Journal of Clinical Epidemiology 72 (2016) 16–26

Journal of Clinical Epidemiology

REVIEW ARTICLES

Systematic reviews experience major limitations in reporting absolute effects

Pablo Alonso-Coello[a,b,c,*], Alonso Carrasco-Labra[c,d], Romina Brignardello-Petersen[d,e], Ignacio Neumann[f], Elie A. Akl[c,g], Robin W.M. Vernooij[a], Brad C. Johnston[c,e,h], Xin Sun[i], Matthias Briel[c,j], Jason W. Busse[c,k,l], Shanil Ebrahim[c,m], Carlos E. Granados[n], Alfonso Iorio[o], Affan Irfan[p], Laura Martínez García[a], Reem A. Mustafa[c,q], Anggie Ramírez-Morera[r], Anna Selva[a], Ivan Solà[a], Andrea Juliana Sanabria[a], Kari A.O. Tikkinen[c,s], Per Olav Vandvik[t,u], Oscar E. Zazueta[a], Yuqing Zhang[c,u], Qi Zhou[c], Holger Schünemann[c,o], Gordon H. Guyatt[c,o]

[a]*Clinical Epidemiology and Public Health Department, Iberoamerican Cochrane Centre, Biomedical Research Institute Sant Pau (IIB Sant Pau), C/ Sant Antoni Maria Claret, 167, Pavelló 18, planta 0, 08025 Barcelona, Spain*

[b]*Centro de Investigación Biomédica en Red de Epidemiología y Salud Pública (CIBERESP)*

[c]*Department of Clinical Epidemiology & Biostatistics, McMaster University, 1280 Main Street West, Hamilton, Ontario L8S 4K1, Canada*

[d]*Evidence-Based Dentistry Unit, Faculty of Dentistry, Universidad de Chile, Sergio Livingstone 943, Santiago, Chile*

[e]*Institute of Health Policy, Management & Evaluation, University of Toronto, 4th Floor, 155 College St, Toronto, Ontario M5T 3M6, Canada*

[f]*Department of Medicine, Pontificia Universidad Católica de Chile, Avenida Libertador Bernardo O Higgins 340, Santiago, Región Metropolitana, Chile*

[g]*Department of Internal Medicine, American University of Beirut Medical Center, P.O. Box: 11-0236, Riad-El-Solh, Beirut 1107 2020*

[h]*Department of Anesthesia & Pain Medicine, The Hospital for Sick Children, 555 University Ave., Toronto, Ontario M5G 1X8, Canada*

[i]*Chinese Evidence-based Medicine Center, West China Hospital, Sichuan University, 37 Guo Xue Xiang, Chengdu 610041, China*

[j]*Basel Institute for Clinical Epidemiology and Biostatistics, Department of Clinical Research, University Hospital Basel, Spitalstrasse 12, 4031 Basel, Switzerland*

[k]*Department of Anesthesia, McMaster University, 1280 Main Street West, Hamilton, Ontario L8S 4K1, Canada*

[l]*Michael G. DeGroote Institute for Pain Research and Care, McMaster University, 1280 Main Street West, Hamilton, Ontario L8S 4K1, Canada*

[m]*Meta-Research Innovation Center at Stanford (METRICS), Stanford University, 1265 Welch Road, Stanford, CA 94305, USA*

[n]*Área de investigaciones, Facultad de Medicina, Universidad de La Sabana, Campus del Puente del Común Km, 7 Autopista Norte, Chía, Colombia*

[o]*Department of Medicine, McMaster University, 1280 Main St W, Hamilton, Ontario L8S 4L8, Canada*

[p]*Department of Cardiology, University of Louisville, Rudd Heart & Lung Building, 201 Abraham Flexner Way, Suite 600, Louisville, KY 40292, USA*

[q]*Departments of Medicine and Biomedical & Health Informatics, University of Missouri-Kansas City, School of Medicine, M4-303, 2411 Holmes St, Kansas City, MO, USA*

[r]*CCSS Permanent Medical Advisor, Health Care Development Division, IHCAI Foundation & Central America Cochrane, 1st Ave., 35th and 37th St, Number 3530, Barrio Escalante, San José, Costa Rica*

[s]*Departments of Urology and Public Health, Helsinki University Hospital and University of Helsinki, Haartmaninkatu 4, 00029 Helsinki, Finland*

[t]*Institute of Health and Society, Faculty of Medicine, University of Oslo, P.O. Box 1130 Blindern, 0318 Oslo, Norway*

[u]*Guang'anmen Hospital China Academy of Chinese Medical Science, No. 5 Bei Xian Ge St, Xuan Wu District, Beijing 100053, China*

Accepted 3 November 2015; Published online 10 November 2015

Funding: P.A-.C. is funded by a Miguel Servet research contract from the Instituto de Salud Carlos III (CP09/00137); L.M.G. and A.J.S. are funded by a Río Hortega research contract from the Instituto de Salud Carlos III (CM10/00014 and CM12/00168); M.B. is supported by santésuisse and the Gottfried and Julia Bangerter-Rhyner Foundation; J.W.B. is supported by a New Investigator Award from the Canadian Institutes of Health Research and Canadian Chiropractic Research Foundation; and the work of K.A.O.T. was supported by the Academy of Finland (#276046), Competitive Research Funding of the Helsinki and Uusimaa Hospital District, Finnish Cultural Foundation, Finnish Medical Foundation, Jane and Aatos Erkko Foundation, and Sigrid Juselius Foundation.

Conflict of interest: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: no support from any organization for the submitted; no financial relationships with any organizations that might have an interest in the submitted work in the previous 3 years; and no other relationships or activities that could appear to have influenced the submitted work.

* Corresponding author. Iberoamerican Cochrane Centre, Institute of Biomedical Research (IIB Sant Pau), Barcelona, Hospital de la Santa Creu i Sant Pau, Pavelló 18, C/ Sant Antoni M[a] Claret 167, 08025 Barcelona, Spain. Tel.: +34 935537814; fax: +34 935537809.

E-mail address: PAlonso@santpau.cat (P. Alonso-Coello).

http://dx.doi.org/10.1016/j.jclinepi.2015.11.002

0895-4356/© 2016 Elsevier Inc. All rights reserved.

Abstract

Objectives: Expressing treatment effects in relative terms yields larger numbers than expressions in absolute terms, affecting the judgment of the clinicians and patients regarding the treatment options. It is uncertain how authors of systematic reviews (SRs) absolute effect estimates are reported in. We therefore undertook a systematic survey to identify and describe the reporting and methods for calculating absolute effect estimates in SRs.

