BMJ Open

How are risk ratios reported in orthopaedic surgery journals? A descriptive study of formats used to report absolute risks

Dafina Petrova,[1,2] Alexander Joeris,[3] María-José Sánchez,[1,2,4] Elena Salamanca-Fernández,[1,2,4] Rocio Garcia-Retamero[5,6]

To cite: Petrova D, Joeris A, Sánchez M-J, *et al*. How are risk ratios reported in orthopaedic surgery journals? A descriptive study of formats used to report absolute risks. *BMJ Open* 2018;**8**:e025047. doi:10.1136/bmjopen-2018-025047

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2018-025047).

Received 27 June 2018
Revised 18 October 2018
Accepted 19 October 2018

For numbered affiliations see end of article.

Correspondence to
Dr Alexander Joeris;
alexander.joeris@aofoundation.org

ABSTRACT

Purpose The numerical format in which risks are communicated can affect risk comprehension and perceptions of medical professionals. We investigated what numerical formats are used to report absolute risks in empirical articles, estimated the frequency of biasing formats and rated the quality of figures used to display the risks.

Design Descriptive study of reporting practices.

Method We randomly sampled articles published in seven leading orthopaedic surgery journals during a period of 13 years. From these, we selected articles that reported group comparisons on a binary outcome (eg, revision rates in two groups) and recorded the numerical format used to communicate the absolute risks in the results section. The quality of figures was assessed according to published guidelines for transparent visual aids design.

Outcome measures Prevalence of information formats and quality of figures.

Results The final sample consisted of 507 articles, of which 14% reported level 1 evidence, 13% level 2 and 73% level 3 or lower. The majority of articles compared groups of different sizes (90%), reported both raw numbers and percentages (64%) and did not report the group sizes alongside (50%). Fifteen per cent of articles used two formats identified as biasing: only raw numbers (8%, '90 patients vs 100 patients') or raw numbers reported alongside different group sizes (7%, '90 out of 340 patients vs 100 out of 490 patients'). The prevalence of these formats decreased in more recent publications. Figures (n=79) had on average two faults that could distort comprehension, and the majority were rated as biasing.

Conclusion Authors use a variety of formats to report absolute risks in scientific articles and are likely not aware of how some formats and graph design features can distort comprehension. Biases can be reduced if journals adopt guidelines for transparent risk communication but more research is needed into the effects of different formats.

Strengths and limitations of this study

- We randomly sampled and analysed a broad selection of studies published over 13 years in several leading orthopaedic surgery journals.
- This is the first study to describe the formats used to communicate absolute risks and estimate the prevalence of these formats, including formats previously identified as biasing.
- We assessed the quality of figures used to report absolute risks according to guidelines for quality of visual aids.
- This study was limited to publications regarding orthopaedic surgery from a selected set of journals.
- Additional information about the absolute risks sampled was not considered (eg, what additional effect sizes were reported and whether the selected absolute risks were part of the main study outcomes).

INTRODUCTION

The majority of American surgeons consider scientific articles published in peer-reviewed journals as their main source of information.[1]

To facilitate good science and evidence-based practice, it is essential that the results in scientific articles are described transparently and comprehensively. Studies in surgery often compare two or more groups on binary outcome variables such as mortality, morbidity or treatment success. This type of data allows us to estimate treatment risk reduction or risk increase associated with influential factors. Because relative risks can make small differences appear larger,[2 3] standard guidelines generally recommend that relative risks are accompanied by absolute risks and that group sizes are mentioned.[4] However, research shows that about 35% of articles in leading medical journals do not report absolute risks.[5]

In addition, many of the standard guidelines for communicating results in scientific research do not specify *how* exactly absolute risks should be reported.[4 6] Other sources such as the International Committee of Medical Journal Editors (ICMJE) guidelines recommend that percentages be accompanied by

the raw numbers they were derived from.[7] However, it is not clear to what extent authors follow these guidelines. This is important because extensive research has documented that even small changes in the numerical format used to report risks can have profound effects on comprehension in both laymen and medical professionals.[2 3 8–13]

