How well is the clinical importance of study results reported? An assessment of randomized controlled trials

Karen B.Y. Chan,* Malcolm Man-Son-Hing,*† Frank J. Molnar,*† Andreas Laupacis*‡

Abstract

Background: The interpretation of the results of randomized controlled trials (RCTs) has traditionally emphasized statistical significance rather than clinical importance. Our aim was to assess the quality of reporting of factors related to clinical importance in a sample of published RCTs.

Methods: A random sample of 27 (of a total of 266) RCTs published in 5 major medical journals over a 1-year period were reviewed by 4 independent reviewers for factors considered important in the interpretation of the clinical importance of study results: identification of a clearly defined primary outcome, reporting of the expected difference between groups used in the calculation of sample size (the delta value) and whether it was based on the minimal clinically important difference of the intervention, the statistical significance of the results, presentation of pertinent confidence intervals, and the authors’ interpretation of the clinical importance of the results.

Results: Twenty-two of 27 (81%) articles explicitly reported a single primary outcome. Of the 20 articles that included a sample size calculation, 18 (90%) reported a delta value. Two of the 18 (11%) articles explicitly stated that the delta value was chosen to reflect the minimal clinically important difference of the intervention. For the primary outcomes, confidence intervals surrounding the point estimates of the efficacy of the interventions were reported in 11 of 27 (41%) studies. The study results were interpreted from the perspective of clinical importance in 20 of 27 (74%) of the articles. Of these 20 reports, 5 (25%) provided justification for their clinical interpretation of the results.

Interpretation: Authors of RCTs published in major general medical and internal medicine journals do not consistently provide their own interpretation of the clinical importance of their results, and they often do not provide sufficient information to allow readers to make their own interpretation.

The interpretation of the results of randomized controlled trials (RCTs) has emphasized statistical significance rather than clinical importance. For example, the revised CONSORT statement,1 which is a widely adopted, recently updated series of recommendations designed to improve the quality of reporting of RCTs, failed to recommend specifically that authors discuss the clinical importance of their results. The lack of emphasis on clinical importance has led to frequent misconceptions and disagreement regarding the interpretation of the results of clinical trials and a tendency to equate statistical significance with clinical importance. In some instances, statistically significant results may not be clinically important and, conversely, statistically insignificant results do not rule out completely the possibility of clinically important effects.2,3

The minimal clinically important difference (MCID) of a therapy is defined as the smallest treatment effect that would result in a change in patient management, given its side effects, costs and inconveniences.4 It is a key concept in the design of random allocation in clinical trials.5

© 2001 Canadian Medical Association or its licensors
clinical trials. Sample size calculations for prospective trials require determination of the magnitude of difference in outcomes between treatment groups that the study can reliably detect (often called the delta value). In order for clinical trials to have the best chance of detecting clinically important effect sizes, their delta values should reflect the MCIDs of the study interventions.

The MCID is also a key concept in the interpretation of clinical trial results. For example, in individuals without previous myocardial infarction (MI) or stroke, the regular use of ASA will reduce the incidence of MI by 0.2% per year (from a baseline rate of 0.7%/year to 0.5%/year, which is a relative risk reduction of about 25%), but this benefit is possibly offset by a concomitant absolute increase in the chance of stroke of 0.02% per year (from about 1%/year to about 2%/year) (from a baseline rate of 0.7%/year to 0.5%/year, which is a relative increase of about 10%/year) and of gastrointestinal bleeding of about 1% per year (from about 1%/year to 2%/year). After weighing the advantages and disadvantages of ASA use in this clinical situation, an influential expert panel did not recommend its use, reasoning that the efficacy of ASA in preventing MI was not sufficient to overcome the increased incidence of stroke and gastrointestinal bleeding in this low-risk group. That is, the efficacy of ASA in this clinical situation was insufficient to meet or exceed its MCID.

The comparison of actual study results (including the point estimates and the accompanying confidence intervals) with MCID values can also give an indication of the clinical importance of the study results. If the MCID estimate is less than the value of the lower limit of the 95% confidence interval, the study results are statistically significant and very likely to be clinically important. Alternatively, if the MCID value is greater than the upper limit of the 95% confidence interval, the results of the study are very likely to be clinically unimportant. Finally, for study results in which the MCID values are contained within the 95% confidence intervals, their clinical importance is less clear.

To allow readers to interpret the clinical importance of trial results from their own perspective, specific information must be reported, including a clearly defined primary outcome, a delta value or MCID, statistical significance and confidence intervals around the point estimates of the efficacy of the intervention. A previous review in 1987 found that such information was often underreported. The objective of this study was to assess the quality of reporting with respect to the concept of clinical importance in a random sample of RCTs recently published in high-impact general medical and internal medicine journals.

