Better Reporting of Harms in Randomized Trials: An Extension of the CONSORT Statement

John P.A. Ioannidis, MD; Stephen J.W. Evans, MSc; Peter C. Gøtzsche, MD, DrMedSci; Robert T. O’Neill, PhD; Douglas G. Altman, DSc; Kenneth Schulz, PhD; and David Moher, PhD, for the CONSORT Group*

In response to overwhelming evidence and the consequences of poor-quality reporting of randomized, controlled trials (RCTs), many medical journals and editorial groups have now endorsed the CONSORT (Consolidated Standards of Reporting Trials) statement, a 22-item checklist and flow diagram. Because CONSORT primarily aimed at improving the quality of reporting of efficacy, only 1 checklist item specifically addressed the reporting of safety.

Considerable evidence suggests that reporting of harms-related data from RCTs also needs improvement. Members of the CONSORT Group, including journal editors and scientists, met in Montebello, Quebec, Canada, in May 2003 to address this problem. The result is the following document: the standard CONSORT checklist with 10 new recommendations about reporting harms-related issues, accompanying explanation, and examples to highlight specific aspects of proper reporting.

We hope that this document, in conjunction with other CONSORT-related materials (www.consort-statement.org), will help authors improve their reporting of harms-related data from RCTs. Better reporting will help readers critically appraise and interpret trial results. Journals can support this goal by revising Instructions to Authors so that they refer authors to this document.


For author affiliations, see end of text.
For definitions of terms, see Glossary.
*For a list of members of the CONSORT Group, see Appendix 1, available at www.annals.org.

Improving Patient Care is a special section within Annals supported in part by the U.S. Department of Health and Human Services (HHS) Agency for Healthcare Research and Quality (AHRQ). The opinions expressed in this article are those of the authors and do not represent the position or endorsement of AHRQ or HHS.

© 2004 American College of Physicians

© 2004 American College of Physicians

Table 1

<table>
<thead>
<tr>
<th>Recommendation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>If the study collected data on harms and benefits, the title or abstract should state.</td>
</tr>
</tbody>
</table>

The title should mention harms if the study of harms was a key trial objective. Many phase I and phase II trials, some phase II/III trials, and most phase IV trials target harms as primary outcomes. Yet, the title and abstract seldom contain the word “harm.” Among 375 143 entries in the Cochrane Central Register of Controlled Trials (Cochrane Library, issue 3, 2003), searching titles with the search terms harm or harms yielded 337 references (compared with 55 374 for efficacy and 23 415 for safety). Of the 337, excluding several irrelevant articles on self-harm or harm reduction, only 3 trial reports and 2 abstracts contained the word “harm” in their titles.
Improving Patient Care

Improve the Reporting of Harms

Glossary

Adverse events: Side effects that are harmful. However, side effects suggest causality (effects caused by the tested intervention). Some authors use the term “adverse effects” synonymously with “side effects.” In the typical randomized trial, it is difficult to know whether an observed event is partially or entirely due to the intervention or whether it is totally unrelated to the intervention (for example, a consequence of the underlying disease process). The purpose of a trial is to collect and appropriately report good and bad events and outcomes so that they may be compared across treatment groups. In this regard, the term “adverse events” is probably better to describe harmful events that occur during a trial.

Adverse reaction and adverse drug reaction (ADR): Events for which a causality link to the tested intervention is well established and strong enough (sensitive and specific) to warrant attribution of the event to the intervention (for details, see definitions proposed in references 4 and 5). Attribution of causality in the setting of clinical trials may be misleading.

Harms: The totality of possible adverse consequences of an intervention or therapy; they are the direct opposite of benefits, against which they must be compared.

Passive surveillance of harms: The recorded adverse events are those that the study participants spontaneously report on their own initiative. In active surveillance of harms, participants are asked about the occurrence of specific adverse events in structured questionnaires or interviews or predefined laboratory or other diagnostic tests are performed at prespecified time intervals.

Risk–benefit ratio: The most common expression for the comparison of harms and benefits. It is a technical term that assumes that a ratio can indeed be calculated. Because the benefits and harms of an intervention are often so different in character or are measured on different scales, the term “risk–benefit ratio” has no literal meaning. In addition, there may be several distinct benefits and harms. We advocate using “balance of benefits and harms” rather than “risk–benefit ratio.”

Safety: Substantive evidence of an absence of harm. The term is often misused when there is simply absence of evidence of harm.

Serious adverse events: As defined by the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use, document E2A (available at www.ich.org/UriGrpServer.jsp?_ID=2766&_TEMPLATE=254): “During clinical investigations, adverse events may occur which, if suspected to be medicinal product-related (adverse drug reactions), might be significant enough to lead to important changes in the way the medicinal product is developed (e.g., change in dose, population, needed monitoring consent forms). This is particularly true for reactions which, in their most severe forms, threaten life or function. Such reactions should be reported promptly to regulators.”

Side effects: Unintended drug effects. The term, however, does not necessarily imply harm, as some side effects may be beneficial. Furthermore, it tends to understate the importance of harms because “side” may be perceived as denoting secondary importance.

Toxicity: Describes drug-related harms. The term may be most appropriate for laboratory-determined measurements, although it is also used in relation to clinical events. Abnormal laboratory values may be described as laboratory-determined toxicity. The disadvantage of the term “toxicity” is that it implies causality. If authors cannot prove causality, the terms “abnormal laboratory measurements” or “laboratory abnormalities” are more appropriate to use.