Research shows that people often pay more attention to the number of individuals affected by a risk in a ratio (ie, the numerator) and ignore or pay less attention to the overall number of people at risk (ie, the denominator).[14–18] This effect is called *denominator neglect* (or ratio bias).[14–18] To illustrate it, consider the following example: 90 individuals recovered in a group of 340 individuals who received an old treatment, and 100 individuals recovered in a group of 490 individuals who received a new treatment. Neglecting the denominator can result in distorted risk perceptions more in line with the raw numbers (90<100) rather than the actual proportions (26%>20%).[14–18] In addition, reporting only raw numbers and denominators (without percentages) might also be user-unfriendly. Readers would be required to make a mental calculation to derive the relevant proportions when the denominators are different, which can result in *mental computation errors*. For instance, a recent study showed that reporting only raw numbers in the numerator (eg, 90 vs 100) or raw numbers alongside different denominators (eg, 90 out of 340 vs 100 out of 490) strongly biased surgeons' risk judgements; in addition, surgeons rated these formats as unclear and confusing.[19]

Denominator neglect and mental computation errors can be avoided with the use of transparent visual aids—graphs and figures that display risk information in a way that clarifies key data points and makes part-to-whole relations in the data visually available.[20] For instance, well-designed visual aids were found to significantly improve risk reduction estimates of highly experienced orthopaedic surgeons.[19 21] In contrast, poorly designed visual aids can be confusing or misleading.[20] For instance, displaying absolute risks in two groups in a bar graph with a truncated axis (eg, starting at 50% and not at 0%) can make the difference between the groups appear larger and bias perceptions of risk reduction.

Although there is research showing that certain formats to communicate absolute risks can bias comprehension and judgement, no research to our knowledge has estimated how frequently they are used in practice. Hence, the purpose of this research was threefold. First, we investigated what numerical formats are most frequently used to report absolute risks in leading surgery journals, focusing on orthopaedic surgery. We expected to find a variety of formats, the majority including raw numbers and percentages consistent with the ICMJE guidelines.[7] Second, we estimated the frequency of biasing formats that can give rise to denominator neglect and/or mental computation errors. In particular, we estimated the prevalence of articles that used: (A) raw numbers only (eg, 90 patients vs 100 patients) or (B) raw numbers alongside different denominators (eg, 90 patients in a group of 340 vs 100 patients in a group of 490). Third, we rated the quality of visual aids used to display absolute risks following published evidence-based guidelines for design of effective visual aids.[20 22–26] We focused on orthopaedic surgery because of the following reasons: compared with other medical specialties (eg, pharmacology), publications in orthopaedic surgery are especially heterogeneous when it comes to study design.[27 28] In addition, low level of evidence and studies with small and unbalanced number of subjects are very common, suggesting that the prevalence of biasing formats as described above is of high relevance.[27 29 30]

METHOD

We conducted a descriptive study of formats used to report absolute risks in group comparisons across orthopaedic surgery journals. We started by selecting a diverse sample of leading orthopaedic surgery journals. Journal selection was guided by the following criteria: (1) the journals predominantly published empirical studies (ie, research articles and not reviews or other types of publications without primary data); (2) articles published in the journals made a substantial contribution to the area of orthopaedic surgery according to the opinion of three experienced surgeons, (3) the journals were included in international indexes and databases, and (4) the journals represented a mixture of high and moderate impact factor outlets according to Journal Citations Reports (average impact in 2012, 2013 and 2014 between 1.7 and 4.0).

To provide a broad and representative sample of reporting practices, using Web of Science, we searched the titles and abstracts of all articles published in the selected journals in the previous 13 years (ie, 2005–2017, n=23 508). We decided not to sample articles published earlier because they are more likely to represent outdated practices that have changed as a result of updated journal or reporting guidelines (eg, ICMJE guidelines published in 2004[7]). To reduce the study sample to a manageable number for data collection, we selected the articles published every even (vs odd) year (ie, 2006, 2008, 2010, 2012, 2014 and 2016; n=11 692). Using the RAND() command in Excel, we then randomly selected 25% of the articles of each journal for review (n=2.923) (see refs 31 and 32 for a similar method). To further select articles containing relevant data, we screened the abstracts of all articles against prespecified inclusion criteria: articles proceeded to full-text review if, according to the abstract, the article reported results of an empirical study and mentioned a group comparison on a binary outcome (eg, revision rates in two groups). The agreement on abstract selection between two independent raters was 95%. In order not to overlook relevant data, in case of doubt whether the group comparison met the inclusion criteria, the article proceeded to full-text review.

