Misleading reporting and interpretation of results in major infertility journals

Demian Glujovsky, M.D., M.S.C.,[a,b] Carlos E. Sueldo, M.D.,[b,c] Carolina Borghi, M.D.,[b] Pamela Nicotra, M.D.,[b] Sara Andreucci, M.D.,[b] and Agustín Ciapponi, M.D., M.S.C.[a]

[a] Argentine Cochrane Center, Institute for Clinical Effectiveness and Health Policy; and [b] Center for Studies in Genetics and Reproduction, Buenos Aires, Argentina; and [c] Department of Obstetrics and Gynecology, University of California San Francisco-Fresno, San Francisco, California

Objective: To evaluate the proportion of randomized controlled trials (RCTs) published in top infertility journals indexed on PubMed that reported their results with proper effect estimates and their precision estimation, while correctly interpreting both measures.
Design: Cross-sectional study evaluating all the RCTs published in top infertility journals during 2014.
Setting: Not applicable.
Patient(s): Not applicable.
Intervention(s): Not applicable.
Main Outcome Measure(s): Proportion of RCTs that reported both relative and absolute effect size measures and its precision.
Result(s): Among the 32 RCTs published in 2014 in the top infertility journals reviewed, 37.5% (95% confidence interval [CI], 21.1–56.3) did not mention in their abstracts whether the difference among the study arms was statistically or clinically significant, and only 6.3% (95% CI, 0.8–20.8) used a CI of the absolute difference. Similarly, in the results section, these elements were observed in 28.2% (95% CI, 13.7–46.7) and 15.6% (95% CI, 5.3–32.8), respectively. Only one study clearly expressed the minimal clinically important difference in their methods section, but we found related proxies in 53% (95% CI, 34.7–70.9). None of the studies used CIs to draw conclusions about the clinical or statistical significance. We found 13 studies where the interpretation of the findings could be misleading.
Conclusion(s): Recommended reporting items are underused in top infertility journals, which could lead to misleading interpretations. Authors, reviewers, and editorial boards should emphasize their use to improve reporting quality. (Fertil Steril® 2016;■:■–■. ©2016 by American Society for Reproductive Medicine.)
Key Words: Reporting quality, confidence intervals, *P* value, absolute difference, minimal clinically important difference

Discuss: You can discuss this article with its authors and with other ASRM members at http://fertstertforum.com/glujovskyd-misleading-outcome-reporting-infertility/

Use your smartphone to scan this QR code and connect to the discussion forum for this article now.*

* Download a free QR code scanner by searching for "QR scanner" in your smartphone's app store or app marketplace.

To correctly interpret the findings in a given study, authors should use the most relevant measures to report them. Although there are several items recommended by reporting guidelines (1–3), some of them may specifically highly impact key messages for the reader. Omitting these items could generate an incomplete or distorted overview of the clinical scenario.

A very popular statistical element is the *P* value. The *P* value divides statistically significant associations from those that are not; however, overall it provides scarce information (4). The usual cutoff at .05 means that the probability of having a random error in a specific association between an independent variable and an outcome is at least 5%. In other words, a small *P* value indicates that the observed effect is very unlikely to be generated purely by chance. Although its meaning is important, when it appears by itself, it does not show the association's strength, direction, or imprecision of the measure. Besides, sometimes readers arrive at the wrong conclusion when they see a *P* value greater than .05, as they confuse "no evidence of association" with "evidence of no association," which could be a type II error. In 2014, Hilton published an editorial showing the problem of a *P* value cutoff at .05, if the confidence interval (CI) is not considered (5). He remarked on the importance of having a threshold for what we considered to be an important effect, often called "minimal clinically important difference" (MCID), and

Received November 3, 2015; revised December 16, 2015; accepted December 21, 2015.
D.G. has nothing to disclose. C.E.S. has nothing to disclose. C.B. has nothing to disclose. P.N. has nothing to disclose. S.A. has nothing to disclose. A.C. has nothing to disclose.
Reprint requests: Demian Glujovsky, M.D., M.S.C., Center for Studies in Genetics and Reproduction, Viamonte 1432, Buenos Aires (C1055ABB), Argentina (E-mail: glujovsky@cegyr.com).

Fertility and Sterility® Vol. ■, No. ■, ■ 2016 0015-0282/$36.00

http://dx.doi.org/10.1016/j.fertnstert.2015.12.134

checking whether the CI approaches this value or not. Although the CONSORT statement failed to recommend specifically that authors discuss the clinical importance of their results, it is a key concept for sample size calculations of clinical trials in order for clinical trials to have the best chance of detecting clinically important effect sizes. Therefore, the MCID is also a key concept in the interpretation of clinical trial results (6).

