HEALTH CARE REFORM

Adverse Events in Randomized Trials

Neglected, Restricted, Distorted, and Silenced

ACCURATE INFORMATION ON HARMS OF medical interventions is essential for evidence-based practice. Most newly introduced treatments usually have small, incremental benefits, if any, against already available interventions, and differences in the profile of harms should play a key role on treatment choice. Randomized trials offer an excellent opportunity to evaluate harms of interventions using the most robust experimental design available in clinical research. However, several empirical evaluations (**Table**) have shown that many trials do not report harms or report them in a fragmented or suboptimal way. In this issue, an excellent study by Pitrou et al[1] adds more evidence on this issue.

See also page 1756

Pitrou et al[1] evaluated 133 trials published recently in 6 high–impact factor journals. Most previous empirical studies had examined older trials and typically selected articles based on topic; only 2 other empirical studies (Table) also selected specifically high–impact factor journals. Pitrou et al[1] excluded 159 trials published in these high-profile journals owing to multiple (>2) arms, comparison of public health interventions, factorial, or cluster randomization, noninferiority and/or equivalence designs, or identical safety and efficacy outcome. The exclusions are unfortunate because in all these excluded trials, harms are clearly important to report accurately and in proper detail, as in any trial.

Acknowledging these caveats, Pitrou et al[1] convincingly demonstrate that suboptimal reporting of adverse events continues to plague randomized trials, even after the dissemination of the extension for harms of the CONSORT statement,[2] and even in the top medical journals. Much in line with previous evaluations (Table), the study[1] found that some trials gave absolutely no information on harms, severity was often undefined or vaguely defined, and half the trials reported no information on withdrawal of patients owing to harms. Only 13% reported the reasons why patients withdrew owing to adverse events, information that is of prime clinical relevance.

Pitrou et al[1] explored not only the lack of reporting of harms but also various ways in which reporting of adverse events is restricted and distorted. A third of the trials had restricted the reporting to the most common or the most severe harms, or to statistically significant differences. However, the most common events are usually mild and clinically inconsequential, whereas the most severe events are usually rare given the limited sample size of most trials. Much of the burden of toxic effects is often conferred by events of moderate severity. Reporting only the statistically significant results is also a misconception. Single trials are usually underpowered to detect differences in harms[3] for rare severe events and for most relatively low-rate, moderately severe events. Thus, automatic reassurance statements that "no significant differences were found" are misleading. Concurrently, multiple testing in an underpowered setting is a recipe for obtaining false-positive findings[4] and causing needless anxiety. Nevertheless, the authors of half of the trials analyzed by Pitrou et al[1] succumbed to the temptation of statistical testing for differences in harms between compared arms. A quarter of them combined information on different harms per organ in composite outcomes. This is a stretched effort to gain power, but construct validity is impaired. For example, a composite of neurological toxic effects may include harms as disparate as occasional numbness in the fingertips and deep coma.[5]

Avoidance of spurious selection filters and spurious mergers of data would allow the evaluation of the cumulative picture emerging from multiple trials through meta-analyses. Meta-analyses of adverse events have several caveats.[6] However, their conduct is not a utopian wish, given that for many new interventions the research agenda includes dozens of trials and their combined harms data could often yield adequately powered, conclusive results if they are collected and analyzed in a standardized fashion.

Reporting of trials is their public face; information that reaches peers, clinicians, and the public should be accurate. Why is it then that reporting of adverse events is often insufficient or misleading? There are several explanations that reflect diverse motives:

- A study design that ignores or undervalues adverse events
- Neglected collection of adverse events during the trial contact
- Lack of reporting of adverse events
- Restricted reporting of adverse events
- Distorted reporting of adverse events in the trials and accompanying literature
- Silencing the evidence on harms

Poor reporting may sometimes reflect that collection of information on harms was not included in study design or was neglected during study conduct. Trials for

Table. Empirical Studies Evaluating the Reporting of Harms in Randomized Trials[a]

Study Citation	Clinical Trials Evaluated: No.	Published	Field/Selection	Reporting of Information on Adverse Events, %: Any	Severity Reporting	Withdrawals Due to Harms per Arm
JAMA. 2001;285(4):437-443	192	1967-1999	7 Areas	90	Adequate for clinical events in 39 and for LTE in 29	75 (46 also gave numbersper reason)
BMC Clin Pharmacol. 2001;1:3	185	1997	7 Eminent journals	86	23 Defined severity	ND
Ann Intern Med. 2005;143(1):20-25	193	1999-2005	Rheumatic diseases	72	43 But only 9 deemed appropriate	65 (28 also gave numbers per reason)
Am J Psychiatry. 2004;161(9):1692-1697	142	1959-2002	Mental health	45	Adequate for clinical events in 16 and for LTE in 12	40 (30 also gave numbers per reason)
Otolaryng Head Neck Surg. 2009;140(2):241-244	576[b]	1996 and 2006	Otolaryngology	65	ND	7
J Clin Epidemiol. 2008;61(11):1152-1160	33	1996-2005	Dementia	55	27 Reported grading	ND
Pharmacoepidemiol Drug Saf. 2007;16(3):349-351	521	2000-2003	5 Highly cited journals	89	ND	ND
Ann Rheum Dis. 2007;66(1):124-127	70	Until 2005	Rheumatology	79	ND	ND
J Endocrinol. 2002;175(2):545-552	17	1990-1999	Growth hormone replacement	65	ND	59
Ann Surg. 2002;235(6):803-812	119[c]	1975-2005	Surgery	ND	20 Reported grading	NP
Pitrou et al[d]	133	Jan 1, 2006, to Jan 1, 2007	6 Highly cited journals	89	57 Given per arm but only 16 gave grading scale	53 (13 also gave numbers per reason)

