REVIEW

Determination of the Clinical Importance of Study Results

A Review

Malcolm Man-Son-Hing, MD, MSc, Andreas Laupacis, MD, MSc, Keith O'Rourke, MBA, Frank J. Molnar, MD, MSc, Jeffery Mahon, MD, MSc, Karen B. Y. Chan, MD, George Wells, PhD

Formal statistical methods for analyzing clinical trial data are widely accepted by the medical community. Unfortunately, the interpretation and reporting of trial results from the perspective of clinical importance has not received similar emphasis. This imbalance promotes the historical tendency to consider clinical trial results that are statistically significant as also clinically important, and conversely, those with statistically insignificant results as being clinically unimportant. In this paper, we review the present state of knowledge in the determination of the clinical importance of study results. This work also provides a simple, systematic method for determining the clinical importance of study results. It uses the relationship between the point estimate of the treatment effect (with its associated confidence interval) and the estimate of the smallest treatment effect that would lead to a change in a patient's management. The possible benefits of this approach include enabling clinicians to more easily interpret the results of clinical trials from a clinical perspective, and promoting a more rational approach to the design of prospective clinical trials.

KEY WORDS: **clinical importance; clinical significance; study results; minimal clinically important difference; review.**
J GEN INTERN MED 2002;17:469–476.

THE CONCEPT OF CLINICAL IMPORTANCE

Given the side effects, costs, and inconveniences of a therapy, the minimal clinically important difference (MCID) is the smallest treatment efficacy that would lead to a change in a patient's management.[1] The concept of the MCID is crucial in both the planning of clinical trials and the interpretation of their results.

For example, persons with atrial fibrillation have a risk of stroke that can vary from 1%[2] to 15%[3] per year, depending on the presence of certain risk factors (previous history of stroke or transient ischemic attack, history of hypertension, left ventricular dysfunction, diabetes, and increasing age).[4,5] Warfarin is more efficacious than aspirin (~65% vs 22% relative risk reduction) in the prevention of stroke in these persons.[2,6–8] However, compared with aspirin, warfarin use has a 1% to 2% per year increased risk of major hemorrhage, and is more costly and inconvenient to use.[2] Given the greater efficacy of warfarin, an expert panel[4] recommends that most patients with atrial fibrillation should receive long-term warfarin therapy. However, for atrial fibrillation patients with very low risk of stroke (1% to 2% per year), the panel recommends aspirin therapy. In these patients, they believe that the absolute incremental stroke protection benefit (<1% per year) that warfarin provides over aspirin is not sufficient to offset the increased risk of major bleeding, costs, and inconveniences associated with warfarin therapy.[4] In other words, for low-stroke-risk patients with atrial fibrillation, the panel does not believe that the absolute efficacy of warfarin to prevent stroke is sufficient to meet its MCID.

Determining the MCID of a therapy is also a necessary part of the clinical trial planning process. In general, sample size calculations are based upon 4 important parameters; the chance of concluding the treatments are different when they truly are not (alpha [α] error), the chance of concluding that treatments are the same when they truly are not (beta [β] error), either the estimated probability of the outcome event in the control group (if outcomes are binary) or the estimated standard deviation (if outcomes are continuous), and the delta (Δ) or the amount of difference the trial is trying to detect.[9] In order for clinical trials to have the best opportunity to detect clinically important treatment effects and subsequently change clinical practice, Δ values should be based on the MCIDs of the interventions.

Although agreement on appropriate tolerance levels for both α and β error has evolved over time,[10] the process of determining Δs for prospective clinical trials has not been standardized. Feinstein[11] in 1977 wrote, "[j]udgements

Received from the Clinical Epidemiology Unit, Ottawa Health Research Institute, Ottawa Hospital (MM-S-H, KO, FM, KC, GW); the Institute on Health of the Elderly, Sisters of Charity, Ottawa Health Service (MM-S-H, FM); the Division of Geriatric Medicine (MM-S-H, FM), and the Department of Community Medicine and Epidemiology (GW), University of Ottawa, Ottawa; the Institute for Clinical Evaluative Sciences and Faculty of Medicine, University of Toronto (AL), Toronto; and the University of Western Ontario (JM), London, Ontario, Canada.

Address correspondence to Dr. Man-Son-Hing: Geriatric Assessment Unit, Ottawa Hospital–Civic Campus, 1053 Carling Ave., Ottawa, ON, Canada K1Y 4E9 (e-mail: mhing@ottawahospital.on.ca).

about the proper size of Δs have received almost no concentrated attention via symposia, workshops, or other enclaves of experts assembled to adjudicate matters of clinical importance. ...In the absence of established standards, the clinical investigator picks what seems like a reasonable value. If the sample size that emerges is unfeasible, the Δ gets adjusted accordingly ... until the n comes out right." It is our opinion that little has changed since 1977, with the Δs used in sample size calculations presently based as much on feasibility issues as upon MCIDs. Given the critical nature of the concept of clinical importance in the design and interpretation of clinical trials, it is of concern that methodological standards have not been developed for determining and reporting the clinical importance of therapies.

