# RESEARCH

# Comparison of evidence on harms of medical interventions in randomized and nonrandomized studies

Panagiotis N. Papanikolaou, Georgia D. Christidi, John P.A. Ioannidis

## Abstract

**Background:** Information on major harms of medical interventions comes primarily from epidemiologic studies performed after licensing and marketing. Comparison with data from large-scale randomized trials is occasionally feasible. We compared evidence from randomized trials with that from epidemiologic studies to determine whether they give different estimates of risk for important harms of medical interventions.

**Methods:** We targeted well-defined, specific harms of various medical interventions for which data were already available from large-scale randomized trials (> 4000 subjects). Nonrandomized studies involving at least 4000 subjects addressing these same harms were retrieved through a search of MEDLINE. We compared the relative risks and absolute risk differences for specific harms in the randomized and nonrandomized studies.

**Results:** Eligible nonrandomized studies were found for 15 harms for which data were available from randomized trials addressing the same harms. Comparisons of relative risks between the study types were feasible for 13 of the 15 topics, and of absolute risk differences for 8 topics. The estimated increase in relative risk differed more than 2-fold between the randomized and nonrandomized studies for 7 (54%) of the 13 topics; the estimated increase in absolute risk differed more than 2-fold for 5 (62%) of the 8 topics. There was no clear predilection for randomized or nonrandomized studies to estimate greater relative risks, but usually (75% [6/8]) the randomized trials estimated larger absolute excess risks of harm than the nonrandomized studies did.

**Interpretation:** Nonrandomized studies are often conservative in estimating absolute risks of harms. It would be useful to compare and scrutinize the evidence on harms obtained from both randomized and nonrandomized studies.

*CMAJ* 2006;174(5):635-41

Evidence of the harms of medical interventions is important when weighing the benefits and risks of treatments in clinical decision-making. However, such evidence is often suboptimal.[1,2] For many treatments the safety profile is incomplete, and some treatments are occasionally withdrawn from the market because of the emergence of toxic side effects that were either missed or suppressed during clinical development.[3,4]

Considerable evidence on the harms of medical interventions is accumulated through epidemiologic studies performed after licensing and marketing.[5–8] Recently, there has been an effort to improve the recording and reporting of information on harms derived from clinical trials.[2,9] Although single trials are usually underpowered to address adequately the absolute and relative risks of adverse events, especially uncommon ones, large trials or meta-analyses may achieve adequate power for this purpose. Previously we examined large trials and meta-analyses in the Cochrane Library and found that it is occasionally feasible to generate evidence on specific, well-defined harms from systematic reviews of large-scale randomized trials.[10] As data on harms become available from randomized trials, a major question is whether these data agree with information obtained from traditional observational studies. For most harms, we cannot compare evidence from randomized trials with that from nonrandomized studies because of insufficient data. However, whenever such comparative data exist, we should gain insights from this evidence. Randomized trials and epidemiologic studies may sometimes disagree even on efficacy outcomes,[11] whereas disagreements in the magnitude of a treatment effect may occur even between large trials.[12]

In theory, randomized trials may have better experimental control and protection against bias than observational studies have, and they may use more active surveillance for recording adverse events. This might translate into more accurate results and higher absolute risks of harms estimated in randomized trials than in observational studies. However, these generalizations may not apply in all circumstances. Randomized trials may sometimes report small, fragmented pieces of evidence of harms that are not primary outcomes, whereas large-scale observational studies may be primarily devoted to assessing specific adverse events. Thus, empirical evidence is needed to complement these theoretical speculations.

In this study, we aimed to determine whether randomized controlled trials and observational studies give different estimates of risk for important harms of medical interventions and, if so, in which direction.

## Methods

For our analysis we targeted specific harms of various medical interventions for which data were already available from large-scale randomized controlled trials. These data were obtained from a previous study in which 1754 systematic reviews in the Cochrane Database of Systematic Reviews (issue 3, 2003) were screened to identify quantitative information on at least 4000 subjects for 66 well-defined, specific harms of various interventions.[10] With 4000 subjects equally allocated, there is 80% power ($\alpha$ = 0.05) to detect a difference of 1% in the compared arms for an otherwise uncommon event (risk < 0.25% in the control group). The identified harms had been clearly defined on the basis of clinical and laboratory criteria and with explicit grading of severity and seriousness.[10] For each harm, we searched for respective published nonrandomized studies that had at least 4000 subjects. Single randomized trials with such a sample size are difficult to conduct; thus, we accepted the possibility that a meta-analysis of several trials may pass the sample size threshold. Because nonrandomized studies with this sample size are more readily feasible, each eligible observational study had to have at least 4000 subjects.

