Received: 12 June 2020 | Revised: 10 July 2020 | Accepted: 13 July 2020

DOI: 10.1111/jep.13455

ORIGINAL PAPER

Journal of Evaluation in Clinical Practice
International Journal of Public Health Policy and Health Services Research

WILEY

Statements considering intervention effects in Finnish clinical practice guidelines: Recommending interventions with non-numeric effect-sizes or unspecified outcomes

Eero Raittio DDS, PhD[1,2] | Lauri Raittio MD, BSc[3]

[1]The University of Eastern Finland, Institute of Dentistry, Kuopio, Finland

[2]City of Tampere, Oral Health Care, Tampere, Finland

[3]Tampere University, The Faculty of Medicine and Health Technology, Tampere, Finland

Correspondence
Lauri Raittio, Tampere University, The Faculty of Medicine and Health Technology, Arvo Ylpön katu 34, 33520, Tampere, Finland.
Email: lauri.raittio@tuni.fi

Abstract

Rationale, Aims and Objectives: Representation of benefits and harms associated with specific interventions in an understandable and comparable way is crucial for informed decision making that clinical practice guidelines (CPGs) aim to enhance. Therefore, we investigated how statements concerning the effects of interventions considered and described benefits and harms, magnitude of effect and its uncertainty, numeric and non-numeric information, and outcomes in Finnish CPGs.

Methods: We selected 10 CPGs on common diseases and risk factors published by The Finnish Medical Society, Duodecim. All the statements which were graded with the level of evidence from high to very low (levels A-D) were included in analyses. From these statements, assessments were made regarding whether the statement considered benefits or harms, whether relative or absolute numeric measures were shown, whether the statement supported or was against the intervention considered, and what outcome was reported.

Results: Of the 10 CPGs, 448 statements were assessed. Most of the statements of effects considered intervention benefits (87%) rather than harms. Half of the statements considering harms were represented in a way that supported the intervention. Most of the statements (94%) did not include numeric estimates of magnitude of the effect. When numeric estimates of magnitude of the effect were present, they were most frequently relative measures and were typically placed in a statement considering (a) intervention benefits with a primary outcome, (b) given the grade of A for level of evidence, and (c) that supported the use of intervention.

Conclusions

In the Finnish CPGs, the statements were rarely framed with both absolute and relative numeric measures of an intervention's effect. Harms were rarely reported with a grade indicating the level of evidence. The users of CPGs would benefit from more

consistent and understandable framing of statements considering both benefits and harms of interventions.

KEYWORDS
clinical practice guideline, comprehension, data visualization, decision making, guideline adherence, health communication, health literacy

1 | INTRODUCTION

Informed decision making in clinical practice requires the knowledge of benefits and harms in the context of the practice and the patient. Clinical practice guidelines (CPGs) aim to synthesize evidence and provide a summary of recommendations to users of the guidelines. To achieve this goal, the recommendations should be communicated as clearly and as simply as possible. The quality of evidence and its level of uncertainty comprise the recommendation that should present the size of the effect and the level of certainty (or confidence in) of the effect.

Effects of interventions are most often framed in relative numeric measures (ie, risk ratio, hazard ratio), although patients experience the benefits and harms of the interventions in absolute numeric measures.[1,2] Even high relative intervention effects (ie, a 50% relative risk reduction) can translate into minor absolute risk differences (ie, 1 or 2 per 100 patients). However, relative effects often are similar regardless of baseline risk, whereas absolute effects are more dependent on the baseline risk. Thus, relative numeric measures may be more misleading than absolute ones.[3] In addition, only verbal descriptions of intervention effects may be misleading or ambiguous.[4]

The chosen framing influences the views and judgement of clinicians and patients. In general, presentation of relative risks of benefits exaggerates the pros and absolute risks underrate the cons of the interventions. A decade ago, Cochrane and non-Cochrane systematic reviews reported absolute treatment effects only in one third of the reviews.[1] Currently, systematic reviews published by Cochrane are required to report the results of intervention effects both with absolute and relative numeric measures.[5,6]

The low level of numeracy and knowledge related to health statistics among doctors and patients are barriers to informed decision making in clinical practice.[7-10] However, by choosing the right illustrations of the information, understanding evidence can be substantially improved by doctors and patients.[10-13] The appropriately chosen figures and tables, such as the icon arrays and the fact boxes displaying the evidence, were shown to increase the level of comprehension of the evidence and are probably applicable to many common real-world settings.[12-17] The CPGs do not provide only the objective evidence but combine this information with applicable statements of recommendations to users and also influence whether or not certain treatments are used. Morgott et al found a low level of use of absolute effect size measures in the drug therapy recommendations for two common diseases in the CPGs published in English or German languages.[18] Otherwise, is it not well known how statements related to intervention benefits or harms are framed with numerical or non-numerical information or whether figures tables or other illustrations have been used to illustrate these effects in the CPGs.