APPROACHES TO DETERMINING CLINICAL IMPORTANCE

The Efficacy of Medical Interventions

Naylor and Llewellyn-Thomas[12] have outlined some previous approaches used to determine the clinical importance (MCIDs) of therapies. The subjective opinions of clinician experts are the traditional basis of clinical practice,[13] sometimes using consensus development.[14] For example, despite the slightly greater efficacy of clopidogrel, standard clinical practice is to prescribe aspirin as initial therapy for the secondary prevention of stroke.[15] The consensus of most expert opinion[16] is that the greater cost and side-effects of clopidogrel outweigh its slightly greater efficacy in the treatment of this condition. Thus, according to these experts, when compared with the efficacy of aspirin, the additional stroke protection offered by clopidogrel is not clinically important.

Others have suggested that the perspective of patients is also important in the determination of clinical importance. They have developed systematic approaches to determine the clinical importance of therapies from the perspective of patients.[12,17–19] For example, after explaining the pertinent benefits and risks of warfarin therapy to patients with atrial fibrillation, Man-Son-Hing et al.[19] judged that the known efficacy of warfarin to prevent stroke was clinically important in persons with atrial fibrillation who have an average risk of stroke.

Because the concept of the MCID is of crucial importance in determining the Δs for clinical trial sample size calculations, some have suggested rigorous approaches to their determination.[12,20] Naylor and Llewellyn-Thomas[12] advocated that patient-oriented MCIDs be systematically determined and then used as Δs in study sample size calculations before trials commence. Alternatively, van Walraven et al.[21] suggested the use of physician surveys to determine appropriate MCIDs for sample size Δs. Detsky[20] also developed a method of using economic analysis to determine sample size MCIDs (Δs) for clinical trials. He hoped that this process would enable funding agencies to rationally allocate monies to trials with the greatest chance of detecting interventions with clinically important effects. While methodologically attractive, these approaches add additional steps to the clinical trial design process and, according to recent clinical trial reports, have not been widely implemented.[22]

Minimally Detectable Differences

In addition to determining the clinical importance of the efficacy of different therapies, the concept of the MCID can also be applied to detect clinically important changes of clinical rating scales.[23,24] For example, Redelmeier and Guyatt[25] developed 2 methodologies using patient perceptions to quantitatively determine MCID values for different rating scales. The clinical importance of changes in scores on these scales is often difficult to judge. For example, is a 1-point change on a 10-point scale that measures patients' subjective feeling of anxiety a clinically important or only a trivial change? They suggested that "a difference in functional status scores is clinically important if it is associated with a difference that a typical patient can notice."[26] They equated a "minimally detectable difference" on a scale with its MCID. In association with Wells et al.,[27] Redelmeier attempted to determine the MCID of the Health Assessment Questionnaire (HAQ), a rheumatoid arthritis scale that uses patients' symptoms and physical examination findings to measure their disease severity. First, patients' disease severity was assessed using the HAQ. Then a structured interview was arranged between 2 of the assessed patients. After the interview, using a 7-point Likert scale, patients rated their own disease severity in relation to their interview partner. The difference in score on the HAQ that corresponded to a difference that the patients were just able to notice on interview was deemed the MCID of the HAQ.

Juniper et al.[28] used a slightly different approach to this problem. They examined the relationship between changes on a scale measuring global well-being and changes in the score of the outcome measure of interest. In order to determine the MCID of the Chronic Heart Failure (CHF) scale designed to measure the severity of shortness of breath (SOB) in patients with congestive heart failure, patients were serially asked to rate both their global level of disability and the severity of their SOB. The amount of change in the severity of the patients' shortness of breath that they noticed in a global sense was deemed the MCID of the CHF scale.

Social Science Literature

The effect size index (defined as Δ/standard deviation) is an inter-related concept that has been mainly used by social scientists to judge the clinical importance of interventions.[29] Small, medium, and large treatment effects correlate with effect size indices of 0.2, 0.5, and 0.8 respectively.[29] Although this approach is usually employed to judge the clinical importance of interventions after completion of clinical studies, it also can conceivably

be used in sample size determination as long as the Δ values are based on the MCIDs of the interventions.

Psychotherapy Literature

From the psychology literature, Jacobson et al.[30] argued that psychotherapy makes a clinically important difference if, after treatment, patients could not be differentiated from the normal population. Along a similar vein, Kendall et al.[31] defined clinical importance as "end-state functioning that falls within a normative range on important measures." Considering that these definitions of the MCID do not include patients who improve but are not fully recovered, they set a higher standard for clinical importance.

In summary, different disciplines have taken divergent approaches to determining the clinical importance of their interventions. There is also no consensus on the appropriate method(s) of determining the clinical importance of therapies.