We considered all nonrandomized controlled studies in which the comparison (intervention v. no treatment, or intervention v. other intervention) was similar to that in the respective randomized trial(s), with "no treatment" corresponding to "placebo." We excluded noncontrolled studies (the absolute and relative risk conferred by the intervention per se cannot be estimated), unless the specific harm was so rare in the control population that randomized trials had recorded no events in control subjects (in which case the absolute risk among treated subjects would still be meaningful to compare between study designs).

We included nonrandomized studies that had selected participants with the same, overlapping or wider indications for the intervention as those used in the respective randomized trials; we excluded nonrandomized studies in which the indications differed entirely from those in the randomized trial populations.

We accepted risk estimates in nonrandomized studies regardless of whether they were crude or adjusted, were based on numbers of people or person-years, or were directly given or indirectly estimated from the published data.

Two of us (P.P. and G.C.) searched MEDLINE (PubMed)

**Table 1:** Details of eligible nonrandomized studies of 15 harms of medical interventions for which data from randomized trials were available

| Randomized trials | | | Nonrandomized studies | | | |
|---|---|---|---|---|---|---|
| Topic no. | Harm* (intervention) | No. of trials and patients | Topic no. (ref) | Design/funding | Sample size | No. of adjusting covariates |
| 1 | Encephalopathy (acellular v. whole-cell pertussis vaccine) | 9 N = 113 762 | 1a (Geier et al[15]) | Vaccine surveillance program/PuD | 62 916 460† | 0 |
| 2 | Convulsions (acellular v. whole-cell pertussis vaccine) | 15 N = 124 387 | 2a (Geier et al[15]) | Vaccine surveillance program/PuD | 62 916 460† | 0 |
| | | | 2b (Rosenthal et al[16]) | Vaccine surveillance program/PuA | 32 000 000† | 0 |
| 3 | Hypotonic hyporesponsiveness (acellular v. whole-cell pertussis vaccine) | 11 N = 121 573 | 3a (Le Saux et al[17]) | Vaccine surveillance program/Pu | NED | 0 |
| 4 | Major extracranial bleed (oral anticoagulant therapy) | 16 N = 22 049 | 4a (Kaplan et al[18]) | Prospective cohort‡/Pu | 5 888 | 2 |
| | | | 4b (Johnsen et al[19]) | Population-controlled‡/Pu | 271 486 | 3 |
| | | | 4c, 4d (Jick et al[20])§ | Inpatient cohort‡/Pu | 11 060 | 0 |
| | | | 4e (Quilliam et al[21]) | Case-control‡/Pu | 11 009 | 6 |
| | | | 4f (Shorr et al[22]) | Retrospective cohort/Pu | 103 954 | 8 |
| 5 | Symptomatic intracranial bleed (oral anticoagulant therapy) | 15 N = 22 794 | 5a (Själander et al[23]) | Population registry‡/Pu | 244 434 | 2 |
| | | | 5b (Fogelholm et al[24]) | Population-controlled registry‡/NR | 29 288 | 1 |
| 6 | Major extracranial bleed (anticoagulant v. antiplatelet therapy) | 6 N = 11 721 | 6a (Quilliam et al[21]) | Case-control‡/Pu | 4 249 | 0 |
| 7 | Major extracranial bleed (antiplatelet therapy) | 9 N = 41 399 | 7a, 7b (Jick et al[20])§ | Inpatient cohort‡/Pu | 12 718 | 0 |
| | | | 7c (Quilliam et al[21]) | Case-control‡/Pu | 12 248 | 6 |
| 8 | Symptomatic intracranial bleed (antiplatelet therapy [ASA]) | 9 N = 41 399 | 8a, 8b (Iso et al[25])¶ | Prospective cohort‡/Pu | 31 827 | 17 |
| | | | 8c (Kronmal et al[26]) | Prospective cohort‡/Pu | 4 363 | 3 |
| 9 | Hepatitis-related death (isoniazid chemoprophylaxis) | 2 N = 25 714 | 9a (Nolan et al[27]) | Prospective cohort/PuA | 11 141 | NP |
| | | | 9b (Salpeter[28]) | Survey/NR | 182 285 | NP |

independently (last search October 2004) for qualifying non-randomized studies that would correspond to each of the 66 harms.[10] Whenever there was disagreement on whether a study qualified, consensus was reached with a third investigator (J.I.). For our search strategies, we used the names of the interventions, including the drug-specific names when a class of drugs was involved. There is relatively limited evidence on how best to search for harms-related literature, and thus we started with broad search strategies to maximize sensitivity and avoid missing eligible studies. We also used various terms representing the harms and settings under study, whenever deemed appropriate for improving search specificity without compromising sensitivity. We focused on English-language studies involving humans. The search strategy is described in online Appendix 1 (www.cmaj.ca/cgi/content/full/cmaj.050873/DC1). We also checked the reference lists of retrieved studies for additional pertinent reports.