Study Design and Setting: Two reviewers independently screened title, abstract, and full text and extracted data from a sample of Cochrane and non-Cochrane SRs. We used regression analyses to examine the association between study characteristics and the reporting of absolute estimates for the most patient-important outcome.

Results: We included 202 SRs (98 Cochrane and 104 non-Cochrane), most of which (92.1%) included standard meta-analyses including relative estimates of effect. Of the 202 SRs, 73 (36.1%) reported absolute effect estimates for the most patient-important outcome. SRs with statistically significant effects were more likely to report absolute estimates (odds ratio, 2.26; 95% confidence interval: 1.08, 4.74). The most commonly reported absolute estimates were: for each intervention, risk of adverse outcomes expressed as a percentage (41.1%); number needed to treat (26.0%); and risk for each intervention expressed as natural units or natural frequencies (24.7%). In 12.3% of the SRs that reported absolute effect estimates for both benefit and harm outcomes, harm outcomes were reported exclusively as absolute estimates. Exclusively reporting of beneficial outcomes as absolute estimates occurred in 6.8% of the SRs.

Conclusions: Most SRs do not report absolute effects. Those that do often report them inadequately, thus requiring users of SRs to generate their own estimates of absolute effects. For any apparently effective or harmful intervention, SR authors should report both absolute and relative estimates to optimize the interpretation of their findings. © 2016 Elsevier Inc. All rights reserved.

Keywords: Absolute measures; Absolute effect estimates; Systematic review; Reporting; Framing; Risk difference; Decision making

1. Introduction

Informed clinical decision making requires knowledge of the magnitude of the desirable and undesirable effects of treatment alternatives [1,2]. Investigators may express the impact of an intervention for dichotomous outcomes in either relative terms (i.e., risk ratio, odds ratio, or hazard ratio) or in absolute terms [absolute risk reduction (ARR), also known as risk difference (RD), or as the number needed to treat (NNT)] [3].

There are benefits, as well as downsides, to presentation of treatment effects using either relative or absolute measures. For example, exclusive use of relative effects is likely to be misleading. Expressing treatment effects in relative terms yields apparently larger treatment effects than if absolute terms are used (e.g., a 50% relative risk reduction can mean an ARR of 1%—i.e., 2% to 1%), and this difference influences the judgment of the clinicians and patients regarding the treatment options [4−8]. Relative effect estimates, however, are usually (though not always) similar across populations and subgroups [9], whereas absolute effect estimates typically vary with the baseline risk [8,10−12]. Therefore, expressing a treatment effect estimate as a single ARR is misleading because it will underestimate the effect for patients at high baseline risk and overestimate the effect for those at low baseline risk. As a result, in the context of meta-analysis, pooling absolute risks often results in large variability in effect across studies [8,10−12].

Despite these limitations, applying results from systematic reviews (SRs) in clinical decision making requires an understanding of absolute effects. To obtain this information, one needs to apply the relative effect estimate to a range of baseline risks typically seen in the population of interest. This may require ascertaining clinically identifiable risk groups and clarifying the period over which the associated baseline risk applies [8,11,13].

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement does not address the reporting of absolute estimates [14]. The Grading of Recommendations Assessment, Development and Evaluation (GRADE) working group, which represents an emerging consensus for rating confidence in estimates of intervention effects [15,16], suggests that SR authors present the estimates of absolute risks in intervention and control groups. Presentation should include the difference in the two risks and confidence intervals (CIs) that convey the precision of effect estimates for all important benefits and harms. Consistent with this approach, the Cochrane Handbook provides guidance on how to calculate and report absolute estimates in a prespecified table (Summary of Findings table) [17]. Not all Cochrane SRs, however, include this information.

Individual studies typically report absolute effect estimates poorly [18−20]. Current evidence suggests that approximately 50% of SRs include absolute effects and that one-third do not report benefits and harms using the same metric (mismatched framing) [21]. Studies reporting these findings, however, have been carried out in a relatively limited sample of journals or included only information from abstracts [20,21].

Given the limited information and the potential implications for decision making in health care [22], we systematically evaluated the extent to which SRs report absolute effect estimates and methods used for calculation and reporting.

What is New?

Key finding

- Absolute effect estimates for the most patient-important outcomes are reported seldom and inadequately in systematic reviews.

What is already known on this subject?

- Expressing the same treatment effect in relative terms yields larger estimates than in absolute terms.
- Clinicians and patients are more inclined to use drugs when presented with relative than absolute measures of effect.
- Individual studies often fail to report absolute effects optimally.
- To trade of benefits and risks, clinicians need absolute effect estimates.

What this study adds?

- Systematic reviews seldom report absolute effect estimates for the most patient-important outcomes.
- Absolute effect estimates, when reported, are most frequently expressed as percentages, natural units, or natural frequencies, but are often not optimally presented.

2. Methods

We have previously published a detailed protocol for this study [23].

2.1. Search strategy

We conducted a systematic search of MEDLINE and the Cochrane database of SRs limited to reviews published in the year 2010. The search strategy is available as Supplementary Data (Additional File 1)/Appendix A at www.jclinepi.com. We included studies described as "systematic review" or "meta-analysis" that (1) were published in English; (2) reported a search strategy in at least one database; (3) included a comparison of at least two interventions in humans; (4) reported at least one dichotomous outcome; (5) included at least one randomized controlled trial (RCT); and (6) included only RCTs as primary studies.

2.2. Study selection and data extraction

We stratified SRs as Cochrane and non-Cochrane SRs and repeatedly randomly sampled within each stratum in a 1:1 ratio (sampled 2,328) until we achieved our target sample size of approximately 200 SRs. Reviewers, in pairs, independently screened titles and abstracts for eligibility and, if potentially eligible, reviewed the full text and, for eligible articles, extracted data. Discrepant judgments were resolved by consensus or, if necessary, by a third reviewer.

For all included SRs, we extracted information regarding the population, the intervention and control of interest, the credibility of the SRs using the A Measurement Tool to Assess Systematic Reviews (AMSTAR) instrument [24], and whether the SRs included an absolute effect estimate comparison of intervention and control of primary interest. We identified the most patient-important outcome (which we will call the "outcome of interest") using a hierarchical approach (Additional File 2/Appendix B at www.jclinepi.com).

For SRs that reported an absolute effect estimate for the outcome of interest, we collected information regarding the type of effect estimate. We explored how authors calculated these absolute effect estimates, whether they made estimates for more than one estimate of baseline risk, and whether they specified the source of the baseline risk estimate(s). Finally, we recorded whether authors included a discussion of the likelihood that baseline risk, and therefore RDs, vary across subpopulations. We defined the NNT or number needed to harm as the number of patients who must receive an intervention required to result in one additional beneficial or harmful outcome over a predefined period of time [3,25,26]. ARR or RD is defined as the difference between two event frequencies [3].