THE HISTORICAL APPROACH TO THE REPORTING OF CLINICAL TRIAL RESULTS

Historically, despite the critical nature of the role of clinical importance in the interpretation of clinical trial results, journal editors have typically prompted authors to emphasize statistical significance over clinical importance when reporting their results. For example, the *New England Journal of Medicine* once prohibited the use of the word "significance" in its general circulation, only allowing its use in a statistical context.[32] Also, other editors have stated that they would only accept papers that have "super-significance," which was defined as *P* values less than .01.[33] This lack of emphasis with respect to clinical importance led to frequent misconceptions and disagreement regarding the interpretation of clinical trial results, and the tendency to equate statistical significance with clinical importance.[34] In many studies, statistically significant results may not be clinically important.[34] Alternatively, Freiman et al.[35] systematically documented that many studies that reported statistically insignificant results in fact did not fully exclude the possibility that clinically important differences existed.

More recently, with the publication of influential papers[36] including the evidence-based medicine series in *JAMA*,[37] there has been a clear shift toward greater emphasis on clinical importance when deciding whether the results of clinical trials should change patients' management. Nevertheless, issues still exist. For example, the 2001 revision of the CONSORT statement[38] (a set of widely adopted recommendations designed to improve the quality of reporting of randomized controlled trials) does not specifically recommend that authors discuss the clinical importance of their results. Our recent review[22] of high-impact medical journals found a less than uniform approach regarding the reporting of a priori sample size Δs (or MCIDs) of clinical trials, values that may help readers interpret the clinical importance of study results. Also, the published clinical trial reports did not consistently discuss their results in a clinical context.[22] Thus, there still needs to be greater emphasis on the reporting and interpretation of the clinical importance of clinical trial results.

APPROACHES TO COMBINING CLINICAL IMPORTANCE WITH STATISTICAL SIGNIFICANCE

Other investigators have published methods of relating the clinical importance of trial results to their statistical significance.[36,37] In the context of making treatment recommendations, Cook et al.[36] proposed that stronger recommendations can be made when the entire confidence interval of the efficacy of a therapy is greater than its MCID. Alternatively, inferences for treatment are weaker when the confidence interval overlaps the MCID. Using a slightly different approach based on the number needed to treat, Guyatt et al.[37] proposed a threshold approach to treatment recommendations. Again, if the confidence interval was completely on one side of the threshold, stronger recommendations for treatment could be made than if the confidence interval overlapped the threshold. For patients with chronic airflow obstruction and with use of an n-of-1 design, Mahon et al.[39] used the relationship between the statistical confidence limits of symptom scores for individual patients and a predetermined MCID to assess whether individual patients should continue with theophylline therapy. Other authors including Detsky and Sackett[40] and Goodman and Berlin[10] have outlined the role of confidence intervals in the interpretation of clinical trials without statistically significant results. They suggest that the results of such trials are not definitively "negative" until the upper limit of the confidence interval is smaller than the predetermined MCID. Jones et al.[41] and Dunnett and Gent[42] have performed similar work in the interpretation of clinical trial results designed to demonstrate equivalence between therapies. They stated that for therapies to be considered equivalent, not only should the comparison of the efficacies of the two interventions not reach statistical significance, but also the upper limit of the confidence interval should be smaller than the predetermined MCID. Unfortunately, in the reporting of clinical trial results, these approaches have not been widely adopted.

A SYSTEMATIC METHOD OF INTERPRETING STUDY RESULTS FROM THE PERSPECTIVE OF CLINICAL IMPORTANCE

Study results are regarded as statistically significant ($P < .05$) when values indicating the lower limit of the 95% confidence interval (95% CI) are greater than the null effect. Judging the clinical importance of study results can be more difficult. Fortunately, the previous work reviewed above provides a systematic method by which clinical importance can be determined (Fig. 1). Clinical importance

