Received: 11 December 2022 | Accepted: 8 November 2023

DOI: 10.1111/eos.12962

European Journal of Oral Sciences | NOF Scandinavian Division IADR | WILEY

FOCUS ARTICLE

A case study evaluating the effect of clustering, publication bias, and heterogeneity on the meta-analysis estimates in implant dentistry

Clovis Mariano Faggion Jr[1] | **Momen A. Atieh**[2,3,4] | **Michail Tsagris**[5] | **Jadbinder Seehra**[6] | **Nikolaos Pandis**[7]

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

[2]Mohammed bin Rashid University of Medicine and Health Sciences, Hamdan bin Mohammed College of Dental Medicine, Dubai, United Arab Emirates

[3]Sir John Walsh Research Institute, School of Dentistry, University of Otago, Dunedin, New Zealand

[4]School of Dentistry, University of Jordan, Amman, Jordan

[5]Department of Economics, University of Crete, Rethimnon, Greece

[6]Centre for Craniofacial Development & Regeneration, Faculty of Dentistry, Oral & Craniofacial Sciences, King's College London, London, UK

[7]Department of Orthodontics and Dentofacial Orthopedics, Dental School/Medical Faculty, University of Bern, Bern, Switzerland

Correspondence
Clovis Mariano Faggion 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

Meta-analyses may provide imprecise estimates when important meta-analysis parameters are not considered during the synthesis. The aim of this case study was to highlight the influence of meta-analysis parameters that can affect reported estimates using as an example pre-existing meta-analyses on the association between implant survival and sinus membrane perforation. PubMed was searched on 7 July 2021 for meta-analyses comparing implant failure in perforated and non-perforated sinus membranes. Primary studies identified in these meta-analyses were combined in a new random-effects model with odds ratios (ORs), confidence intervals (CIs), and prediction intervals reported. Using this new meta-analysis, further meta-analyses were then undertaken considering the clinical, methodological, and statistical heterogeneity of the primary studies, publication bias, and clustering effects. The meta-analyses with the greatest number and more homogeneous studies provided lower odds of implant failure in non-perforated sites (OR 0.49, 95 % CI = [0.26, 0.92]). However, when considering heterogeneity, publication bias, and clustering (number of implants), the confidence in these results was reduced. Interpretation of estimates reported in systematic reviews can vary depending on the assumptions made in the meta-analysis. Users of these analyses need to carefully consider the impact of heterogeneity, publication bias, and clustering, which can affect the size, direction, and interpretation of the reported estimates.

KEYWORDS

implant survival, meta-analysis, methods, research design, sinus perforation

INTRODUCTION

Systematic reviews with meta-analyses provide a high level of scientific evidence; however the quality of the scientific evidence in the dental literature is low despite the exponential increase in the publication of systematic reviews [1-3]. There is evidence in the literature that meta-analyses can often include heterogeneous studies of varying methodological quality which can have an impact on our confidence in the results [4].

Including heterogeneous studies of high risk of bias in the meta-analysis, neglecting publication bias, and ignoring clustering effects can bias meta-analytical estimates and therefore clinical decisions. Different types of heterogeneity have been described in the literature and they include statistical, methodological, and clinical heterogeneity [5]. Clinical heterogeneity refers to differences in the characteristics of participants, interventions, and outcomes. Methodological heterogeneity refers to the different designs of the studies, and statistical heterogeneity refers to the dissimilarity of the individual study estimates and their range. Often heterogeneity is measured statistically and interpreted, incorrectly, using the I^2 value. When these values are low, some consider that there is no apparent heterogeneity among trials, which would suggest that combining individual studies is reasonable and that the meta-analytic estimate can be trusted [6]. However, low I^2 values do not give an indication of the level of methodological and clinical heterogeneity, and following only statistical parameters to determine the robustness of a meta-analysis might be misleading [6, 7]. Biased results may also occur in the presence of publication bias [8], which is defined as the publication of studies based on the size and direction of the effect. In essence, publication bias means that from the population of all conducted studies, only a subsample with statistically significant results is published, whereas studies with smaller or non-significant effects are not published, biasing the meta-analytical estimates.

