Do clinicians understand the size of treatment effects? A randomized survey across 8 countries

Bradley C. Johnston PhD, Pablo Alonso-Coello MD PhD, Jan O. Friedrich MD, Reem A. Mustafa MD PhD, Kari A.O. Tikkinen MD PhD, Ignacio Neumann MD, Per O. Vandvik MD PhD, Elie A. Akl MD PhD, Bruno R. da Costa PhD, Neill K. Adhikari MD, Gemma Mas Dalmau MD, Elise Kosunen MD PhD, Jukka Mustonen MD PhD, Mark W. Crawford MD, Lehana Thabane PhD, Gordon H. Guyatt MD

See also www.cmaj.ca/lookup/doi/10.1503/cmaj.151254

Abstract

Background: Meta-analyses of continuous outcomes typically provide enough information for decision-makers to evaluate the extent to which chance can explain apparent differences between interventions. The interpretation of the magnitude of these differences — from trivial to large — can, however, be challenging. We investigated clinicians' understanding and perceptions of usefulness of 6 statistical formats for presenting continuous outcomes from meta-analyses (standardized mean difference, minimal important difference units, mean difference in natural units, ratio of means, relative risk and risk difference).

Methods: We invited 610 staff and trainees in internal medicine and family medicine programs in 8 countries to participate. Paper-based, self-administered questionnaires presented summary estimates of hypothetical interventions versus placebo for chronic pain. The estimates showed either a small or a large effect for each of the 6 statistical formats for presenting continuous outcomes. Questions addressed participants' understanding of the magnitude of treatment effects and their perception of the usefulness of the presentation format. We randomly assigned participants 1 of 4 versions of the questionnaire, each with a different effect size (large or small) and presentation order for the 6 formats (1 to 6, or 6 to 1).

Results: Overall, 531 (87.0%) of the clinicians responded. Respondents best understood risk difference, followed by relative risk and ratio of means. Similarly, they perceived the dichotomous presentation of continuous outcomes (relative risk and risk difference) to be most useful. Presenting results as a standardized mean difference, the longest standing and most widely used approach, was poorly understood and perceived as least useful.

Interpretation: None of the presentation formats were well understood or perceived as extremely useful. Clinicians best understood the dichotomous presentations of continuous outcomes and perceived them to be the most useful. Further initiatives to help clinicians better grasp the magnitude of the treatment effect are needed.

Competing interests: Bradley Johnston, Bruno da Costa and Gordon Guyatt conducted methodological work in the development of minimal important difference units as applied to meta-analyses, and the conversion of continuous meta-analytic data to binary. Jan Friedrich and Neill Adhikari conducted methodological work in the development of the ratio of means approach as applied to meta-analyses. These authors have no financial interests in the work related to this manuscript. No competing interests were declared by the remaining authors.

This article has been peer reviewed.

Accepted: Sept. 9, 2015
Online: Oct. 26, 2015

Correspondence to: Bradley Johnston, bradley.johnston@sickkids.ca

CMAJ 2016. DOI:10.1503/cmaj.150430

Health professionals increasingly rely on summary estimates from systematic reviews and meta-analyses to guide their clinical decisions and to provide information for shared decision-making. Meta-analyses of clinical trials typically provide the information necessary for decision-makers to evaluate the extent to which chance can explain apparent intervention effects (i.e., statistical significance). However, interpreting the magnitude of the treatment effect — from trivial to large — particularly for continuous outcome measures, can be challenging.

Such challenges include decision-makers' unfamiliarity with the instruments used to measure the outcome. For instance, without further information, clinicians may have difficulty grasping the importance of a 5-point difference on the Short-Form Health Survey-36 (SF-36) or a 1-point difference on a visual analogue scale for pain.[1] Second, trials often use different instruments to measure the same construct. For instance, investigators may measure physical function among patients with arthritis using 1 of 5 instruments (the Western Ontario and McMaster Universities Arthritis Index using either a visual analogue or Likert scale; the Arthritis Impact Measurement Scale; the SF-36 Physical Function; or the Lequesne index).[2,3]

Authors have several options for pooling results of continuous outcomes. When all trials have used

the same instrument to measure outcomes such as physical function or pain, the most straightforward method is to present the mean difference in natural units between the intervention and control groups. When trialists have used different instruments to measure the same construct, authors of systematic reviews typically report differences between intervention and control groups in standard deviation units, an approach known as the standardized mean difference (SMD). This approach involves dividing the mean difference in each trial by the pooled standard deviation for that trial's outcome.[4]