Abbreviations: LTE, laboratory toxic effects; ND, no data; NP, not pertinent.

[a] Only studies focusing on reporting of adverse events are shown in this table, excluding reviews where the primary aim was to synthesize information and obtain estimates of risk for harms for a specific topic. Exact definitions may vary across these empirical studies.

[b] Only 10% of the trials were level-1 evidence (randomized trials).

[c] Only 42 of the trials were randomized trials.

[d] Current study (by Pitrou et al[1]).

some interventions (eg, psychotherapies) almost never report any harms. Then, information on toxic effects may be explicitly collected as planned, but authors may not report it or may report it in a fragmented or restricted fashion. The usual argument in such a case is printed space limitations. However, Web supplements can cater to any amount of information. It should also be acknowledged that safety data may sometimes cause information overload from routine but unnecessary measurements.[7] Nevertheless, careful consideration of what data are useful to collect may avoid this deluge and eliminate the need for spurious post hoc restrictions in presenting results.[7]

Distorted reporting goes a step further than restricted reporting. In some cases, distortion may engage typical data dredging and manipulation to find statistically significant harms that cause unnecessary panic. Probably more common is distortion in the opposite direction, where clear conflicts of interest operate, trying to hide bothersome risks under the carpet. Distortion can happen not only in the trial reports themselves but also in the accompanying literature that comments on the trial results, through editorials, expert reviews, and even biased guidelines that focus on effectiveness and neglect or distort harms. Distortion can proceed even to orchestrated silencing of the evidence as in the case of Vioxx[8] or Neurontin.[9] In these cases, marketing needs prevail over scientific accuracy and clinical prudence.

Perhaps conflicts of interest and marketing rather than science have shaped even the often accepted standard that randomized trials study primarily effectiveness, whereas information on harms from medical interventions can wait for case reports and nonrandomized studies. Nonrandomized data are very helpful,[10] but they have limitations, and many harms will remain long undetected if we just wait for spontaneous reporting and other nonrandomized research to reveal them. In an environment where effectiveness benefits are small and shrinking, the randomized trials agenda may need to reprogram its whole mission, including its reporting, toward better understanding of harms.

John P. A. Ioannidis, MD

Correspondence: Dr Ioannidis, Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, University Campus, Ioannina 45110, Greece (jioannid@cc.uoi.gr).

Financial Disclosure: None reported.

REFERENCES

1. Pitrou I, Boutron I, Ahmad N, Ravaud P. Reporting of safety results in published reports of randomized controlled trials. *Arch Intern Med.* 2009;169(19):1756-1761.
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(10):781-788.

3. Tsang R, Colley L, Lynd LD. Inadequate statistical power to detect clinically significant differences in adverse event rates in randomized controlled trials. *J Clin Epidemiol*. 2009;62(6):609-616.
4. Ioannidis JPA. Why most published research findings are false. *PLoS Med*. 2005; 2(8):e124.
5. Ferreira-Gonzalez I, Permanyer-Miralda G, Busse JW, et al. Composite outcomes can distort the nature and magnitude of treatment benefits in clinical trials. *Ann Intern Med*. 2009;150(8):566-567.
6. Hernandez AV, Walker E, Ioannidis JP, Kattan MW. Challenges in meta-analysis of randomized clinical trials for rare harmful cardiovascular events: the case of rosiglitazone. *Am Heart J*. 2008;156(1):23-30.
7. Mahoney MR, Sargent DJ, O'Connell MJ, Goldberg RM, Schaefer P, Buckner JC. Dealing with a deluge of data: an assessment of adverse event data on North Central Cancer Treatment Group trials. *J Clin Oncol*. 2005;23(36):9275-9281.
8. Jüni P, Nartey L, Reichenbach S, Sterchi R, Dieppe PA, Egger M. Risk of cardiovascular events and rofecoxib: cumulative meta-analysis. *Lancet*. 2004; 364(9450):2021-2029.
9. Landefeld CS, Steinman MA. The Neurontin legacy: marketing through misinformation and manipulation. *N Engl J Med*. 2009;360(2):103-106.
10. Papanikolaou PN, Christidi GD, Ioannidis JP. Comparison of evidence on harms of medical interventions in randomized and nonrandomized studies. *CMAJ*. 2006; 174(5):635-641.