For each eligible nonrandomized study, we extracted information on the authors, the year of publication, the study design, the setting, the total sample size, and the estimates of absolute and relative risks conferred by the interventions under study. We captured both crude and adjusted estimates when both were available, and we recorded the covariates used for adjustment.

For the respective evidence from the randomized trials, we retrieved information already recorded[10] on the authors and the year of publication of each trial, the setting, the total sample size, and the absolute risk difference and risk ratios with 95% confidence intervals (CIs).

Risk ratios were always available for evidence from the randomized trials. For evidence from the nonrandomized studies, estimates of relative risk included risk ratios, incidence rate ratios, relative hazard ratios from Cox models, and odds ratios from case–control studies (proxies of population-level risk ratios). All results were expressed as point estimates with 95% CIs. When available, adjusted estimates were preferred over crude ones. When several adjusted estimates existed, we selected the one that considered the largest number of covariates. Analyses using the crude estimates yielded qualitatively similar results (data not shown).

For each harms topic, we compared the risk estimates from the randomized trial (or from the meta-analysis of several trials) with the risk estimates from the respective nonrandomized study (or the meta-analysis of several studies).

**Table 1** continued

| Randomized trials | | | Nonrandomized studies | | | |
|---|---|---|---|---|---|---|
| Topic no. | Harm* (intervention) | No. of trials and patients | Topic no. (ref) | Design/funding | Sample size | No. of adjusting covariates |
| 10 | Visceral or vascular injury (laparoscopy v. open surgery for inguinal hernia) | 22 N = 4 914 | 10a (Hair et al[29]) | Prospective cohort/Pu | 5 506 | 0 |
| 11 | Wound infection (laparoscopy v. open surgery for appendicitis) | 34 N = 4 324 | 11a (Koch et al[30]) | Prospective cohort/NR | 4 727 | 0 |
| | | | 11b (Guller et al[31]) | Inpatient cohort/Pu | 145 290 | 9 |
| 12 | Spontaneous miscarriage (folate supplementation**) | 3 N = 7 600 | 12a (Gindler et al[32]) | Pregnancy monitoring registry/Pu | 15 365 | 1 |
| | | | 12b (Windham et al[33]) | Prospective pregnancy cohort/PuA | 5 144 | 0 |
| 13 | Multiple gestation (folate supplementation**) | 3 N = 6 241 | 13a (Li et al[34]) | Pregnancy monitoring registry/Pu | 242 015 | 1 |
| | | | 13b (Werler et al[35]) | Case-control birth defect registry/NR | 5 474 | 0 |
| | | | 13c (Ericson et al[36]) | Population registry/PuA | 442 906 | 3 |
| 14 | Major bleed (platelet glycoprotein IIB/IIIA blocker therapy in PCI) | 12 N = 17 469 | 14a (Horwitz et al[37]) | Registry of procedures/NR | 18 821 | 25 |
| 15 | Acute myocardial infarction (rofecoxib v. naproxen therapy) | 1 N = 8 076 | 15a, 15b (Ray et al[38])¶ | Retrospective cohort‡/Pu | 90 629 | 14 |

Note: PCI = percutaneous coronary intervention, Pu = public funding, PuA = first author affiliated with public institute, PuD = database supported by public funding, NR = funding not reported, NED = no exact data, but pertains to population vaccinated in 6 years in Canada (periods using different vaccines; relative rate estimated with Poisson regression), NP = not pertinent (no relative risks estimated for uncontrolled studies).
*Harm definitions were similar in randomized and nonrandomized studies, although for major extracranial bleeding, we included nonrandomized studies that addressed specifically gastrointestinal bleeding or even only upper gastrointestinal bleeding.
†Data pertain to doses, not patients.
‡The indications for use of the intervention in the nonrandomized studies were either overlapping (but not identical to) or were wider than the indications for use of the intervention in the respective randomized trials.
§Separate data were available for 2 different definitions of the harm, both compatible with the definition used in the randomized trials.
¶Separate data were available for 2 different doses of the intervention, both compatible with the doses used in the randomized trials.
**Evidence from randomized trials addressed primarily periconceptional use; in the nonrandomized studies, we used either data on periconceptual supplementation or data on supplementation during pregnancy, when eligible data were not specifically available for periconceptual use; some nonrandomized studies also had data on folate or multivitamin supplementation rather than specific data on folate supplementation alone.