For meta-analyses of outcomes measured using different instruments, presenting results as an SMD is the longest standing and most widely used approach and is recommended in the Cochrane handbook for systematic reviews of interventions.[4] Limitations of this approach include, however, statistical bias toward decreased treatment effects,[5,6] the possibility that decision-makers will find the measure difficult to interpret[7,8] and the possibility that the same treatment effect will appear different depending on whether the study population had similar results in the measure of interest (i.e., if homogeneous, a small standard deviation) or varied greatly in the measure of interest (i.e., if heterogeneous, a large standard deviation).[9,10]

Several research groups have proposed alternative statistical formats for presenting continuous outcomes from meta-analyses that they postulate clinicians will more easily interpret.[6–8,11–16] The Grading of Recommendations Assessment, Development and Evaluation (GRADE) Working Group recently provided an overview of methods for presenting pooled continuous data.[9,10] These alternatives (Appendix 1, available at www.cmaj.ca/lookup/suppl/doi:10.1503/cmaj.150430/-/DC1), although intuitively compelling, have seen limited use.

We conducted a survey to determine clinicians' understanding of the magnitude of treatment effect for 6 approaches to the presentation of continuous outcomes from meta-analyses, as well as their perceptions of the usefulness of each approach for clinical decision-making. We also evaluated whether their understanding and perceptions of usefulness were influenced by country, medical specialty, clinical experience or training in health research methodology.

Methods

Design and study sample

We administered a paper-based survey to medical staff and trainees attending internal and family medicine educational sessions in 8 countries (Canada, Chile, Finland, Lebanon, Norway, Spain, Switzerland and the United States). We chose to survey clinicians in internal medicine and family medicine because these groups constitute a large proportion of physicians who treat chronic pain in their practice,[17] and because educational programs in both areas often emphasize using the medical literature to improve clinical decision-making.

To be eligible for inclusion, participants had to be competent in English, or the local language (at the University of Tampere, the survey was administered in Finnish; in all other centres, it was administered in English); had to attend educational rounds hosted by family medicine or internal medicine at one of the participating centres; and had to agree to participate. Medical staff and trainees who were not in an internal medicine or family medicine program but who attended educational sessions hosted by these training programs were also eligible. Clinicians were free to decline participation. Acceptance of the invitation to participate in the survey was deemed informed consent.

We collected data on the following characteristics of study participants: country, sex, specialty, professional status (medical student, resident, clinical fellow, or attending or staff physician), year of graduation from medical school and training in health research methodology.

Questionnaire development

The 6 statistical formats for presenting continuous outcomes from meta-analyses (described later in the section) were chosen on the basis of a review of the literature[16] and consensus among members of the GRADE Working Group.[9,10]

We pilot-tested the questionnaire in 2 centres among 20 clinicians not affiliated with our research team. Based on the feedback from each centre, we modified the survey to improve the clarity of the questions. The final version consisted of 5 demographic questions, 12 questions addressing clinician understanding of each presentation format and 6 questions addressing clinician preferences among the presentation formats.

Participants were given summary estimates of hypothetical interventions versus placebo for chronic pain. The estimates showed either a small or a large treatment effect for each of the 6 presentation formats. For each format, a core research team consisting of study authors, including clinical epidemiologists (B.C.J., J.O.F., P.O.V., B.R.D., N.K.A. and G.H.G.) and a senior statistician (L.T.), came to consensus on effect sizes corresponding to small and large treatment effects (Appendix 2, available at www.cmaj.ca/lookup/suppl/doi:10.1503/cmaj.150430/-/DC1). The 6 presentation formats were as follows: SMD, minimal important difference units, mean difference in natural units of the most familiar instrument (in

our example, pain on a 10-point numeric rating scale), ratio of means, relative risk and risk difference (Appendix 1). Participants were asked to focus on the magnitude of the treatment effect estimate and not on other issues such as potential adverse events, cost, inconvenience, systematic error (bias) or random error (precision).

For each question, we indicated that pooled estimates were precise with narrow confidence intervals (CIs) and very low p values. We developed 4 versions of the survey, each with a different combination by effect size (large or small) and, to address potential order effects, question order (1 to 6, or 6 to 1, for the 6 approaches), and arranged these versions in random order before distribution.

We used multiple-choice questions to assess clinicians' understanding of the magnitude of the treatment effect for each of the 6 presentation formats. Each question had a single correct answer. Responses were coded as correct (effect size correctly identified) or incorrect (effect size incorrectly identified). For example, the question addressing understanding of the SMD approach for a small treatment effect, presented as 0.20 standard deviation units, asked clinicians if the magnitude of difference was (a) trivial, probably not important; (b) small, but important; (c) moderate, surely important; or (d) large, very important.