Finally, we reviewed the full text of each selected article and extracted data about the reported group comparison

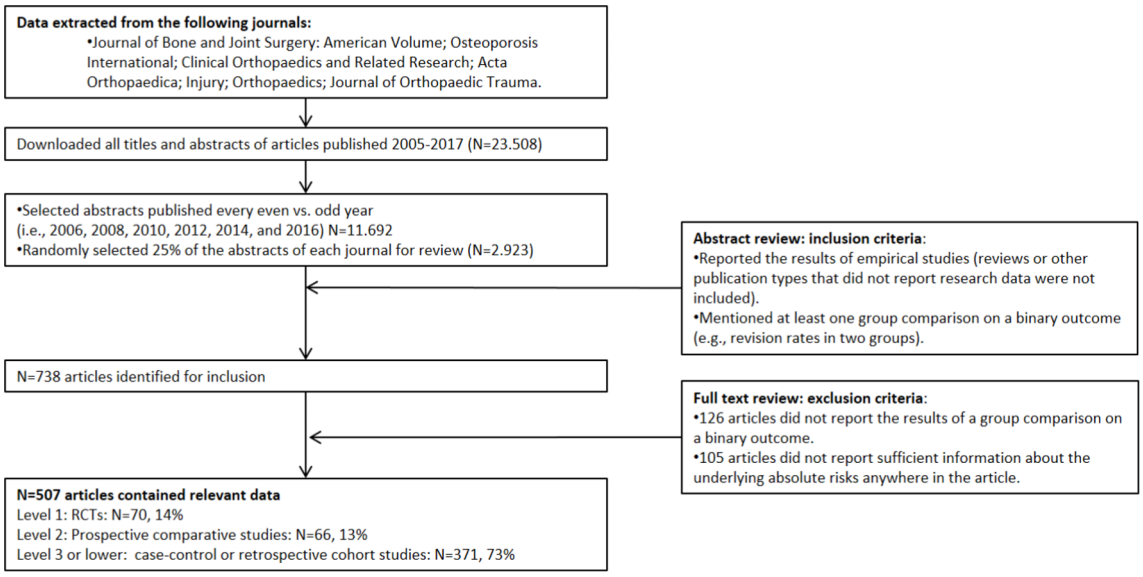

Figure 1 Flow chart showing the review process. RCTs, randomised controlled trials.

on a binary outcome measure. Data from each article were extracted using a prepiloted form and were double-checked for accuracy. We selected the first group comparison reported in the abstract when multiple comparisons were reported. Articles were reviewed in a random order for each year. After reviewing the full text, we excluded 126 articles (17%) because after examination of the full text, it became clear that they do not report the results of a group comparison on a binary outcome. Of the remaining 612 articles, 105 (17%) did not report any raw data for the underlying absolute risks (ie, number of events in each group) but only reported a standardised measure of effect size for the relevant comparison (eg, OR, HR, p values or words only). Hence, the final article sample size was n=507.

Because absolute risks are relevant for a variety of study designs, we did not restrict the sample to specific study designs; instead, we sampled a broad selection of designs including RCTs, prospective studies, case–control studies and retrospective cohort studies, among others. Figure 1 displays a detailed flow chart of the article selection process, inclusion/exclusion criteria and results.

Measures

For each article, we recorded the level of evidence: '1' (randomised controlled trial (RCT)), '2' (prospective comparative study) or '3 or lower' (eg, case–control study or retrospective cohort study) as characterised by the journal/authors in the title, abstract or text of the study; we also recorded the total sample size of the study, the number of groups compared, whether group sizes of the compared groups (ie, denominators) were equal; and whether the group comparison was significant.

We recorded how data were reported in the results section of the article. We recorded the presentation mode (ie, the mode in which the data were reported): whether all or part of the information about the size of the numerators/denominators was reported in the text, a table or a figure. Out of these three modes, we also defined the main presentation mode: the mode in which most information was communicated (see online supplementary appendix 1 for a more extensive definition). For each presentation mode, we recorded whether numerators were reported in raw numbers (eg, five patients), and if that was the case, whether denominators were reported alongside (eg, 5 in 250 patients); whether results were reported in percentages (eg, 2%), and if that was the case, whether denominators were reported alongside (eg, 2% of 250 patients). We defined 'alongside' as in the same sentence in text, not further than three rows or columns away in a table, and anywhere in a figure.

From these, we identified nine different information formats of risk communication, which were a function of the size of the groups compared (equal or different denominators) and the type of information used to communicate the group differences (raw numbers, percentages or both). We additionally grouped these formats into three categories based on the available evidence for their effects on comprehension and judgement: *biasing*, *not biasing* and *evidence needed*. Based on previous research including nationally representative samples and samples of highly qualified medical professionals, two formats were identified as *biasing*: raw numbers with different denominators based on converging evidence from multiple studies[14–19] and raw numbers without denominators based on evidence and the logic that no valid inference could be made without knowledge about the group sizes.[19] In contrast, one format was identified as *not biasing*: communicating rates using raw numbers with the same denominators was found to result in good comprehension and judgement.[14–19] There was no sufficient evidence for the remaining formats, and they were labelled as *evidence needed* (although see ref 19 for some initial results on those formats).