In Finland, developers of the CPGs provide synthesis of evidence in line with the recommendations of the Grading of Recommendations Assessment, Development and Evaluation (GRADE) working group.[19] The statements of evidence (eg, statins reduce mortality) should be consistent with the certainty of the evidence (grade categories from A to D) which should be based on deliberate overall consideration of beneficial or harmful effects of the intervention and also take into account the importance of outcomes used in the literature.[20] However, the developers of the Finnish CPGs have a considerable degree of freedom for presenting the benefits and harms of the interventions. Therefore, the Finnish CPGs can provide fruitful insight into the ways of framing the benefits and harms of treatments in the CPG setting.

Our aim was to investigate how statements concerning effects of interventions considered (a) benefits and harms, (b) magnitude of effect and its uncertainty, (c) numeric and non-numeric information, (d) outcomes, (e) strength of recommendation, and (f) tables and figures to communicate the effects of interventions in 10 Finnish CPGs on common diseases and risk factors.

2 | METHODS

2.1 | The context of the CPGs in the assessed data

We selected 10 CPGs published by the Finnish Medical Society Duodecim which are publicly available in Finnish and of which few are translated into English (https://www.kaypahoito.fi/en/guidelines). Both of the authors are native speakers of Finnish and fluent in English. The developmental process of the CPGs does not have direct financial ties to a medical industry and the process involves a multidisciplinary team of medical professionals with a lead author, who is often a professor in the specialty area of the CPG topic. However, the intellectual and financial ties of the authors of the CPG of the last 3 years are reported in the end of the CPG document. The published CPGs are publicly available and they are developed for the applicability of the clinical practitioners in mind. However, the CPGs are also accessed by the greater public as well as patients.

The Finnish Medical Society Duodecim has published CPGs for over 25 years in Finnish and the total number of updated CPGs is 104 as of June 2020. Ten CPGs on topics with a high prevalence of

diseases and risk factors conjointed with well-known effective treatments were selected for this study. We chose CPGs on the following topics: benign prostatic hyperplasia; chronic pulmonary obstructive disease; depression; dyslipidaemia; hypertension; osteoarthritis of knee and hip; osteoporosis; stable coronary artery disease; stroke and transient ischemic attack; type 2 diabetes mellitus. These CPGs were last updated between 2014 and 2020 and the length of the PDF files of these CPGs range from 17 to 60 pages. The English translations show the structure used in the CPGs published by the Finnish Medical Society Duodecim (link above). In short, CPGs consist of a long list of bullet points arrayed to two levels of hierarchy associated with a few tables and figures.

The main texts of the CPGs produced by the Finnish Medical Society Duodecim are supplemented by the systematic literature reviews of the level of evidence of the recommendations outlined in the CPG. These systematic literature reviews are conducted by the CPG authors with the help of informatician and library secretary in the Medline and Cochrane library. The authors select the most suitable articles of the question in hand (prefer the most updated and rigorous meta-analysis if extant) and compose narrative synthesis of evidence that are linked to the recommendations of the CPG electronically as supplementary material. The level of evidence of the recommendations derived from these narrative synthesis of evidence is categorized by the approach of the GRADE working group into four categories: A, B, C, and D, in which A represents 'High' and D represents 'Very low' confidence in the recommendation. All the selected CPGs contain dozens of these systematic literature reviews on various topics conducted by 5 to 10 authors in theirs' respective field of expertise.

In the guidelines of the CPG itself, the Finnish Medical Society Duodecim refers to the GRADE working group by linking the wording of the recommendations to the level of evidence: that is, 'X intervention may reduce mortality [B]'. Nevertheless, the CPGs are not homogenous in this regard and variability exists in interpretations of the evidence and representation of intervention outcomes, effect size, and uncertainty.