Interpretation of meta-analysis estimates can also be influenced by multiple measurements associated with clustering, a term used to describe aggregates of individuals, or a collection of multiple measurements taken from the same person, such as implants in the same jaw. When multiple implants are used within the same patient the outcomes are not independent and clustering effects arise, which reduces the effective sample size of a study [9]. Treating those dependent within-patient outcomes as independent falsely increases precision and reduces *p*-values, and the *p*-values may not be correct and can lead to misinterpretation of study findings depending on whether the *p*-value crosses the threshold of statistical significance [10]. Interpretation based solely on *p*-values is problematic [11, 12], and at the trial level is compounded in the presence of clustering effects. The correlations arising from the multiple implants may be measured using the intra-cluster correlation coefficient (ICC), which ranges from 0 to 1. An ICC value of 0 implies independence and an ICC value of 1 implies perfect correlation within patients [13]. More specifically, for ICC = 0 the multiple measurements within a cluster are effectively independent and each one of the multiple measurements contributes just as much information as if each one of the within-cluster multiple measurements were taken from different patients without loss of information. By contrast, for ICC = 1, the multiple within-cluster measurements are equivalent to a single measurement resulting in a substantial loss of information. The decrease in the effective sample size in clustered designs can be determined by the design effect, which is related to the ICC according to the formula $D = 1 + (m - 1)\cdot r$, where m is the number of implants per patient and r = ICC. Please note that the above formula assumes constant cluster size, which, however, is not always the case. From the design effect formula, higher ICC values and larger clusters necessitate an increase in the required sample size in a clustered design to maintain the desired precision and power. In the context of meta-analysis, the inclusion of studies with multiple observations can potentially have a similar effect by producing (too) small *p*-values for the pooled effect, which is not genuine.

An area of particular interest in implant dentistry is when implant insertion must be combined with sinus augmentation, with either autogenous bone or bone substitutes which are inserted into the maxillary sinus through a lateral window [14]. This procedure requires a degree of operator skill if the most frequent complication of perforation of the Schneiderian membrane is to be avoided [15]. The placement of multiple implants in a patient's jaw or quadrant following sinus augmentation forms a cluster.

Despite the possible issues, the main question relevant for any meta-analysis is whether the impact of bias is trivial, modest, or substantial and whether meta-analysis findings are robust. Thus, the aim of this case study was to demonstrate, by re-evaluating pre-existing meta-analyses related to a specific clinical question (sinus membrane perforation and implant survival), the impact of a number of parametrers. In particular, we evaluate the influence of clinical, methodological, and statistical heterogeneity of the primary studies, publication bias, and the effect of clustering (multiple implants within patients) on the reported estimates and their precision, and hence on the clinical interpretation of the reported findings.

MATERIAL AND METHODS

Search strategy and data selection

Two authors (CMF, MAA) independently searched for meta-analyses of studies assessing implant survival in

perforated and non-perforated maxillary membrane sinuses. The unit of analysis was meta-analyses included in systematic reviews. Therefore, we focused on the search of systematic reviews and excluded any other primary studies such as randomized and non-randomized clinical studies. We used a pre-defined search strategy applied to the MEDLINE via PubMed database to identify potential meta-analyses (Table S1). The search was conducted on 7 July 2021 and included articles published from database inception to July 2021. We screened the reference lists of meta-analyses retrieved from the electronic search for potentially relevant meta-analyses. The selection of meta-analyses was based on the pre-defined research question: P = human participants, I = implant placement in perforated membrane sinuses, C = implant placement in non-perforated membrane sinuses, and O = implant failure. Implant failure in the context of the present work means implant loss.

The PICO (patient, intervention, comparison, and outcome) concept is widely used for the planning and assessing of methodological quality for systematic reviews [16, 17]. Any disagreement in the inclusion/exclusion of studies was discussed between the two authors until consensus was achieved. It should be appreciated that the aim of this study was not to undertake a new systematic review with a quantitative analysis on this clinical question. With this in mind, the search was limited to a single electronic database, the grey literature was not searched, reporting was not undertaken in relation to PRISMA, and the protocol was not registered. Furthermore, we did not update the search for relevant studies as the premise of this study was to identify existing meta-analyses meeting the pre-defined question and using data extracted from these studies to inform the planned re-analysis, rather than generating updated estimates on the effectiveness of the interventions.