As we previously mentioned, the CI is another statistical measure that provides critical information but is often not reported. CIs not only identify statistical significance if one exists (when the interval touches or not the null effect point), but they also add important information about imprecision and effect direction (7–9).

However, to appreciate the clinical effect, we need to measure the effect size. Both the relative and absolute measures are useful. In binary outcomes, some of the most commonly used relative effect measures include the relative risk or risk ratio (RR), the odds ratio (OR), and the hazard ratio (HR), which indicate how many times more or less frequent is one event in the intervention group in comparison with a control group. Less popular, complementary measures are the absolute risk reduction (ARR), which shows the absolute difference of the effect, and the number needed to treat (NNT), which is the inverse of ARR. Numeracy has clinical implications, therefore, having both relative and absolute size effects helps to fulfill the results dashboard and assists clinicians in better decision making. There are a lot of data describing the best way to report results, and although there is no general consensus about the effectiveness of each measure, most investigators agree that both relative and absolute measures are needed (2).

The objective of our study is to evaluate the proportion of randomized controlled trials (RCTs) published in top infertility journals indexed in PubMed that reported their results with proper effect estimates and precision evaluation, while correctly interpreting both measures.

MATERIALS AND METHODS

In this cross-sectional study, we reported using the STROBE statement (10). In January 2015, we ran a search strategy to retrieve all potential RCTs published in three major infertility journals (*Fertility and Sterility*, *Human Reproduction*, and *Reproductive Biomedicine Online*) that publish clinical studies with the highest impact factors, according to the 2014 impact factor (Institute for Scientific Information) and H index (from SciMagO) (11–13).

As performed in our previous studies about quality research and reporting quality (14, 15), we ran an initial search on PubMed using the following strategy: limits, type of article: Randomized Controlled Trial, year: 2014. We analyzed each potential retrieved RCT by using pairs of independent reviewers, who evaluated the titles and abstracts of identified articles, according to prespecified criteria, using EROS software (16). Next, two randomly selected independent reviewers (out of D.G., C.B., P.N., S.A., and A.C.) assessed potentially eligible studies to finally include them in the analysis and to perform the data extraction. Discrepancies were resolved by consensus.

Additionally, we present a descriptive analysis of the results interpretation in the Discussion and Conclusion sections.

We analyzed separately in the abstract and in the full text whether the authors mentioned the *P* value and a CI for the main outcome. For binary outcomes, we evaluated whether any relative measures (i.e., RR, OR) or any absolute measures (i.e., ARR) were used. For continuous measures, we evaluated whether the mean difference and its CI were used. We analyzed whether the MCID or other proxy, such as the expected difference used for the sample size calculation, were mentioned in Materials and Methods. We also evaluated in Discussion and Conclusion whether the interpretation that the authors arrived at was based on the results published.

We used proportion and 95% CI to describe each of the evaluated parameters.

RESULTS

Of the 58 studies published in 2014 from the above-mentioned journals that were found in our search strategy, 18 were excluded by title and abstract evaluation and eight more were excluded by full-text assessment because they were not RCTs. We finally included 32 studies.

In the abstracts, which were structured in 84% of the cases, 12 out of 32 (37.5%, 95% CI, 21.1–56.3) did not mention whether the difference found between the intervention group and the comparison group was statistical or clinically significant (see Table 1). Among the other 62.5% that found a statistically or clinically significant difference, one fifth expressed this concept using a *P* value, one fifth used only CIs, and the rest used the words "significant" or "nonsignificant."

In the abstract, imprecision of the effect estimate of the main outcome was reported by nine of the 32 studies (28.1%; 95% CI, 13.7–46.7) using a CI, but only two of the 32 studies (6.3%; 95% CI, 0.8–20.8) used a CI of the absolute observed difference, among the trial arms.

Finally, also in the abstract, main outcomes were displayed with relative measures (RR or OR) in four of the 32 studies (12.5%; 95% CI, 3.5–29.0), and with absolute risk differences in two of the 32 studies (6.3%; 95% CI, 0.8–20.8). In all cases, authors who used RR, OR, or absolute risks used CIs too.

TABLE 1

Proportion of key items reported in the 32 RCTs.