2.2 | Data assessment, extraction, and analysis

The 10 selected and above-mentioned CPGs were screened for the statements concerning the effects of intervention(s) which were provided with the associated systematic literature review and category of level of evidence (A-D, Table 1). For every statement, we assessed the outcome category (primary/secondary/unspecified, Table 1). If a statement included several outcomes, the statement was assessed based on the most essential (primary or secondary) outcome. For instance, a statement 'Teriparatide presumably increases bone density and reduced vertebra fractures among those on glucocorticoid therapy' was considered as a statement with a primary outcome (ie, fracture). We also assessed the representation of the magnitude of the effect (relative numeric measures/absolute numeric measures/both absolute and relative numeric measures/non-numeric measures) and whether the statement considered benefits or harms of the intervention (nature of statement) (Table 1). If a statement included both benefits and harms, the statement was assessed twice. For instance, the statement 'Theophylline presumably improves the lung function in chronic obstructive pulmonary disease, but then it increases the adverse effects' was analysed separately for 'Theophylline presumably improves the lung function in chronic obstructive pulmonary disease' and 'but then it increases the adverse effects'. Also, if a statement

TABLE 1 Examples of outcomes, representation of magnitude of effect, and representation of intervention effectiveness or safety

Outcomes	Examples
Primary	Mortality, survival, stroke, myocardial infarction, fractures, suicide, gastric perforation
Secondary	Quality of life, laboratory measures, diabetes incident, coronary disease, hyper/hypoglycaemia, COPD exacerbation
Unstated	Reference to prognosis, unspecific symptom relief, 'is recommended', 'is beneficial', 'with no harms', 'is efficient', 'better results'
Representation of magnitude of effect	
Non-numeric	Electroconvulsive therapy is the most effective, is a safe, and recommendable treatment modality.
Relative	The high-dose statin therapy reduced myocardial infarctions and strokes 16% to 18% compared to the regular dose.
Absolute	Ambulatory measured daytime blood pressure reduces by 3/3 mm Hg for hypertensive and normotensive patients.
Both relative and absolute	Acetylsalicylic acid and dipyridamole joint use reduces risk of stroke by 37% to 38% in patients with a history of stroke or transient ischemic attack compared to the placebo (absolute risk reduction of 5.9% in the next 2 years, numbers needed to treat 17).
Representation of intervention effectiveness (benefit-statements) or safety (harm-statements)	
Support	Strontium ranelate increases bone mineral density and decreases the incidence of the clinical and the radiological vertebra fractures on women with postmenopausal osteoporosis.
Weakly support	Mindfulness-based therapy is presumably effective acute treatment for adolescent depression.
Neutral	There is not enough evidence on the effects of speech therapy on dysartria.
Weakly against	The prolonged therapeutic exercise after the hospitalization does not seem to increase the range of motion after the total arthroplasty of the knee.
Against	A transurethral resection of the prostate causes the ejaculation dysfunction, retrograde ejaculation.

contained several references to different levels of evidence reviews, the statement was assessed separately for each level of evidence. For instance, the statement: 'The restriction of the natrium intake improves the effect of the hypertensive drugs, especially of angiotensin converting enzyme inhibitors [B], angiotensin II receptor blocker [C], beta blockers [B] and diuretics [B]' was assessed as four different statements, separately for each level of evidence (intervention).

In addition, we evaluated in what light statements represented an intervention; was it against an intervention or supporting its usage? We categorized statements to five levels: support, weakly support, neutral, weakly against, or against. If a statement referred to an unconditional effect of an intervention or lack of adverse effects, it was considered to support the intervention (ie, its effectiveness or safety), while if a statement mentioned adverse effects or lack of favourable effects it was deemed to be against the intervention (ie, its effectiveness or safety). If a statement showed some uncertainty, it was considered as a weaker statement. Reference to 'lack of evidence' or statements that did not consider any effects were interpreted to be neutral statements. For

TABLE 2 Number of statements in clinical practice guidelines stratified by benefit- and harm-statements