2.3. Analysis

We calculated frequencies and proportions, including the measures of statistical dispersion, for all items, stratified by Cochrane and non-Cochrane SRs. We calculated chance-corrected agreement between reviewers' judgments of whether the investigators reported an absolute effect estimate for the outcome of interest and interpreted the results according to Landis and Koch guidelines (kappa values of 0–0.20 represent slight agreement, 0.21–0.40 fair agreement, 0.41–0.60 moderate agreement, 0.61–0.80 substantial agreement, and greater than 0.80 almost perfect agreement) [27].

We conducted multivariable logistic regression analyses examining the association between study characteristics and whether authors reported an absolute estimate of effect for the outcome of interest. The independent variables in our regression analyses were as follows: Cochrane SRs vs. non-Cochrane SRs, AMSTAR score, significant effect vs. nonsignificant effect (threshold $P \leq 0.05$), pharmacologic intervention vs. others, and whether authors reported the source of funding. Data analysis was performed using SPSS statistical software, version 18.0 (SPSS Inc., Chicago, IL, USA).

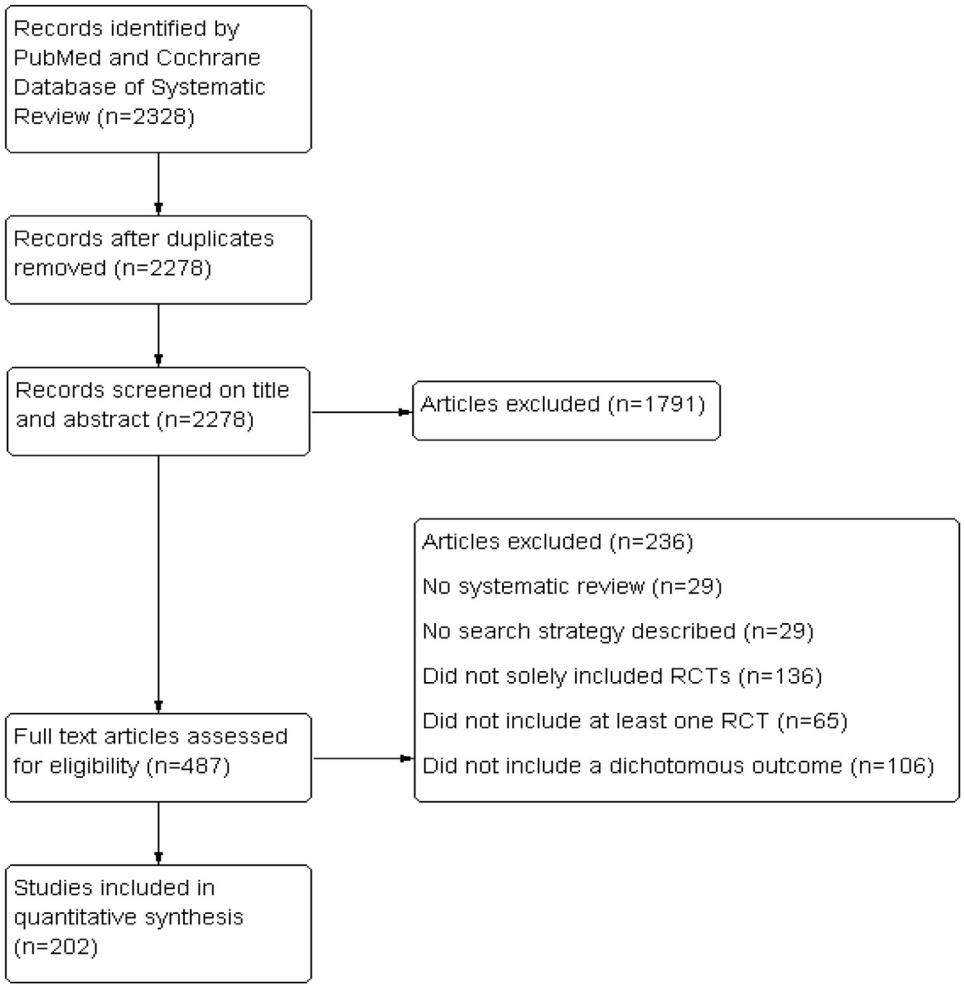

Fig. 1. Flow chart of the screening literature process. RCT, randomized controlled trial.

3. Results

We screened the titles and abstract of 2,278 articles, examined 438 full-text articles, and included 202 SRs (Fig. 1). Reasons for exclusion on full-text screening were as follows: not an SR (n = 29), did not include at least one electronic database in the search strategy (n = 29), did not report a dichotomous outcome (n = 106), did not include at least one RCT (n = 65), included non-RCTs (n = 136), or a combination of several of these reasons.

The 202 eligible SRs (98 Cochrane and 104 non-Cochrane) included a median of five RCTs [interquartile range (IQR), 2–9], a median number of 851 patients (IQR, 258–2108), and a median of 109 events for the outcome of interest (IQR, 30–388; Table 1). Of the 202 SRs, 164 (81.2%) focused on a medical area and 57 (28.2%) on a surgical area. The primary outcome represented a benefit in 87.1% of the SRs and harm in the remainder; in 68.3% of the SRs, investigators chose a patient-important, primary outcome.

Table 2 presents the absolute effect estimates for the dichotomous outcome of greatest importance to patients reported among the included SRs. Agreement regarding whether the investigators reported an absolute estimate for the most patient-important outcome was excellent (K = 0.72; 95% CI: 0.59, 0.86).

Of the 202 SRs and for the outcome of interest, 36.1% provided an absolute effect estimate, none of which were directly calculated. In the subgroup of reviews with a statistically significant result for this type of outcome, 25/54 (46.3%) provided an absolute estimate [24/86 (27.9%) in the nonsignificant SRs subgroup]. In the studies that presented an absolute risk effect estimate, the most common presentations were the risk in intervention and control groups expressed as a percentage (41.1%), the NNT (26.0%), and the risk in intervention and control groups expressed as natural units or natural frequencies (24.7%; Table 2).