Section	Statistics	n (%, 95% CI)
Abstract	*P* value	20 (62.5, 43.7–78.9)
	Absolute risk differences	2 (6.3, 0.8–20.8)
	95% CI of the absolute difference	2 (6.3, 0.8–20.8)
Methods	MCID	0
	MCID proxies	17 (53, 34.7–70.9)
Results	*P* value	23 (71.8, 53.3–86.3)
	Absolute risk differences	2 (6.3, 0.8–20.8)
	95% CI of the absolute difference	5 (15.6, 5.3–32.8)

Glujovsky. Misleading outcome reporting in infertility. Fertil Steril 2016.

In Materials and Methods, MCID was explicitly expressed in only one case. However, in 17 of the 32 studies (53%; 95% CI, 34.7–70.9), the MCID was inferred from the sample size calculation. MCID was not supported by citations in any case.

In the results section, the *P* value was used in 23 of the 32 studies (71.8%; 95% CI, 53.3–86.3). In 25% of them, this concept was not mentioned in the abstract. Regarding the CI, 28.1% of the studies used it in both the abstract and the results section. Absolute differences were used in five of the 32 studies (15.6%; 95% CI, 5.3–32.8): two were mean differences, and three were ARR. Relative effect estimates were used in 11/22 binary outcomes (50.0%; 95% CI, 28.2–71.8).

In the Discussion and Conclusion sections, none of the studies used CIs to draw conclusions about the clinical and statistical significance. Specifically, there was no mention of the implications of the CI limits. We found that in 13 of the 32 studies (40.6%; 95% CI, 23.7–59.3) the interpretation could be potentially misleading: two with error type I, five with error type II, three without a statistical test, one with a wrong comparator to draw conclusions, and 10 where no CI was used to evaluate whether its limits crossed the MCID (Table 2).

DISCUSSION

This study shows that a large proportion of the RCTs published in major infertility journals not only report incomplete data according to the reporting guidelines (17, 18) but also that some of the reported information is improperly interpreted and could be misleading to readers. In more than one third of the abstracts, the reader cannot identify whether there is a clinical or statistically significant difference among the intervention groups. Some of the authors expressed the relevance of this finding according to subjective interpretations, not considering the clinical and statistical significance of the observed differences, which are essential to interpret results. Only one of every four articles showed a CI, a crucial statistical element to make a correct interpretation of the precision of the central effect estimate. The absence of these intervals does not allow the reader to understand and analyze the potential best and worst scenarios with the observed effect estimates and whether they cross the null effect point, but more importantly, the MCID.

Another relevant problem identified was that very few studies used absolute differences to show results, which is considered to be one of the best measures to interpret differences between interventions (19). A third of those studies that found a statistically significant difference used a *P* value with multiple zeros after the period, without using the effect estimates; this could mislead the reader to think that the observed difference is large (while the correct interpretation is that the observed difference has a very low probability of happening by chance). Finally, in the Conclusion section, none of the authors used the CIs and the MCID to discuss the observed results. The conclusion was not obtained as a function of clinically significant results but was based only on the presence or absence of statistically significant differences; in some cases, authors used the word "trend" in the discussion, favoring only one direction (i.e., describing trends favoring an intervention but not the trends favoring the control), omitting the fact that the real result could be in the opposite direction, considering the limits of the CIs.

The most important limitations of our study are the small number of RCTs evaluated and that they were all published in just one year. Although reporting quality in other journals or years could be different, we assume that the most recent RCTs published in journals of reproductive medicine with the highest impact factor (IF) index are those that should have more control of their reporting format, and one could extrapolate that reporting is not any better in other years or similar journals. The large CIs in our results show high imprecision in our observations, nevertheless, they are worrisome enough in both limits of CIs to be called to the attention of readers, authors, and editorial boards. We cannot state with great precision that RCT reports are incorrect in many aspects, but our study shows that we should pay more attention to the writing and editorial process to assure a correct interpretation of the research findings. However, a strength of this study is the methodology that we used, with two independent reviewers analyzing in depth the results and interpretations and solving discrepancies by consensus.

Zipkin et al. showed in a systematic review that the use of absolute risks improves the understanding and maximizes accuracy (20). Fagerlin et al. published a review where the evidence suggested that using RRR makes risks seem larger than what they really are (21). Akl et al. published another systematic review where RRR and ARR are equally well understood, but RRR is perceived to be larger and more persuasive than ARR; some studies showed that NNT does not increase the understanding of the results (22). Ahmed et al. concluded that ARR is a more balanced and understandable representation of risk reduction for patients and clinicians than RRR (19). Finally, Johnston et al. recently published a randomized survey where clinicians better understood risk difference than some other formats of presenting outcomes (23). In other health specialties, a similar approach has been made; Kloukos et al. found that only 14% of the journals in prosthodontics and implant dentistry reported CIs (24). Although we found some more frequent use of CIs, we should remark that we only evaluated RCTs, while Kloukos et al. also evaluated other study designs. In a previous study published by our group, we showed that in infertility journals, 42% of the RCTs did not mention the effect size and its precision, and the relative and absolute effect size were missing in 72% (15).