Authors should present information on harms in the abstract. If no important harms occurred, authors should state. Explicit reference to the reporting of adverse events in the title or abstract is also important for appropriate database indexing and information retrieval (19).

Introduction

Background

Recommendation 2. If the trial addresses both harms and benefits, the introduction should so state.

The introduction states the scientific background and rationale of an RCT. This requires a balanced presentation whereby the possible benefits of the intervention under investigation are outlined along with the possible harms associated with the treatment. Randomized, controlled trials that focus primarily on harms should clearly state this interest when describing the study objectives in the Introduction and in defining these objectives in the Methods.

Methods

Outcomes

Recommendation 3. List addressed adverse events with definitions for each (with attention, when relevant, to grading, expected vs. unexpected events, reference to standardized and validated definitions, and description of new definitions).

The Methods section should succinctly define the recorded adverse events (clinical and laboratory). Authors should clarify whether the reported adverse events encompass all the recorded adverse events or a selected sample. They should explain how, why, and who selected adverse events for reporting. In trials that do not mention harms-related data, the Methods section should briefly explain the reason for the omission (for example, “the design did not include the collection of any information on harms”).

Authors should also be explicit about separately reporting anticipated and unexpected adverse events. Expectation may influence the incidence of reported or ascertained adverse events. Making participants aware in the consent form of the possibility of a specific adverse event (“priming”) may increase the reporting rate of the event (20). Another example of priming is the finding that the rates of withdrawals due to adverse events and the rates of specific adverse events were significantly higher in trials of aspirin, diclofenac, or indomethacin with comparator drugs compared with placebo-controlled trials (21). Presumably, participants were more eager to come forth and report an adverse event or to withdraw from treatment when they knew they could not be receiving inactive placebo.

Authors should report whether they used standardized and validated measurement instruments for adverse events. Several medical fields have developed standardized scales (22–32). Use of nonvalidated scales is common. The source document for well-established definitions and scales should be referenced. New definitions for adverse events should be explicit and clear. Authors should describe how they developed and validated new scales.

For interventions that target healthy individuals (for example, many preventive interventions), any harm, however minor, may be important to capture and report because the balance of harms and benefits may easily lean toward harms in a low-risk population. For other popula-
Recommendation 4. Clarify how harms-related information was collected (mode of data collection, timing, attribution methods, intensity of ascertainment, and harms-related monitoring and stopping rules, if pertinent).

It is important to describe the questionnaires, interviews, and tests used to collect information on harms, as well as their timing during follow-up. Passive surveillance of harms leads to fewer recorded adverse events than active surveillance (4). Open-ended questions may yield different information, both quantitatively and qualitatively, than structured questionnaires (33). Studies of nonsteroidal, anti-inflammatory drugs (NSAIDs) exemplify how data...

Table 1. Original CONSORT Checklist

<table>
<thead>
<tr>
<th>Paper Section and Topic</th>
<th>Item Number</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title and abstract</td>
<td>1</td>
<td>How participants were allocated to interventions (e.g., “random allocation”, “randomized”, or “randomly assigned”).</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Introduction</td>
<td>2</td>
<td>Scientific background and explanation of rationale.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Methods</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participants</td>
<td>3</td>
<td>Eligibility criteria for participants and the settings and locations where the data were collected.</td>
</tr>
<tr>
<td>Interventions</td>
<td>4</td>
<td>Precise details of the interventions intended for each group and how and when they were actually administered.</td>
</tr>
<tr>
<td>Objectives</td>
<td>5</td>
<td>Specific objectives and hypotheses.</td>
</tr>
<tr>
<td>Outcomes</td>
<td>6</td>
<td>Clearly defined primary and secondary outcome measures and, when applicable, any methods used to enhance the quality of measurements (e.g., multiple observations, training of assessors).</td>
</tr>
<tr>
<td>Sample size</td>
<td>7</td>
<td>How sample size was determined and, when applicable, explanation of any interim analyses and stopping rules.</td>
</tr>
<tr>
<td>Randomization</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sequence generation</td>
<td>8</td>
<td>Method used to generate the random allocation sequence, including details of any restriction (e.g., blocking, stratification).</td>
</tr>
<tr>
<td>Allocation concealment</td>
<td>9</td>
<td>Method used to implement the random allocation sequence (e.g., numbered containers or central telephone), clarifying whether the sequence was concealed until interventions were assigned.</td>
</tr>
<tr>
<td>Implementation</td>
<td>10</td>
<td>Who generated the allocation sequence, who enrolled participants, and who assigned participants to their groups.</td>
</tr>
<tr>
<td>Blinding (masking)</td>
<td>11</td>
<td>Whether or not participants, those administering the interventions, and those assessing the outcomes were blinded to group assignment. If done, how the success of blinding was evaluated.</td>
</tr>
<tr>
<td>Statistical methods</td>
<td>12</td>
<td>Statistical methods used to compare groups for primary outcome(s); Methods for additional analyses, such as subgroup analyses and adjusted analyses.</td>
</tr>
<tr>
<td>Results</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participant flow</td>
<td>13</td>
<td>Flow of participants through each stage (a diagram is strongly recommended). Specifically, for each group report the numbers of participants randomly assigned, receiving intended treatment, completing the study protocol, and analyzed for the primary outcome. Describe protocol deviations from study as planned, together with reasons.</td>
</tr>
<tr>
<td>Recruitment</td>
<td>14</td>
<td>Dates defining the periods of recruitment and follow-up.</td>
</tr>
<tr>
<td>Baseline data</td>
<td>15</td>
<td>Baseline demographic and clinical characteristics of each group.</td>
</tr>
<tr>
<td>Numbers analyzed</td>
<td>16</td>
<td>Number of participants (denominator) in each group included in each analysis and whether the analysis was by “intention-to-treat”. State the results in absolute numbers when feasible (e.g., 10/20, not 50%).</td>
</tr>
<tr>
<td>Outcomes and estimation</td>
<td>17</td>
<td>For each primary and secondary outcome, a summary of results for each group, and the estimated effect size and its precision (e.g., 95% confidence interval). Address multiplicity by reporting any other analyses performed, including subgroup analyses and adjusted analyses, indicating those pre-specified and those exploratory.</td>
</tr>
<tr>
<td>Ancillary analyses</td>
<td>18</td>
<td>All important adverse events or side effects in each intervention group.</td>
</tr>
<tr>
<td>Discussion</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interpretation</td>
<td>20</td>
<td>Interpretation of the results, taking into account study hypotheses, sources of potential bias or imprecision and the dangers associated with multiplicity of analyses and outcomes.</td>
</tr>
<tr>
<td>Generalizability</td>
<td>21</td>
<td>Generalizability (external validity) of the trial findings.</td>
</tr>
<tr>
<td>Overall evidence</td>
<td>22</td>
<td>General interpretation of the results in the context of current evidence.</td>
</tr>
</tbody>
</table>
collection methods can affect the detection and reporting of harms. When selective NSAIDs with fewer gastrointestinal adverse events became available, trials reported more than 10 times as many ulcers when comparing these drugs with older NSAIDs as when older NSAIDs were compared with placebo. In the newer trials, more ulcers were detected because participants had regular endoscopy, and the case definition of ulcers was more sensitive (34).