To assess clinicians' perceptions of the usefulness of each presentation format for clinical decision-making, respondents were asked to use a 7-point Likert scale, with response options ranging from "not useful in understanding the size and importance of the effect" (1 point) to "extremely useful in understanding the size and importance of the effect" (7 points) (Appendix 3, available at www.cmaj.ca/lookup/suppl/doi:10.1503/cmaj.150430/-/DC1).

The complete survey for a small treatment effect with formats ordered from 1 to 6 is provided in Appendix 3. The answer key for the questions on understanding small treatment effects is shown in Appendix 4 (available at www.cmaj.ca/lookup/suppl/doi:10.1503/cmaj.150430/-/DC1).

Statistical analysis

Of our 2 primary objectives (understanding of the magnitude of treatment effect, and usefulness of each of the presentation formats), we based the sample size estimation for CIs for proportions on the first objective. Our outcome of interest for this objective was the proportion of correct responses. We chose to focus on this objective because a larger sample size would be required since the outcome is binary. To estimate the proportion of correct answers, we conducted a pilot study of 20 clinicians in Toronto at The Hospital for Sick Children and St. Michael's Hospital; the proportion of correct responses for the 6 presentation formats ranged from 21% to 50%. We chose 50% as the most conservative value, since it has the largest standard deviation, and applied this value and our desired CI width of 10% (margin of error ± 0.05) to estimate the number of survey participants needed. The required sample size was 384.

We performed descriptive analyses to summarize participant characteristics. For the analysis of our primary objectives, the estimates were based on available cases (i.e., questionnaires with responses to some or all questions) and were reported as proportions and means with corresponding 95% CIs, including tests of significance addressing differences across statistical presentation formats for each objective. A priori, we hypothesized that dichotomous statistical formats, with which clinicians are most familiar, would be best understood and perceived as most useful.

For our secondary analysis of factors associated with understanding of the 6 statistical formats, we used multiple logistic regression to determine factors associated with correct responses. The factors explored included country, specialty (internal medicine v. family medicine), clinical experience (≤ 10 yr v. > 10 yr) and level of training in health research methodology (none, some training, graduate degree completed). Linear regression was used for perceived usefulness. For the logistic regression analysis, we expressed the results as odds ratios (ORs) with corresponding 95% CIs. For the linear regression analysis, we expressed the results as coefficients with 95% CIs. We used the χ^2 test for contingency tables to compare proportion of correct responses between various subgroups for secondary analysis. We tested the order effect using a χ^2 test of independence for understanding and an independent sample t test for perceived usefulness. The threshold for statistical significance was set at a p value of 0.05.

We performed analyses using SPSS version 21 (IBM SPSS Statistics for Windows) and SAS 9.2 (SAS Institute).

Ethics approval

The study design was approved by the ethics review boards at the study centres in 6 of the 8 countries (Canada: University of Toronto, The Hospital for Sick Children and McMaster University; Lebanon: American University of Beirut; Norway: University of Oslo; Chile: Pontificia Universidad Catolica de Chile; United States: University of Missouri–Kansas City; Spain: Universitat Autonoma de Barcelona). The requirement for ethics approval was waived by the institutional review boards in 2 countries (Finland: University of Tampere; and Switzerland: Cantonal Ethics Committee of Bern).

Results

A total of 610 staff and trainees were present during internal and family medicine rounds when the surveys were introduced. Overall, 531 completed and returned their questionnaire, for a response rate of 87% (range 72%–100% across the participating centres); 482 (79.0%) provided answers to all of the questions. Of the 531 respondents, 45.6% were from Canada or the United States, 56.9% were women, 48.8% were in internal medicine, 43.1% had graduated from medical school within 5 years of completing the survey, and 44.3% had formal graduate training in health research methodology (Table 1).