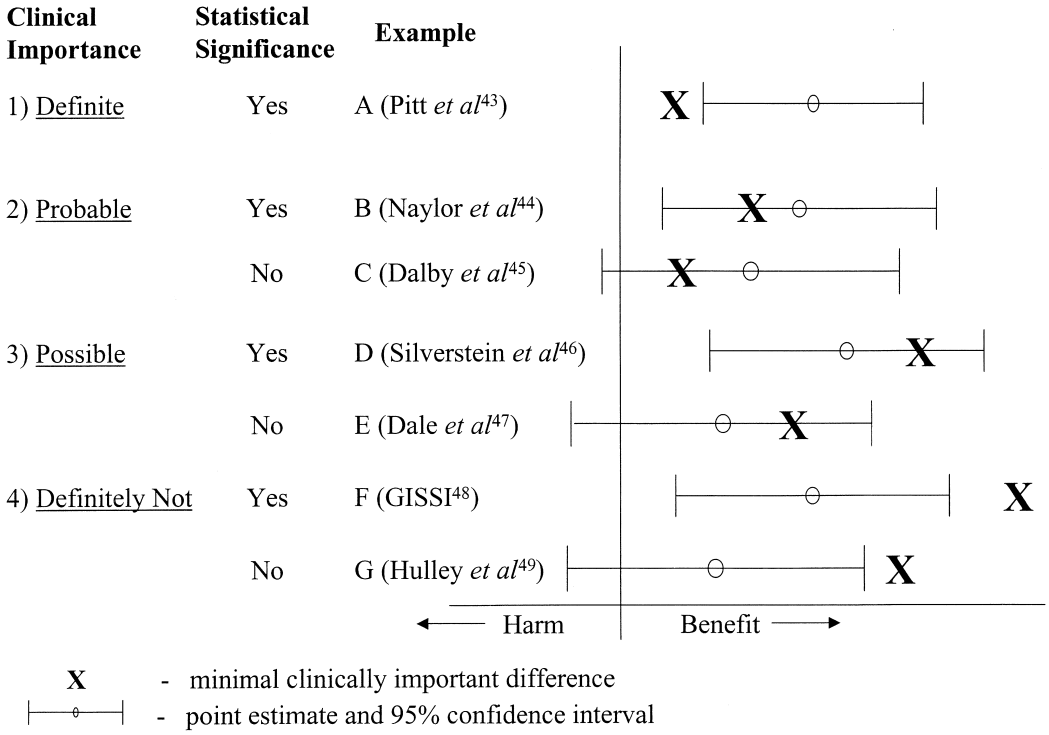


FIGURE 1. Possible combinations for the relationship between clinical importance and statistical significance (see text for details).

can take 4 forms, depending on the relationship of the MCID of the intervention to the point estimate (the best single value of the efficacy of the intervention that has been derived from the study results) and the 95% CI surrounding it:

1. definite – when the MCID is smaller than the lower limit of the 95% CI;
2. probable – when the MCID is greater than the lower limit of the 95% CI, but smaller than the point estimate of the efficacy of the intervention;
3. possible – when the MCID is less than the upper limit of the 95% CI, but greater than the point estimate of the efficacy of the intervention; and
4. definitely not – when the MCID is greater than the upper limit of the 95% CI.

Study Results of Definite Clinical Importance

Study results have definite clinical importance if the lower limit of the 95% CI is greater than the MCID. Given that the 95% CI lies completely above the MCID estimate, studies with definite clinical importance will always be statistically significant ($P < .05$). An example (Fig. 1, example A) is a randomized placebo controlled trial[43] that assessed the efficacy of the addition of spironolactone to the medication regimen of patients with severe heart failure. The results demonstrated that the study intervention reduced the chance of death by an absolute value of 11.4% (calculated 95% CI, 6.7% to 16.1%). The Δ used for sample size calculation in this study was an absolute 6.5% reduction in mortality. If the Δ value is assumed to be the MCID, the results of the study can be interpreted as being both statistically significant and definitely clinically important.

Study Results of Probable Clinical Importance

The results of studies with probable clinical importance have 95% CIs that include the value of the MCID, with the point estimate being greater than the MCID estimate. The study result may or may not be statistically significant. A recent randomized trial[44] (Fig. 1, example B) assessed the efficacy of comprehensive discharge planning and home follow-up of hospitalized elderly persons, and found a statistically significant ($P < .05$) absolute reduction in readmission rates of 16.8% (calculated 95% CI, 7.6% to 25.9%) when comparing the intervention and usual care groups. The sample size Δ was reported as 15%. Thus, the study result is probably clinically important, but there is still a reasonable possibility of a clinically unimportant effect.

An example of a study with results that were not statistically significant but of probable clinical importance is a randomized controlled trial[45] that assessed the efficacy of nursing visits on outcomes of elderly people in the community (Fig. 1, example C). The authors showed a statistically insignificant absolute decrease in the primary outcome events (combined death and hospital admissions) of 4.2% (calculated 95% CI, −4.6% to 13.0%) when

comparing the intervention and control groups. Their sample size Δ was a relative 15% (absolute 1.5%) decrease in primary outcome events. Thus, while the results are not statistically significant, it is probable that the efficacy of the study intervention is clinically important.

Study Results of Possible Clinical Importance

Studies with possible clinical importance have 95% CIs that include the value of the MCID, and an MCID greater than the point estimate of the efficacy. Again, the study result may or may not be statistically significant. An example of a study with statistically significant results but only possible clinical importance is a randomized placebo-controlled trial by Silverstein et al.[46] (Fig. 1, example D). They assessed the efficacy of misoprostol to reduce serious gastrointestinal complications in patients with rheumatoid arthritis receiving nonsteroidal anti-inflammatory drugs. Comparing the misoprostol and placebo groups, the results showed a statistically significant 0.38% absolute reduction (calculated 95% CI, 0.02% to 0.74%) in primary outcome events in favor of the misoprostol group. In their sample size calculation, the authors used a 0.5% absolute reduction in primary events as their Δ. Using this value as the MCID, it is possible that the efficacy of misoprostol is or is not clinically important. Even though the result was statistically significant (i.e., the lower limit of the 95% CI did not cross the point of no effect), the difference may or may not be clinically important. Therefore, the clinical importance of the efficacy of the study intervention still needs confirmation.