Of studies reporting absolute risks, the majority did not report the source of the baseline risk estimate for the absolute difference (56.2%). Of those that did report the source

Table 1. Characteristics of the included systematic reviews

Characteristics of reviews	Cochrane (n = 98) (%)	Non-Cochrane (n = 104) (%)	Overall (n = 202) (%)	P-value
Type of analysis developed by authors[a]				
Standard meta-analysis	88 (89.8)	98 (94.2)	186 (92.1)	0.3
Meta-regression	3 (3.1)	15 (14.4)	18 (8.9)	**0.006**
Individual patient data meta-analysis	2 (2.0)	5 (4.8)	7 (3.5)	0.45
Network meta-analysis or multiple treatment comparison	1 (1.0)	0 (0.0)	1 (0.5)	0.49
Not applicable	6 (6.1)	3 (2.9)	9 (4.5)	0.32
Characteristics of meta-analysis				
Median number of trials included (IQR)	3 (2–8)	6 (4–9)	5 (2–9)	<**0.0001**
Median number of patients included (IQR)	59 (17–185)	299 (62–811)	109 (30–388)	<**0.0001**
Median number of events included (IQR)	567 (189–1,391)	1,306 (449–3,769)	851 (258–2,108)	**0.003**
Characteristics of meta-analysis[a]				
Medical area	71 (72.4)	93 (89.4)	164 (81.2)	**0.002**
Surgical area	36 (36.7)	21 (20.2)	57 (28.2)	**0.01**
Intervention for comparison of interest				
Pharmacologic	70 (71.4)	60 (57.7)	130 (64.4)	**0.04**
Nonpharmacologic	28 (29.6)	44 (42.3)	72 (35.6)	**0.04**
Control for comparison of interest				
Pharmacologic	36 (36.7)	33 (31.7)	69 (34.2)	0.45
Nonpharmacologic	33 (33.7)	38 (36.5)	71 (35.1)	0.67
Placebo	29 (29.6)	33 (31.7)	62 (30.7)	0.74
Is the most patient-important outcome the primary outcome for the authors?				
Yes	65 (66.3)	49 (47.1)	114 (56.4)	**0.01**
No, but some other outcome specified as primary	23 (23.5)	14 (13.5)	37 (18.3)	0.07
Authors did not specify a primary outcome	10 (10.2)	41 (39.4)	51 (25.3)	<**0.0001**
Is the primary outcome of the authors patient-important?				
Yes	81 (82.7)	57 (54.8)	138 (68.3)	<**0.0001**
No	7 (7.1)	6 (5.8)	13 (6.4)	0.69
Authors did not specify a primary outcome	10 (10.2)	41 (39.4)	51 (25.3)	<**0.0001**
Most patient-important outcome one of benefit or harm?				
Beneficial outcome	85 (86.7)	91 (87.5)	176 (87.1)	0.87
Harm outcome	13 (13.3)	13 (12.5)	26 (12.9)	0.87
Relative effect estimates for the most patient-important outcome				
Relative risk	59 (60.2)	59 (56.7)	118 (58.4)	0.62
Odds ratio	24 (24.5)	30 (28.9)	54 (26.7)	0.48
Hazard ratio	9 (9.2)	4 (3.9)	13 (6.4)	0.16
Rate ratio	1 (1.1)	1 (1.0)	2 (1.0)	1.00
Other	5 (5.1)	11 (9.6)	16 (7.9)	0.15
Use of GRADE				
Yes	27 (27.6)	5 (4.8)	32 (15.8)	<**0.0001**
No	70 (71.4)	98 (94.2)	168 (83.2)	<**0.0001**
Evaluation of risk of bias				
Risk of bias tool	84 (85.7)	10 (9.6)	94 (46.5)	<**0.0001**
By dimensions	8 (8.2)	31 (29.8)	39 (19.3)	<**0.0001**
Jadad or other scales	4 (4.1)	33 (31.7)	37 (18.3)	<**0.0001**
Not evaluated	0 (0.0)	18 (17.7)	18 (8.9)	<**0.0001**
Other	2 (2.0)	9 (8.7)	11 (5.5)	0.06
Funding				
For profit	2 (2.0)	5 (4.8)	7 (3.5)	0.45
Not for profit	65 (66.3)	37 (35.6)	102 (50.5)	<**0.0001**
Not funded	9 (9.2)	14 (13.5)	23 (11.4)	0.34
Not reported	23 (23.5)	49 (47.1)	72 (35.6)	**0.0005**
Did authors report ties to industry?				
Yes	19 (19.4)	18 (17.3)	37 (18.3)	**0.70**
No	60 (61.2)	35 (33.7)	95 (47.0)	<**0.0001**
Not reported	19 (19.4)	49 (47.1)	68 (33.7)	<**0.0001**
Unclear	0 (0.0)	2 (2.0)	2 (1.0)	0.50
Mean quality score on the AMSTAR instrument (SD)	9.7 (0.9)	6.8 (2.2)	8.2 (2.3)	<**0.0001**

The values that are bold are all P-values that are statistically significant (<0.05).

Abbreviations: IQR, interquartile range; GRADE, Grading of Recommendations Assessment, Development and Evaluation; SD, standard deviation.

[a] The questions allow multiple selection; therefore, sum of percentages might exceed 100%.

Table 2. Absolute effect estimates frequency and, when reported, characteristics of the report