Table 1: Characteristics of 531 clinicians who participated in the survey

Characteristic	No. (%) of respondents
Sex	
Female	302 (56.9)
Male	228 (42.9)
Not specified	1 (0.2)
Country	
Canada	179 (33.7)
Chile	27 (5.1)
Finland	57 (10.7)
Lebanon	26 (4.9)
Norway	27 (5.1)
Spain	129 (24.3)
Switzerland	23 (4.3)
United States	63 (11.9)
Specialty	
Internal medicine	259 (48.8)
Family medicine	225 (42.4)
Other	47 (8.8)
Professional status	
Resident	343 (64.6)
Attending/staff	110 (20.7)
Clinical fellow	20 (3.8)
Medical student	19 (3.6)
Other	39 (7.3)
Year of graduation or expected graduation from medical school	
Before 1990	50 (9.4)
1990–1999	41 (7.7)
2000–2009	205 (38.6)
2010 or later	229 (43.1)
Not specified	6 (1.1)
Training in health research methodology	
Never completed a formal course	293 (55.2)
Completed formal courses, but no masters or doctorate in health research methodology	219 (41.2)
Has masters or doctorate in health research methodology	16 (3.0)
Not specified	3 (0.6)

Understanding of statistical formats

Risk difference proved to be the best understood statistical format for presenting continuous outcomes from meta-analyses (correct response given by 40.0%); however, correct responses were submitted by less than 50% for each of the 6 formats (Figure 1). Relative risk (34.9%) and ratio of means (33.0%) were similarly understood, followed by SMD (29.6%). Minimal important difference units and mean difference in natural units were the least well understood formats. The χ^2 test showed a significantly higher level of understanding for dichotomous approaches (37.3% correct) than for continuous approaches (24.9% correct) ($p < 0.001$). When we explored presentation approach and used risk difference (the best understood approach) as the reference category, we found that respondents were less likely to understand the continuous approaches: SMD ($p = 0.002$), MID units ($p < 0.001$), mean difference in natural units ($p < 0.001$) and ratio of means ($p < 0.03$).

Perceived usefulness of formats

Respondents found continuous outcomes that were dichotomized (i.e., risk difference and relative risk) as moderately useful for clinical decision-making; mean difference in natural units and ratio of means still somewhat useful; and MID units and SMD of limited use (Figure 2). A paired-samples t test showed a significantly higher level of perceived usefulness for dichotomous approaches compared with continuous approaches ($p < 0.001$). When exploring presentation approach and using risk difference (the most useful approach) as the reference category, participants were less likely to perceive each of the remaining approaches as useful (all comparison p values < 0.002).

Factors associated with understanding statistical formats

We found no significant association between country, clinical experience or level of graduate training in health research methodology and the understanding of any of the 6 statistical formats ($p > 0.2$ for each). Respondents whose specialty was internal medicine had a better understanding of the various presentation formats compared with those in family medicine (33% v. 25%, $p = 0.004$) (Figure 3). We also conducted a post hoc logistic

regression analysis with understanding as the dependent variable and included the following interaction terms: country and clinical experience; country and training in health research methodology; country and specialty. None of the interactions were statistically significant (data not shown).

The order in which the statistical formats were presented did not affect the clinicians' accuracy of understanding. Aside from relative risk, the respondents randomly assigned to the group shown a small treatment effect consistently had more correct responses than those allocated to the group shown a large treatment effect, and the difference was statistically significant ($p < 0.01$) (Appendix 5, available at www.cmaj.ca/lookup/suppl/doi:10.1503/cmaj.150430/-/DC1).

Factors associated with perceived usefulness

With Canadian clinicians as the reference category, we found that respondents from the US and Spain reported a lower perceived usefulness of all presentation formats ($p < 0.007$), whereas those from Chile reported higher usefulness ($p = 0.001$). Compared with respondents who had no training in research methodology, those with some training or a graduate degree in research methodology reported higher ratings of usefulness ($p < 0.001$). Compared with respondents in family medicine, those in internal medicine reported higher perceived usefulness ($p < 0.001$). Clinical experience was not a significant factor (Appendix 6, available at www.cmaj.ca/lookup/suppl/doi:10.1503/cmaj.150430/-/DC1).

The order in which the statistical formats were presented did not affect perceived usefulness except for SMD ($p < 0.04$) (data not shown). Clinicians randomly assigned to the group shown a large treatment effect perceived SMD and ratio of means to be more useful than the remaining formats ($p < 0.02$) (Appendix 7, available at www.cmaj.ca/lookup/suppl/doi:10.1503/cmaj.150430/-/DC1).

Interpretation

Respondents understood magnitude of effect best when presented as a risk difference; relative risk and ratio of means were the next best understood formats, although no approach was well understood. Respondents found the dichotomous presentations of continuous outcomes (relative risk and risk difference) more useful than the continuous approaches. Of the continuous presentation methods, mean difference in natural units was perceived to be the most useful. This was followed by ratio of means, which was found to be the most useful of the continuous methods required when results among trials are reported in different units (which precludes the use of mean difference), followed by minimal important difference units. The method perceived to be the least useful of all 6 formats was SMD, although no approach was perceived as very useful.