An example of a study with results that were not statistically significant but had possible clinical importance is a recent randomized double-blind placebo controlled trial[47] that assessed the efficacy of pentoxifylline in the treatment of venous leg ulcers (Fig. 1, example E). The investigators found that compared with 53% of patients in the placebo group, complete healing occurred in 64% of the patients in the pentoxifylline group (an absolute difference of 11%), but this result was not statistically significant (95% CI, −2% to 25%). The authors based their sample size on the ability to detect a 20% absolute difference between the 2 groups. Thus, assuming the sample size Δ represents the MCID, the true efficacy of pentoxifylline in this clinical situation is possibly clinically important.

Study Results that Are Definitely Not Clinically Important

Studies with results that are definitely not clinically important have the upper limit of the 95% CI that is below the value of the MCID. Again, the study results may or may not be statistically significant. An example of a study result that is statistically significant but definitely not clinically important is a randomized placebo controlled trial[48] assessing the effect of dietary supplementation with polyunsaturated fats (Fig. 1, example F). It demonstrated an absolute decrease of 1.3% (95% CI, 0.1% to 2.6%) in the primary outcome (combined endpoint of death, nonfatal myocardial infarction, and stroke). In the sample size calculation, the Δ was a 4% absolute difference between groups over a 3.5-year period. If one assumes that this value represents the MCID of the intervention, the magnitude of the intervention's effect is definitely not clinically important. However, because the authors concluded that their study result provided both "a clinically important and statistically significant benefit," their estimation of the MCID of the intervention is clearly not reflected in the sample size Δ. Thus, this example also demonstrates the need for greater methodological rigor in determining Δ values in sample size calculations.

An example of a study with results that are not statistically significant ($P > .05$) and definitely not clinically important is a recent randomized controlled trial[49] that assessed the efficacy of estrogen plus progestin for the secondary prevention of coronary heart disease in postmenopausal women (Fig. 1, example G). It did not find a statistically significant benefit to the intervention (relative risk, 0.99; 95% CI, 0.80 to 1.22). The sample size Δ was reported as a 24% relative reduction (1.2% absolute reduction) in primary coronary events. Therefore, if one accepts this Δ as the benchmark for the MCID, a clinically important effect was ruled out. The lack of both statistical significance and clinical importance suggests that no further studies are required to test the efficacy of this intervention in similar clinical situations.

BENEFITS OF A SYSTEMATIC METHOD OF DETERMINING CLINICAL IMPORTANCE

Table 1 shows the possible benefits of this method of determining clinical importance. Clarifying the concept of clinical importance and the relationship between statistical significance and clinical importance will lead to improvements in both the design and interpretation of clinical trials. Widespread use of the proposed method will likely lead to standardized reporting of study sample size Δ values that are based on the MCIDs of the intervention. In turn, this will enhance the ability of readers to interpret the results of studies from the perspective of clinical importance. When differences in opinion arise regarding the clinical importance of study results, this method may allow the

Table 1. Benefits of a Simple, Standardized Approach to Determining Clinical Importance

1. Clarifies the relationship between statistical significance and clinical importance.
2. Promotes balanced interpretation of trial results from both a statistical and a clinical perspective.
3. Forces designers of prospective clinical trials to more carefully choose Δ values for their sample size calculations.
4. Can be combined with cumulative meta-analysis to decide when no further studies are needed.
5. May help reduce publication bias.
6. Can provide guidance to trialists regarding stopping rules.
7. Promotes further methodologic work to determine clinically important differences.

reasons for these differences to be more easily articulated (e.g., differences in opinion about actual MCID values). Thus, more explicit, orderly, and focused discussion regarding the clinical importance of study results would be promoted.

Systematic determination of the clinical importance of the results of studies that use sample size Δs as the MCID benchmarks may also encourage more responsible study sample size determination. Efforts to reduce sample sizes by choosing relatively large Δs that do not accurately reflect the MCIDs of interventions would be counterbalanced by decreasing the likelihood that the results of these studies would be judged as clinically important when compared to these Δs. Thus, there would be a disincentive to inappropriately increasing the sample size Δs in an attempt to decrease the number of study patients. Additionally, because some trials can be statistically insignificant yet possibly or even probably clinically important, the greater emphasis on the clinical interpretation of trial results would further encourage trialists to choose sample size Δs that more closely reflect the MCIDs of their interventions.

Other methodological issues in the design and conduct of clinical trials could be addressed. For example, should trials be powered to reliably determine whether an intervention produces statistically significant ($P < .05$) or definitely clinically important benefits? Clearly, such issues have serious implications for the sample sizes (and thus feasibility) of prospective clinical trials. Further debate regarding this issue is necessary. Also, at present, stopping rules for trials are based primarily on the statistical significance of the interim results.[50] Greater emphasis on the clinical importance of trial results may encourage the development of stopping rules that better balance the emphasis between statistical significance and clinical importance. Hence, trials would not be stopped until more reliable evidence of clinically important effects are found.