Data extraction and analysis

All primary studies included in the forest plots were further evaluated. Subsequently, a new meta-analysis using the identified primary studies and three additional simulation meta-analyses, based on clinical homogeneity and study design, were conducted. To assess the degree of homogeneity of the clinical outcomes (clinical heterogeneity) and to determine if it was appropriate for those studies to be combined in a meta-analysis [18], the primary studies were evaluated by two clinical assessors who are specialists in periodontology and experienced in dental implant therapy (CMF, MAA). Because there is no clear methodology to evaluate clinical heterogeneity among studies, we applied the PICO concept [17]. We assessed each study and conducted intensive discussions on how the studies differ from the PICO perspective before a consensus was reached on the studies that should be included in the new meta-analyses.

Methodological heterogeneity may arise as studies with different designs might influence the size of the meta-analytic estimate [5, 19, 20]. To account for this, the design of the primary study was assessed and classified as retrospective or prospective. The methodological quality of the systematic reviews that included the three meta-analyses was also assessed using the AMSTAR-2 checklist [16]. The AMSTAR-2 checklist is one the most widely used tools to assess the methodological quality of systematic reviews and it has been validated [21]. The tool includes 16 items in the form of questions which are answered with *yes*, *partial yes*, and *no* to address important domains of a systematic review. The selection, data extraction, and methodological assessment were conducted independently and in duplicate by two assessors (CMF, MAA) and disagreements were resolved by discussion and consensus.

We implemented random-effects meta-analyses and calculated odds ratios (ORs) with the respective 95% confidence intervals (CIs) and 95% prediction intervals (PIs). A PI is defined as the interval within which the effect size of a new study would fall if this study was selected at random from the same population of the studies already included in the meta-analysis [22, 23]. Therefore, reporting of the 95% PI provides a more clinically meaningful assessment of the between-trial heterogeneity in random effects meta-analyses [24]. Random effects meta-analysis accounts for the heterogeneity among the included studies because it assumes that there is no single joint effect but rather the effect follows a distribution. The restricted maximum likelihood estimator was used to calculate the heterogeneity variance τ^2 and the Knapp-Hartung adjustments [25], to reduce false positive results [26–28], and to calculate the CI around the pooled effect. It has been shown that the Knapp-Hartung adjustment can reduce in a relatively large range of cases the chance of false positives since it uses the *t*-distribution rather than the *z*-distribution [26–28]. For the main meta-analysis (meta-analysis 1), we examined the influence of individual studies, small study effects, and publication bias and their effect on the estimates and the heterogeneity. A simulation study was conducted to assess the effect of multiple implants (clustering effect) on the conclusions of the meta-analyses. In the simulation study, we varied the number of implants from 2 to 5 and the ICC from 0 to 0.5 to assess the possible effects on the precision of the estimates and corresponding *p*-values. The choice of the implant range was based on the range of the number of implants inserted in the included studies. The ICC range was more arbitrary as no relevant ICC values are reported in the literature and we had no access to any of the raw datasets to calculate the ICC. Since clustering reduces the amount of information, the original sample size is not the effective sample size and

TABLE 1 Results of the three meta-analyses on the association between non-perforated/perforated sinus membranes and implant loss.

Meta-analysis	Number of studies	Estimate (95% CI)	*p*-value
Al-Dajani 2016	7	OR 0.39 (0.8, 1.87)	0.24
Al-Moraissi et al. 2018	13	RR 2.17 (1.52, 3.11)	0.01
Kim et al. 2019	4	OR 1.76 (0.73, 4.28)	0.21

Abbreviations: CI, confidence interval; OR, odds ratio; RR, risk ratio.

should be adjusted downward for correct inferences. We adjusted the sample size by dividing the number of events and the number of implants by the corresponding design effect calculated by varying the number of implants and the ICC according to the Cochrane recommendations for the inclusion of clustered designs with binary outcomes in the meta-analysis [29].