Each statistical format was accompanied by 4 response options, so by chance alone, respondents should provide the correct answer 25% of the time. Responses for mean difference in natural units and minimal important difference units were below what would be expected by chance (19%). Although respondents performed above chance for the dichotomous formats (35% chose the correct response for relative risk and 40% for risk difference), no format was well understood by the respondents.

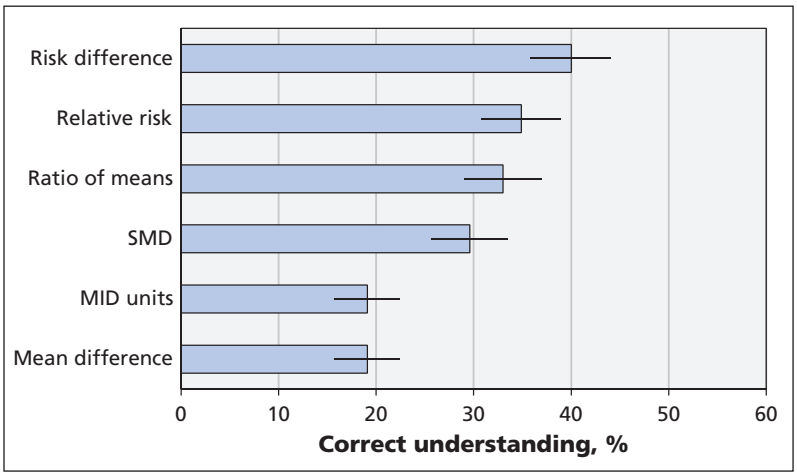

Figure 1: Respondents' understanding of the magnitude of the treatment effect for each of 6 statistical formats used to present continuous outcomes from meta-analyses. Higher percentages represent greater understanding; error bars = 95% confidence intervals. Mean difference = mean difference in natural units, MID = minimal important difference, SMD = standardized mean difference.

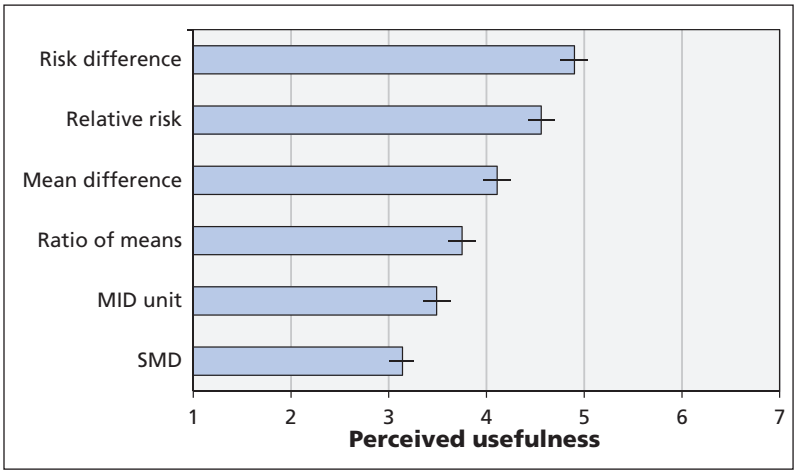

Figure 2: Perceived usefulness of each statistical format for clinical decision-making. Higher scores represent higher perceived usefulness; error bars = 95% confidence intervals. Mean difference = mean difference in natural units, MID = minimal important difference, SMD = standardized mean difference.

Earlier studies have shown that doctors have limitations in their understanding of statistics.[20] Many studies have shown that clinicians presented with relative effects, rather than the corresponding absolute effects, will perceive the effects as larger and will be more inclined to recommend treatments.[21,22] Studies have not previously addressed the questions of accuracy and usefulness that we investigated.

Strengths and limitations

Strengths of our study include a comprehensive choice of presentation approaches based on prior methodologic work.[9,10,16] Participation by 531 clinicians recruited from 11 academic and non-academic centres in 8 countries enhances the generalizability of our findings and allowed us to compare differences across countries, clinical experience and level of graduate training in health research methodology. We also surveyed both internal medicine and family medicine clinicians, which allowed us to compare across these specialties. Rather than conducting an online survey that would likely have resulted in low response rates,[18,19] we chose to canvas staff and trainees in person at the beginning of regularly scheduled educational or grand rounds, which achieved a response rate of 87%. Finally, we were able to prove our a priori hypotheses: dichotomous measures would be best understood and perceived as most useful.