There are other potential benefits. Researchers could use this approach in combination with cumulative meta-analyses[51] to provide clearer indication regarding when no further studies are needed to assess the clinical importance of therapies. Also, greater emphasis on reporting the clinical importance encourages more rigorous interpretation of the results of trials without statistically significant results. This, in turn may help reduce publication bias—the tendency to differentially publish papers on the basis of the direction and/or magnitude of effect estimate, which is usually based on the presence of statistical significance. For example, journals may recognize studies that reliably exclude clinically important treatment effects (Fig. 1, example G) as "definitive negative" trials. In order to discourage other investigators from embarking on similar studies, editors may be more willing to publish such manuscripts. Finally, this method of clinical importance determination may also encourage investigators to develop methodologies to empirically determine the clinically important differences of therapies.

POTENTIAL LIMITATIONS

Clearly, a major methodological limitation of the above approach is the uncertainty in the judgment and determination of MCIDs. MCID values for specific interventions will vary from person to person depending on their own values and the perspective (individual, professional, or societal) from which they view the intervention. Also, without the widespread use of sample size Δs that accurately reflect the MCIDs of the interventions, is it reasonable to use sample size Δs as benchmarks for determining the clinical importance of therapies? However, by comparing the relationship of other possible MCID values to the point estimate of the efficacy of the intervention and its confidence interval, readers can easily determine the level of clinical importance of the study result on the basis of their own or any other estimate of the MCID. Interestingly, the systematic determination of clinical importance of study results can help to resolve this issue. With the more careful approach to sample size calculations that is likely to result from this approach, sample size Δs would represent more accurate reflections of the investigators' beliefs about the magnitude of clinically important effect sizes. In turn, this would allow greater confidence in the clinical interpretation based on these benchmarks.

Caution must also be taken to use this approach only for the interpretation of the clinical importance of individual studies and not with recommendations for treatment. For example, it is tempting to suggest that therapies should be introduced into clinical practice if study results suggest both statistical significance (P value $<.05$) and definite clinical importance. In most instances, this is not appropriate until the totality of pertinent evidence is considered, such as whether efficacy translates into effectiveness and whether the intervention is cost-effective. Only then can treatment recommendations be made with confidence. Guyatt et al.[52] have written extensively regarding the process of the development of treatment recommendations.

An inherent limitation of this approach is that it does not reflect the methodologic quality of studies. There is continued need to judge the validity and generalizability of study results independent of clinical importance. Just as the statistical significance of the results of methodologically flawed studies should be viewed with scepticism, so should their clinical importance.

We also recognize that the choice of values to indicate definite/probable/possible/definitely not clinically important results is arbitrary and based solely on convention. Depending on many factors, including the seriousness of the condition, the outcome being measured, the potential of the intervention to cause adverse consequences, and the availability of alternative treatments, these standards may be too strict or not strict enough. For example, suppose an initial study investigating a possible cure for Creutzfeld-Jakob disease (a rapidly fatal condition with no known effective treatment) found a beneficial effect that was statistically significant but did not meet the criteria for

definite or even probable clinical importance. In such situations, implementation into clinical practice may be warranted even though the clinical importance of the intervention's efficacy is not fully confirmed. Alternatively, for conditions in which extremely effective therapies exist, very rigorous standards for the clinical importance of new therapies should be set. For example, if a medication is to become an effective alternative to corticosteroids in the treatment of temporal arteritis, extreme confidence in the clinical importance of its efficacy is necessary. Rather than using cut-points to indicate definitely, probably, possibly, and definitely not clinically important, an alternative approach would be to simply present the relationship between the point estimate of the efficacy and its confidence intervals and the MCID. This approach would avoid an important issue that presently plagues the statistical interpretation of clinical trial results—the artificial dichotomization of results into those that are statistically significant and those that are statistically insignificant. However, we believe that our proposed levels for cut-off values resulting in 4 levels of clinical importance provides clinicians with the added benefit of general guidance in interpreting clinical importance. Further debate is necessary to resolve this issue.

Finally, we recognize that further development of methods to determine the clinical importance of study results is necessary. For example, based on the relationship of the confidence interval to the MCID, it is possible to determine the probability that the true value of the efficacy of the intervention meets or exceeds an a priori determined MCID, and thus quantify the concept of clinical importance. However, the intention of this paper is to summarize previous work on the rationale and framework for the concept of clinical importance. This will hopefully encourage further theoretical and statistical developments.

SUMMARY

Systematic determination of the clinical importance of study results will increase discussion about the concept of the MCID. Thus, it has the potential of improving the methodological rigor with which clinical trials are designed, conducted, reported, and interpreted. These potential benefits may provide greater clarity for the process in which interventions with putative clinical benefits are incorporated into or rejected from the standard care of patients.

