

FOCUS ARTICLE

Prediction intervals should be included in meta-analyses published in dentistry

Faggion CM Jr¹  | Max Clemens Menne¹  | Nikolaos Pandis² 

¹ Department of Periodontology and Operative Dentistry, Faculty of Dentistry, University Hospital Münster, Münster, Germany

² Department of Orthodontics and Dentofacial Orthopedics, Dental School/Medical Faculty, University of Bern, Bern, Switzerland

Correspondence

Faggion CM Jr, Department of Periodontology and Operative Dentistry, Faculty of Dentistry, University Hospital Münster, Waldeyerstraße 30, Münster 48149, Germany.

Email: clovisfaggion@yahoo.com; Clovis.Faggion@ukmuenster.de

Abstract

This focus article aims to highlight the value of reporting prediction intervals (PIs) in random effects meta-analysis and to assess the prevalence of PI-reporting in periodontology and implant dentistry meta-analyses. We searched in the PubMed database for meta-analyses published in the fields of periodontology and implant dentistry. We selected meta-analyses related to primary outcomes with at least three trials. Additionally, we extracted information on the type of the meta-analysis model (fixed or random) and whether the random effects meta-analyses included PIs in addition to the 95% confidence intervals. Three-hundred and forty-nine meta-analyses were found in 94 systematic reviews. Two-hundred and sixty-three (75.4%) subgroup and full meta-analyses used the random-effects model, 81 (23.2%) used fixed-effect methods, and 5 (1.4%) did not specify the model used. In 75 systematic reviews, we found 231 meta-analyses with three or more trials (173 full meta-analyses and 58 subgroup meta-analyses). Only one systematic review reported PIs. Interpretation of the results of random effects meta-analyses which ignore heterogeneity can be misleading. Heterogeneity should be explored, and two common approaches include subgroup analyses and meta-regression. Random effects meta-analyses should include PIs because they convey the extent of heterogeneity in treatment effects across studies in a clinically relevant context.

KEYWORDS

confidence intervals, meta-analysis, methods, prediction intervals, research design

INTRODUCTION

Systematic reviews command the highest level of the evidence pyramid because they aim, ideally, to combine the totality of the high-quality evidence in a particular field. The combination of the evidence, where applicable, is expressed via the mathematical synthesis, termed meta-analysis, of the results of the included individual trials. The two most common meta-analyses models are the fixed effect and the random effects models [1].

The fixed effect model assumes that there is a common effect with no heterogeneity between studies and only within-study variability due to random error, whereas the random effects model assumes a distribution of effects and consequently between-study heterogeneity in the treatment effects. The focus of the interpretation in a meta-analysis is the pooled estimate, the corresponding confidence interval, the *p*-value and for the random effects model in addition the measures of heterogeneity such as I^2 (the proportion of the observed variance attributed to the true effects rather than to sampling

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2021 The Authors. *European Journal of Oral Sciences* published by John Wiley & Sons Ltd on behalf of Scandinavian Division of the International Association for Dental Research.

error) and τ^2 (defined as the between-study variance or the variance of the distribution of the effect sizes across the population of studies). The I^2 has been often reported as a between-study variance metric; however, this interpretation is incorrect [2].

The pooled estimate in fixed effect and random effects meta-analysis represents different metrics, and the pooled estimate in random effects meta-analyses is often misinterpreted as the overall effect, just as in the fixed effect model. This naive interpretation ignores the fact that a single parameter cannot adequately summarize heterogeneous effects [3], and so this approach is usually insufficient. This problem is nicely illustrated by Higgins et al. [4], where the interpretation of two meta-analyses that produced the same pooled estimates and were interpreted using only the pooled estimate failed to highlight important differences between the two datasets that could radically alter the conclusions. Furthermore, the I^2 and τ^2 random effects metrics are difficult to interpret in a clinical context. The interpretation of I^2 is not straightforward due to its dependency on the sample size; this can result in a low and non-statistically significant I^2 when there are few studies, and vice versa. The τ^2 , although a better measure, is again not easy to interpret because it is not expressed in the same metric as the effect estimates since it is squared. By contrast, the prediction interval (PI) [4] metric captures the between-trial heterogeneity and can be readily interpreted and communicated with the clinician. The PI is defined as the interval within which the effect size of a new study would fall if this study were selected at random from the same population of studies already included in the meta-analysis [5]. The PI can be calculated when there are at least three trials in the meta-analysis, and depending on the extent of heterogeneity, the PI is wider than the 95% CI of the pooled estimate, unless τ^2 is zero, reflecting the added uncertainty of a future trial. A simple $100(1-\alpha)\%$ PI can be calculated as follows [4]:

$$M \pm t_{k-2}^{\alpha} \sqrt{\tau^2 + SE(M)^2}$$

where M is the summary mean (pooled estimate) from the random effects meta-analysis, t_{k-2}^{α} is the $100(1 - \frac{\alpha}{2})\%$ percentile of a *t-distribution* with $k-2$ degrees of freedom, k is the number of studies, $\hat{\tau}^2$ is the between-study heterogeneity, and $SE(M)$, the standard error of the summary mean.

In random effects meta-analysis, the confidence interval quantifies the precision of the mean effect size, and the PI reflects the dispersion of the true effect sizes of the new studies. For example, a 95% confidence interval indicates that, in 95% of the cases, the mean effect size will fall within the random effects meta-analysis diamond, whereas a 95% PI indicates that, in 95% of the cases, the true effect size of a new study will fall within the meta-analysis PI.

Despite the differences between fixed effect and random effects meta-analysis, findings are often interpreted in the same manner [6]. A confidence interval may indicate a clinical effect, which may not corroborate with the corresponding PI. The confidence interval is a measure of the precision of the average effect across the distribution of all effects. The average effect may indicate that the treatment, on average, works. The PI indicates the variability and the range of effects, and it could very well be that some effects are beneficial, while others are trivial or even harmful. Thus, the interpretation of the result from a random effects meta-analysis solely based on the confidence interval may convey an incorrect message about the range of the plausible effects of an intervention.

Despite the advantages of PIs [6,7], they are not often reported for the random effects meta-analyses. To the best of our knowledge, no systematic evaluation of the reporting of PIs in the field of dentistry (and specifically in periodontology and implant dentistry) in random effects meta-analyses has been undertaken. Accordingly, the aim of this focus article was to assess the use of PIs in random effects meta-analyses in periodontology and implant dentistry.

MATERIAL AND METHODS

We included meta-analyses of interventional clinical studies conducted in the fields of periodontology and implant dentistry. Other types of meta-analyses (e.g., on prevalence of disease), network-meta-analyses, meta-analyses of the association of risk factors and periodontal and dental implant outcomes, and meta-analyses from other dental fields were excluded. This sample of meta-analyses is derived from a sample from another study [8] that addressed a different research question.

In the PubMed database, we searched for meta-analyses published between August 2015 and August 2020 by using specific keywords (see Table S1 in the Supporting Information). We selected the forest plots directly related to the primary outcomes. In the event of no report of primary outcomes, we selected the forest plots with the greatest number of primary studies. If multiple examples of the selected forest plots included the same number of studies, we included all of them. Forest plots of meta-analyses were identified, and only meta-analyses using the random-effects approach with at least three primary studies were selected. From the selected forest plot, we recorded the reporting or not reporting of PIs and whether authors interpreted the meta-analyses results based on the PIs.

RESULTS

The search in PubMed initially generated 282 potential reviews; after screening and full-text assessment, 94

TABLE 1 Outcomes with corresponding 95% confidence and predictions intervals^a

Outcome	Number of studies	95% Confidence interval	<i>p</i> -value	95% Prediction interval
Marginal bone level	11	1.27, 2.71	<0.05	-0.4, 4.4
Clinical attachment level gain	8	1.32, 1.22	<0.05	0.6, 3.0
Recession level	10	-0.97, -0.33	<0.05	-1.7, 0.4
Pocket depth reduction	21	2.32, 3.37	<0.05	0.4, 5.3
Bleeding on probing at implants	5	0.23, 0.75	<0.05	0.1, 2.3
Bleeding on probing at implant sites	18	0.15, 0.37	<0.05	0.0, 1.7