Our study has limitations. We presented clinicians with only one clinical scenario, chronic pain. Generalizing our findings to other continuous outcomes is therefore open to question. Chronic pain is, however, a common condition familiar to both family medicine and internal medicine staff and trainees and to many other specialties (e.g., anesthesiologists, pediatricians, obstetricians, oncologists, general surgeons and orthopedic surgeons). Generalizability to community-based settings may be limited because we conducted our study primarily within academic medical centres.

What constitutes a small and large effect for the 6 measures we chose is a matter of judgment. To the extent that our choices are open for debate, one could argue that alternative answers may have been correct. This possibility gains credence in that, for 5 of the 6 presentations of effect, respondents understood small treatment effects better than large ones. This suggests we may have been too conservative in choosing our large treatment effects. To the extent that alterna-

Factor	Odds ratio (95% CI)
Country (v. Canada)	
Switzerland	0.73 (0.46–1.14)
Finland	0.78 (0.57–1.08)
Lebanon	0.81 (0.55–1.21)
United States	0.84 (0.63–1.12)
Chile	0.87 (0.57–1.33)
Spain	0.92 (0.73–1.17)
Norway	1.03 (0.69–1.54)
Specialty (v. internal medicine)	
Family medicine	0.74 (0.61–0.91)
Training in health research methodology (v. none)	
Some training	0.93 (0.77–1.12)
Masters or doctorate	1.18 (0.74–1.90)
Clinical experience (v. ≤ 10 yr)	
> 10 yr	0.93 (0.75–1.16)
Presentation format (v. risk difference)	
SMD	0.65 (0.49–0.86)
MID units	0.34 (0.25–0.47)
Mean difference	0.33 (0.24–0.45)
Ratio of means	0.72 (0.55–0.94)
Relative risk	0.79 (0.61–1.04)

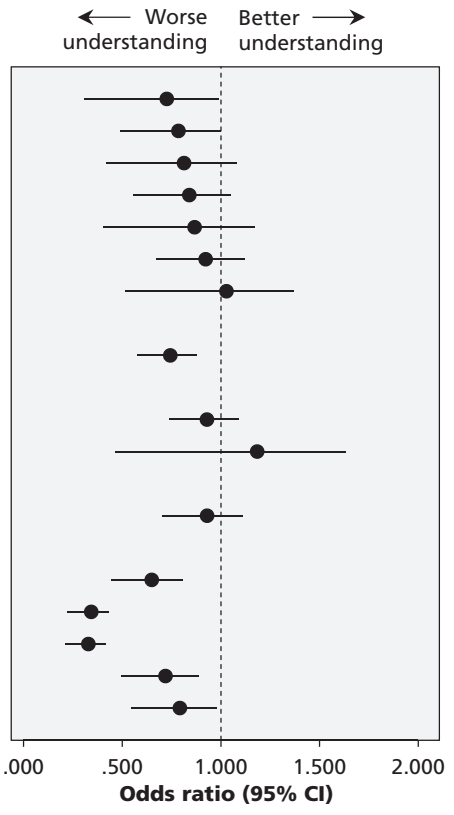

Figure 3: Factors associated with respondents' understanding of the statistical formats. An odds ratio below 1.0 indicates a worse understanding than the reference category. CI = confidence interval, mean difference = mean difference in natural units, MID = minimal important difference, SMD = standardized mean difference.

tive answers may have been correct (e.g., moderate rather than large for effects we designated as large), our results may underestimate clinicians' understanding.

Implications for practice and research

Every day, clinicians and patients discuss trade-offs between alternative management strategies. Optimal shared decision-making requires that clinicians communicate the magnitude of benefits and harms: at the very least, they must help patients understand whether the effects are large or small. This is true for all patient-important outcomes, including those measured as continuous variables such as pain and quality of life. Clinicians must therefore be trained to understand the magnitude of effect, and researchers must help clinicians by presenting results in an optimally interpretable way.[9,10]

On the basis of experience with teaching evidence-based medicine to clinicians, our a priori hypothesis was that dichotomous measures with which clinicians are most familiar would be best understood and perceived as most useful. Moreover, there is a large body of research on how to make these measures best understood by patients.[21,23] Research involving both clinicians and patients has shown that, of the binary presentation measures, risk difference is most useful for clinical decision-making.[23] Our results are consistent with this observation and suggest that researchers presenting results using continuous outcomes with which clinicians may not be familiar should also present absolute effects, and likely relative effects as well. The methodology for transforming continuous outcomes and presenting them as dichotomous outcomes is well established.[7,11,16]

Troubling, however, was the low rate of correct responses in our study, even for the dichotomous presentations. Even if one considers the underestimation of clinicians' understanding because of the debatable categorization of large and small effects, our results suggest important limitations in understanding. The results therefore highlight the need for enhanced education in interpreting treatment effects in both undergraduate and postgraduate medical programs.

