported a higher rate of complications with the former procedure than that reported in the respective RCTs. Besides differences in the selection of patients, the reason for this finding might be the learning curve that naturally occurs with new surgical procedures. Surgeons in the randomized trials may have started the trial once they felt sufficiently secure about their laparoscopy skills, whereas the community data still included early mishaps.

Far from detracting from the value of observational research, these findings reinforce the idea that we need several sources of data for harm assessment.<sup>8</sup> The comparison of adverse effect information from systematic reviews of randomized trials with that from observational studies may lead to more complete insight. Thus, systematic reviews on the effect of therapy should include the best evidence on benefits, from RCTs, as well as the best evidence on harms, which will often come from nonrandomized studies.

The conclusion that observational studies can estimate the harms of treatment as well as, and maybe even more comprehensively than, randomized trials should not lead us to throw caution to the wind. Clinicians should beware of observational studies that claim *beneficial* effects, even if unanticipated.<sup>3</sup> For example, several classes of drugs that are used for long periods (NSAIDs, hormone replacement therapy and statins) were at one time thought to have a possible protective effect against dementia. All of these associations have subsequently been refuted by randomized trials.<sup>9-11</sup> The original associations arose most likely because patients who were taking the drugs were sufficiently well to continue using them over a long period. The difference lies again in the allocation to treatment, which for adverse effects is more likely to be unbi-

ased, because adverse effects will not have similar clinical correlates that are intrinsic to the patients who continue using a drug. The best studies on harms not only have sufficient patient numbers, and strict and verifiable definitions, they also seek a comparison that comes close to an unbiased allocation (i.e., a treatment allocation that has nothing to do with risk factors for the adverse effect). For example, to determine the late adverse effects of radiation therapy on coronary artery disease, women who had right-sided breast cancer were compared with those who had left-sided breast cancer, since irradiation of the left breast would result in a larger dose of radiation to the coronary arteries than would right-sided irradiation.<sup>12</sup>

Even with all of these precautions, observational studies of adverse effects are never foolproof, but neither is any research undertaking.

Jan Vandenbroucke is with the Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, The Netherlands.

**Competing interests:** None declared.

**Acknowledgement:** The author thanks Stephen Evans and John Ioannidis for their critical review of an earlier version of the article.

## REFERENCES

1. Cuervo LG, Aronson JK. The road to health care. *BMJ* 2004;329:1-22.
2. Glasziou P, Vandenbroucke JP, Chalmers I. Assessing the quality of research. *BMJ* 2004;328:39-41.
3. Vandenbroucke JP. When are observational studies as credible as randomized trials? *Lancet* 2004;363:1728-31.
4. Miettinen OS. The need for randomization in the study of intended effects. *Stat Med* 1983;2:267-71.
5. Jick H, Vessey MP. Case-control studies in the evaluation of drug-induced illness. *Am J Epidemiol* 1978;107:1-7.
6. Papanikolaou PN, Christidi GD, Ioannidis JPA. Comparison of evidence on harms of medical interventions in randomized and nonrandomized studies. *CMAJ* 2006;174(5):635-41.
7. Ioannidis JP, Lau J. Improving safety reporting from randomized trials. *Drug Saf* 2002;25:77-84.
8. Loke YK, Derry S, Aronson JK. A comparison of three different sources of data in assessing the frequencies of adverse reactions to amiodarone. *Br J Clin Pharmacol* 2004;57:616-21.
9. Shumaker SA, Legault C, Kuller L, et al. Conjugated equine estrogens and incidence of probable dementia and mild cognitive impairment in postmenopausal women: Women's Health Initiative Memory Study. *JAMA* 2004;291:2947-58.
10. Shepherd J, Blauw GJ, Murphy MB, et al. PROspective Study of Pravastatin in the Elderly at Risk. Pravastatin in elderly individuals at risk of vascular disease (PROSPER): a randomized controlled trial. *Lancet* 2002;360:1623-30.
11. Aisen PS, Schafer KA, Grundman M, et al. Alzheimer's Disease Cooperative Study. Effects of rofecoxib or naproxen vs placebo on Alzheimer disease progression: a randomized controlled trial. *JAMA* 2003;289:2819-26.
12. Darby S, McGale P, Peto R, et al. Mortality from cardiovascular disease more than 10 years after radiotherapy for breast cancer: nationwide cohort study of 90 000 Swedish women. *BMJ* 2003;326:256-7.

**Correspondence to:** Dr. Jan P. Vandenbroucke, Department of Clinical Epidemiology, Leiden University Medical Center, Building 1, PO Box 9600, 2300 RC Leiden, The Netherlands; [j.p.vandenbroucke@lumc.nl](mailto:j.p.vandenbroucke@lumc.nl)