Authors should specify the time frame of surveillance for adverse events. Some investigators stop recording adverse events at the end of the intervention period or a certain number of days afterward (for example, 30 days after discontinuation of drug therapy) and miss events with long latency (35). Surgical trials often capture only the adverse events that occur intraoperatively. Several important surgical complications are likely to occur later. Finally,

### Table 2. Checklist of Items To Include When Reporting Harms in Randomized, Controlled Trials*

<table>
<thead>
<tr>
<th>Standard CONSORT Checklist: Paper Section and Topic</th>
<th>Standard CONSORT Checklist: Item Number</th>
<th>Descriptor</th>
<th>Reported on Page Number</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Title and abstract</strong></td>
<td>1</td>
<td>If the study collected data on harms and benefits, the title or abstract should so state.</td>
<td></td>
</tr>
<tr>
<td><strong>Introduction</strong></td>
<td>2</td>
<td>If the trial addresses both harms and benefits, the introduction should so state.</td>
<td></td>
</tr>
<tr>
<td><strong>Methods</strong></td>
<td></td>
<td>List addressed adverse events with definitions for each (with attention, when relevant, to grading, expected vs. unexpected events, reference to standardized and validated definitions, and description of new definitions). Clarify how harms-related information was collected (mode of data collection, timing, attribution methods, intensity of ascertainment, and harms-related monitoring and stopping rules, if pertinent).</td>
<td></td>
</tr>
<tr>
<td><strong>Sample size</strong></td>
<td>7</td>
<td>Describe plans for presenting and analyzing information on harms (including coding, handling of recurrent events, specification of timing issues, appropriate metrics for recurrent events, continuous variables, and scale variables, whenever pertinent.†</td>
<td></td>
</tr>
<tr>
<td><strong>Randomization</strong></td>
<td></td>
<td>Describe for each arm the participant withdrawals that are due to harms and their experiences with the allocated treatment.</td>
<td></td>
</tr>
<tr>
<td><strong>Recruitment</strong></td>
<td>13</td>
<td>Provide the denominators for analyses on harms.</td>
<td></td>
</tr>
<tr>
<td><strong>Baseline data</strong></td>
<td>14</td>
<td>Present the absolute risk per arm and per adverse event type, grade, and seriousness, and present appropriate metrics for recurrent events, continuous variables, and scale variables, whenever pertinent.†</td>
<td></td>
</tr>
<tr>
<td><strong>Numbers analyzed</strong></td>
<td>15</td>
<td>Describe any subgroup analyses and exploratory analyses for harms.†</td>
<td></td>
</tr>
<tr>
<td><strong>Outcomes and estimation</strong></td>
<td>16</td>
<td>Provide a balanced discussion of benefits and harms with emphasis on study limitations, generalizability, and other sources of information on harms.‡</td>
<td></td>
</tr>
<tr>
<td><strong>Adverse events</strong></td>
<td>17</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Discussion</strong></td>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Interpretation</strong></td>
<td>21</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Generalizability</strong></td>
<td>22</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Overall evidence</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* This proposed extension for harms includes 10 recommendations that correspond to the original CONSORT checklist.
† Descriptors refer to items 17, 18, and 19.
‡ Descriptor refers to items 20, 21, and 22.
in crossover trials, delayed events might occur while the patient is taking a subsequent assigned treatment.

Attribution is the process of deciding whether an adverse event is related to the intervention. Whenever authors filter events through an attribution process, they should state who makes the attribution (investigators, participants, sponsors, or combinations), whether the process is blinded to assigned treatment, and what definitions of adverse events they use (4).