Conclusion

None of the presentation formats was well understood by the respondents or perceived as extremely useful in clinical decision-making. Clinicians best understood the dichotomous presentation of continuous outcomes (relative risk and risk difference) and perceived these formats to be the most useful. We found that the presentation of results of continuous outcomes as an SMD, the longest standing and most widely used statistical approach, was poorly understood and perceived as the least useful format. Further initiatives that guide clinicians to better grasp the magnitude of treatment effects are needed.

References

1. Guyatt GH, Osoba D, Wu AW, et al. Methods to explain the clinical significance of health status measures. *Mayo Clin Proc* 2002;77:371-83.
2. Fransen M, McConnell S. Exercise for osteoarthritis of the knee. *Cochrane Database Syst Rev* 2008;(4):CD004376.
3. Juhl C, Lund H, Roos EM, et al. A hierarchy of patient-reported outcomes for meta-analysis of knee osteoarthritis trials: empirical evidence from a survey of high impact journals. *Arthritis* 2012;2012:136245.
4. Higgins JPT, Green S, editors. *Cochrane handbook for systematic reviews of interventions.* Version 5.1.0 [updated March 2011]. London (UK): Cochrane Collaboration; 2011. Available: www.cochrane-handbook.org (accessed 2015 May 29).
5. Van Den Noortgate W, Onghena P. Estimating the mean effect size in meta-analysis: bias, precision, and mean squared error of different weighting methods. *Behav Res Methods Instrum Comput* 2003;35:504-11.
6. Friedrich JO, Adhikari NK, Beyene J. The ratio of means method as an alternative to mean differences for analyzing continuous outcome variables in meta-analysis: a simulation study. *BMC Med Res Methodol* 2008;8:32.
7. da Costa BR, Rutjes AW, Johnston BC, et al. Methods to convert continuous outcomes into odds ratios of treatment response and numbers needed to treat: meta-epidemiological study. *Int J Epidemiol* 2012;41:1445-59.
8. Johnston BC, Thorlund K, Schunemann HJ, et al. Improving the interpretation of quality of life evidence in meta-analyses: the application of minimal important difference units. *Health Qual Life Outcomes* 2010;8:116.
9. Guyatt GH, Thorlund K, Oxman AD, et al. GRADE guidelines: 13. Preparing summary of findings tables and evidence profiles-continuous outcomes. *J Clin Epidemiol* 2013;66:173-83.
10. Johnston BC, Patrick DL, Thorlund K, et al. Patient-reported outcomes in meta-analyses — part 2: methods for improving interpretability for decision-makers. *Health Qual Life Outcomes* 2013;11:211.
11. Anzures-Cabrera J, Sarpatwari A, Higgins JP. Expressing findings from meta-analyses of continuous outcomes in terms of risks. *Stat Med* 2011;30:2967-85.
12. Friedrich JO, Adhikari NK, Beyene J. Ratio of means for analyzing continuous outcomes in meta-analysis performed as well as mean difference methods. *J Clin Epidemiol* 2011;64:556-64.
13. Furukawa TA. From effect size into number needed to treat. *Lancet* 1999;353:1680.
14. Hasselblad V, McCrory DC. Meta-analytic tools for medical decision making: a practical guide. *Med Decis Making* 1995;15:81-96.
15. Johnston BC, Thorlund K, da Costa BR, et al. New methods can extend the use of minimal important difference units in meta-analyses of continuous outcome measures. *J Clin Epidemiol* 2012;65:817-26.
16. Thorlund K, Walter SD, Johnston BC, et al. Pooling health-related quality of life outcomes in meta-analysis — a tutorial and review of methods for enhancing interpretability. *Res Synth Methods* 2011;2:188-203.
17. Upshur CC, Luckmann RS, Savageau JA. Primary care provider concerns about management of chronic pain in community clinic populations. *J Gen Intern Med* 2006;21:652-5.
18. Kongsved SM, Basnov M, Holm-Christensen K, et al. Response rate and completeness of questionnaires: a randomized study of Internet versus paper-and-pencil versions. *J Med Internet Res* 2007;9:e25.
19. Yetter G, Capaccioli K. Differences in responses to Web and paper surveys among school professionals. *Behav Res Methods* 2010;42:266-72.
20. Martyn C. Risky business: doctors' understanding of statistics. *BMJ* 2014;349:g5619.
21. Akl EA, Oxman AD, Herrin J, et al. Using alternative statistical formats for presenting risks and risk reductions. *Cochrane Database Syst Rev* 2011;16:CD006776.
22. Montori VM, Jaeschke R, Schünemann HJ, et al. Users' guide to detecting misleading claims in clinical research reports. *BMJ* 2004;329:1093-6.
23. Zipkin DA, Umscheid CA, Keating NL, et al. Evidence-based risk communication: a systematic review. *Ann Intern Med* 2014; 161:270-80.