Finally, because figures are meant to be stand-alone communications, we rated their quality following published evidence-based guidelines.[20 22–26] In particular, for each figure, we computed the number of eight possible predefined *comprehension faults* (ie, features of the figure that can confuse or mislead readers; features that do not facilitate comprehension or that can significantly influence the perception of differences between the groups). We considered figures with no comprehension faults as *transparent visual aids*; figures with one comprehension fault as *biasing visual aids*; and figures with two or more comprehension faults as *very biasing visual aids*. These cut-offs were based on our expert judgement regarding the potential impact of comprehension faults.

To account for the way different reporting practices could affect readers varying in habits, knowledge and motivation, we analysed the frequency of information formats using three different analytical strategies based on the exact location of the information (detailed in the online supplementary appendix 1). Because these strategies produced very similar results, here we summarise the most important findings considering information reported across the entire results section. Detailed results from the other analytical strategies are found in online supplementary appendix 1.

Patient involvement

This is a descriptive study of reporting practices in specialised medical journals. No individual patient data were used, and no patients were otherwise involved in the research.

RESULTS

What type of data did the articles report?

The most common outcomes were rates of survival/mortality (12%), complications (9%), infections (5%), failure (5%), fractures (5%), revisions (4%), (non) unions (3%), blood transfusions (3%), injury (3%) and other very diverse outcomes (50%). Figure 1 shows the distribution of articles according to level of evidence. Three-hundred and twenty-eight (65%) articles reported a significant effect of the selected comparison. The median number of groups compared was 2 (min=2, max=8). The median sample size of the studies was 183 (IQR: 81–689).

Only 52 (10%) of the articles compared groups of equal sizes (ie, equal denominators) on the selected data comparison; the remaining 455 (90%) articles compared groups of different sizes (ie, different denominators). The median difference between the size of the denominators was 30 (IQR: 4–197).

Two-hundred and seventy-three articles (54%) reported the data using only one presentation mode; 223 (44%) used two presentation modes, and 11 (2%) presented the data using the three presentation modes. In the majority or articles (486, 96%) the main presentation mode was numerical (text for 282 (56%) and tables for 204 (40%)). The main presentation mode was figures for only 21 (4%) articles.

Prevalence of information formats

The main results are summarised in table 1. Regarding numerators, the majority of articles (323, 63%) reported raw numbers and percentages. Regarding denominators, half of the articles (253, 50%) did not report the denominators alongside raw numbers and/or percentages, and 234 (46%) reported denominators of different sizes alongside the raw numbers and/or percentages. The remaining 18 articles (4%) reported denominators of equal sizes alongside the raw numbers and/or percentages. Importantly, considering information reported anywhere in the results section, 77 (15%) articles reported the data using the two biasing formats: only raw numbers (8%) or raw numbers reported alongside different denominators (7%). Considering information reported in the main presentation mode of this article, this number was 86 (17%): only raw numbers (8%) or raw numbers reported alongside different denominators (9%).

We next explored the prevalence of information formats as a function of publication year. Data are displayed in table 2, where a decrement in the prevalence of the two biasing formats can be appreciated in the most recent years. In order to analyse whether there was a significant tendency, we created two groups: a biasing group that consisted of the articles that used one of the two biasing formats (as per information reported anywhere in the article) and another group with the articles using any of the other formats. We then fitted a logistic regression model using *glm* in *R* with dependent variable group (biasing vs other) and independent variable year of publication, which showed that more recent publications were less likely to use one of the biasing formats, OR_{2years}=0.79, 95% CI 0.68 to 0.91, p=0.001.

Quality of figures

Eighty-two (16%) articles communicated the absolute risks in a figure. Three of the 82 figures were study flowcharts and could not be rated using the rating criteria. Of the remaining 79 figures, 25 (32%) were bar graphs, 53 (67%) were survival graphs and 1 (1%) was a line graph.