Discontinuations and withdrawals due to adverse events are especially important because they reflect the ultimate decision of the participant and/or physician to discontinue treatment. Although treatment may occasionally be discontinued for mild or moderate adverse events, attributing discontinuation to a specific reason (to toxicity, lack of efficacy, other reasons, or combinations of reasons) may be difficult. For example, in psychopharmacology, dropouts may reflect treatment ineffectiveness as much as toxicity-related intolerance (36). Trial reports should specify who gave the reasons for discontinuation (participants or physicians) and whether attribution was blinded to the assigned treatment. For example, even in blinded trials, participants and their clinicians are often unblinded before they decide whether to discontinue the intervention. It is important to report participants who are nonadherent or lost to follow-up because their actions may reflect their inability to tolerate the intervention. Moreover, authors should specify how they handled withdrawals in the analyses of the data.

Randomized, controlled trials should report any plan for monitoring for harms and rules for stopping the trial because of harms (37). They should clarify whether stopping guidelines examine benefits and harms separately or evaluate a composite measure that reflects the trade-off between benefits and harms (38).

**Statistical Methods**

Recommendation 5. Describe plans for presenting and analyzing information on harms (including coding, handling of recurrent events, specification of timing issues, handling of continuous measures, and any statistical analyses).

Using only descriptive statistics to report harms is perfectly appropriate in most RCTs because most trials lack power to test harms-related hypotheses and indeed have no explicit prespecified harms-related hypotheses. If investigators combine data for different adverse events into 1 outcome measure, they should describe each combination, cite the dictionary that lists the definitions of the adverse events, and state whether they decided the grouping of events post hoc or a priori.

The distributions of adverse events over the follow-up period can pose problems for analysis of the data. When pertinent, authors should specify whether they count recurrent events (events that occur more than once in the same participant) as separate events or as 1 event. For trials with longitudinal follow-up, specifying the timing of the events may be important (for example, to separate early from late toxicity). Incidence rates, period prevalence rates, and point prevalence rates may provide complementary information about the occurrence of an adverse event. Kaplan–Meier curves showing cumulative incidence of important adverse events can be helpful. Simple summaries with person-time denominators (for example, median months after treatment) can be misleading if the event occurs only after extended treatment and long follow-up, and most participants had short follow-up and therefore no events.

For continuous variables (such as reported for most laboratory tests), means and SDs or medians and interquartile ranges may provide an aggregate picture, but they may not convey information on extreme values that correspond to severe toxicity. Means and medians may be useful in informing participants and clinicians about expected, relatively minor changes.

Scales are increasingly used for measuring quality of life in RCTs. These measures are composite outcomes that reflect both benefits and harms (39). Authors should describe the development of these instruments, their validity and sensitivity to detect change, and whether they assumed interval scaling in order to use the scale as a continuous variable.

When harms are major primary or secondary outcomes of a trial, the authors should describe plans to perform any formal statistical analyses and inferences. They should separate prespecified statistical analyses from post hoc analyses (40) and address common problems: low power for uncommon events, adjustment for multiple outcomes, composite outcomes, regression to the mean (for example, for laboratory tests that are also used for screening for study eligibility), and heterogeneity of treatment effects across prespecified subgroups (41).

**Results**

**Participant Flow**

Recommendation 6. Describe for each arm the participant withdrawals that are due to harms and the experience with the allocated treatment.

Authors should describe the reasons for discontinuations and reductions in dosage of the allocated treatment and withdrawals from the study. They should emphasize harms-related reasons and acknowledge the caveats noted under Recommendation 6. Authors should always report deaths in each study group during a trial, regardless of whether death is an end point and regardless of whether attribution to a specific cause is possible (42).

Randomized, controlled trials with prolonged follow-up should report the timing of allocated treatment received, dose reductions and discontinuations, and study withdrawals. The cause of early withdrawals may differ from that of late withdrawals; separate descriptions of each may enhance the accuracy of information on the tolerability profile of an intervention. Kaplan–Meier plots of the
Table 3. Common Poor Reporting Practices for Harms-Related Data

| 1. Using generic or vague statements, such as “the drug was generally well tolerated” or “the comparator drug was relatively poorly tolerated.” |
| 2. Failing to provide separate data for each study arm. |
| 3. Providing summed numbers for all adverse events for each study arm, without separate data for each type of adverse event. |
| 4. Providing summed numbers for a specific type of adverse event, regardless of severity or seriousness. |
| 5. Reporting only the adverse events observed at a certain frequency or rate threshold (for example, >3% or >10% of participants). |
| 6. Reporting only the adverse events that reach a P value threshold in the comparison of the randomized arms (for example, \( P < 0.05 \)). |
| 7. Reporting measures of central tendency (for example, means or medians) for continuous variables without any information on extreme values. |
| 8. Improperly handling or disregarding the relative timing of the events, when timing is an important determinant of the adverse event in question. |
| 9. Not distinguishing between patients with 1 adverse event and participants with multiple adverse events. |
| 10. Providing statements about whether data were statistically significant without giving the exact counts of events. |
| 11. Not providing data on harms for all randomly assigned participants. |

Moreover, authors should state whether they use the same type of analysis for both efficacy end points and harms.

Rates of Outcomes and Ancillary Analyses for Adverse Events

Recall the absolute risk of each adverse event (specifying type, grade, and seriousness per arm), and present appropriate metrics for recurrent events, continuous variables and scale variables, whenever pertinent.

Authors should present results separately for each study group of a trial. For each type of adverse event, they should offer appropriate metrics of absolute risk (for example, frequency or incidence), with separate information about the severity grade of the event, if relevant. Serious events should be reported separately for each type of event. Recurrent events and timing of events need appropriate reporting, as discussed in Recommendation 5. For events with many recurrences, it is useful to provide both the number of affected participants and the number of events for each study group and rate (events per unit of person-time). Occasionally, a graphical representation of the distribution of number of events per patient or time-to-event analyses may be informative.