	Total		Benefit-statements			Harm-statements		
	n	Col %	n	Col %	Row %	n	Col %	Row %
Clinical practice guideline								
Osteoarthritis of knee and hip	75	17	62	16	83	13	22	17
Stroke and transient ischemic attack	90	20	75	19	83	15	26	17
Stable coronary artery disease	10	2	10	3	100	0	0	0
Chronic pulmonary obstructive disease	35	8	32	8	91	3	5	9
Depression	93	21	90	23	97	3	5	3
Type 2 diabetes mellitus	21	5	15	4	71	6	10	29
Dyslipidaemia	23	5	20	5	87	3	5	13
Hypertension	34	8	32	8	94	2	3	6
Osteoporosis	22	5	22	6	100	0	0	0
Benign prostatic hyperplasia	45	10	32	8	71	13	22	29
Nature of statement								
Benefit	390	87	390	100	100	0	0	0
Harm	58	13	0	0	0	58	100	100
Grade of evidence								
A (high)	214	48	182	47	85	32	55	15
B (moderate)	135	30	120	31	89	15	26	11
C (low)	87	19	77	20	89	10	17	11
D (very low)	12	3	11	3	92	1	2	8
Outcome in statement								
Primary	79	18	68	17	86	11	19	14
Secondary	190	42	157	40	83	33	57	17
Unspecific	179	40	165	42	92	14	24	8
Representation of magnitude of effect								
Non-numeric	419	94	361	93	88	53	100	12
Relative	20	4	20	5	100	0	0	0
Absolute	6	1	6	2	100	0	0	0
Both relative and absolute	3	1	3	1	100	0	0	0
Representation of intervention effects or safety								
Support	229	51	208	53	91	21	36	9
Weakly support	131	29	125	32	95	6	10	5
Neutral	31	7	27	7	87	4	7	13
Weakly against	28	6	18	5	64	10	17	36
Against	29	6	12	3	41	17	29	59

[Correction added on 15 August 2020, after first online publication: In Table 2, data under the heading 'Outcome in statement' have been corrected. The affected columns are Total (n, Col%) and Harm Statements (n).]

instance, a statement 'With small doses all primary drugs are well tolerated' was interpreted to support the intervention despite considering adverse effects or harms, whereas the statement 'Acupuncture does not improve motor recovery among stroke patients' was interpreted to be against the intervention despite that it considered intervention benefits (see other examples in Table 1).

We also screened and assessed figures, tables, and other illustrations concerning intervention effects in the selected CGPs. As only 16 tables or figures of these kinds of illustrations were detected, we decided to describe them verbally and categorize them based on whether they considered benefits or harms or both or non-numeric, numeric or symbolic intervention effect measures.

Basic tabulations of the assessments of statements are presented by benefits and harm-statements and by the magnitude of effect measures.

3 | RESULTS

In total, 448 assessments of statements were made. Of those, 58 (13%) considered harms and 390 (87%) benefits of an intervention. Almost half (47%) of the statements were graded as the highest A-level of evidence and 30% as the B-level of evidence. The D-level of evidence occurred only in 12 cases (3%). Primary outcomes were less common (18%) than were secondary (42%) or unspecified (40%) ones (Tables 1 and 2).

A vast majority (92%) of statements did not mention the magnitude of effect of an intervention in numerical terms. When numbers were represented, they were more frequently relative rather than absolute measures. Only three cases reported the magnitude of effect in both absolute and relative numbers (Table 2).

Most of the statements represented an intervention in a supportive (53%) or weakly supportive (32%) frame. One hundred and eighteen statements represented an intervention in a neutral way (7%) or were weakly against (5%) or more clearly against (3%) it (Table 2).

Most statements that considered intervention benefits reported favourable findings of interventions in a supportive or weakly supportive way. Only 12 statements considered intervention as not being effective or recommended. If a statement represented the magnitude of the effect of an intervention in numerical format, relative measures were more frequently represented than absolute ones (Table 2).

Almost half of the statements that considered intervention adverse effects or harms were represented in a way that supported

TABLE 3 Number of statements considering intervention benefits in clinical practice guidelines stratified by representation of magnitude of effect of an intervention

	Non-numeric	Relative	Absolute	Both
Total	361	20	6	3
Clinical practice guideline				
Osteoarthritis of knee and hip	62	0	0	0
Stroke and transient ischemic attack	73	0	0	2
Stable coronary artery disease	10	0	0	0
Chronic pulmonary obstructive disease	31	1	0	0
Depression	89	1	0	0
Type 2 diabetes mellitus	14	0	1	0
Dyslipidaemia	11	9	0	0
Hypertension	25	3	4	0
Osteoporosis	19	3	0	0
Benign prostatic hyperplasia	27	3	1	1
Grade of evidence				
A (high)	160	18	2	2
B (moderate)	113	2	4	1
C (low)	77	0	0	0
D (very low)	11	0	0	0
Outcome in statement				
Primary	54	12	0	2
Secondary	143	7	6	1
Unspecific	164	1	0	0
Representation of intervention effects				
Support	183	19	3	3
Weakly support	121	1	3	0
Neutral	27	0	0	0
Weakly against	18	0	0	0
Against	12	0	0	0

use of the intervention (Table 2). No statement indicated adverse effects with relative or absolute risk increase measures and only five statements represented prevalence or incidence of adverse effects or harms numerically (not shown in the tables).