The pitfalls described in this study have relevant implications when readers try to understand the content of RCTs. An in-depth reading of the methods and results could help readers avoid the subjectivity that authors insert in the rest of the manuscript. Optimal results reporting helps readers reach a correct interpretation. Conclusions based on the results, to answer the objectives established, are a must in any publication, and their absence could be misleading.

In the abstract, which is the only segment read by most people (25), it is important for the reader to clearly identify whether the observed difference is statistically and/or clinically significant, which could be expressed using a CI and accompanied by a *P* value. When the author fails to discuss

TABLE 2

Potentially misleading interpretations.

Flaw	Study	Reported results	Reported interpretation	Potential weakness
Type II error and absence or no interpretation of CI	1	RR 0.90 (95% CI, 0.39–2.1)	INT 1 does not affect OUTCOME A	A $P>.05$ does not discard a potential not observed difference
	2	OR 0.56 (95% CI, 0.22–1.35)	INT 1 does not result in a significant reduction of OUTCOME A	
	3	OR 1.04 (95% CI, 0.82–1.33)	OUTCOME A comparable between POP A and POP B	
	4	RR 1.13 (95% CI, 0.93–1.38)	INT 1 does not improve OUTCOME A	
	5	There were no statistically significant differences between the groups	INT 1 does not significantly improve OUTCOME A	
Type I error	6	The trend persisted in CONDITION 1 ($P=.08$)	Our data suggest a beneficial role for CONDITION 1 in OUTCOME A	A $P>.05$ would not discard an observed difference by chance
	7	INT 1: 56 events INT 2: 36 events INT 3: 44 events INT 4: 30 events Unadjusted $P<.05$ Adjusted $P>.05$[a]	Important difference in OUTCOME A	
No test was used	8	INT 1: 24% INT 2: 21%	In POPULATION X, INT 1 was equally effective as INT 2	There is no test (P value), but even a $P>.05$ would not discard a potential not observed difference (type II error)
	9	POP X (INT 1): 43.4% POP Y (INT 1): 38.8%	OUTCOME was comparable in POPULATION X and POPULATION Y for INT 1	
	10	INT 1: 87% INT 2: 94% INT 3: 97% INT 1 vs. 3: $P<.05$ INT 1 vs. 2: $P<.05$ INT 2 vs. 3: $P=$ NS	There was a higher trend towards OUTCOME A with INT 1, 2, and 3	There is no a test to evaluate a trend
No absolute risks	11	INT 1: HR 1.09 (95% CI, 1.03–1.15) INT 2: HR 1.00 (95% CI, 0.95–1.05)	Better OUTCOME A with INT 1 in SUBPOP A	No absolute estimates and absolute differences are reported No mention of the INT 2 results. Both positive and negative findings should be reported
Wrong comparison	12	COMPARATOR demonstrated lower values than INT 1. But the INT 1 arm did not differ from PLACEBO	In the article title: INT 1 stimulates OUTCOME A	The title should explain that COMPARATOR is deleterious in comparison to INT 1 or PLACEBO, rather than saying that INT 1 is beneficial (because it is not better than PLACEBO)
No CI of the absolute effect size	13	INT 1: 4.4 ± 3.0 INT 2: 2.7 ± 1.6 $P<.01$ MCID: 3 to 5	OUTCOME A was higher with INT 1 compared to INT 2	There is no CI of the absolute difference allowing the appreciating of the clinical impact of its limits. A CI that crosses both the harm and beneficial MCID does not support a definitive conclusion
	1–5,8–11	–	–	

Note: INT = intervention.

[a] Adjusted P values are used to correct for multiple testing. Therefore, conclusions should be based on them.

Glujovsky. Misleading outcome reporting in infertility. Fertil Steril 2016.

TABLE 3

Key messages.