Overall, the Results section should report on what the Methods section promises (43). Any break from this symmetry requires explanation. If no adverse events of a specific type and severity occurred, authors should state so in the Results section (44). Table 3 shows common reporting practices to avoid.

Recommendation 9. Describe any subgroup analyses and exploratory analyses for harms.

Reporting of adverse events for different participant subgroups follows the same principles that govern the reporting of subgroup analyses for efficacy. Authors should avoid overstating the significance of false-positive subgroup findings (45). Authors should state how, why, and when they planned subgroup analyses (a priori or post hoc). Regulatory agencies increasingly require subgroup analyses by age, sex, and race for license applications. However, these variables rarely show any significant effect modification for efficacy outcomes (45, 46) and may be equally low-yield (or even misleading) for harms.

Discussion

Recommendation 10. Provide a balanced discussion of benefits and harms with emphasis on study limitations, generalizability, and other sources of information on harms.

The Discussion is typically the most poorly structured section of an RCT report and could be modified to gain space for reporting harms. Authors may curtail the length of the Discussion section to gain space for appropriate reporting of harms.

In summarizing the key findings of an RCT, the Discussion section should pay attention to the original trial objectives and provide a balanced view that puts the benefits and harms into perspective. Authors should avoid overinterpretation of the findings. Limitations are probably
the most important part of the Discussion section. Common limitations in studies that report harms include inconclusive findings, lack of power, multiplicity of comparisons, post hoc analyses, and short duration of exposure to the allocated treatment, especially for treatments of chronic diseases. Generalizability is often a problem for harms. The frequency and severity of adverse events may depend on the clinical setting and participants. Often, clinical trials enroll participants who have the disease of interest but are otherwise healthy and do not have comorbid conditions. Once licensed, however, most approved interventions are used in individuals who have several comorbid conditions and who are taking several other drugs with potentially additive or synergistic toxicity.

The Discussion section should also appraise emerging data on benefits and harms. Authors should systematically integrate prior evidence on harms, whenever possible (47). If a systematic review of previous studies of harms is not possible, authors should state, perhaps to stimulate someone to correct the deficiency in the future. Authors should contrast the trial results on harms with other sources of information on harms, including observational data from spontaneous reporting, automated databases, case–control studies, and case reports.

**Manuscript Length**

Improved reporting of harms need not lead to longer manuscripts. In the Methods section, adopting standard definitions with appropriate references may actually save space. Tables may summarize key results on harms. Graphs may convey important time-to-event outcomes or repeated measurements of adverse events. Finally, it is possible to write short Discussion sections by using an appropriate structure (48).

Occasionally, investigators present adverse events separately in another paper. This practice denies both author and reader the opportunity to formulate the balance between benefits and harms. Therefore, authors should report harms and benefits together in the same manuscript. Authors and journal editors should publish additional trial information on the Web, as an adjunct to the main results. For example, Web material could include enhanced graphical displays that present individual participants’ experiences (49).

Investigators may sparingly use single patient reports embedded in the results of an RCT to describe severe, serious, and previously unreported adverse events. More comprehensive case reports may require a separate paper. In the RCT report, a single sentence may suffice to describe an unusual adverse event by succinctly summarizing the type of adverse event, when it occurred, the type of patient, the management of the adverse event, and the outcome of the adverse event.

**Conclusions**

This extension of the main CONSORT statement to include harms is a work in progress. We therefore invite readers to submit comments, critique, and suggestions for improvement through www.consort-statement.org. We also hope that journals and editorial groups will support our efforts to improve the reporting of harms. We ask journals that endorse the CONSORT reporting requirement to include a reference to this document in their Instructions to Authors section. Adherence to reporting standards for harms should help to inform readers and the public on the harms of interventions.

From University of Ioannina School of Medicine and Biomedical Research Institute, Foundation for Research and Technology–Hellas, Ioannina, Greece; London School of Hygiene and Tropical Medicine, London, United Kingdom; The Nordic Cochrane Centre, Copenhagen, Denmark; U.S. Food and Drug Administration, Rockville, Maryland; Cancer Research UK/National Health Service Centre for Statistics in Medicine, Oxford, United Kingdom; Family Health International, Research Triangle Park, North Carolina; and Children’s Hospital of Eastern Ontario Research Institute, Ottawa, Ontario, Canada.

**Disclaimer:** Dr. O’Neill is an employee of the U.S. Food and Drug Administration. The views expressed are those of the author and do not necessarily represent those of the U.S. Food and Drug Administration.

**Grant Support:** Abbott Laboratories, Children’s Hospital of Eastern Ontario Research Institute, Health Canada, The Lancet, and Merck and Co., Inc. provided funding for the Montebello meeting. Dr. Altman is supported by Cancer Research UK.

**Potential Financial Conflicts of Interest:** None disclosed.

**Requests for Single Reprints:** David Moher, PhD, Chalmers Research Group, Children’s Hospital of Eastern Ontario Research Institute, 401 Smyth Road, Ottawa, Ontario K1H 8L1, Canada; e-mail, dmoher@uottawa.ca.

Current author addresses are available at www.annals.org.