Whenever 2 or more nonrandomized studies were available for a specific topic, we used a random-effects summary estimate,[13] using the inverse variance method and estimating the variance from the 95% CIs of each relative risk. Random-effects syntheses were already available for the randomized evidence, as previously described.[10] Between-study heterogeneity was estimated using the Q statistic (considered significant at $p < 0.10$).[13]

We examined whether the estimated increase in relative risk (relative risk – 1) with the harmful intervention (v. no treatment or a less harmful intervention) differed more than 2-fold between the randomized and nonrandomized studies and, if so, in which direction. The 2-fold threshold has been used previously to compare efficacy data from observational and randomized studies.[11] We also tested secondarily whether the differences in relative risk between the randomized and nonrandomized studies were beyond chance ($p < 0.05$), based on a z statistic estimated from the difference of the natural logarithms of the relative risks and the variance of this difference.[14]

Finally, we examined whether the differences in absolute risk between the randomized and nonrandomized studies were greater than 2-fold in their point estimates and, secondarily, whether these differences were beyond chance ($p < 0.05$). We excluded from our analyses topics for which the absolute risk of harm would depend on the available follow-up, since this could vary in different studies.

## Results

We screened 18 198 articles. Of these, 82 were screened as full articles: 25 articles were found to contain eligible data from nonrandomized studies[15–39] that could be compared with evidence from large-scale randomized trials. In total, data from nonrandomized studies could be juxtaposed against data from randomized trials for 15 of the 66 harms (Table 1). All of the studied harms were serious and clinically relevant.

The interventions included drugs, vitamins, vaccines and surgical procedures. A large variety of prospective and retrospective approaches were used in the nonrandomized studies, including both controlled and uncontrolled designs (Table 1). Of the 25 eligible nonrandomized studies, 16 received funding from a public source, another 4 had first authors from public institutions, 1 used a database that was funded by public sources, and 4 did not state any information on funding.

Comparison of the estimated increases in relative risk between the randomized and nonrandomized studies was feasible for 13 of the 15 harms topics (Fig. 1); a comparison was not feasible for the remaining 2 topics because no encephalopathy events were reported in the randomized trials of acellular versus whole-cell pertussis vaccination (topic 1), and there was no control group in the observational studies of hepatitis-related deaths with isoniazid chemoprophylaxis (topic 9). Comparison of differences in absolute risk between the randomized and nonrandomized studies was feasible for 8 of the 15 harms topics (Table 2).

For 5 (38%) of the 13 topics for which estimated increases in relative risk could be compared, the increase was greater in the nonrandomized studies than in the respective randomized trials; for the other 8 topics (62%), the increase was greater in the randomized trials. The estimated increase in relative risk differed more than 2-fold between the randomized and nonrandomized studies for 7 (54%) of the 13 topics (symptomatic intracranial bleed with oral anticoagulant therapy [topic 5], major extracranial bleed with anticoagulant v. antiplatelet therapy [topic 6], symptomatic intracranial bleed with ASA [topic 8], vascular or visceral injury with laparoscopic v. open surgical repair of inguinal hernia [topic 10], major bleed with platelet glycoprotein IIb/IIIa blocker therapy for percutaneous coronary intervention [topic 14], multiple gestation with folate supplementation [topic 13], and acute myocardial infarction with rofecoxib v. naproxen therapy [topic 15]).

Differences in relative risk beyond chance between the randomized and nonrandomized studies occurred for 2 of the 13 topics: the relative risks for symptomatic intracranial bleed with oral anticoagulant therapy (topic 5) and for vascular or visceral injury with laparoscopic versus open surgical repair of inguinal hernia (topic 10) were significantly greater in the nonrandomized studies than in the randomized trials.

Between-study heterogeneity was more common in the syntheses of data from the nonrandomized studies than in the syntheses of data from the randomized trials. There was significant between-study heterogeneity ($p < 0.10$ on the Q statistic) among the randomized trials for 2 data syntheses (topics 3 and 14) and among the nonrandomized studies for 5 data syntheses (topics 4, 7, 8, 13 and 15). The adjusted and unadjusted estimates of relative risk in the nonrandomized studies were similar (see online Appendix 4, available at www.cmaj.ca/cgi/content/full/cmaj.050873/DC1).

In terms of absolute risk, the estimated increases in risk were greater in the randomized trials than in the nonrandomized studies for 6 (75%) of the 8 topics for which a comparison was feasible (Table 2). The estimated increases differed more than 2-fold between the randomized and nonrandomized studies for 5 (62%) of the 8 topics. The difference was beyond chance for only 1 topic (convulsions with acellular v. whole-cell pertussis vaccination [topic 2]): the estimated increase in absolute risk was almost 40-fold greater in the randomized trials than in the nonrandomized studies.

## Interpretation

We found no clear predilection for randomized trials or for nonrandomized studies to estimate greater relative risks of harms of medical interventions. Disagreements in the estimated magnitude of the absolute risk were common between the randomized and nonrandomized studies. The randomized trials usually estimated larger absolute risks of harms than the nonrandomized studies did; for 1 topic, the difference was almost 40-fold.