Thirteen percent of the statements which supported an intervention or which were graded as the A-level of evidence were represented with numbers. None of the statements against the intervention or those graded at the C- or D-level of evidence were represented with numbers. A clearly higher proportion of the statements with the primary outcomes (17%) were represented with numbers than those with secondary outcomes (5%, Table 3). The most common relative numeric measure of magnitude of effect was the relative risk reduction.

In total, 16 tables or figures considered the benefits or harms of interventions. Of those, four considered harms, eight benefits, and four both benefits and harms. The most common purpose of these tables or figures was to guide drug therapies based on drug properties, dosage, benefits, or harms (n = 6). Only three represented intervention effects numerically and six used symbols: traffic lights (four), arrows (one), or plus or minus signs (one). All three numeric representations were from the dyslipidaemia CGP, the first reported prevalences of three adverse effects; the second illustrated the correlation between relative risk reduction and baseline risk of a cardiovascular event; the third illustrated relative reduction in low-density lipoprotein with different statins and dosages.

4 | DISCUSSION

In the 10 selected CGPs published by the Finnish Medical Society Duodecim, most of the statements of effects considered intervention benefit rather than harms or adverse effects. Furthermore, approximately half of the statements considering adverse effects or harms were represented in a way that supported the intervention. Most statements did not include numeric estimates of magnitude of effect. When numeric estimates of magnitude of effect were present, they were most frequently relative measures and were typically placed in a statement (a) considering intervention benefits with a primary outcome, (b) graded as an A-level of evidence, and (c) that supported intervention usage. Tables and figures were rarely used to communicate the effects of interventions.

Numeric representations of the effects of the interventions were rare, while verbal statements about the effects were ambiguous and potentially misleading. For example, a verbal statement in the stable coronary artery disease CPG states that 'statin therapy improves prognosis among patients with coronary heart disease [A]' without providing any statement on the magnitude. When important adverse outcomes are associated with the intervention, judgements in favour or against the use of the intervention are left solely to the authors of the CPG, as the associated benefits and harms are derivable only from the original research articles. The kind of statements mentioned above are of minimal utility for informed decision making in a clinical practice where doctors have little or no time to read actual research articles.

On the level of the outcomes, the statements were weighted on the unspecified and secondary outcomes instead of the primary outcomes. This classification was done retrospectively after reading all the statements and the examples of the outcomes in each category are provided in Table 1. The primary outcomes were more often framed with the numerical measure of the beneficial effects of the intervention than the secondary and unspecified outcomes. This finding may reflect the selection of the 10 CPGs that focused on common diseases and risk factors (ie, osteoporosis, dyslipidaemia, hypertension) and the finding could have been different if other topics would have been chosen. However, we argue that busy clinicians could benefit more from the CPGs if the ratio of the outcomes would be balanced between primary and secondary outcomes in these 10 selected CPGs. At least in our view, ambiguous expressions like 'is effective' and 'is beneficial' should be avoided whenever possible.

A substantial majority (92%) of the statements omitted the magnitude of the intervention's effect in numerical terms. Presented numbers of the intervention effects were frequently presented in relative measures and rarely in absolute measures. Three cases (1%) reported the magnitude of effect in both absolute and relative measures. Recently, after receiving feedback on the guidelines of interpretations of the evidence, the GRADE working group uniformly proposed incorporating the size and certainty of the intervention effect in the statements of evidence.[21] This guideline developed by the GRADE working group shows that this information is able to be integrated into CPGs without substantially increasing the burden of the readers or guideline developers, especially if the level of evidence is classified by the GRADE categories (from A to D). Studies have shown that even specialized clinicians have wide-ranging beliefs on the effects of the most common interventions in their area of specialization.[7,22-24] The preferences of the patients and the clinicians with regard to the benefits and harms of the common preventive interventions are shown to be extremely divergent between persons, raising the need for assessment of the numerical information associated with the effects of the interventions.[23-25]