**References**


Improving the Reporting of Harms

44. Jovanovic BD, Zalenski RJ. Safety evaluation and confidence intervals when the number of observed events is small or zero. Ann Emerg Med. 1997;30:301-6. [PMID: 9287891]
APPENDIX 1: MEMBERS OF THE CONSORT GROUP
Kamran Abbasi, MB (British Medical Journal); Douglas Altman, DSc (Centre for Statistics in Medicine, Oxford, United Kingdom); Matthias Egger, MD (Bristol University, Bristol, United Kingdom, and University of Bern, Bern, Switzerland; International Journal of Epidemiology); Diana Elbourne, PhD (London School of Hygiene and Tropical Medicine, London, United Kingdom); Stephen J.W. Evans, MSc (London School of Hygiene and Tropical Medicine, London, United Kingdom); Joel Gagnier, ND (University of Toronto, Toronto, Ontario, Canada); Peter Gøtzsche, MD (Nordic Cochrane Centre, Copenhagen, Denmark); David Grimes, MD (Family Health International, Research Triangle Park, North Carolina); Jaques, MD (Memorial Sloan-Kettering Cancer Center, New York, New York); Tom Lang, MA (Tom Lang Communications, Lakewood, Ohio), Joseph Lau, MD (Tufts-New England Medical Center, Boston, Massachusetts); Yoon Kong Loke, MD (University of Oxford, Oxford, United Kingdom); Faith McLellan, PhD (The Lancet); Siddika Mithani, MD (Health Canada – Clinical Trials, Ottawa, Ontario, Canada); David Moher, PhD (Chalmers Research Group and University of Ottawa, Ottawa, Ontario, Canada); Cynthia Mulrow, MD (Annals of Internal Medicine and University of Texas Health Sciences Center, San Antonio, Texas); Robert O’Neill, PhD (Center for Drug Evaluation and Research, U.S. Food and Drug Administration, Rockville, MD); Gilda Piaggio, PhD (World Health Organization, Geneva, Switzerland); Drummond Rennie, MD (University of California at San Francisco, San Francisco, California; Journal of the American Medical Association); David Riley, MD (University of New Mexico Medical School, Santa Fe, New Mexico; EXPLORE: The Journal of Science and Healing); David Sackett, MD (Trout Research and Education Centre, Irish Lake, Ontario, Canada); Roberta Scherer, PhD (University of Maryland School of Medicine, Baltimore, Maryland); Kenneth Schulz, PhD (Family Health International, Chapel Hill, North Carolina); James Scott, MD (University of Utah, Salt Lake City, Utah; Obstetrics and Gynecology). Observers: Don Husereau, BScPharm, MSc (Canadian Coordinating Office for Health Technology Assessment, Ottawa, Ontario, Canada); Andra Morrison, BSc (Chalmers Research Group, Ottawa, Ontario, Canada); Isabelle French, BSc (Dalhousie University, Halifax, Nova Scotia, Canada); Gabriela Lewin, MD (Chalmers Research Group, Ottawa, Ontario, Canada).

Note: Laurence Hirsch, MD, and Beate Stych, MD (Merck), participated in the meetings as observers.

APPENDIX 2: EXAMPLES OF RECOMMENDATIONS

Recommendation 1

Title with benefits and harms:
The safety and efficacy of prophylactic ondansetron in patients undergoing modified radical mastectomy (50).

Recommendation 2

Title of trial with primary harms outcome:
An assessment of the safety of pediatric ibuprofen. A practitioner-based randomized clinical trial (17).

Title with emphasis on harms:
No benefit, but increased harm from high dose (100 μg) misoprostol for induction of labour: a randomised trial of high vs. low (50 μg) dose misoprostol (51).

Abstract:
There were two uterine ruptures and four intrapartum stillbirths in the high misoprostol group. There was no difference in postpartum haemorrhage, 9.5% vs. 7.9% (P = 1.00) and admissions to the neonatal unit 18.8% vs. 17.0% (P = 0.980) in the low- and high-groups respectively. . . . the higher dose had an increased risk of serious complications (51).

Abstract conclusion:
CONCLUSION: In postmenopausal women with coronary disease, neither hormonal replacement therapy nor antioxidant vitamin supplements provide cardiovascular benefit. Instead, a potential for harm was suggested with each treatment (52).

Recommendation 3

Comprehensive list:
. . . assessments included minimum systolic blood pressure, minimum pressure of oxygen, maximum concentration of delivered nitrogen dioxide, and maximum concentration of methemoglobin (54).

Comprehensive list with definitions:
Bleeding complications and other adverse events were documented by interview or were reported by the pa-
patients during the study. Major bleeding was defined as any clinically apparent bleeding associated with a decrease of at least 2.0 g per dL in the hemoglobin level, requirement for transfusion of at least 2 units of packed red cells, or retroperitoneal or intracranial bleeding or other bleeding that the investigators decided required permanent discontinuation of treatment. Bleeding that did not meet this definition was considered minor. An adverse event was considered serious if it was fatal or life-threatening, caused permanent disability, or required hospitalization or prolonged hospitalization (55).

Definitions and grading (referral to established system):

... using the AIDS Clinical Trials Group adverse event grading scheme (56).

New definition:

In June 1998, the protocol was amended to include laboratory monitoring, toxicity management, and dose reduction of adefovir dipivoxil for proximal renal tubular dysfunction (PRTD), because of new information about the toxicity provided by Gilead Sciences. The protocol definition of PRTD was as follows: serum creatinine 0.5 mg/dL above baseline and serum phosphate < 2.0 mg/dL, or 1 of these abnormalities plus 2 of the following: proteinuria (2+), glycosuria (1+) in the absence of hyperglycemia, hypokalemia (<3.0 mEq/L), or serum bicarbonate <19 mEq/L (57).