^aData extracted from Tomasi et al. [9]

systematic reviews produced 349 meta-analyses. Figure S1 reports the search and selection of studies. The median number of trials included in the random-effects meta-analyses was 5 (range 2–31). Of the 349 meta-analyses, 263 (75.4%) belonging to 75 systematic reviews used the random-effects model and 81 (23.2%) used the fixed-effect model. Two systematic reviews produced five (1.4%) meta-analyses not reporting the meta-analysis model. Seventy-five systematic reviews included 231 meta-analyses with three or more trials (173 full meta-analyses and 58 subgroup meta-analyses, respectively). Only one systematic review [9] reported PIs in one meta-analysis related to the primary outcome (change of marginal bone level [MBL]). That systematic review also reported PIs for the following secondary outcomes: clinical attachment level gain, recession, pocket depth reduction, and bleeding on probing at implants and implant sites. Table 1 presents the outcomes with the corresponding 95% confidence intervals and the 95% PIs. The interpretation of all six outcomes using the confidence interval shows a statistically significant treatment effect, whereas, when using the PI, for four out the six outcomes the results suggest that no effect is also possible in a future trial using this intervention in a similar population mix. For the primary outcome (MBL), the 95% CI (1.27–2.71 mm) suggests (on average) clinically relevant improvements; however, the 95% PI (-0.4 to 4.4 mm) also includes trivial effects or effects that do not preclude reduction in the MBL. In fact, the PIs conveyed a different result from the confidence intervals in four of six meta-analyses. This indicates the uncertainty of the estimates and in the conclusions, given the observed between-study heterogeneity.

DISCUSSION

Our study found that most meta-analyses in periodontology and implant dentistry used the random-effects model, whereas PIs, despite their advantages, are rarely reported. However, calculation of the PIs would temper the conclusions drawn from meta-analyses of most outcomes. This change in interpretation has been reported elsewhere on a large sam-

ple of Cochrane systematic reviews [5]. In that study, the PI indicated that the range of the intervention effect could be null or could even include negative effects in 72.4% of 479 statistically significant (in terms of the average effect) random effects meta-analyses. The oversimplistic interpretation of the random effects meta-analysis based only on the mean of the random-effects distribution can be quite deceptive. Heterogeneity across trials, expressed via τ^2 , communicates the degree of inconsistency among the included trials, which is an integral part of random effects meta-analysis. A key element of random-effects meta-analysis is to provide inferences for trials not included in the meta-analysis (generalization) by considering the uncertainty of treatment effects over different settings.

On the other hand, true heterogeneity in treatment effects can be confounded by bias in the included trials, and in the presence of bias, interpretation of random-effects meta-analysis can be difficult. Meta-analyses with a small number of trials, which is a common theme in oral health [10,11], pose additional problems in precisely estimating heterogeneity and treatment effects. Statistical assessment of heterogeneity in meta-analyses with few trials has very low power and interpretation of heterogeneity based on *p*-values can be misleading. This occurs because *p*-values are very sensitive to sample size. In the same way, a meta-analysis with many trials can have high power and a highly statistically significant result on the heterogeneity estimate when, in fact, true heterogeneity is low [12].

A possible explanation for the low use of PIs in this sample of meta-analyses is lack of awareness [6]. The precision of the estimate derived from the PI is dependent on the number and size of trials included in the meta-analysis [7,13]. The fewer and smaller the studies, the less precise is this estimate. However, if the small number of trials (fewer than three) negates the use of PIs, resorting to the naive interpretation of the random-effects meta-analytic estimate is not appropriate, as explained earlier. Accordingly, despite the imprecision of the PI due to the limited number of trials, its use and appropriate interpretation should be encouraged.

Although our assessment was focused on two specific fields, we believe that there is no evidence to suggest that this scenario is different in other dental fields. We examined forest plots related only to primary outcomes; however, we do not expect that forest plots for any other outcomes in the same meta-analysis will include PIs.

In conclusion, we find that PIs are important for the proper interpretation of treatment effects estimated from random-effects meta-analyses. Unfortunately, PIs are rarely provided in random-effects meta-analyses in periodontology and implant dentistry. Journals should include recommendations for authors submitting systematic reviews with random-effects meta-analyses to report PIs together with confidence intervals.