For efficacy outcomes, data from observational studies have sometimes been overtly refuted on important medical questions; data from randomized trials have also been refuted, but to a lesser extent.[40,41] Although some empirical evaluations have suggested that data from observational studies typically agree with the results of randomized trials on the same topic

for efficacy outcomes,[42–44] one evaluation showed significant differences in the relative risks between randomized and non-randomized studies in one-fifth of the tested interventions.[11] Harms may occur far less frequently than efficacy-related events. For uncommon outcomes, considerable differences in risk estimates may not reach formal statistical significance. Moreover, harms are usually evaluated as secondary outcomes. Within clinical trials, discrepancies have been documented more frequently for secondary outcomes than for primary ones.[45,46] Thus, disagreements in the risk estimates of harms with different study designs should not be surprising.

We found that differences in the absolute risk estimates between the 2 study types were common. Such differences could have considerable clinical impact, since physicians and patients are interested more in absolute than in relative risks.[47] The differences in absolute risk between the study designs may have been due in part to differences in the exact definitions and the level of severity of the harms that were
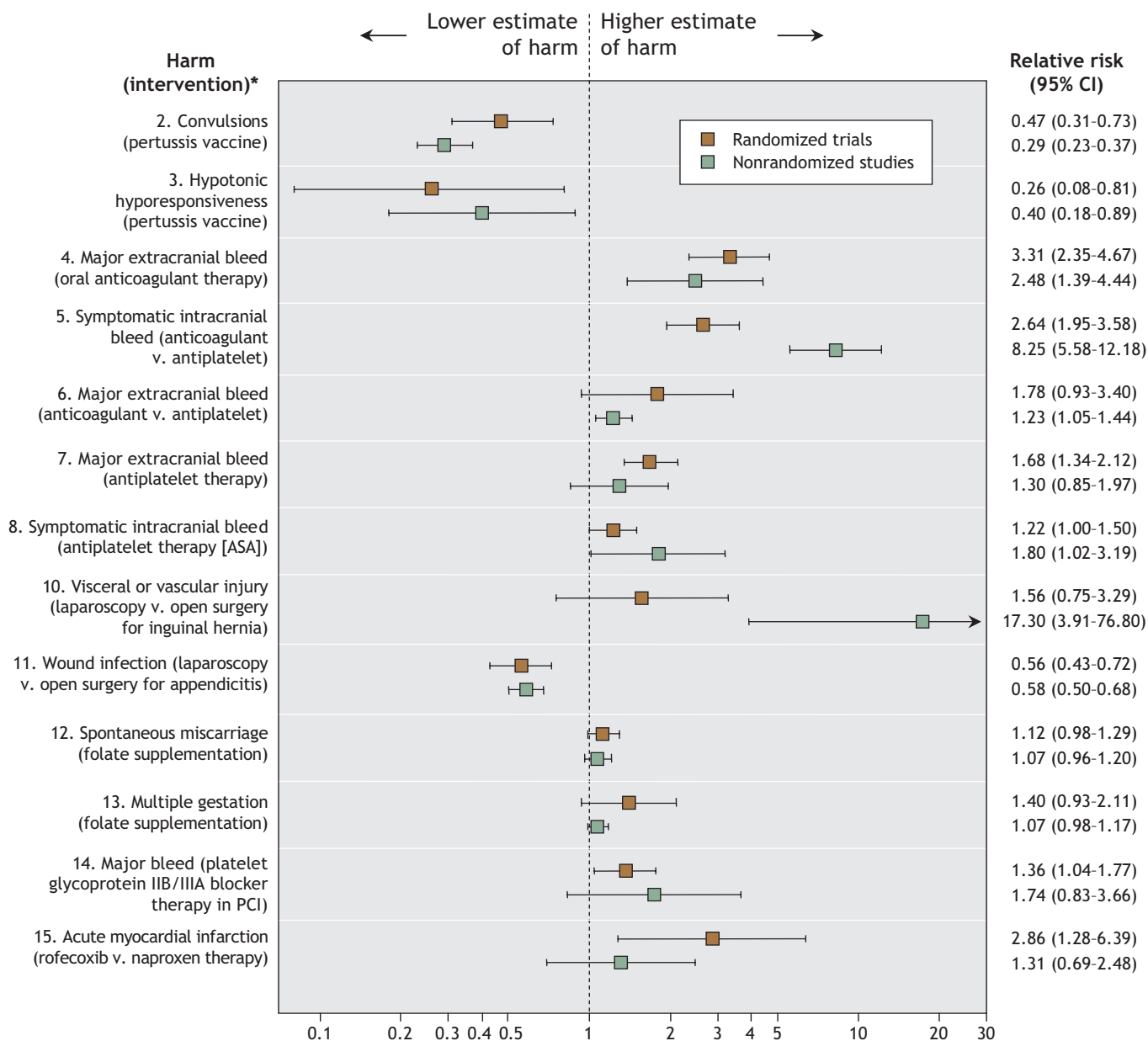


**Fig. 1:** Comparison of relative risk estimates for specific harms of medical interventions from randomized and nonrandomized studies. Error bars represent 95% confidence intervals. (A random-effects summary estimate of the relative risk is provided for topics for which 2 or more nonrandomized studies were available. The risk estimates for the individual nonrandomized studies appear in a detailed version of this figure [see online Appendix 2, available at www.cmaj.ca/cgi/content/full/cmaj.050873/DC1].) *The topic numbers correspond to those in Table 1, which contains details about the nonrandomized studies.

evaluated. Also, adverse events are recorded more thoroughly in randomized trials, owing to regulatory requirements, than in observational databases. Several important harms in everyday medical practice may be more common than what observational studies suggest.