Recommendation 4

Mode of data collection:

At each semiannual contact, a standardized interview collected information on designated symptoms and [harms] concerns, and initial reports of outcome events were obtained using a self-administered questionnaire (58).

Timing:

Adverse experiences and toxic effects were assessed every 4 weeks until 12 weeks after the discontinuation of the study drugs (59).

Attribution methods:

Causality was assessed by the investigator at the time of the event, using a modified version of Karch and Lasagna’s 5-point scale (60).

Harms-related stopping rules:

Trial monitoring guidelines for early stopping considerations were based on O’Brien-Fleming boundaries using asymmetric upper and lower boundaries: a 1-sided, .05-level upper boundary for benefit and 1-sided, .05-level lower boundaries for adverse effects. The adverse-effect boundaries were further adjusted with a Bonferroni correction for the 7 major outcomes other than breast cancer that were specifically monitored (58).

Recommendation 5

Coding:

More than 200 distinct types of clinical, laboratory or ECG events were noted using ... literal description, but the FDA coding symbols for the Thesaurus of Adverse Reaction Terms (COSTART), used to report adverse events, reduced that number to 110. Since some adverse events were similar, i.e. fatigue, asthenia, feeling uncase, they were also regrouped, resulting in ... 37 [descriptions] ... (60)

Timing issues—early vs. late events:

Early reactions. Early reactions were recorded by the nursing staff in the cardiac catheter laboratory and on the cardiology ward after the patient left the catheter laboratory. ... Late reactions. Each patient was asked to complete a simple questionnaire after discharge from the hospital, on which to record any adverse reactions occurring within 1 week of the cardiac catheterization. ... The analysis of patients with late skin reactions was confined to patients with reactions that had clearly started after hospital discharge and therefore were not a [Note: use of the term “reactions” may not be optimal, since it implies causality, but the example is appropriate for timing issues].

Continuous measures (mean estimates and serious extremes):

... assessed by changes in vital signs (summarized as a mean [SE] change from baseline) and by reports of adverse events with onset within 8 weeks of randomization. All reports of adverse events were included whether or not they were deemed by the investigator to be related to treatment. An adverse event was defined in the study protocol as serious if it was fatal or life-threatening, required or prolonged hospitalization, or resulted in persistent or significant disability or incapacity (62).

Statistical analyses:

Patients reported occurrence and severity of 18 side effects. ... Between-group differences were compared by using the Wilcoxon test (63).

Recommendation 6

Withdrawals and harm-related reasons for withdrawals per arm:

Eight other patients (6 treated with fluconazole and 2 treated with itraconazole) were withdrawn from the
study because of mild to moderate symptoms, such as rash (fluconazole group), dry skin (fluconazole group), nausea (fluconazole group), or difficulty concentrating (itraconazole group) (64).

No withdrawals due to adverse events:

No patients receiving trimethoprim–sulfamethoxazole or ciprofloxacin discontinued therapy with the drug because of side effects (65).

_Treatment experience over time:_

A substantial number of women had stopped taking study drugs at some time (42% of estrogen plus progestin and 38% of placebo) [Figure shows cumulative dropout rates over time] (58).

_Treatment exposure:_

The study drug exposure was as follows: 15% received 1 dose; 31% received 2 doses; 37% received 3 doses; and 17% received 4 doses (66).

**Recommendation 7**

All randomly assigned and treated patients (n = 68) were included in the . . . [harms] . . . analysis (67).

**Recommendation 8**

Absolute risks for binary events per arm and per type and grade (follow-up/exposure time approximately comparable for all participants):

See Appendix Table 1.

Absolute risks for binary events per arm and per type and grade (follow-up/exposure time is differential and not comparable for all participants):

45 patients (46 events) in the rofecoxib group and 20 patients (20 events) in the naproxen group were adjudicated to have serious thrombotic cardiovascular adverse events (myocardial infarction, unstable angina, cardiac thrombus, resuscitated cardiac arrest, sudden or unexplained death, ischemic stroke, and transient ischemic attacks). Event-free survival analysis of these 66 patients showed that the RR (95% confidence interval [CI]) of developing a cardiovascular event in the rofecoxib treatment group was 2.38 (1.39-4.00), P < .001 [Figure shows the Kaplan–Meier plots for time to cardiovascular adverse event in each arm] (69).

Recurrent events expressed with person-time denominator:

The frequency of hypoglycemia at 3:00 a.m. was greater in the mixed-treatment period than in the split-treatment period (0.28 [SD 0.04] episode/patient-day vs. 0.10 [SD 0.02] episode/patient-day, respectively;

**Appendix Table 1. Adverse Events among Human Immunodeficiency Virus-Infected Patients Whose Aphthous Ulcers Had Healed Previously When Treated with Thalidomide and Who Then Were Treated with Thalidomide or Placebo in a Maintenance-Phase Study**