ACKNOWLEDGMENTS

The authors did not receive any financial support for this study.

CONFLICT OF INTEREST

The authors declare no potential conflicts of interest with respect to the authorship and/or publication of this article.

AUTHOR CONTRIBUTIONS

Conceptualization: Nikolaos Pandis and Faggion CM Jr. **Methodology:** Nikolaos Pandis and Faggion CM Jr. **Validation:** Nikolaos Pandis, Max Clemens Menne, and Faggion CM Jr. **Formal analysis:** Nikolaos Pandis. **Investigation:** Nikolaos Pandis, Max Clemens Menne, and Faggion CM Jr. **Writing—original draft preparation:** Faggion CM Jr and Nikolaos Pandis. **Writing—review and editing:** Faggion CM Jr, Nikolaos Pandis, and Max Clemens Menne. **Visualization:** Faggion CM Jr and Nikolaos Pandis. **Supervision:** Nikolaos Pandis.

ORCID

Faggion CM Jr  <https://orcid.org/0000-0001-8323-909X>

Max Clemens Menne  <https://orcid.org/0000-0002-9175-0688>

Nikolaos Pandis  <https://orcid.org/0000-0003-0258-468X>

REFERENCES

1. Borenstein M, Hedges LV, Higgins JP, Rothstein HR. A basic introduction to fixed-effect and random-effects models for meta-analysis. *Res Synth Methods*. 2010;1:97-111.

2. Borenstein M, Higgins JPT, Hedges LV, Rothstein HR. Basics of meta-analysis: I2 is not an absolute measure of heterogeneity. *Res Synth Methods*. 2017;8:5-18.
3. Raudenbush SW, Bryk AS. Empirical Bayes meta-analysis. *J Educ Statist*. 1985;10:75-98.
4. Higgins JP, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. *J R Stat Soc Ser A Stat Soc*. 2009;172:137-59.
5. Ades AE, Lu G, Higgins JP. The interpretation of random-effects meta-analysis in decision models. *Med Decis Making*. 2005;25:646-54.
6. Riley RD, Higgins JPT, Deeks JJ. Interpretation of random effects meta-analyses. *BMJ*. 2011;342:d549. <https://doi.org/10.1136/bmj.d549>
7. Int'Hout J, Ioannidis JPA, Rovers MM, Goeman JJ. Plea for routinely presenting prediction intervals in meta-analysis. *BMJ Open*. 2016;6:e010247. <https://doi.org/10.1136/bmjopen-2015-010247>
8. Menne MC, Seitidis G, Faggion CM Jr, Mavridis D, Pandis N. Early optimistic effect in periodontology and implant dentistry trials. *J Dent Res*. 2021. <https://doi.org/10.1177/00220345211025242>
9. Tomasi C, Regidor E, Ortiz-Vigón A, Derks J. Efficacy of reconstructive surgical therapy at peri-implantitis-related bone defects. A systematic review and meta-analysis. *J Clin Periodontol*. 2019;46:340-56.
10. Koletsi D, Fleming PS, Michelaki I, Pandis N. Heterogeneity in Cochrane and non-Cochrane meta-analyses in orthodontics. *J Dent*. 2018;74:90-4.
11. Saltaji H, Cummings GG, Armijo-Olivo S, Major MP, Amin M, Major PW, et al. A descriptive analysis of oral health systematic reviews published 1991–2012: cross sectional study. *PLoS One*. 2013;8:e74545. <https://doi.org/10.1371/journal.pone.0074545>
12. Higgins JPT, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. *J R Stat Soc Ser A Stat Soc*. 2009;172:137-59.
13. Nagashima K, Noma H, Furukawa TA. Prediction intervals for random-effects meta-analysis: a confidence distribution approach. *Stat Methods Med Res*. 2019;28:1689-702.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Faggion CM Jr, Menne MC, Pandis N. Prediction intervals should be included in meta-analyses published in dentistry. *Eur. J. Oral Sci*. 2021;1–4. <https://doi.org/10.1111/eos.12827>