Overall, the quality of figures was poor. On average, figures had a median of two comprehension faults (IQR: 2–3). Only one figure (1%) had no comprehension faults and was hence considered *a transparent visual aid;* 15 (19%) had one comprehension fault and were considered biassing; the remaining 63 (80%) were *very biasing:* 41 (52%) had two comprehension faults, and 22 (28%) had three comprehension faults. The most common comprehension faults were related to communicating the reliability of the data and the documented differences: the majority of the figures did not include the number of individuals in each group or did not display confidence intervals or error bars (see table 3 for details). Truncated axes were also very common.

Table 1 Percentage of articles reporting data as a function of information format and presentation mode

Information format	Effect on comprehension and judgements	Example	All data reported anywhere in the results section	By presentation mode		
				Main presentation mode (across text, tables and figures)	Numerical main presentation mode (text or table, n=486)	Graphical main presentation mode (figure, n=21)
			% out of 507	% out of 507	% out of 507	% out of 507
Raw numbers: same denominators	Not biasing	64 in 160 versus 40 in 160	1	1	1	0
Raw numbers: different denominators	Biasing	64 in 160 versus 12 in 48	7	8	8	0
Raw numbers: no denominators	Biasing	64 versus 12	8	9	9	0
Percentages: same denominators	Evidence is needed	40% in 160 versus 25% in 160	0	0	0	0
Percentages: different denominators	Evidence is needed	40% in 160 versus 25% in 48	5	4	2	2
Percentages: no denominators	Evidence is needed	40% versus 25%	15	16	15	2
Raw numbers and percentages: same denominators	Evidence is needed	64 in 160 (40%) versus 40 in 160 (25%)	2	2	2	0
Raw numbers and percentages: different denominators	Evidence is needed	64 in 160 (40%) versus 12 in 48 (25%)	35	33	33	0
Raw numbers and percentages: no denominators	Evidence is needed	64 (40%) versus 12 (25%)	26	25	25	0
			100	100	96	4

Percentages may not sum up to 100 due to rounding.
The effects of the information formats on comprehension and judgements are based on several available studies, including representative national samples and samples of highly qualified surgeons.[14–19] The main presentation mode was defined as the mode in which authors communicated most information (see online supplementary appendix 1 for details).

DISCUSSION

Authors use diverse information formats to communicate absolute risks, the majority of which are a combination of raw numbers and percentages. The current study shows that about 15%–17% of the articles in the surveyed orthopaedic surgery journals used a format that can bias surgeons' estimates: these articles reported absolute risks using raw numbers alongside denominators of different sizes or did not report the denominators alongside at all. This relatively frequent occurrence is due to both the type of data analysed (ie, 90% of articles compared groups of different sizes) and the reporting practices of authors (eg, reporting group sample sizes in the method section only and not in the results). As this is a recent and active research topic, authors and reviewers may not be aware of how these and other information formats can influence comprehension and perceptions. Biasing risk communication formats may be more dangerous in reports of RCTs (that are more likely to be directly used for treatment decision making) rather than level 3 or lower studies, where absolute risks may be less important.

On a more positive note, we found an encouraging trend that the prevalence of biasing formats is decreasing, which may be a result of authors following general reporting guidelines.[7] Nevertheless, journals should adopt policies regarding the reporting of such statistics in order to prevent bias and facilitate comprehension as much as possible. Whereas the current study established the prevalence of information formats, in order to fully understand the implications of current practices and formulate specific reporting guidelines—regarding what formats should be avoided and what formats should be used—more research is needed, especially on the effects of the most commonly used formats.

A recent survey of a diverse international sample of 292 surgeons showed that 40% of them could not correctly estimate relative risk reduction from raw numbers and denominators even when the groups compared were of equal sizes.[21] In addition, when the groups compared are of different sizes, denominator neglect can strongly bias perceptions.[14–19] Whereas communicating percentages alongside the raw numbers should in theory eliminate any resulting biases, recent results from our lab indicate that this practice is not preferred by surgeons who perceive it as unclear and think it is too much information.[19] Given the very high frequency of use of formats that combine raw numbers and percentages (63% of articles in the current review), further research into their effects is needed.