Variable of interest[a]	Cochrane (*n* = 98) (%)	Non-Cochrane (*n* = 104) (%)	Overall (*n* = 202) (%)	*P*-value
Absolute effect estimates reported	36 (36.8)	37 (35.6)	73 (36.1)	0.86
Expressions used for effect estimates[a]	(*n* = 36)	(*n* = 37)	(*n* = 73)	
Risk difference or absolute risk reduction	2 (5.6)	4 (10.8)	6 (8.2)	0.67
Number needed to treat (NNT) or number needed to treat to benefit	9 (25.0)	10 (27.0)	19 (26.0)	0.79
Number needed to harm (NNH) or number needed to treat to harm	1 (2.8)	0 (0)	1 (1.4)	0.49
Risk in one, risk in another expressed as a percentage	8 (22.2)	22 (59.5)	30 (41.1)	**0.002**
Risk in one, risk in another expressed as a proportion	3 (8.3)	1 (2.7)	4 (5.5)	0.61
Risk in one, risk in another expressed as natural units or natural frequency	15 (41.7)	3 (8.1)	18 (24.7)	**0.002**
Risk difference expressed as a percentage	1 (2.8)	3 (8.1)	4 (5.5)	0.61
Risk difference expressed as a proportion	0 (0.0)	1 (2.7)	1 (1.4)	1
Risk difference expressed as natural units or natural frequency	1 (2.8)	2 (5.4)	3 (4.1)	1
Other	1 (2.8)	3 (8.1)	4 (5.5)	0.61
Methods				
What is the source of the baseline risk estimate?	(*n* = 36)	(*n* = 37)	(*n* = 73)	
Using the median of the control group from the included studies in the meta-analysis	7 (19.4)	1 (2.7)	8 (11.0)	0.06
Using the mean of the control group from the included studies in the meta-analysis	1 (2.8)	2 (5.4)	3 (4.1)	1
Using one or more representative control groups (not the median nor the mean) from the included studies in the meta-analysis	3 (8.3)	0 (0.0)	3 (4.1)	0.24
Not applicable—the authors mention the term ''pooled risk difference'' and not the use of a determined baseline risk	5 (13.9)	7 (18.9)	12 (16.4)	0.53
Other (e.g., observational studies)	4 (11.1)	2 (5.4)	6 (8.2)	0.67
Not reported	16 (44.4)	25 (67.6)	41 (56.2)	**0.03**
Reporting				
Where are the absolute numbers reported?[a]	(*n* = 36)	(*n* = 37)	(*n* = 73)	
Abstract	10 (27.8)	14 (37.8)	24 (32.9)	0.32
Results	22 (61.1)	32 (86.5)	54 (74.0)	**0.01**
Discussion	10 (27.8)	13 (35.1)	23 (31.5)	0.45
Tables and figures	26 (72.2)	17 (45.9)	43 (58.9)	**0.03**
Different absolute values for patients with different baseline risks?	(*n* = 36)	(*n* = 37)	(*n* = 73)	
Yes	9 (25.0)	1 (2.7)	10 (13.7)	**0.01**
No	27 (75.0)	35 (94.6)	62 (84.9)	**0.01**
Reviews that included a Summary of Findings table	29 (29.6)	3 (2.9)	32 (15.9)	**<0.0001**
What absolute effect estimates are reported in Summary of Findings tables?[a]	(*n* = 29)	(*n* = 3)	(*n* = 32)	
Rate of events per group (with and without intervention)	7 (24.1)	1 (33.3)	8 (25.0)	**0.03**
Rate difference	1 (3.4)	0 (0.0)	1 (3.1)	0.49
Risk in one, risk in another expressed as a percentage	1 (3.4)	0 (0.0)	1 (3.1)	0.49
Risk in one, risk in another expressed as a proportion	3 (3.1)	0 (0.0)	3 (1.5)	0.11
Risk in one, risk in another expressed as natural units or natural frequency	18 (62.1)	1 (33.3)	19 (59.4)	**<0.0001**
Risk difference expressed as a percentage	1 (3.4)	2 (66.6)	3 (9.4)	**0.02**
Risk difference expressed as natural units or natural frequency	1 (3.4)	0 (0.0)	1 (3.1)	0.49
NNT/NNH	1 (3.4)	1 (33.3)	2 (6.3)	0.17
Other	1 (3.4)	1 (33.3)	2 (6.3)	0.17

The values that are bold are all *P*-values that are statistically significant (<0.05).

[a] The questions allow multiple selection; therefore, sum of percentages might exceed 100%.

of the baseline risk, the median of the control group from the included studies in the meta-analysis was more frequently used as a source (57.1%), rather than the mean of one or more representative control groups from other studies (21.4%; Table 2). Most SRs that included absolute effect estimates reported these in the Section 3 (74.0%). Few SRs (13.7%) presented different absolute values for patients with different baseline risks; all but one of these were Cochrane SRs. The 32 SRs that included Summary of Findings tables all reported absolute risks, the majority as risks in intervention and control groups expressed as natural units or natural frequencies (59.4%).

Of the SRs that included absolute effect estimates, 57.5% reported both benefits and harms for the comparison of interest (Table 3). Studies including absolute effects of both benefit and harm outcomes described reported harms

Table 3. Reporting of harm and benefit outcomes

Characteristics of reviews[a]	Cochrane (%)	Non-Cochrane (%)	Overall (%)	*P*-value
In reviews reporting benefit and harm patient-important dichotomous outcomes, any harm outcome reported only in absolute terms?	(*n* = 26)[a]	(*n* = 22)[a]	(*n* = 48)[a]	
Yes	3 (8.3)	6 (16.2)	9 (12.3)	0.28
No	23 (63.9)	16 (43.2)	39 (53.4)	0.08
In reviews reporting benefit and harm patient-important dichotomous outcomes reported, any benefit outcome reported only in absolute terms?	(*n* = 26)[a]	(*n* = 22)[a]	(*n* = 48)[a]	
Yes	2 (5.5)	3 (8.1)	5 (6.8)	0.33
No	24 (66.7)	19 (48.6)	43 (57.5)	0.12

Abbreviation: SR, systematic review.

[a] The number of SRs included in this table is the amount of SRs that included both a benefit and harm patient-important dichotomous outcome and absolute effect estimates.

only as absolute effects in 12.3% (i.e., relative effect estimates unreported). Beneficial outcomes were reported only as absolute effect estimates in 6.8% of SRs.

Our logistic regression showed one significant predictor of presenting absolute effect estimates for the most patient-important outcome (Table 4): Studies reporting a statistically significant result were more likely to report an absolute effect estimate (OR, 2.26; 95% CI: 1.08, 4.74; *P*-value: 0.03). Results suggested the possibility that studies of a lower quality according to the AMSTAR score were less likely to report absolute effect estimates (OR, 0.79; 95% CI: 0.62, 1.01, *P*-value: 0.06).

4. Discussion

Our methodological survey of 202 SRs found that the majority did not report absolute effect estimates for the outcome of interest. Most of those that did report absolute risks failed to report the source of the baseline risk estimate used to calculate the absolute effect and seldom provided different absolute values for patients with different baseline risks. Studies that reported absolute effect estimates most frequently presented risks in intervention and control groups expressed as percentages, as natural units, or natural frequencies (Table 2).

4.1. Interpretation of findings

Our results highlight the limitations in the reporting of results for SRs. Expressing results exclusively in relative terms, as most current SRs—both Cochrane and non-Cochrane—are doing, will result in health care professionals overestimating the magnitude of treatment effects [3–6,8,28,29]. Optimal medical decision making requires knowledge of the absolute effect of interventions on both benefits and harms in judging the trade-off between the two. SRs that fail to provide these estimates will be less useful to their target audiences [30].

None of the SRs included a pooled analysis for the reported absolute estimates—they rather elected to pool relative effects. This is consistent with the generally accepted guidance of not pooling RDs in meta-analyses because relative effects tend to be consistent across baseline risk, resulting in larger heterogeneity when pooling absolute effects, and difficulty interpreting results of such pooling [10–12].

Therefore, absolute estimates should, in most cases, be calculated by explicitly using baseline risks, either selected from a cohort study that enrolled a representative population-based sample (Box 1). Investigators can then apply pooled relative risk estimates for the meta-analysis to the baseline risks to calculate RDs. If, as is often the case, high-quality observational studies are lacking, simulation studies suggest an optimal approach is to use the median risk among the control groups of the included studies [31]. On the rare occasions when RDs, rather than relative risks, are similar across risk groups, authors can pool RDs directly.