Using the above assumptions, meta-analysis 1 was conducted for all possible combinations of the number of implants (2–5) and the ICC values (0–0.5, in 0.01 increments), resulting in 200 estimates for meta-analysis 1. The pooled estimates, the corresponding CIs, PIs, and *p*-values of those 200 meta-analyses were collected and the evolution of the standard error (as a measure of precision) over the ICC range for different numbers of implants were plotted. All analyses were conducted with R SOFTWARE VERSION 4.0.3 (R Foundation for Statistical Computing) using the packages *metafor* ver. 3.4-0, *dmetar* and *meta* ver. 5.5-0.

RESULTS AND DISCUSSION

The search yielded five potential meta-analyses that fulfilled the pre-defined research question. Following further review, two meta-analyses [30, 31] were excluded as the reported results were not directly related to the research question. The remaining three meta-analyses [32–34] were deemed appropriate and analysed. The results of the three meta-analyses on the association between non-perforated/perforated sinus membranes and implant failure are reported in Table 1. The list of the 16 studies included in these three meta-analyses is reported in the supplementary file in the Supporting Information.

Methodological quality of the systematic reviews

In relation to the AMSTAR-2 checklist, no systematic review had more than 50% of items answered with *yes* or *partial yes*. Al-Moraissi et al. [33] had eight items (50%), Kim et al. [34] had five items (31.3%), and Al-Dajani [32] had three items (18.8%) answered with *yes* or *partial yes* among a total of 16 items.

Overlap and methodological and clinical heterogeneity of primary studies

The total number of primary studies included in the three original meta-analyses was 16, of which 12 were included in the new meta-analysis. As expected, some level of overlapping of primary studies was evident when the respective forest plots of the three original meta-analyses were assessed. Six studies were included in two meta-analyses and one study was included in the three selected meta-analyses (Table 1). Clinical heterogeneity was observed mainly in terms of the type of instruments used to open the lateral bone wall and the time of the placement of implants (during or after sinus lift procedures). Table S2 reports the characteristics of primary studies included in the three meta-analyses.

New meta-analysis 1, based on twelve primary studies

Meta-analysis 1 (Figure 1) included 12 studies and reported a statistically significant difference favouring implant survival when sinus membrane was not perforated (OR 0.49, 95% CI = [0.26, 0.92], $p < 0.001$) (Table 2). The pooled estimate indicates an average 51% lower odds of failure for non-perforated sites. However, the 95% PI for the OR [0.18, 1.37], which provides a range of the OR estimate for a future similar study, includes a mixture of associations in both directions suggesting uncertainty in the findings as it also includes worse survival (37% higher odds) for the non-perforated sites compared to the perforated sites. Some studies initially selected were excluded from the final meta-analysis 1 (12 included from 16) due to the great clinical heterogeneity among studies that did not allow a reasonable comparison or due to patient overlap/duplication in the studies selected (Table S2).

We further examined how robust the results were by exploring publication bias and heterogeneity and their influence on the stability of the estimates. Funnel plots (Figure 2) show some evidence of small study effects (a possible reason for publication bias) and funnel plot asymmetry; however, this assessment is subjective [35] and may be due to other reasons [35]. Peter's test [36] did not provide enough evidence to reject the null hypothesis for the absence of small study effects