Table 2 Information formats (according to all data reported anywhere in the results section) as a function of publication year

Information format	Example	Publication year						
		2006 % of 68	2008 % of 45	2010 % of 72	2012 % of 97	2014 % of 124	2016 % of 101	Total % of 507
Raw numbers: same denominators	64 in 160 versus 40 in 160	2	0	0	1	1	1	1
Raw numbers: different denominators*	64 in 160 versus 12 in 48	9	13	10	4	6	3	7
Raw numbers: no denominators*	64 versus 12	12	9	12	11	6	3	8
Percentages: same denominators	40% in 160 versus 25% in 160	0	0	0	1	0	1	0
Percentages: different denominators	40% in 160 versus 25% in 48	6	2	6	4	2	7	5
Percentages: no denominators	40% versus 25%	19	22	10	13	14	18	15
Raw numbers and percentages: same denominators	64 in 160 (40%) versus 40 in 160 (25%)	4	2	1	2	3	1	2
Raw numbers and percentages: different denominators	64 in 160 (40%) versus 12 in 48 (25%)	35	33	46	31	32	36	35
Raw numbers and percentages: no denominators	64 (40%) versus 12 (25%)	13	18	15	32	36	31	26
Total		100	100	100	100	100	100	100

Percentages may not sum up to 100 due to rounding. An asterisk (*) marks formats previously found to be biasing.

Both denominator neglect and mental computation errors would be more likely to occur when group sizes are not reported alongside (eg, but are reported in another section of the article) and when readers are quickly scanning versus carefully examining a research article. This is important to consider because many surgeons would lack the time (or training) to read carefully and critically the ever-increasing number of new articles pertinent to their specialty. This often leads to skimming articles or reading only the most important sections.[33 34]

One way to facilitate risk comprehension is with the use of well-designed visual aids: figures that display the information in a way that clarifies key data points and make part-to-whole relations in the data transparent and visually available.[20] For instance, displaying absolute risks in a transparent visual aid helped surgeons correctly infer

Table 3 Number and percentage of figures that displayed each of the assessed comprehension faults

	N	% of 79 figures
1. The figure type is inappropriate (eg, a line graph implying change over time is used when a bar graph is the appropriate figure).	1	1
2. The y-axis is inverted.	0	0
3. The y-axis is truncated at the start (not starting at the true minimum value, for example, starting at 50% and not at 0%).	17	22
4. The y-axis is truncated at the end (not ending with the true maximum value, for example, ending at 50% and not at 100%).	17	22
5. The number of participants per group is not mentioned.	63	80
6. There are no error bars or confidence intervals displayed.	61	77
7. In cases where multiple figures are used to display the outcomes in the compared groups: the graph features are not consistent (eg, y-axis from 0% to 100% for group A but y-axis from 0% to 50% for group B).	1	1
8. In case of bar graphs, stacked bar graphs are used in a way that does not facilitate the comparison between the different categories.	3	4

the risk reduction associated with the use of a new type of anaesthesia: 85% answered correctly with the visual aid versus only 60% without it.[21] Transparent visual aids can also eliminate biases when denominators of the compared groups are different.[19] However, we found that only a small proportion of the articles in orthopaedic surgery journals use figures to display the absolute risks. This low frequency of use can be explained by the fact that authors use figures to depict other types of results (eg, the central outcome of the research, complex patterns or continuous outcomes). In our study, we selected the first group comparison that was reported in the abstract, which does not always coincide with the central outcome of the study.

The generally poor quality of the figures in the reviewed research could be explained by the fact that authors were probably not aware of the potential negative effects of some graph design features on comprehension. They could have also considered these figures as complementary rather than as stand-alone communications (ie, 82/507 articles included figures, but the majority provided detailed information in the text or in tables). In addition, we applied strict criteria for evaluating the quality of figures. We do believe that strict criteria are necessary because figures are powerful communication tools that attract readers' attention. If figures are not properly designed, they could easily mislead readers.[35 36] For instance, truncating the y-axis can make the differences between the compared groups appear much larger—an effect similar to that of communicating relative versus absolute risks and that was present in about half of the figures we evaluated. Similar distortions were documented in a review of the quality of graphs used in pharmaceutical advertisement in leading medical journals.[37] Our results suggest that there is room for improvement in the quality of figures used to communicate absolute risks. Adding specific guidelines for transparent figure design to the instructions for authors in scientific journals could be useful.

Fourteen per cent (105/738) of the articles selected for the current research did not report the underlying absolute risks of the data comparison that was selected. This estimate is lower than the one obtained by Schwartz *et al*[5] 10 years ago. In their review of RCTs and cohort studies in leading medical journals, 35% of articles did not report the absolute risks anywhere in the article.

Finally, we did not consider whether the risks communicated were a central finding of the article or not or what additional measures of effect size were reported. Future studies should address these important aspects, especially having in mind that effect size measures such as OR, relative risk or number-needed-to treat can strongly affect comprehension and perceptions[13] and can have diverse effects in relation to reports of the underlying absolute risks. Our study was restricted to orthopaedic surgery only, in which our team has expertise, and did not cover all journals that publish studies relevant to orthopaedic surgery. We have no reason to believe that reporting practices would be very different in other surgical specialties; however, future research should address this issue. Although in reporting our study we followed the STROBE checklist for cross-sectional studies, no specific guidelines exist to inform reporting practices studies that mix primary research and review methods.