If investigators find important variation among control group risks, they should consider presenting a range of RDs, including those for higher and lower risk patients. Finally, if investigators calculate odds ratios rather than relative risks, they need to use the appropriate conversions to

Table 4. Logistic regression of the association between characteristics of the SR and presentation of absolute effect estimates

Comparisons	Frequency	Frequency	OR	95% CI	*P*-value
Cochrane vs. non-Cochrane	73 (Cochrane)	64 (Non-Cochrane)	1.74	0.60–5.04	0.31
AMSTAR score, median (Q1–Q3)	9 (7–10)		0.79	0.62–1.01	0.06
Significant effect vs. nonsignificant effect	54 (Significant effect)	83 (Nonsignificant)	2.26	1.08–4.74	**0.03**
Pharmacologic vs. other	94 (Pharmacologic)	43 (Other)	1.19	0.53–2.71	0.67
Funding reported vs. not reported	89 (Funding reported)	48 (Funding not reported)	0.89	0.39–2.04	0.78

The values that are bold are all *P*-values that are statistically significant (<0.05).

Abbreviations: SR, systematic review; OR, odds ratio; CI, confidence interval.

Box 1 Choosing a baseline risk to calculate risk differences

To calculate risk differences, systematic review authors should ideally apply a relative estimate (typically a relative risk) to baseline risk from well-designed observational studies. If high-quality observational studies are not available, authors need to consider variation in the baseline risk among included studies in the systematic review:

- If little variation: investigators can use the median control group risk from the included studies.
- If large variation: investigators can consider using two or more representative baseline risks from the included studies.

Whenever an OR is available, to generate an estimate of risk difference involves converting baseline risk to odds, multiplying by the OR, and converting the resulting odds back to risks. Alternatively, one can use the following formula (where RC is the risk in the control group) (+):

$$\text{Risk difference per } 1,000 = 1,000 \times \text{RC} - \left(\frac{\text{OR} \times \text{RC}}{1 - \text{RC} + (\text{OR} \times \text{RC})}\right)$$

+ Adapted from GRADE JCE series (article 12).
Abbreviation: OR, odds ratio.

Box 2 Different risk across different patient groups (+)

Authors of a systematic review (*) addressing air travel and the risk of venous thromboembolism identified risk factors for asymptomatic DVT (previous episodes of DVT, coagulation disorders, severe obesity, limited mobility because of bone or joint problems, cancer, and large varicose veins) that, when considered together, more than tripled the risk of thrombosis. Applying the RR of 10% allowed calculation of expected event rates for the high- and low-risk populations using prophylactic stockings. In the low-risk population, applying the RR of 10% to the risk without the intervention of 5 per 10,000 generates a risk of 0.5 per 10,000 with the intervention. In the higher-risk population, the corresponding numbers are 18 and 1.8 per 10,000 (Table 5).

+ Adapted from GRADE JCE series (article 12).

(*) Philbrick JT, Shumate R, Siadaty MS, Becker DM. Air travel and venous thromboembolism: a systematic review. J Gen Intern Med. 2007; 22:107–114.

Abbreviations: RR, relative risk; DVT, deep vein thrombosis.

calculate RDs (Box 1). Box 2 and Table 5 present examples of observational studies identifying a gradient of identifiable risk based on the Framingham score for cardiovascular risk.

The underreporting of absolute effect estimates is not surprising: instruments, such as PRISMA, that provide guidance for reporting SRs, do not specify the need to report both absolute and relative effect estimates [14,32,33]. The handbook of the Cochrane Collaboration, however, includes guidance on how to estimate absolute effects and encourages authors to include a Summary of Findings table that always includes absolute effects(Table 5) [11,17]. It is likely that in the near future, most Cochrane SRs will include Summary of Findings tables [34], greatly ameliorating the problem we have identified in the prior literature.

Our results highlight the extent to which authors are not yet following this guidance and thus the need for making the use of Summary of Findings a required standard for Cochrane reviews. For illustrative purposes, we include an example of Summary of Findings tables, both conventional (Table 5) and interactive (http://isof.epistemonikos.org/#finding/54c23176f30d0c2002a2efe4).

There is an exception to our guidance regarding the importance of including RDs in SRs. If results, in relative terms, are nonsignificant, then the addition of absolute effects becomes uninformative and could be subject to misinterpretation. When CIs include no effect but the point estimate suggests an important benefit and the least favorable CI boundary is very near no effect, the possible usefulness of absolute effect estimates is much greater than if the CI includes large benefit and large harm. In the latter situation, not reporting absolute effects, but simply noting nonsignificant results, represent a reasonable approach.

4.2. Our study in relation to previous research

Our study highlights a problem that, despite the available guidance in the STrengthening the Reporting of OBservational studies in Epidemiology and Consolidated Standards of Reporting Trials checklists [32,33], is also present in individual randomized trials and observational studies. Surveys of primary studies in leading medical journals, both cohort studies and RCTs, have reported a low frequency of absolute effect estimates [18]. In the field of health equity, only 7% of all articles of any design reported both relative and absolute effect estimates in the full text [19].

Returning to SRs, our results are, overall, less optimistic than those of Sedrakyan and Shih [21] who reported use of absolute effects in 50% of the SRs. The difference is likely due to differences in the eligibility criteria between their survey and ours; Sedrakyan et al. restricted inclusion to SRs published in ‘‘top journals’’. Although we did not find a significant difference between top and nontop journals in

Table 5. Summary of findings table regarding the use of aspirin compared with no aspirin

Outcomes	No of participants (studies) follow-up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects*	
				Risk without aspirin	Risk difference with aspirin (75–100 mg) (95% CI)
Total mortality	100,076 (9 studies) 3.8–10 y	Moderate due to imprecision	RR 0.94 (0.88–1.00)	60-y-old man[a] 100 deaths per 1,000	6 fewer per 1,000 (from 12 fewer to 0 fewer)
Nonfatal myocardial infarction (MI)	95,000 (6 studies) 3.8–10 y	High	RR 0.77 (0.69–0.86)	Low cardiovascular–risk population[b] 27 MI per 1,000[c]	6 fewer per 1,000 (from 8 fewer to 4 fewer)
				Moderate cardiovascular–risk population[b] 83 MI per 1,000[c]	19 fewer per 1,000 (from 26 fewer to 12 fewer)
				High cardiovascular–risk population[b] 136 MI per 1,000[c]	31 fewer per 1,000 (from 42 fewer to 19 fewer)
Stroke includes nonfatal ischemic and hemorrhagic strokes	95,000 (6 studies) 3.8–10 y	Moderate due to imprecision	RR 0.95 (0.85–1.06)	Low cardiovascular–risk population[b] 23 strokes per 1,000[c]	1 fewer per 1,000 (from 3 fewer to 1 more)
				Moderate cardiovascular–risk population[b] 65 strokes per 1,000[c]	3 fewer per 1,000 (from 10 fewer to 4 more)
				High cardiovascular–risk population[b] 108 strokes per 1,000[c]	5 fewer per 1,000 (from 16 fewer to 8 more)
Major extracranial bleed	95,000 (6 studies) 3.8–10 y	High	RR 1.54 (1.30–1.82)	Low cardiovascular–risk population[b] 8 bleeds per 1,000[c]	4 more per 1,000 (from 2 more to 7 more)
				Moderate cardiovascular–risk population[b] 24 bleeds per 1,000[c]	16 more per 1,000 (from 7 more to 20 more)
				High cardiovascular–risk population[b] 40 bleeds per 1,000[c]	22 more per 1,000 (from 12 more to 33 more)

Abbreviations: GRADE, Grading of Recommendations Assessment, Development and Evaluation; CI, confidence interval; RR, risk ratio.