Observational studies to date have been the standard for identifying serious, yet uncommon harms of medical interventions.[5,7,8] Our results suggest that, if a nonrandomized study finds harm, chances are that a randomized study would find even greater harm in terms of the magnitude of absolute risk. It may be unfair to invoke bias and confounding to discredit observational studies as a source of evidence on harms. Moreover, almost all of the large-scale observational studies that we identified were supported by public funding, whereas the randomized trials, in particular those of drugs and technologies, were typically supported by the manufacturers, with unknown potential for bias.

Differences in data between observational and randomized trials may be due in part to differences in study populations; however, this information is often difficult to dissect, and important details about patient populations may not be transparent in published reports.

Our study had limitations. First, we arbitrarily selected nonrandomized studies with at least 4000 subjects. There is no consensus on what constitutes a large epidemiologic study. For consistency, we selected the same cutoff for sample size as we had previously used for randomized trials.[10] For serious harms with a frequency of less than 1%, even studies involving several thousand subjects would be grossly underpowered to detect, let alone quantify, risk with any accuracy. Second, because there is no standardized and widely accepted method for searching PubMed for harms-related articles, we may have missed some eligible studies. However, the problem is unlikely to be major, since we targeted only studies with large samples. The same caveat also applies to unpublished data. Third, a large number of studies of harms are uncontrolled. However, unless the adverse event is extremely rare among untreated subjects, absolute estimates of risk are difficult to interpret and relative risk cannot be estimated in uncontrolled studies. Finally, we included only randomized trials from Cochrane reviews, which do not cover every medical intervention. However, although other topics may be available for comparing evidence on harms from large-scale trials, the Cochrane database is still the most comprehensive systematic source of evidence available.

There is a need to enhance the use of information on harms from long-term prospective studies and randomized trials. Data from randomized trials, even if limited, may provide some insights into important harms and can add to the accumulating evidence from observational studies. Currently, we have estimates of risk for only a minority of harms of commonly used medical interventions, and even these estimates have considerable uncertainty. The small number of eligible pairs of randomized trials and observational studies in our analysis was due to the fact that we set stringent criteria for the definition of harms and the availability of evidence from large-scale trials. However, our original search for evidence from randomized trials[10] was of the entire Cochrane database, which covers a sizeable portion of clinical medicine. Therefore, it is fair to say that evidence on harms from concurrent, large-scale randomized and observational data are rarely encountered in the medical literature. This is a deficit that needs to be remedied. More unbiased, quantitative information should be actively sought for major harms of medical interventions.

**Table 2:** Differences in absolute risks of harms between randomized and nonrandomized studies

| Topic no.* | Harm (intervention) | Study type; risk difference (95% CI), % | |
|---|---|---|---|
| | | Randomized | Nonrandomized† |
| 1 | Encephalopathy (acellular v. whole-cell pertussis vaccine) | 0.00 (-0.01 to 0.01) | -0.0001 (-0.0001 to 0.0000) |
| 2 | Convulsions (acellular v. whole-cell pertussis vaccine) | -0.04 (-0.07 to -0.01) | -0.00103 (-0.0012 to -0.00086) |
| 9 | Hepatitis-related death (isoniazid chemoprophylaxis) | 0.02 (0.00 to 0.05) | 0.00109 (0.00009 to 0.00209) |
| 10 | Visceral or vascular injury (laparoscopy v. open surgery for inguinal hernia) | 0.02 (-0.28 to 0.31) | 1.22 (-0.25 to 2.69) |
| 11 | Wound infection (laparoscopy v. open surgery for appendicitis) | -2.96 (-4.26 to -1.66) | -2.13 (-3.57 to -0.69) |
| 12 | Spontaneous miscarriage (folate supplementation) | 1.13 (-0.20 to 2.47) | 0.73 (-0.32 to 1.78) |
| 13 | Multiple gestation (folate supplementation) | 0.49 (-0.11 to 1.08) | 0.35 (-0.52 to 1.22) |
| 14 | Major bleed (platelet glycoprotein IIB/IIIA blocker therapy in PCI) | 0.96 (0.09 to 1.83) | 0.25 (-0.02 to 0.52) |

Note: CI = confidence interval, PCI = percutaneous coronary intervention.
*The topic numbers correspond to those listed in Table 1.
†Whenever 2 or more nonrandomized studies were available for a specific topic (topics 2, 9, 11, 12 and 13), we used a random-effects summary estimate of the risk difference.[13] Random-effects syntheses were already available for the randomized evidence, as previously described.[10] See online Appendix 3 for the individual risk differences in each of the nonrandomized studies (available at www.cmaj.ca/cgi/content/full/cmaj.050873/DC1).