The current review demonstrates that authors use a variety of formats to report absolute risks in scientific articles and a significant proportion of them have previously been identified as biasing. Authors are likely not aware of how some formats and graph design features can distort comprehension. More research is needed into the effects of the most frequently used formats in order to make specific recommendations for policies aiming to standardise the reporting of such data.

Author affiliations
[1]Escuela Andaluza de Salud Pública, Granada, Spain
[2]Instituto de Investigación Biosanitaria ibs.GRANADA, Universidad de Granada, Granada, Spain
[3]AO Clinical Investigation and Documentation, Duebendorf, Switzerland
[4]CIBER de Epidemiología y Salud Pública (CIBERESP), Madrid, Spain
[5]Mind, Brain, and Behavior Reseach Center (CIMCYC), Department of Experimental Psychology, University of Granada, Granada, Spain
[6]Harding Center for Risk Literacy, Max Planck Institute for Human Development, Berlin, Germany

Contributors DP, AJ and RG-R contributed to the design of the study. DP, ES-F and RG-R collected and analysed the data. All authors interpreted the results and revised the manuscript (DP produced the first draft). All authors approved the final version of the manuscript.

Funding DP is supported by a Juan de la Cierva Fellowship (FJCI-2016-28279) from the Spanish Ministry of Economy, Industry, and Competitiveness. The presented study was partially funded by the AO Foundation via the AO Clinical Investigation and Documentation network and by the Spanish Ministry of Economy and Competitiveness (Spain) (PSI2011-22954 and PSI2014-51842-R).

Disclaimer The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests None declared.

Patient consent Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement Data are publicly available on the Open Science Framework: https://osf.io/g7df4/.

REFERENCES

1. Schein M, Paladugu R, Sutija VG, *et al*. What American surgeons read: a survey of a thousand Fellows of the American College of Surgeons. *Curr Surg* 2000;57:252–8.
2. Covey J. A meta-analysis of the effects of presenting treatment benefits in different formats. *Med Decis Making* 2007;27:638–54.
3. Nexøe J, Gyrd-Hansen D, Kragstrup J, *et al*. Danish GPs' perception of disease risk and benefit of prevention. *Fam Pract* 2002;19:3–6.
4. Schulz KF, Altman DG, Consort MD. statement: updated guidelines for reporting parallel group randomised trials. *BMC medicine* 2010;2010:18.
5. Schwartz LM, Woloshin S, Dvorin EL, *et al*. Ratio measures in leading medical journals: structured review of accessibility of underlying absolute risks. *BMJ* 2006;333:1248.