The representation of the intervention has been found to be crucial and has shown to influence patients' and clinicians' informed decision making.[11,12,26,27] The studies also show that numerical information is often easier to understand from the tables compared to text.[13] Graphical representations may also be illustrative and effective tools to highlight important information.[13-15,28] For example, the Cochrane group has developed the Summary of Findings tables for this purpose and has mandated reporting of absolute and relative measures of the intervention effects.[6,29-31] Especially for people with lower levels of numeracy,[8,32,33] absolute measures of effects are more understandable if presented in natural frequencies, for example, '5 out of 1000 patients' and by keeping the denominator (1000 in this case) as a constant.[8,12,27,34] This way of representation of the intervention effects has shown to increase the understanding of the effects compared to the numerical representations alone.[12,17]

A small minority (13%) of the statements were considered as harms or adverse effects of the interventions. No statement reported

harms with relative or absolute risk increase measures and incidence or prevalence of adverse events were reported only in five statements and in one table. Half of the statements on harms or adverse effects were framed in non-numerical language in strong or weak support of the intervention. For instance, 'For symptomatic patients with over 50% stenosis of the carotid artery, the carotid endarterectomy surgery was safer than endovascular treatment and it is associated with fewer strokes and deaths in the 30-day follow-up'. The infrequency of statements of harms is probably partly explained by the deficiencies in reporting of the original publications of trials and studies on harms and adverse effects of the interventions.[35,36] Moreover, we focused on the statements that were based on the systematic literature review (the level of evidence statement accompanied the statement) and the authors of the selected CPGs may have prioritized the systematic reviews on the benefits of the interventions. Nevertheless, as it goes with the numerical information and size of the effect, informed decision making requires knowledge about the harms as well as the benefits of the intervention.[37]

This study has several limitations. We focused only on the statements that had a level of evidence mark associated with the statement, and therefore limited our assessment to the statements that are based on the systematic literature review by the authors of the respective CPG. The subjectivity of the assessment varied and was probably the greatest in the classification of the statement to strongly supportive and strongly against the intervention. The classification of the outcomes was somewhat arbitrary and subjective but the authors solved borderline cases by discussion. In quite many cases, the specification of the outcomes and the statements of the effects were obscure enough to hinder the classification. The 10 selected CPGs compose one tenth of all CPGs published by the Finnish Medical Society Duodecim and these may not be a representative sample of the CPGs. Furthermore, the statements of the effects of the interventions in these 10 CPGs are clearly not generalizable to other CPGs produced elsewhere on the same diseases and risk factors. Nevertheless, we believe the challenges of the statements and the recommendations of the interventions are universal and in need of much broader attention.

The CPGs currently used for clinical practice in Finland lack the numerical information on magnitude of the benefits and harms of the interventions' effects. The selected 10 CPGs on common diseases and risk factors were weighted towards secondary outcomes and ambiguous non-numeric frames of magnitude of intervention effects, hindering the prospects for informed decision making by busy clinicians and their patients. There is a room for improvement in representations of the magnitude and the uncertainty of the intervention effects in an understandable way where outlining absolute measures of the effects of the primary outcomes of interest could be seen as a starting point. Other representations such as natural frequencies and infographics of the intervention effects would decrease the burden of information gathering, especially among those with low numeracy skills.[10,14,15,28] Frankly speaking, there is an obvious and compelling need for more detailed information on risks and the effects of medical interventions, presented in a frame that is both sensible and understandable, to inform clinical decision making.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ORCID