### Editor's take

- Clinical trials are designed to show benefit of therapy. Evidence on harms of therapy is apparent usually only after larger, after-market, observational studies have been performed.

- In this study, the authors compared risks of 13 major harms of medical interventions using data from both randomized trials and observational studies. Contrary to current belief, the nonrandomized studies were often more conservative in their estimates of risk than the randomized trials were.

**Implications for practice:** Although we need both types of studies, it may be that data from randomized trials provide reasonable estimates of harms that could be used to guide clinical decisions and advise patients.

This article has been peer reviewed.

From the Clinical Trials and Evidence-Based Medicine Unit, Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece (all authors)

### REFERENCES

1. Ioannidis JP, Lau J. Completeness of safety reporting in randomized trials: an evaluation of 7 medical areas. *JAMA* 2001;285:437-43.
2. Ioannidis JP, Evans SJ, Gotzsche PC, et al; CONSORT Group. Better reporting of harms in randomized trials: an extension of the CONSORT statement. *Ann Intern Med* 2004;141:781-8.
3. Topol EJ. Failing the public health — rofecoxib, Merck, and the FDA. *N Engl J Med* 2004;351:1707-9.
4. Lasser KE, Allen PD, Woolhandler SJ, et al. Timing of new black box warnings and withdrawals for prescription medications. *JAMA* 2002;287:2215-20.
5. Jick H. The discovery of drug-induced illness. *N Engl J Med* 1977;296:481-5.
6. Vandenbroucke JP. Benefits and harms of drug treatments. *BMJ* 2004;329:2-3.
7. Stephens MD, Talbot JC, Routledge PA, editors. *The detection of new adverse reactions*. 4th ed. London (UK): Macmillan Reference; 1998.
8. Edwards IR, Aronson JK. Adverse drug reactions: definitions, diagnosis, and management. *Lancet* 2000;356:1255-9.
9. Ioannidis JP, Lau J. Improving safety reporting from randomised trials. *Drug Saf* 2002;25:77-84.
10. Papanikolaou PN, Ioannidis JP. Availability of large-scale evidence on specific harms from systematic reviews of randomized trials. *Am J Med* 2004;17:582-9.
11. Ioannidis JP, Haidich AB, Pappa M, et al. Comparison of evidence of treatment effects in randomized and nonrandomized studies. *JAMA* 2001;286:821-30.
12. Furukawa TA, Streiner DL, Hori S. Discrepancies among megatrials. *J Clin Epidemiol* 2000;53:1193-9.
13. Lau J, Ioannidis JP, Schmid CH. Quantitative synthesis in systematic reviews. *Ann Intern Med* 1997;127:820-6.
14. Cappelleri JC, Ioannidis JP, Schmid CH, et al. Large trials vs meta-analysis of smaller trials: How do their results compare? *JAMA* 1996;276:1332-8.
15. Geier DA, Geier MR. An evaluation of serious neurological disorders following immunization: a comparison of whole-cell pertussis and acellular pertussis vaccines. *Brain Dev* 2004;26:296-300.
16. Rosenthal S, Chen R, Hadler S. The safety of acellular pertussis vaccine vs whole-cell pertussis vaccine: a postmarketing assessment. *Arch Pediatr Adolesc Med* 1996;150:457-60.
17. Le Saux N, Barrowman NJ, Moore DL, et al. Canadian Paediatric Society/Health Canada Immunization Monitoring Program–Active (IMPACT). Decrease in hospital admissions for febrile seizures and reports of hypotonic-hyporesponsive episodes presenting to hospital emergency departments since switching to acellular pertussis vaccine in Canada: a report from IMPACT. *Pediatrics* 2003;112:e348.
18. Kaplan RC, Heckbert SR, Koepsell TD, et al. Risk factors for hospitalized gastrointestinal bleeding among older persons. Cardiovascular Health Study Investigators. *J Am Geriatr Soc* 2001;49:126-33.
19. Johnsen SP, Sørensen HT, Mellemkjoer L, et al. Hospitalisation for upper gastrointestinal bleeding associated with use of oral anticoagulants. *Thromb Haemost* 2001;86:563-8.
20. Jick H, Porter J. Drug-induced gastrointestinal bleeding. *Lancet* 1978;2:87-9.
21. Quilliam BJ, Lapane KL, Eaton CB, et al. Effect of antiplatelet and anticoagulant agents on risk of hospitalization for bleeding among a population of elderly nursing home stroke survivors. *Stroke* 2001;32:2299-304.