What we found	Suggestions to be used
No mention of clinical or statistical significance	*P* value, CIs, phrases referring to both statistical and clinical significance
Just *P* value	Both *P* value and CI (or just the CI)
RR, OR, or HR	Both relative measures and ARR (or just ARR)
CI of the absolute risk	CI of the absolute difference
Only interpretation of the average estimated point	Interpreting CI assuming the best and worst scenario and mentioning if there is high or low imprecision
No mention of MCID	The evidence or at least the justification supporting the MCID
Conclusions are drawn considering the observed difference and/or if there was any statistically significant difference	CIs and clinically significant differences should also be considered for the interpretations of the results

Glujovsky. Misleading outcome reporting in infertility. Fertil Steril 2016.

clinically relevant differences, any statistically significant difference could make the reader think that the intervention is beneficial, even when the real improvement is clinically irrelevant. The use of the *P* value alone is insufficient for a complete understanding by the reader, and CIs add relevant information to help the reader fully appreciate the message of any study. They should not be included only in the results but also analyzed in the conclusions. The limits of the CIs are relevant to learn, not only whether they cross the null effect line, but also whether they cross the MCID. When a CI is entirely over the MCID, the conclusions are definitive, while if the CI crosses the MCID but not the null effect line, we should then be more cautious, as the difference may be statistically significant but not clinically relevant. Regarding the effect estimates, when using just a relative effect measure, the reader may overestimate the real impact of the intervention, which could be very small in absolute terms. Therefore, absolute risks and absolute differences with their CIs are important (see Table 3). Authors sometimes do not use the most relevant statistics, which could be a source of several biases, or they may have conflicts of interests that could push them to commit reporting biases (26, 27).

Another issue is about the wording in interpreting results. It is important to use a consistent language, such as the one described in Table 4 (adapted from suggestions for Cochrane plain language summaries). Although it was conceived to describe a body of evidence, it could be applied to a single study as a way to avoid overestimating or underestimating the conclusions.

This study shows that although reporting guidelines are widely available, and their use is suggested by editorial boards, authors are not accustomed to use them by following a checklist, and reviewers do not emphasize enough their utilization. CONSORT guidelines state in item 17 that both relative and absolute effect size, as well as their precision, should be reported. If we want our readers to correctly interpret the articles published, then authors, peer reviewers, and editorial boards should continue to emphasize reporting in the abstract relevant information, using CIs and absolute differences, as part of a long list of improvements (15). Mandatory checklists for authors and reviewers could have a major role in changing the reporting culture and helping to improve the correct interpretation of the RCTs results.

TABLE 4

Matrix for reporting results according to the guidelines for preparing SUPPORT Summaries (27).

Quality or certainty of the evidence	Magnitude of the effect size		
	Large	Small (may not be important)	Little or no difference
High	Improves/decreases/prevents/leads to [outcome]	Improves slightly/decreases slightly/ leads to slightly fewer (more) [outcome]	Results in little or no difference in [outcome]
Moderate	Probably improves/decreases/prevents/ leads to [outcome]	Probably improves slightly/decreases slightly/leads to slightly fewer (more) [outcome]	Probably leads to little or no difference in [outcome]
Low	May improve/decrease/prevent/lead to [outcome]	May slightly improve/slightly decrease/ lead to slightly fewer (more) [outcome]	May lead to little or no difference in [outcome]
Very low	It is uncertain whether [intervention] improves, decreases, prevents, leads to [outcome] because the quality of the evidence is very low		
No data	[Outcome] was not measured		

Note: Adapted from the guidelines for preparing SUPPORT summaries www.supportsummmaries.org, March 18, 2013. To grade the quality of evidence consider [1] limitations in the design (according to risk of bias); [2] indirectness of evidence (no direct evidence about population, intervention, comparison, or outcome); [3] imprecision of results (wide CIs crossing not only the null effect but also the MCID).

Glujovsky. Misleading outcome reporting in infertility. Fertil Steril 2016.