**Methods**

For the period from Dec. 1, 1998, to Nov. 30, 1999, 5 high-impact general medical and internal medicine journals (Annals of Internal Medicine, BMJ, JAMA, The New England Journal of Medicine and the Lancet) were manually searched by 2 independent reviewers for RCTs. From each journal, a random sampling of 10% of the identified articles were selected for review. These articles were evaluated by 4 independent reviewers using a standardized data collection sheet developed to allow assessment of key components in the interpretation of study results from the perspective of clinical importance (Table 1). Any disagreements were resolved by collaborative review. The standardized data collection sheet evaluated the reporting of 5 factors:

1. **Identification of a clearly defined primary outcome.** The Methods section of each report was reviewed for an explicitly stated primary outcome. The primary outcome was defined as the one upon which the sample size calculation was based. If no sample size calculation was reported, the rest of the report was reviewed for an explicitly stated primary outcome. If an explicitly stated primary outcome could not be identified, the outcome that was deemed to be of greatest clinical relevance was chosen.

2. **Reporting of the sample size delta value.** Sample size calculations were evaluated for the reporting of a delta value, that is, the magnitude of difference in outcomes between treatment groups that the trial was attempting to detect. It was recorded whether the delta value was chosen to reflect the authors’ perception of the MCID of the study intervention, and whether it was reported in absolute or relative terms. In order to prevent MCID values from changing with the baseline rate in the control groups (which will occur if MCID values are reported as relative risk reductions), we believe that MCID values should always be stated in absolute terms.

3. **Documentation of the statistical significance of the results.** With respect to the primary outcome, the statistical significance (p < 0.05) of the efficacy of the study intervention was recorded.

4. **Presentation of the pertinent confidence intervals.** For the primary outcome, we recorded whether the 95% confidence intervals surrounding the point estimates of the efficacy of the interventions were reported.

5. **Authors’ interpretation of the clinical importance of their results.** The Discussion sections were reviewed to determine whether the clinical importance of the study results was discussed. If this was discussed, we judged whether the discussion of clinical importance was explicit or implicit. The discussion was classified as explicit if there was direct comment.

<table>
<thead>
<tr>
<th>Table 1: Methodological attributes important in the interpretation of study results from a clinical perspective</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Section</strong></td>
</tr>
<tr>
<td>Methods</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Results</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Discussion</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

Note: MCID = minimal clinically important difference.
RCTs, a total of 27 trials were selected for review. The on a random 10% sampling of each journal’s published term follow-up data from previously published trials. Based Two reports were excluded because they presented long-

Results

From the 5 journals, 266 eligible RCTs were identified. Two reports were excluded because they presented long-term follow-up data from previously published trials. Based on a random 10% sampling of each journal’s published RCTs, a total of 27 trials were selected for review. The overall results of our study are shown in Table 3.

A total of 22 of 27 (81%) articles explicitly reported a single primary outcome. The remaining 5 articles measured multiple outcomes without identifying one as the primary outcome. For these articles, the primary outcome was chosen to be the one we judged most clinically relevant.

Twenty of 27 (74%) articles included a sample size calculation. Of these 20, a delta value was reported in 18 (90%). Two of the 18 (11%) articles explicitly stated that the delta value was chosen to reflect the MCID of the intervention. The delta values were reported as absolute values in 13 of 18 (72%) studies. In 2 of the 5 studies that did not report absolute delta values, the anticipated baseline event rates in the control groups were given and, therefore, absolute delta values could be calculated.

Of the 27 RCTs, 17 (63%) studies had primary outcomes that were reported as being statistically significant ($p < 0.05$).

For the primary outcome, confidence intervals surrounding the point estimates of the efficacy of the intervention were reported in 11 of 27 (41%) of the studies.

The study results were interpreted from the perspective of clinical importance in 20 of 27 (74%) articles, with 10 of 20 (50%) articles having an explicit discussion and 10 of 20 (50%) articles having an implicit discussion.

Table 3: Reporting characteristics of published randomized controlled trials

<table>
<thead>
<tr>
<th>Justification</th>
<th>Primary outcome explicitly reported</th>
<th>Multiple outcomes measured without a clear primary outcome being reported</th>
<th>Sample size calculation reported</th>
<th>Delta value reported in sample size calculation</th>
<th>Delta value explicitly stated as MCID</th>
<th>MCID is an absolute value†</th>
<th>Primary outcome statistically significant</th>
<th>Confidence intervals reported for primary outcomes</th>
<th>Clinical importance mentioned</th>
<th>Explicit discussion</th>
<th>Implicit discussion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1 A–B</td>
<td>22/27 (81%)</td>
<td>5/27 (19%)</td>
<td>20/27 (74%)</td>
<td>18/20 (90%)</td>
<td>2/18 (11%)</td>
<td>13/18 (72%)</td>
<td>17/27 (63%)</td>
<td>11/27 (41%)</td>
<td>20/27 (74%)</td>
<td>10/20 (50%)</td>
<td>10/20 (50%)</td>
</tr>
<tr>
<td>Level 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Justification</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level 1 A–B</td>
<td>0/20 (0%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level 2</td>
<td>1/20 (5%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level 3</td>
<td>4/20 (20%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level 4</td>
<td>15/20 (75%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Levels of justification for the discussion of the clinical importance of study results