<table>
<thead>
<tr>
<th>Adverse Event</th>
<th>Patients Taking Thalidomide</th>
<th>Patients Taking Placebo</th>
</tr>
</thead>
<tbody>
<tr>
<td>(n = 23), n (%)</td>
<td>n (%)</td>
<td>n (%)</td>
</tr>
<tr>
<td>Grade 3 or higher</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elevated absolute neutrophil count</td>
<td>5 (22)</td>
<td>0</td>
</tr>
<tr>
<td>Fatigue</td>
<td>3 (13)</td>
<td>0</td>
</tr>
<tr>
<td>Weakness</td>
<td>1 (4)</td>
<td>0</td>
</tr>
<tr>
<td>Nausea or vomiting</td>
<td>1 (4)</td>
<td>0</td>
</tr>
<tr>
<td>Elevated triglyceride level</td>
<td>1 (4)</td>
<td>0</td>
</tr>
<tr>
<td>Ataxia</td>
<td>1 (4)</td>
<td>0</td>
</tr>
<tr>
<td>Fever</td>
<td>0</td>
<td>1 (4)</td>
</tr>
<tr>
<td>Seizures</td>
<td>0</td>
<td>1 (4)</td>
</tr>
<tr>
<td>Total</td>
<td>12 (52)</td>
<td>2 (8)</td>
</tr>
<tr>
<td>Any grade</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rash</td>
<td>5 (22)</td>
<td>2 (8)</td>
</tr>
<tr>
<td>Peripheral neuropathy</td>
<td>3 (13)</td>
<td>2 (8)</td>
</tr>
<tr>
<td>Somnolence</td>
<td>0</td>
<td>1 (4)</td>
</tr>
<tr>
<td>Total</td>
<td>8 (35)</td>
<td>5 (19)</td>
</tr>
</tbody>
</table>


P = 0.0022 [table also shown]. On average, each patient experienced 2.8 (95% CI, 1.9 to 3.7) fewer episodes of hypoglycemia with split than with mixed dosing [Figure also shown] (70).

Continuous measures presented with information on both averages and extremes:

There were small changes in mean (SE) systolic blood pressure (placebo group, −2.0 [0.5] mm Hg and carve-

**Appendix Table 2. Adverse Events during the First 8 Weeks**

<table>
<thead>
<tr>
<th>Adverse Event</th>
<th>Patients Taking Placebo n (n = 1133), n (%)</th>
<th>Patients Taking Carvedilol n (n = 1156), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bradyarrhythmia</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Serious adverse event</td>
<td>1 (0.1)</td>
<td>10 (0.9)</td>
</tr>
<tr>
<td>Trial drug decreased</td>
<td>3 (0.3)</td>
<td>27 (2.3)</td>
</tr>
<tr>
<td>Withdrawn due to adverse event</td>
<td>0</td>
<td>4 (0.3)</td>
</tr>
<tr>
<td><strong>Hypotension</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Serious adverse event</td>
<td>2 (0.2)</td>
<td>6 (0.5)</td>
</tr>
<tr>
<td>Trial drug decreased</td>
<td>11 (1.0)</td>
<td>38 (3.3)</td>
</tr>
<tr>
<td>Withdrawn due to adverse event</td>
<td>0 (0.0)</td>
<td>5 (0.4)</td>
</tr>
</tbody>
</table>

dilol group, −3.6 [0.5] mm Hg) and in diastolic blood pressure (placebo group, −1.8 [0.3] mm Hg and carvedilol group, −2.7 [0.3] mm Hg) at the end of 8 weeks. . . . Patients in the carvedilol group were more likely than in the placebo group to report . . . hypotension . . . and bradycardia [see Appendix Table 2] (62).

Recommendation 9

Across the same subgroups shown in Figures 1–3, the relative risk of “any bleeding” with treatment compared with placebo ranged from 0.65 to 1.86 (data not shown). No statistically significant treatment-by-subgroup interactions were noted . . . Using the same consistency criteria employed for mortality, “any bleeding” and “serious bleeding” results for all subgroups were consistent with the overall trial results. For bleeding event end points (any or serious bleeding), there was no statistically significant interaction with predicted risk of mortality (P = .55 and P = .21) (72).

Recommendation 10

Discussion of harms:

[In] our study pretreatment with a small dose of subcutaneous adrenaline significantly reduced the incidence of acute adverse reactions to polyvalent antivenom serum. . . . We did not encounter significant adverse effects attributable to it; there were no cases of acute neurological deficit suggestive of cerebrovascular accidents or patients in whom blood pressure rose significantly (73).

Limitations:

Even this large study was still too small to allow for a rigorous . . . assessment [of harms]. The data on drug related clinical adverse experiences show that etoricoxib may be better tolerated than indometacin [sic], but additional studies are needed to enable any definitive conclusions . . . (74).

Other sources of information:

Adefovir dipivoxil-related PRTD occurred in 1% of study patients during the first 16 weeks. In other studies, 35% of patients taking adefovir dipivoxil developed significant increases in serum creatinine levels, and 50% developed significant hypophosphatemia, laboratory abnormalities consistent with PRTD, by 48 weeks (57).

Tropical Medicine, Keppel Street, London WC1E 7HT, United Kingdom.

Dr. Gøtzsche: The Nordic Cochrane Centre, Rigshospitalet, Department 7112, Blegdamsvej 9, DK-2100 Copenhagen, Denmark.

Dr. O’Neill: Center for Drug Evaluation and Research, U.S. Food and Drug Administration, 5600 Fishers Lane, Rockville, MD 20857.

Dr. Altman: Cancer Research UK/NHS Centre for Statistics in Medicine, Old Road Campus, Old Road, Headington, Oxford OX3 7LF, United Kingdom.

Dr. Schultz: Family Health International, PO Box 13950, Research Triangle Park, NC 27709.

Dr. Moher: Chalmers Research Group, Children’s Hospital of Eastern Ontario Research Institute, 401 Smyth Road, Ottawa, Ontario K1H 8L1, Canada.

Web-Only References