6. von Elm E, Altman DG, Egger M, *et al*. The strengthening the reporting of observational studies in epidemiology (STROBE) Statement: guidelines for reporting observational studies. *Int J Surg* 2014;12:1495–9.
7. International Committee of Medical Journal Editors (ICMJE). International Committee of Medical Journal Editors (ICMJE): Uniform Requirements for Manuscripts Submitted to Biomedical Journals: writing and editing for biomedical publication. *Haematologica* 2004;89:264.
8. Elting LS, Martin CG, Cantor SB, *et al*. Influence of data display formats on physician investigators' decisions to stop clinical trials: prospective trial with repeated measures. *BMJ* 1999;318:1527–31.
9. Gigerenzer G, Gaissmaier W, Kurz-Milcke E, *et al*. Helping doctors and patients make sense of health statistics. *Psychol Sci Public Interest* 2007;8:53–96.
10. Wegwarth O, Schwartz LM, Woloshin S, *et al*. Do physicians understand cancer screening statistics? A national survey of primary care physicians in the United States. *Ann Intern Med* 2012;156:340–9.
11. Petrova D, Kostopoulou O, Delaney BC, *et al*. Strengths and gaps in physicians'risk communication: a scenario study of the influence of numeracy on cancer screening communication. *Med Decis Making* 2018;38:355–65.
12. Anderson BL, Gigerenzer G, Parker S, *et al*. Statistical literacy in obstetricians and gynecologists. *J Healthc Qual* 2014;36:5–17.
13. Akl EA, Oxman AD, Herrin J, *et al*. Using alternative statistical formats for presenting risks and risk reductions. *Cochrane Database Syst Rev* 2011;3:CD006776.
14. Garcia-Retamero R, Galesic M, Gigerenzer G. Do icon arrays help reduce denominator neglect? *Med Decis Making* 2010;30:672–84.
15. Reyna VF, Brainerd CJ. Numeracy, ratio bias, and denominator neglect in judgments of risk and probability. *Learn Individ Differ* 2008;18:89–107.
16. Okan Y, Garcia-Retamero R, Cokely ET, *et al*. Individual differences in graph literacy: overcoming denominator neglect in risk comprehension. *J Behav Decis Mak* 2012;25:390–401.
17. Garcia-Retamero R, Galesic M. Communicating treatment risk reduction to people with low numeracy skills: a cross-cultural comparison. *Am J Public Health* 2009;99:2196–202.
18. Denes-Raj V, Epstein S, Cole J. The generality of the ratio-bias phenomenon. *Personality and Social Psychology Bulletin* 1995;21:1083–92.
19. Garcia-Retamero R, Petrova D, Cokely ET, *et al*. Scientific risk reporting in medical journals can bias expert judgment: comparing surgeons' risk comprehension across reporting formats. Under review.
20. Garcia-Retamero R, Cokely ET. Designing visual aids that promote risk literacy: a systematic review of health research and evidence-based design Heuristics. *Hum Factors* 2017;59:582–627.
21. Garcia-Retamero R, Cokely ET, Wicki B, *et al*. Improving risk literacy in surgeons. *Patient Educ Couns* 2016;99:1156–61.
22. Ancker JS, Senathirajah Y, Kukafka R, *et al*. Design features of graphs in health risk communication: a systematic review. *J Am Med Inform Assoc* 2006;13:608–18.
23. Gillan DJ, Wickens CD, Hollands JG, *et al*. Guidelines for presenting quantitative data in HFES publications. *Hum Factors* 1998;40:28–41.
24. Fischhoff B, Brewer NT, Downs JS. *Communicating risks and benefits: an evidence based user's guide*: Government Printing Office, 2012.
25. Lipkus IM. Numeric, verbal, and visual formats of conveying health risks: suggested best practices and future recommendations. *Med Decis Making* 2007;27:696–713.
26. Trevena LJ, Zikmund-Fisher BJ, Edwards A, *et al*. Presenting quantitative information about decision outcomes: a risk communication primer for patient decision aid developers. *BMC Med Inform Decis Mak* 2013;13 Suppl 2:S7.
27. Obremskey WT, Pappas N, Attallah-Wasif E, *et al*. Level of evidence in orthopaedic journals. *J Bone Joint Surg Am* 2005;87:2632–8.
28. Turaga KK. *Evidential strength of medical specialties: a comparative analysis*, 2003.
29. Okike K, Kocher MS, Torpey JL, *et al*. Level of evidence and conflict of interest disclosure associated with higher citation rates in orthopedics. *J Clin Epidemiol* 2011;64:331–8.
30. Kiter E, Karatosun V, Günal I. Do orthopaedic journals provide high-quality evidence for clinical practice? *Arch Orthop Trauma Surg* 2003;123:82–5.
31. Parsons NR, Hiskens R, Price CL, *et al*. A systematic survey of the quality of research reporting in general orthopaedic journals. *J Bone Joint Surg Br* 2011;93:1154–9.
32. Nuovo J, Melnikow J, Chang D. Reporting number needed to treat and absolute risk reduction in randomized controlled trials. *JAMA* 2002;287:2813–4.
33. Barry HC, Ebell MH, Shaughnessy AF, *et al*. Family physicians' use of medical abstracts to guide decision making: style or substance? *J Am Board Fam Pract* 2001;14:437–42.
34. Saint S, Christakis DA, Saha S, *et al*. Journal reading habits of internists. *J Gen Intern Med* 2000;15:881–4.
35. Okan Y, Garcia-Retamero R, Galesic M, *et al*. When higher bars are not larger quantities: on individual differences in the use of spatial information in graph comprehension. *Spatial Cognition & Computation* 2012;12:195–218.
36. Woller-Carter MM, Okan Y, Cokely ET, *et al*. Communicating and distorting risks with graphs: an eye-tracking study. *Proc Hum Factors Ergon Soc Annu Meet* 2012;56:1723–7.
37. Cooper RJ, Schriger DL, Wallace RC, *et al*. The quantity and quality of scientific graphs in pharmaceutical advertisements. *J Gen Intern Med* 2003;18:294–7.