Aspirin (75–100 mg) compared with no aspirin for primary prevention of cardiovascular disease.

GRADE Working Group grades of evidence.

High quality, further research is very unlikely to change our confidence in the estimate of effect; moderate quality, further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate; low quality, further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate; very low quality, we are very uncertain about the estimate.

Bibliography: Baigent C, Blackwell L, Collins R, et al.; Antithrombotic Trialists' (ATT) Collaboration. Aspirin in the primary and secondary prevention of vascular disease: collaborative meta-analysis of individual participant data from randomized trials. Lancet.2009; 373(9678):1849–1860.

* The basis for the assumed risk (e.g., the median control group risk across studies) is provided in footnotes. The corresponding risk (and its 95% confidence interval) is based on the assumed risk in the comparison group and the relative effect of the intervention (and its 95% CI).

[a] Control group risk estimate for 10-year mortality apply to a 60-year-old man and came from population-based data from Statistics Norway. Mortality increases with age (e.g., 50-year-old man; 50 deaths per 1,000 in 10 years) and is lower in women than in men (e.g., 3% in women aged 50 years vs. 5% in men aged 50 years).

[b] Risk groups correspond to low risk (5%), medium risk (15%), and high risk (25%) according to the Framingham score (or other risk tool) to estimate 10-year risk.

[c] Control group risk estimates in low, moderate, and high cardiovascular–risk groups are based on the Framingham score. As explained in the text, we have used data from an individual patient data meta-analysis to provide estimated risks for patient-important outcomes not covered by the Framingham risk score. We have also adjusted for 20% overestimation associated with Framingham risk score.

the frequency of reporting absolute effects, our results did show that nontop journals report benefits and harms using different metrics (mismatched framing) more often (19.1% vs. 33%) than top journals (Journal of the American Medical Association, The New England Journal of Medicine, Lancet, British Medical Journal, Annals of Internal Medicine, or PLoS Medicine). We were not able to compare our findings according to the statistical significance of the results, as these two authors did not provide this degree of detail [21].

4.3. Strengths and limitations

Strengths of our study include explicit eligibility criteria with independent duplicate adjudication of eligibility,

independent and duplicate data abstraction, and a large and representative sample of both Cochrane and non-Cochrane SRs. In contrast to previous studies, our sample is not restricted to top journals, and we addressed the type of absolute effect estimates reported, stratified our findings by statistical significance, the methods used for the calculation, and factors associated with the reporting of absolute effect estimates.

One limitation of our study is that we sampled only MEDLINE and the Cochrane database of SRs. The reporting of absolute effects in SRs in journals that are not indexed in MEDLINE may differ from our sample. It is likely, however, that nonindexed journals do no better than those we sampled.

Another limitation is the time frame of the sample. It is likely that the inclusion of absolute estimates increases time, although that increase is likely to be slow. Cochrane SRs are increasingly using GRADE and producing associated Summary of Findings tables, which necessarily involves producing estimates of absolute estimates. A cohort from 2012 showed that 45% of the Cochrane SRs included GRADE ratings [34]. If we consider the inclusion of GRADE to be a surrogate for including absolute estimates, this represents approximately a 6% increase in 2 years in the reporting of absolute measures. Given this modest increase and the likelihood that similar changes will be slower in non-Cochrane reviews, any underestimate of use of absolute risk in our data is probably modest.

4.4. Implications of findings

Our findings suggest that major improvements in the reporting of absolute effects in SRs are required. Whenever an intervention demonstrates possible benefit or harm on a particular outcome, authors of SRs should include both absolute and relative effect estimates for that outcome, report the methods for their calculation, and if appropriate, report different absolute values for patients with varying baseline risks.

Leading authorities in SR methodology could facilitate greater and more appropriate presentation of absolute effects in SRs. The next iteration of the PRISMA statement should include an item addressing the reporting of absolute effect estimates, and the Cochrane Collaboration should include more explicit guidance about their reporting in its handbook. Requirements to include Summary of Findings tables that require absolute effects would likely solve the problem. Finally, editors should insist on optimal reporting of absolute effects in SRs published in their journals.

Acknowledgments

The authors would like to thank Jill Hayden for her insightful comments on previous versions of this manuscript.

Authors' contributions: Conception and design was done by P.A-.C., G.H.G., A.C-.L., R.B-.P., I.N., E.A.A., and H.S. Design of search strategy was done by A.C-.L. and P.A-.C. Article selection was done by P.A-.C., A.C-.L., R.B-.P., I.N., E.A.A., R.W.M.V., B.C.J., X.S., M.B., J.W.B., S.E., C.E.G., A.Io., A.Ir., L.M.G., R.A.M., A.R-.M., A.S., I.S., A.J.S., K.A.O.T., P.O.V., and O.E.Z. Data abstraction was done by P.A-.C., A.C-.L., R.B-.P., I.N., E.A.A., R.W.M.V., B.C.J., X.S., M.B., J.W.B., S.E., C.E.G., A.Io., A.Ir., L.M.G., R.A.M., A.R-.M., A.S., I.S., A.J.S., K.A.O.T., P.O.V., R.W.M.V., and O.E.Z., Q.Z. and R.B-.P. analyzed the data. P.A-.C., A.C-.L., R.B-.P., I.N., E.A.A., B.C.J., X.S., M.B., J.W.B., S.E., C.E.G., A.Io., A.Ir., L.M.G., R.A.M., A.R-.M., A.S., I.S., A.J.S., K.A.O.T., P.O.V., R.W.M.V., O.E.Z., and G.H.G. participated in the interpretation of results. P.A-.C., R.W.M.V., and G.H.G. drafted the manuscript. P.A-.C., A.C-.L., R.B-.P., I.N., E.A.A., B.C.J., X.S., M.B., J.W.B., S.E., C.E.G., A.Io., A.Ir., L.M.G., R.A.M., A.R-.M., A.S., I.S., A.J.S., K.A.O.T., P.O.V., R.W.M.V., O.E.Z., Q.Z., H.S., and G.H.G. reviewed and approved the manuscript.

Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.jclinepi.2015.11.002.

References

[1] Laine C, Taichman DB, Mulrow C. Trustworthy clinical guidelines. Ann Intern Med 2011;154:774–5.

[2] Sackett DL, Cook RJ. Understanding clinical trials. BMJ 1994;309:755–6.

[3] Barratt A, Wyer PC, Hatala R, McGinn T, Dans AL, Keitz S, et al. Evidence-Based Medicine Teaching Tips Working Group. Tips for learners of evidence-based medicine: 1. Relative risk reduction, absolute risk reduction and number needed to treat. CMAJ 2004;171:353–8.

[4] Nexoe J, Gyrd-Hansen D, Kragstrup J, Kristiansen IS, Nielsen JB. Danish GPs' perception of disease risk and benefit of prevention. Fam Pract 2002;19:3–6.

[5] McGettigan P, Sly K, O'Connell D, Hill S, Henry D. The effects of information framing on the practices of physicians. J Gen Intern Med 1999;14:633–42.

[6] Hux JE, Naylor CD. Communicating the benefits of chronic preventive therapy: does the format of efficacy data determine patients' acceptance of treatment? Med Decis Making 1995;15:152–7.

[7] Akl EA, Oxman AD, Herrin J, Vist GE, Terrenato I, Sperati F, et al. Framing of health information messages. Cochrane Database Syst Rev 2011;12:CD006777.

[8] Malenka DJ, Baron JA, Johansen S, Wahrenberger JW, Ross JM. The framing effect of relative and absolute risk. J Gen Intern Med 1993;8:543–8.

[9] Mantese VA, Timaran CH, Chiu D, Begg RJ, Brott TG, CREST Investigators. The Carotid Revascularization Endarterectomy versus Stenting Trial (CREST): stenting versus carotid endarterectomy for carotid disease. Stroke 2010;41:S31–4.

[10] Schmid CH, Lau J, McIntosh MW, Cappelleri JC. An empirical study of the effect of the control rate as a predictor of treatment

efficacy in meta-analysis of clinical trials. Stat Med 1998;17:1923−42.
[11] Deeks JJ, Higgins JPT, Altman DG. Chapter 9: Analysing data and undertaking meta-analyses. In: Higgins JPT, Green S, editors. Cochrane handbook for systematic reviews of interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration; 2011. Available at http://handbook.cochrane.org/chapter_9/9_analysing_data_and_undertaking_meta_analyses.htm. Accessed November 30, 2015.
[12] Furukawa TA, Guyatt GH, Griffith LE. Can we individualize the number needed to treat (NNT)? An empirical study of summary effect measures in meta-analyses. Int J Epidemiol 2002;31:72−6.
[13] Smeeth L, Haines A, Ebrahim S. Numbers needed to treat derived from meta-analyses sometimes informative, usually misleading. BMJ 1999;318:1548−51.
[14] Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ 2009;339:b2700.
[15] Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al, GRADE Working Group. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 2008;336:924−6.
[16] Guyatt G, Oxman AD, Kunz R, Brozek J, Alonso-Coello P, Rind D, et al. GRADE guidelines 6. Rating the quality of evidence—imprecision. J Clin Epidemiol 2011;64:1283−93.
[17] Schünemann HJ, Oxman AD, Vist GE, Higgins JPT, Deeks JJ, Glasziou P, et al. Chapter 12: interpreting results and drawing conclusions. In: Higgins JPT, Green S, editors. Cochrane handbook for systematic reviews of interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration; 2011. Available at http://handbook.cochrane.org/chapter_12/12_8_chapter_information.htm. Accessed November 30, 2015.
[18] Schwartz LM, Woloshin S, Dvorin EL, Welch HG. Ratio measures in leading medical journals: structured review of accessibility of underlying absolute risks. BMJ 2006;333:1248−52.
[19] King NB, Harper S, Young ME. Use of relative and absolute effect measures in reporting health inequalities: structured review. BMJ 2012;345:e5774.
[20] Beller EM, Glasziou PP, Hopewell S, Altman DG. Reporting of effect direction and size in abstracts of systematic reviews. JAMA 2011;306:1981−2.
[21] Sedrakyan A, Shih C. Improving depiction of benefits and harms: analyses of studies of well-known therapeutics and review of high-impact medical journals. Med Care 2007;45:523−8.
[22] Gigerenzer G, Wegwarth O, Feufel M. Misleading communication of risk. BMJ 2010;341:c4830.
[23] Alonso-Coello P, Carrasco-Labra A, Brignardello-Petersen R, Neumann I, Akl EA, Sun X, et al. A methodological survey of the analysis, reporting and interpretation of absolute risk reduction in systematic reviews (ARROW): a study protocol. Syst Rev 2013;2:113.
[24] Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Med Res Methodol 2007;7:10.
[25] Cook RJ, Heddle NM. Measuring and reporting the effect of an intervention: absolute measures. Transfusion 2009;49:2022−3.
[26] Citrome L. Compelling or irrelevant? Using number needed to treat can help decide. Acta Psychiatr Scand 2008;117:412−9.
[27] Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159−74.
[28] Forrow L, Taylor WC, Arnold RM. Absolutely relative: how research results are summarized can affect treatment decisions. Am J Med 1992;92:121−4.
[29] Naylor CD, Chen E, Strauss B. Measured enthusiasm: does the method of reporting trial results alter perceptions of therapeutic effectiveness? Ann Intern Med 1992;117:916−21.
[30] Tucker G, Metcalfe A, Pearce C, Need AG, Dick IM, Prince RL, et al. The importance of calculating absolute rather than relative fracture risk. Bone 2007;41:937−41.
[31] Hayden J, Tomlinson G, Ni A, Parker R, Cartwright J, van M Tulder, et al., Approaches to estimate and present baseline risks: recommendations for Cochrane review 'Summary of findings' (SoF) tables. 19th Cochrane Colloquium Madrid 2011.
[32] Rennie D. CONSORT revised—improving the reporting of randomized trials. JAMA 2001;285:2006−7.
[33] von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE)statement: guidelines for reporting observational studies. J Clin Epidemiol 2008;61:344−9.
[34] Hopewell S, Boutron I, Altman DG, Ravaud P. Incorporation of assessments of risk of bias of primary studies in systematic reviews of randomised trials: a cross-sectional study. BMJ Open 2013;3:e003342.