<table>
<thead>
<tr>
<th>Level</th>
<th>Justification</th>
</tr>
</thead>
<tbody>
<tr>
<td>1A</td>
<td>Explicitly discusses the clinical importance of the primary study result in relation to previous empirical work done to determine the MCID of the therapy. For example, “the MCID of the CRDQ (Chronic Respiratory Disease Questionnaire) has been [empirically determined] as a change of 4 points.”⁵⁵</td>
</tr>
<tr>
<td>1B</td>
<td>Explicitly discusses the clinical importance of the primary study result in relation to an accepted, but not empirically determined, MCID for the therapy. For example, “before starting the trial, a MCID in score for each test was defined using a Delphi panel consensus.”¹⁰</td>
</tr>
<tr>
<td>2</td>
<td>Discusses the clinical importance of the primary study result in relation to the reported sample size MCID. For example, “efficacy [did not] reach the level we initially posed as necessary to justify surgery.”¹¹</td>
</tr>
<tr>
<td>3</td>
<td>Discusses the clinical importance of the primary study result in relation to previously performed related studies. For example, “previous prevention interventions [have] been relatively ineffective at reducing body fatness.”¹²</td>
</tr>
<tr>
<td>4</td>
<td>Mentions the clinical importance of the study result but with no accompanying justification. For example, “overall, these results provide evidence of the clinical benefits of celecoxib in the treatment of RA.”¹³</td>
</tr>
</tbody>
</table>

Note: RA = rheumatoid arthritis.
Most of the reports (15 of 20, 75%) provided no justification for the authors’ interpretation of the clinical importance of their results (level 4). In only one article was it judged that the authors discussed the clinical importance of their study results in relation to a reported sample size MCID (level 2). No study authors justified their interpretation of clinical importance in relation to empirically determined or consensus-based MCID values (level 1). No study authors provided justification for the magnitude of their MCID value or commented on the relation of the MCID with the confidence intervals surrounding the point estimates.

**Interpretation**

The results of this study document that many reports of RCTs provide no discussion regarding the clinical importance of the trial results. These results agree with the findings of a previous review of 102 RCTs with statistically insignificant results in which only 20% (20/102) of studies made any statement about clinical importance.8

Although clinical importance was discussed in 74% (20/27) of articles in our study, we were liberal in our categorization. Several studies had only one sentence that implied clinical importance (e.g., “our findings suggest that lorazepam is safe and effective for these patients”). For studies that did provide a discussion of clinical importance, most provided little or no justification for their clinical interpretation. The one report that ranked highest on our categorization. Several studies had only one sentence that implied clinical importance (e.g., “our findings suggest that lorazepam is safe and effective for these patients”). For studies that did provide a discussion of clinical importance, most provided little or no justification for their clinical interpretation. The one report that ranked highest on our categorization. Several studies had only one sentence that implied clinical importance (e.g., “our findings suggest that lorazepam is safe and effective for these patients”). For studies that did provide a discussion of clinical importance, most provided little or no justification for their clinical interpretation. The one report that ranked highest on our categorization. Several studies had only one sentence that implied clinical importance (e.g., “our findings suggest that lorazepam is safe and effective for these patients”). For studies that did provide a discussion of clinical importance, most provided little or no justification for their clinical interpretation. The one report that ranked highest on our categorization. Several studies had only one sentence that implied clinical importance (e.g., “our findings suggest that lorazepam is safe and effective for these patients”). For studies that did provide a discussion of clinical importance, most provided little or no justification for their clinical interpretation. The one report that ranked highest on our categorization. Several studies had only one sentence that implied clinical importance (e.g., “our findings suggest that lorazepam is safe and effective for these patients”). For studies that did provide a discussion of clinical importance, most provided little or no justification for their clinical interpretation. The one report that ranked highest on our categorization. Several studies had only one sentence that implied clinical importance (e.g., “our findings suggest that lorazepam is safe and effective for these patients”). For studies that did provide a discussion of clinical importance, most provided little or no justification for their clinical interpretation. The one report that ranked highest on our categorization. Several studies had only one sentence that implied clinical importance (e.g., “our findings suggest that lorazepam is safe and effective for these patients”).