Eero Raittio https://orcid.org/0000-0002-9258-9355

REFERENCES

1. Alonso-Coello P, Carrasco-Labra A, Brignardello-Petersen R, et al. Systematic reviews experience major limitations in reporting absolute effects. *J Clin Epidemiol*. 2016;72:16-26. https://doi.org/10.1016/j.jclinepi.2015.11.002.
2. Agarwal A, Johnston BC, Vernooij RWM, et al. Authors seldom report the most patient-important outcomes and absolute effect measures in systematic review abstracts. *J Clin Epidemiol*. 2017;81:3-12. https://doi.org/10.1016/j.jclinepi.2016.08.004.
3. Akl EA, Oxman AD, Herrin J, et al. Using alternative statistical formats for presenting risks and risk reductions. *Cochrane Database Syst Rev*. 2011;3. https://doi.org/10.1002/14651858.cd006776.
4. Chalmers I, Oxman AD, Austvoll-Dahlgren A, et al. Key Concepts for Informed Health Choices: a framework for helping people learn how to assess treatment claims and make informed choices. *BMJ Evid Based Med*. 2018;23(1):29-33. https://doi.org/10.1136/ebmed-2017-110829.
5. Higgins JPT, Green S. *Cochrane Handbook for Systematic Reviews of Interventions*. Vol 4. John Wiley & Sons: Chichester (UK); 2011.
6. Yepes-Nuñez JJ, Li SA, Guyatt G, et al. Development of the summary of findings table for network meta-analysis. *J Clin Epidemiol*. 2019;115:1-13. https://doi.org/10.1016/j.jclinepi.2019.04.018.
7. Wegwarth O, Gigerenzer G. The barrier to informed choice in cancer screening: statistical illiteracy in physicians and patients. *Recent Results Cancer Res*. 2018;210:207-221. https://doi.org/10.1007/978-3-319-64310-6_13.
8. Nelson W, Reyna VF, Fagerlin A, Lipkus I, Peters E. Clinical implications of numeracy: theory and practice. *Ann Behav Med*. 2008;35(3):261-274. https://doi.org/10.1007/s12160-008-9037-8.
9. Gigerenzer G, Gray JA. *Better Doctors, Better Patients, Better Decisions: Envisioning Health Care 2020*. The MIT Press: Cambridge, MA; 2011.
10. Gigerenzer G, Gaissmaier W, Kurz-Milcke E, Schwartz LM, Woloshin S. Helping doctors and patients make sense of health statistics. *Psychol Sci Public Interest*. 2007;8(2):53-96. https://doi.org/10.1111/j.1539-6053.2008.00033.x.
11. Stacey D, Légaré F, Lewis K, et al. Decision aids for people facing health treatment or screening decisions. *Cochrane Database Syst Rev*. 2017;2017(4). https://doi.org/10.1002/14651858.CD001431.pub5.
12. Garcia-Retamero R, Galesic M, Gigerenzer G. Do icon arrays help reduce denominator neglect? *Med Decis Making*. 2010;30(6):672-684. https://doi.org/10.1177/0272989X10369000.
13. Brick C, McDowell M, Freeman ALJ. Risk communication in tables versus text: a registered report randomized trial on 'fact boxes'. *R Soc Open Sci*. 2020;7(3):190876. https://doi.org/10.1098/rsos.190876.
14. McDowell M, Gigerenzer G, Wegwarth O, Rebitschek FG. Effect of tabular and icon fact box formats on comprehension of benefits and harms of prostate cancer screening: a randomized trial. *Med Decis Making*. 2019;39(1):41-56. https://doi.org/10.1177/0272989X18818166.
15. McDowell M, Rebitschek FG, Gigerenzer G, Wegwarth O. A simple tool for communicating the benefits and harms of health interventions. *MDM Policy Pract*. 2016;1(1):238146831666536. https://doi.org/10.1177/2381468316665365.
16. Petrova D, Garcia-Retamero R, Cokely ET. Understanding the harms and benefits of cancer screening. *Med Decis Making*. 2015;35(7):847-858. https://doi.org/10.1177/0272989X15587676.
17. Zipkin DA, Umscheid CA, Keating NL, et al. Evidence-based risk communication. *Ann Intern Med*. 2014;161(4):270-280. https://doi.org/10.7326/M14-0295.