REFERENCES

1. Jaeschke R, Singer J, Guyatt GH. Measurement of health status: ascertaining the minimal clinically important difference. Control Clin Trials. 1989;10:407–15.
2. Atrial Fibrillation Investigators. Risk factors for stroke and efficacy of antithrombotic therapy in atrial fibrillation. Arch Intern Med. 1994;154:1449–57.
3. EAFT (European Atrial Fibrillation Study) Group. Secondary prevention in non-rheumatic atrial fibrillation after transient ischemic attack or minor stroke. Lancet. 1993;342:1255–62.
4. Albers GW, Dalen JE, Laupacis A, Manning WJ, Petersen P, Singer DE. Antithrombotic therapy in atrial fibrillation. Chest. 2001;119: 194S–206S.
5. Gage BF, Waterman AD, Shannon W, Boechler M, Rich M, Radford MJ. Validation of clinical classification schemes for predicting stroke. JAMA. 2001;285:2864–70.
6. Hart RG, Benavente O, McBride R, Pearce LA. Antithrombotic therapy to prevent stroke in patients with atrial fibrillation: a meta-analysis. Ann Intern Med. 1999;131:492–501.
7. Atrial Fibrillation Investigators. The efficacy of aspirin in patients with atrial fibrillation. Arch Intern Med. 1997;157:1237–40.
8. Segal JB, McNamara RL, Miller MR, et al. Prevention of thromboembolism in atrial fibrillation. A meta-analysis of trials of anticoagulants and antiplatelet drugs. J Gen Intern Med. 2000; 15:56–67.
9. Young MJ, Bresnitz EA, Strom BL. Sample size nomograms for interpreting negative clinical studies. Ann Intern Med. 1983;99: 248–51.
10. Goodman SN, Berlin JA. The use of predicted confidence intervals when planning experiments and the misuse of power when interpreting results. Ann Intern Med. 1994;121:200–6.
11. Feinstein AR. Clinical Biostatistics. Saint Louis: CV Mosby Company; 1977:333.
12. Naylor CD, Llewellyn-Thomas HA. Can there be a more patient-centred approach to determining clinically important effect sizes for randomized treatment trials. J Clin Epidemiol. 1994;47:787–95.
13. Pauker SG, Kassirer JP. The threshold approach to clinical decision making. N Engl J Med. 1980;302:1109–17.
14. Bellamy N, Carrette S, Ford PM, et al. Osteoarthritis antirheumatic drug trials III. Setting the delta for clinical trials—results of a consensus development (Delphi) exercise. J Rheumatol. 1992;20: 557–60.
15. Gorelick PB, Born GV, D'Agostino RB, Hanley DF Jr, Moye L, Pepine CJ. Therapeutic benefit. Aspirin revisited in light of the introduction of clopidogrel. Stroke. 1999;30:1716–21.
16. Albers GW, Amarenco P, Easton JD, Sacco RL, Teal P. Antithrombotic and thrombolytic therapy for ischemic stroke. Chest. 2001; 119:300S–20S.
17. Llewellyn-Thomas HA, Thiel EC, Clark RM. Patients versus surrogates: whose opinion counts on ethics review panels? Clin Res. 1989;37:501–5.
18. Llewellyn-Thomas HA, McGreal MJ, Thiel EC, Fine S, Erlichman C. Patients' willingness to enter clinical trials: measuring the association with perceived benefit and preference for decision participation. Soc Sci Med. 1991;32:35–42.
19. Man-Son-Hing M, Laupacis A, O'Connor A, et al. Warfarin for atrial fibrillation: the patient's perspective. Arch Intern Med. 1996;156: 1841–8.
20. Detsky AS. Using economic analysis to determine the resource consequences of choices made in planning clinical trials. J Chronic Dis. 1985;38:753–65.
21. van Walraven C, Mahon JL, Moher D, Bohn C, Laupacis A. Surveying physicians to determine the minimal important difference: implications for sample-size calculations. J Clin Epidemiol. 1999;52:717–23.
22. Chan K, Man-Son-Hing M, Molnar FJ, Laupacis A. How well is the clinical importance of study results reported. An assessment of randomized controlled trials. CMAJ. 2001;165:1197–202.
23. Kosinski M, Zhao SZ, Dedhiya S, Osterhaus JT, Ware JE Jr. Determining minimally important changes in generic and disease-specific health-related quality of life questionnaires in clinical trials of rheumatoid arthritis. Arthritis Rheum. 2000;43: 1478–87.
24. Wells G, Beaton D, Shea B, et al. Minimal clinically important differences: review of methods. J Rheumatol. 2001;28:406–12.
25. Redelmeier DA, Guyatt GH, Goldstein RS. Assessing the minimal important difference in symptoms: a comparison of two techniques. J Clin Epidemiol. 1996;49:1215–9.