REFERENCES

1. Moher D, Hopewell S, Schulz KF, Montori V, Gotzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. Br Med J 2010;340:c869.
2. Harbin Consensus Conference Workshop G. Improving the Reporting of Clinical Trials of Infertility Treatments (IMPRINT): modifying the CONSORT statement. Fertil Steril 2014;102: 952–9. e15.
3. Harbin Consensus Conference Workshop G, Conference Chairs, Scientific Committee, Legro RS, Wu X, Barnhart KT, et al. Improving the Reporting of Clinical Trials of Infertility Treatments (IMPRINT): modifying the CONSORT statement. Hum Reprod 2014;29:2075–82.
4. The Cochrane Collaboration. 12.4.2 P values and statistical significance. In: Higgins J, Green S, eds. Cochrane handbook for systematic reviews of interventions. Version 5.1.0. The Cochrane Collaboration, 2011.
5. Hilton J. Results should not be reported as statistically significant or statistically non-significant. In: Cochrane Editorial Unit, 2014.
6. Chan KB, Man-Son-Hing M, Molnar FJ, Laupacis A. How well is the clinical importance of study results reported? An assessment of randomized controlled trials. CMAJ 2001;165:1197–202.
7. Higgins J, Green S. Cochrane handbook for systematic reviews of interventions version 5.1.0. The Cochrane Collaboration. In: Confidence intervals, 2011.
8. du Prel JB, Hommel G, Rohrig B, Blettner M. Confidence interval or P value?: part 4 of a series on evaluation of scientific publications. Dtsch Arztebl Int 2009;106:335–9.
9. Ciapponi A. Clinical significance or statistical significance? EVIDENCIA—Actualización en la Práctica Ambulatoria 2013;16:122–5.
10. von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Lancet 2007;370:1453–7.
11. Web of Science. Journal impact factors. http://wokinfo.com/products_tools/analytical/jcr/. Accessed January 9, 2015.
12. Thomson Reuters. Introducing the impact factor. http://thomsonreuters.com/en/articles/2013/introducing-the-2013-edition-journal-citation-reports.html. Accessed January 10, 2015.
13. Reveiz L, Cortes-Jofre M, Asenjo Lobos C, Nicita G, Ciapponi A, Garcia-Dieguez M, et al. Influence of trial registration on reporting quality of randomized trials: study from highest ranked journals. J Clin Epidemiol 2010;63:1216–22.
14. Glujovsky D, Riestra B, Coscia A, Boggino C, Comande D, Ciapponi A. Assessment of research quality in major infertility journals. Fertil Steril 2012;98:1539–43.
15. Glujovsky D, Boggino C, Riestra B, Coscia A, Sueldo CE, Ciapponi A. Quality of reporting in infertility journals. Fertil Steril 2015;103:236–41.
16. Glujovsky D, Bardach A, García Martí S, Comande D, Ciapponi A. New software for early stage of systematic reviews. In: XVIII Cochrane Colloquium: The Joint Colloquium of the Cochrane & Campbell Collaborations. Colorado: Keystone Resort, 2010.
17. Equator network. Enhancing the quality and transparency of health research. Available at: http://www.equator-network.org. Accessed December 1, 2015.
18. Consolidated Standards of Reporting Trials. CONSORT. Available at: http://www.consort-statement.org. Accessed December 1, 2015.
19. Ahmed H, Naik G, Willoughby H, Edwards AG. Communicating risk. Br Med J 2012;344:e3996.
20. Zipkin DA, Umscheid CA, Keating NL, Allen E, Aung K, Beyth R, et al. Evidence-based risk communication: a systematic review. Ann Intern Med 2014;161:270–80.
21. Fagerlin A, Zikmund-Fisher BJ, Ubel PA. Helping patients decide: ten steps to better risk communication. J Natl Cancer Inst 2011;103:1436–43.
22. Akl EA, Oxman AD, Herrin J, Vist GE, Terrenato I, Sperati F, et al. Using alternative statistical formats for presenting risks and risk reductions. Cochrane Database Syst Rev 2011:CD006776.
23. Johnston BC, Alonso-Coello P, Friedrich JO, Mustafa RA, Tikkinen KA, Neumann I, et al. Do clinicians understand the size of treatment effects? A randomized survey across 8 countries. CMAJ 2016;188:25–32.
24. Kloukos D, Papageorgiou SN, Fleming PS, Petridis H, Pandis N. Reporting of statistical results in prosthodontic and implantology journals: P values or confidence intervals? Intl J Prosthodont 2014;27:427–32.
25. Gotzsche PC. Readers as research detectives. Trials 2009;10:2.
26. Ciapponi A, Reveiz L, Rada G, Glujovsky D, García MS, Rubinstein F, et al. How evidence-based are the pharmaceutical industry's printed promotional material presented to physicians: a multinational study. In: XVIII Cochrane Colloquium: The Joint Colloquium of the Cochrane & Campbell Collaborations. Colorado: Keystone Resort, 2010.
27. Ciapponi A, Reveiz L, Rada G, Glujovsky D, García MS, Rey Ares L, et al. How evidence-based are the pharmaceutical industry's printed promotional material presented to physicians in three Latin American countries. In: XIX Cochrane Colloquium Scientific Evidence for Healthcare Quality and Patient Safety. Madrid, 2011.