18. Morgott M, Heinmüller S, Hueber S, Schedlbauer A, Kühlein T. Do guidelines help us to deviate from their recommendations when appropriate for the individual patient? A systematic survey of clinical practice guidelines. *J Eval Clin Pract*. 2019;26(3):709-717. https://doi.org/10.1111/jep.13187.
19. Guyatt G, Oxman AD, Akl EA, et al. GRADE guidelines: 1. Introduction - GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol*. 2011;64(4):383-394. https://doi.org/10.1016/j.jclinepi.2010.04.026.
20. Guyatt G, Oxman AD, Sultan S, et al. GRADE guidelines: 11. Making an overall rating of confidence in effect estimates for a single outcome and for all outcomes. *J Clin Epidemiol*. 2013;66(2):151-157. https://doi.org/10.1016/j.jclinepi.2012.01.006.
21. Santesso N, Glenton C, Dahm P, et al. GRADE guidelines 26: informative statements to communicate the findings of systematic reviews of interventions. *J Clin Epidemiol*. 2020;119:126-135. https://doi.org/10.1016/j.jclinepi.2019.10.014.
22. Gigerenzer G, Mata J, Frank R. Public knowledge of benefits of breast and prostate cancer screening in Europe. *J Natl Cancer Inst*. 2009;101(17):1216-1220. https://doi.org/10.1093/jnci/djp237.
23. Hoffmann TC, Del Mar C. Patients' expectations of the benefits and harms of treatments, screening, and tests: a systematic review. *JAMA Intern Med*. 2015;175(2):274-286. https://doi.org/10.1001/jamainternmed.2014.6016.
24. Hoffmann TC, Del Mar C. Clinicians' expectations of the benefits and harms of treatments, screening, and tests. *JAMA Intern Med*. 2017;177(3):407-419. https://doi.org/10.1001/jamainternmed.2016.8254.
25. Douglas F, Petrie KJ, Cundy T, Horne A, Gamble G, Grey A. Differing perceptions of intervention thresholds for fracture risk: a survey of patients and doctors. *Osteoporos Int*. 2012;23(8):2135-2140. https://doi.org/10.1007/s00198-011-1823-7.
26. Fagerlin A, Zikmund-Fisher BJ, Ubel PA. Helping patients decide: ten steps to better risk communication. *J Natl Cancer Inst*. 2011;103(19):1436-1443. https://doi.org/10.1093/jnci/djr318.
27. Wegwarth O, Gaissmaier W, Gigerenzer G. Deceiving numbers: survival rates and their impact on doctors' risk communication. *Med Decis Making*. 2011;31(3):386-394. https://doi.org/10.1177/0272989X10391469.
28. Way D, Blazsin H, Löfstedt R, Bouder F. Pharmaceutical benefit–risk communication tools: a review of the literature. *Drug Saf*. 2017;40(1):15-36. https://doi.org/10.1007/s40264-016-0466-1.
29. Yordanov Y, Dechartres A, Ravaud P. Patient-important outcomes in systematic reviews: poor quality of evidence. *PLoS One*. 2018;13(4):e0195460. https://doi.org/10.1371/journal.pone.0195460.
30. Schünemann HJ, Mustafa R, Brozek J, et al. GRADE guidelines: 16. GRADE evidence to decision frameworks for tests in clinical practice and public health. *J Clin Epidemiol*. 2016;76:89-98. https://doi.org/10.1016/j.jclinepi.2016.01.032.
31. 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(2):173-183. https://doi.org/10.1016/j.jclinepi.2012.08.001.
32. Reyna VF, Nelson WL, Han PK, Dieckmann NF. How numeracy influences risk comprehension and medical decision making. *Psychol Bull*. 2009;135(6):943-973. https://doi.org/10.1037/a0017327.
33. Gaissmaier W, Gigerenzer G. Statistical illiteracy undermines informed shared decision making. *Z Evid Fortbild Qual Gesundhwes*. 2008;102(7):411-413. https://doi.org/10.1016/j.zefq.2008.08.013.
34. Yamagishi K. When a 12.86% mortality is more dangerous than 24.14%: implications for risk communication. *Appl Cogn Psychol*. 1997;11(6):495-506. https://doi.org/10.1002/(SICI)1099-0720(199712)11:6<495::AID-ACP481>3.0.CO;2-J.
35. Glasziou P, Altman DG, Bossuyt P, et al. Reducing waste from incomplete or unusable reports of biomedical research. *Lancet*. 2014;383(9913):267-276. https://doi.org/10.1016/S0140-6736(13)62228-X.
36. Gyawali B, Shimokata T, Honda K, Ando Y. Reporting harms more transparently in trials of cancer drugs. *BMJ*. 2018;363:k4383. https://doi.org/10.1136/bmj.k4383.
37. Blalock SJ, Sage A, Bitonti M, Patel P, Dickinson R, Knapp P. Communicating information concerning potential medication harms and benefits: what gist do numbers convey? *Patient Educ Couns*. 2016;99(12):1964-1970. https://doi.org/10.1016/j.pec.2016.07.022.

How to cite this article: Raittio E, Raittio L. Statements considering intervention effects in Finnish clinical practice guidelines: Recommending interventions with non-numeric effect-sizes or unspecified outcomes. *J Eval Clin Pract*. 2020;1–8. https://doi.org/10.1111/jep.13455