22. Shorr RI, Ray WA, Daugherty JR, et al. Concurrent use of nonsteroidal anti-inflammatory drugs and oral anticoagulants places elderly persons at high risk for hemorrhagic peptic ulcer disease. *Arch Intern Med* 1993;153:1665-70.
23. Själander A, Engström G, Berntorp E, et al. Risk of haemorrhagic stroke in patients with oral anticoagulation compared with the general population. *J Intern Med* 2003;254:434-8.
24. Fogelholm R, Eskola K, Kiminkinen T, et al. Anticoagulant treatment as a risk factor for primary intracerebral haemorrhage. *J Neurol Neurosurg Psychiatry* 1992;55:1121-4.
25. Iso H, Hennekens CH, Stampfer MJ, et al. Prospective study of aspirin use and risk of stroke in women. *Stroke* 1999;30:1764-71.
26. Kronmal RA, Hart RG, Manolio TA, et al. Aspirin use and incident stroke in the Cardiovascular Health Study. CHS Collaborative Research Group. *Stroke* 1998;29:887-94.
27. Nolan CM, Goldberg SV, Buskin SE. Hepatotoxicity associated with isoniazid preventive therapy. A 7-year survey from a public health tuberculosis clinic. *JAMA* 1999;281:1014-8.
28. Salpeter SR. Fatal isoniazid-induced hepatitis. Its risk during chemoprophylaxis. *West J Med* 1993;159:560-4.
29. Hair A, Duffy K, McLean J, et al. Groin hernia repair in Scotland. *Br J Surg* 2000;87:1722-6.
30. Koch A, Zippel R, Marusch F, et al. Prospective multicenter study of antibiotic prophylaxis in operative treatment of appendicitis. *Dig Surg* 2000;17:370-8.
31. Guller U, Jain N, Peterson ED, et al. Laparoscopic appendectomy in the elderly. *Surgery* 2004;135:479-88.
32. Gindler J, Li Z, Berry RJ, et al. Folic acid supplements during pregnancy and risk of miscarriage; Jiaxing City Collaborative Project on Neural Tube Defect Prevention. *Lancet* 2001;358:796-800.
33. Windham GC, Shaw GM, Todoroff K, et al. Miscarriage and use of multi-vitamins or folic acid. *Am J Med Genet* 2000;90:261-2.
34. Li Z, Gindler J, Wang H, et al. Folic acid supplements during early pregnancy and likelihood of multiple births: a population-based cohort study. *Lancet* 2003;361:380-4.
35. Werler MM, Cragan JD, Wasserman CR, et al. Multivitamin supplementation and multiple births. *Am J Med Genet* 1997;71:93-6.
36. Ericson A, Kallen B, Aberg A. Use of multivitamins and folic acid in early pregnancy and multiple births in Sweden. *Twin Res* 2001;4:63-6.
37. Horwitz PA, Berlin JA, Sauer WH, et al.; Registry Committee of the Society for Cardiac Angiography Interventions. Bleeding risk of platelet glycoprotein IIb/IIIa receptor antagonists in broad-based practice (results from the Society for Cardiac Angiography and Interventions Registry). *Am J Cardiol* 2003;91:803-6.
38. Ray WA, Stein M, Daugherty JR, et al. COX-2 selective non-steroidal anti-inflammatory drugs and risk of serious coronary heart disease. *Lancet* 2002;360:1071-3.
39. Mamdani M, Rochon P, Juurlink DN, et al. Effect of selective cyclooxygenase 2 inhibitors and naproxen on short-term risk of acute myocardial infarction in the elderly. *Arch Intern Med* 2003;163:481-6.
40. Ioannidis JPA. Why most published research findings are false. *PLoS Med* 2005;2:e124.
41. Ioannidis JPA. Contradicted and initially stronger effects in highly cited clinical research. *JAMA* 2005;294:218-28.
42. Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. *N Engl J Med* 2000;342:1878-86.
43. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *N Engl J Med* 2000;342:1887-92.
44. Ioannidis JP, Haidich AB, Lau J. Any casualties in the clash of randomized and observational evidence? *BMJ* 2001;322:879-80.
45. LeLorier J, Gregoire G, Benhaddad A, et al. Discrepancies between meta-analyses and subsequent large randomized, controlled trials. *N Engl J Med* 1997;337:536-42.
46. Ioannidis JP, Cappelleri JC, Lau J. Issues in comparisons between meta-analyses and large trials. *JAMA* 1998;279:1089-93.
47. Cook RJ, Sackett DL. The number needed to treat: a clinically useful measure of treatment effect. *BMJ* 1995;310:452-4.

***Correspondence to:*** *Dr. John P.A. Ioannidis, Professor and Chairman, Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina 45110, Greece; fax +3026510-97867; jioannid@cc.uoi.gr*