Affiliations: Department of Anesthesia and Pain Medicine (Johnston, Crawford), The Hospital for Sick Children, University of Toronto, Toronto, Ont.; Institute of Health Policy, Management and Evaluation (Johnston), Dalla Lana School of Public Health, University of Toronto, Toronto, Ont.; Child Health Evaluative Sciences (Johnston), The Hospital for Sick Children Research Institute, Toronto, Ont.; Iberoamerican Cochrane Center (Alonso-Coello, Dalmau), Biomedical Research Institute Sant Pau, CIBER Epidemiologiay Salud Publica, Barcelona, Spain; Departments of Critical Care and Medicine (Friedrich), St. Michael's Hospital, Toronto, Ont.; Department of Medicine and Interdepartmental Division of Critical Care (Friedrich), University of Toronto, Toronto, Ont.; Departments of Medicine and Biomedical and Health Informatics (Mustafa), University of Missouri–Kansas City, Kansas City, Mo.; Department of Clinical Epidemiology and Biostatistics (Mustafa, Tikkinen, Neumann, Akl, Thabane, Guyatt), McMaster University, Hamilton, Ont.; Departments of Urology and Public Health (Tikkinen), Helsinki University Central Hospital and University of Helsinki, Helsinki, Finland; Department of Internal Medicine (Neumann), Pontificia Universidad Catolica de Chile, Santiago, Chile; Institute of Health and Society (Vandvik), Faculty of Medicine, University of Oslo, Oslo, Norway; Department of Medicine (Vandvik), Innlandet Hospital Trust, Division Gjövik, Norway; Clinical Epidemiology Unit (Akl), American University of Beirut, Beirut, Lebanon; Institute of Primary Health Care (da Costa), University of Bern, Bern, Switzerland; Department of Critical Care Medicine and Sunnybrook Research Institute (Adhikari), Sunnybrook Health Sciences Centre, Toronto, Ont.; Interdepartmental Division of Critical Care (Adhikari), University of Toronto, Toronto, Ont.; School of Medicine and Centre for General Practice (Kosunen), University of Tampere and Pirkanmaa Hospital District, Tampere, Finland; School of Medicine and Department of Internal Medicine (Mustonen), University of Tampere and Tampere University Hospital, Tampere, Finland; Physiology and Experimental Medicine (Crawford), The Hospital for Sick Children Research Institute, Toronto, Ont.; Biostatistics Unit of the Centre for Evaluation of Medicines (Thabane), McMaster University, Hamilton, Ont.; Population Health Research Unit, Hamilton Health Sciences (Thabane), McMaster University, Hamilton, Ont.; Department of Medicine (Guyatt), McMaster University, Hamilton, Ont.

Contributors: Bradley Johnston, Jan Friedrich, Kari Tikkinen, Per Vandvik, Bruno da Costa, Lehana Thabane and Gordon Guyatt contributed to the study concept and design. Bradley Johnston, Pablo Alonso-Coello, Jan Friedrich, Reem Mustafa, Kari Tikkinen, Ignacio Neumann, Per Vandvik, Elie Akl, Bruno da Costa, Neill Adhikari, Gemma Mas Dalmau, Elise Kosunen, Jukka Mustonen and Mark Crawford contributed to the data collection. Bradley Johnston, Lehana Thabane and Gordon Guyatt contributed to the data analysis. All of the authors contributed to the preparation of the manuscript, approved the final version to be published and agreed to act as guarantors of the work.

Funding: This study was funded in part by the European Union's Seventh Framework Programme for research, technological development and demonstration (grant no. 258583). The funding body had no role in the concept, the collection and analysis of data or the reporting of the final results. Sole responsibility lies with the authors. The European Commission is not responsible for any use that may be made of the information contained herein.

Acknowledgement: We thank our colleagues in Spain who assisted with site coordination and data collection, including Merce Marzo, Javier Zamora, Jose Ignacio Emperanza, Rafael Rotaeche, Eva Muro, Marta Besa, Mariam de la Poza and Eulalia Mariñelarena Mañero.