26. Redelmeier DA, Lorig JP. Assessing the clinical importance of statistical significance: illustration in rheumatology. Arch Intern Med. 1993;153:1337–42.
27. Wells GA, Tugwell P, Kraag GR, Baker P, Groh J, Redelmeier D. Minimum important difference between patients with rheumatoid arthritis: the patient's perspective. J Rheumatol. 1993;20:557–60.
28. Juniper EF, Guyatt GH, Willan A, Griffith LE. Determining a minimal important change in a disease-specific quality of life questionnaire. J Clin Epidemiol. 1994;47:81–7.
29. Cohen J. Statistical Power Analysis for the Behavioral Sciences, 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988:19–27.
30. Jacobson NS, Roberts LJ, Berns SB, McGlinchey JB. Methods for defining and determining the clinical significance of treatment effects: description, application, and alternatives. J Consul Clin Psychol. 1999;67:300–307.
31. Kendall PC, Marrs-Garcia A, Nath SR, Sheldrick RC. Normative comparisons for the evaluation of clinical significance. J Consult Clin Psychol. 1999;67:285–99.
32. Anonymous. Significance of significance. N Engl J Med. 1968;278: 1232–3.
33. Melton AW. Editorial. J Exp Psychol. 1962;64:553–7.
34. Sackett DL, Haynes RB, Guyatt GH, Tugwell P. Clinical Epidemiology. A Basic Science for Clinical Medicine, 2nd ed. Boston: Little, Brown and Company; 1991.
35. Freiman J, Chalmers TC, Smith H, Kaubler R. The importance of beta, the type II error and sample size in the design and interpretation of the randomized control trial. N Engl J Med. 1978;299:690–4.
36. Cook DJ, Guyatt GL, Laupacis A, Sackett DL, Goldberg RJ. Clinical recommendations using levels of evidence for antithrombotic agents. Chest. 1995;108:227S–30S.
37. Guyatt GH, Sackett DL, Sinclair JC, Hayward R, Cook DJ, Cook RJ. Users' guide to the medical literature IX. A method for grading health care recommendations. JAMA. 1995;274:1800–4.
38. Moher D, Schulz KF, Altman DG, for the CONSORT Group. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. Ann Intern Med. 2001;134:657–62.
39. Mahon J, Laupacis A, Donner A, Wood T. Randomised study of n of 1 trials versus standard practice. BMJ. 1996;312:1069–74.
40. Detsky AS, Sackett DL. When was a 'negative' clinical trial big enough. Arch Intern Med. 1985;145:709–12.
41. Jones B, Jarvis P, Lewis JA, Ebbutt AF. Trials to assess equivalence: importance of rigorous methods. BMJ. 1996;313: 36–9.
42. Dunnett CW, Gent M. Significance testing to establish equivalence between treatments, with special reference to data in the form of 2×2 tables. Biometrics. 1977;33:593–602.
43. Pitt B, Zannad F, Remme WJ, et al. The effect of spironolactone on morbidity and mortality in patients with severe heart failure. N Engl J Med. 1999;341:709–17.
44. Naylor MD, Broten D, Campbell R, et al. Comprehensive discharge planning and home follow-up of hospitalized elders. JAMA. 1999; 281:613–20.
45. Dalby DM, Sellors JW, Fraser FD, Fraser C, van Ineveld C, Howard M. Effect of preventive home visits by a nurse on the outcomes of frail elderly people in the community: a randomized controlled trial. CMAJ. 2000;62:497–500.
46. Silverstein FE, Graham DY, Senior JR, et al. Misoprostol reduces serous gastrointestinal complications in patients with rheumatoid arthritis receiving nonsteroidal anti-inflammatory drugs. Ann Intern Med. 1995;123:241–9.
47. Dale JJ, Ruckley CV, Harper DR, Gibson B, Nelson EA, Prescott RJ. Randomised, double blind placebo controlled trial of pentoxifylline in the treatment of venous leg ulcers. BMJ. 1999;319:875–8.
48. GISSI-Prevenzione Investigators. Dietary supplementation with n-3 polyunsaturated fatty acids and vitamin E after myocardial infarction: results of the GISSI-Prevenzione trial. Lancet. 1999; 354:447–55.
49. Hulley S, Grady D, Bush T, et al. Randomized trial of estrogen for secondary prevention of coronary heart disease in postmenopausal women. Heart and Estrogen/progestin Replacement Study (HERS) Research Group. JAMA. 1998;280:605–13.
50. Pocock SJ. When to stop a clinical trial. BMJ. 1992;305:235–40.
51. Lau J, Antman EM, Jimenez-Silva J, Kupelnick B, Mosteller F, Chalmers TC. Cumulative meta-analysis of therapeutic trials for myocardial infarction. N Engl J Med. 1992;327:248–54.
52. Guyatt GH, Sinclair J, Cook DJ, Glasziou P. Users' guide to the medical literature XVI. How to use a treatment recommendation. JAMA. 1999;281:1836–43.

Copyright © 2002 EBSCO Publishing