In this trial that assessed the effect of low-dose hydrocortisone on chronic fatigue syndrome, a 9-point reduction on a fatigue scale was deemed to be clinically important. However, the delta for the sample size calculation was reported as a 4-point reduction on the same scale. Thus, the authors reported that the study result, a statistically significant 4.5-point reduction on the fatigue scale when comparing the intervention group with the control group, was not clinically important.

If delta values are not representative of MCIDs and there is little discussion of the clinical importance of the study results, it is difficult to determine whether the authors perceive the results of their study to be clinically important. Previous work done by Burback and colleagues clearly illustrates this. The clinical importance of the results of several RCTs that assessed the efficacy of tacrine in the treatment of Alzheimer’s disease was unclear, because the reports provided no MCID values and the authors generally failed to comment on the clinical importance of their results. Therefore, Burback and colleagues surveyed a group of practising physicians to determine the MCID for this intervention. The MCID value was then compared with the results of trials under review. Of the 2 of 12 studies that found a statistically significant difference in favour of tacrine therapy, neither study result could be deemed clinically important when compared with the MCID value.

One limitation of our study was that we had access only
to the final published reports of the trials we reviewed. Thus, it is possible that pertinent information regarding sample size calculations, authors’ perceptions of MCID values of the interventions and interpretation of clinical importance did not survive the editorial process. In addition, our assessment of the adequacy of discussion with respect to clinical importance was subjective. Some disagreements did occur among the reviewers. However, consensus was eventually reached in all cases, with a deliberate tendency to err toward accepting ambiguous statements as indicating the presence of a discussion regarding clinical importance. Thus, our study possibly overestimated the percentage of trials that adequately addressed the issue of clinical importance. Another possible limitation is the small number of journals we reviewed.

An argument can be made that the clinical interpretation of RCTs using surrogate outcomes is problematic. However, we believe that the onus is on the authors of such trials to connect their trial results with the potential clinical implications, because not doing so leaves readers with no guidance regarding the clinical relevance of the results. With the increasing prevalence of published meta-analyses, it may be argued that the need for interpretation of the clinical importance of primary study results is superseded once a relevant meta-analysis is published. However, at the time of publication of an RCT with primary data, it is usually not known whether a meta-analysis on a similar topic will subsequently be published. Thus, the need for primary studies to publish the elements shown in Table 1 remains important. Furthermore, publication of these elements can help provide the basis with which meta-analysts (and their readers) can judge the clinical interpretation of meta-analytic results.

Our findings demonstrate that the reports of RCTs published in major general medical and internal medical journals do not provide consistently the authors’ interpretation of the clinical importance of their results. Moreover, they do not contain adequate information to allow readers to assess the clinical importance of the results from their own perspective. Ideally, each published RCT should report an explicit primary outcome, a sample size delta value that reflects the authors’ perception of the MCID of their intervention, the statistical significance of the primary outcome, confidence intervals surrounding the point estimates for the primary outcome and an explicit discussion of the clinical importance of the study results. Many of these criteria are already included in the revised CONSORT statement. However, we also recommend that authors report the MCID of the intervention, justify its magnitude and explicitly discuss the clinical importance of the results in relation to the MCID.

Competing interests: None declared.

Contributors: All authors contributed to the study design, collected and interpreted the data, and participated in the revising of the manuscript. In addition, Drs. Man-Son-Hing and Laupacis proposed the study. Dr. Chan wrote the initial draft of the manuscript.

References


Correspondence to: Dr. M. Man-Son-Hing, Geriatric Assessment Unit, Ottawa Hospital, Civic Campus, 1053 Carling Ave., Ottawa ON K1Y 4E9; fax 613 761-5334; mhing@ottawahospital.on.ca

“I learned about epidemiology, statistics, scientific and copy editing, the whole world of medical journalism, right on the job. It was a special year, a year about meeting mentors and finding focus.”
— Dr. Caralee Caplan, 1998 Fellow

“It provides a unique opportunity to sharpen research and writing skills and to establish relationships with physicians across the country. It lets you build a platform for issues and perspectives pertinent to new physicians and physicians in training.”
— Dr. Erica Weir, 1999 Fellow

The CMAJ Editorial Fellowship: Come for a year, learn for a lifetime

Since CMAJ launched its 1-year Editorial Fellowship program in 1998, both sides have learned from the experience: editors have gained insights from doctors at the start of their careers, while the fellows have gained from authors and helping to decide which manuscripts are published to soliciting commentaries. They are also expected to write extensively, and the first 2 fellows have taken lead roles in developing theme issues. Interested? We’re considering applications for the next fellowship, which will begin July 1, 2002.

For information, contact Dr. John Hoey, the editor-in-chief, at johnhoey@cma.ca; 800 663-7336 x2118; 1867 Alta Vista Dr., Ottawa ON K1G 3Y6. The deadline for applications is Jan. 7, 2002.