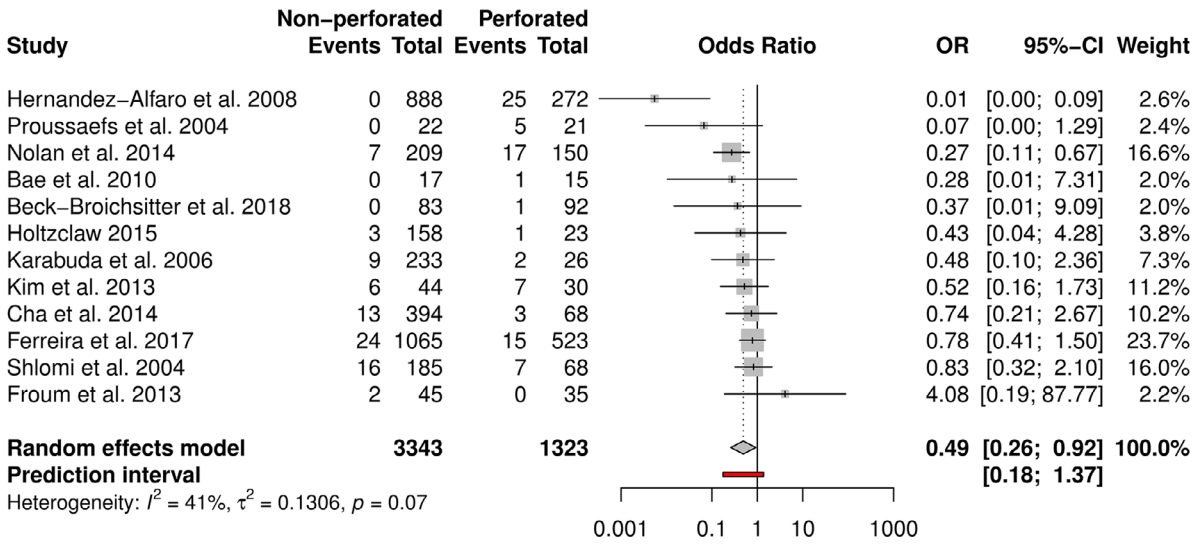

FIGURE 1 Meta-analysis 1, with all studies included.

TABLE 2 Main meta-analysis 1, meta-analysis 1 with trim-and-fill sensitivity, and outlier removal analysis.

	Number of studies	OR (95% CI)	*p*-value	I^2	tau	PI	Q test for heterogeneity
Meta-analysis 1	12	0.49 (0.26, 0.92)	0.05	41.0% (0.0%, 70.1%)	0.36 (0.00, 2.34)	0.18, 1.37	0.07
Trim-and-fill	16 (4 added)	0.71 (0.27, 1.82)	0.44	62.6% (35.8%, 78.2%)	1.16 (0.71, 2.93)	0.05, 10.09	<0.001
Outliers removed	Hernandez-Alfaro et al. 2008	0.57 (0.38, 0.85)	0.01	0.0% (0.0%, 51.5%)	0.14 (0.00, 1.07)	0.34, 0.96	0.61

Abbreviations: OR, odds ratio; CI, confidence interval; PI, prediction interval.

($p = 0.78$). However, such tests have low power, and a non-significant result does not necessarily exclude the possibility that some studies have not been published and thus are not available for inclusion in the meta-analysis. We also applied the trim-and-fill method [37] in order to estimate the number of missing studies under the assumption that in the absence of publication bias, studies should be symmetrically distributed around the pooled effect. The pooled estimate is recomputed after the data augmentation process and its robustness can be examined by comparing the estimates in the original and augmented plots. This method added four studies to account for funnel plot asymmetry (Figure 2). The estimates became more conservative, and heterogeneity increased dramatically (Table 2). It should be made clear that the trim-and-fill method is only a sensitivity analysis and not a method to compute a more valid pooled estimate.

The heterogeneity measured by $\tau = 0.36$, 95% CI = [0.00, 2.34], indicates that some between-study heterogeneity exists in our data. The $I^2 = 41\%$ indicates that close to half of the variation in our data is estimated to come from true effect size differences. The PI, which incorporates the between-study heterogeneity, includes values below and above 1.0, suggesting that it is possible that some future studies of the same population mix of studies can favour perforated sites in terms of implant survival compared to non-perforated sites.

There was weak evidence of significant heterogeneity based on the Q test ($p = 0.07$); however, given the limitations of this test [38], decisions on heterogeneity should not be based on the Q test alone. The results suggest that we have low to substantial heterogeneity, indicating some differences in the true effect sizes between studies, and it would be worth investigating this further to see if any studies have much higher/lower effects pulling the estimates up or down. Outlier studies and studies with high influence that differ substantially from the rest and increase the between-study heterogeneity and estimates should be identified [39]. The outlier analysis identified the study by Hernandez-Alfaro et al. [15] and the meta-analysis without this study seem to be more stable (Table 2) with decreased heterogeneity and range of effects in favour of non-perforated sites both for the 95% CI and the 95% PI.

Influence diagnostics examines the impact of a study or studies on the effect and heterogeneity estimates. Several diagnostic approaches are available [39, 40], and the Baujat plot [41] (not shown) indicates the contribution of each study to the overall heterogeneity and its influence on

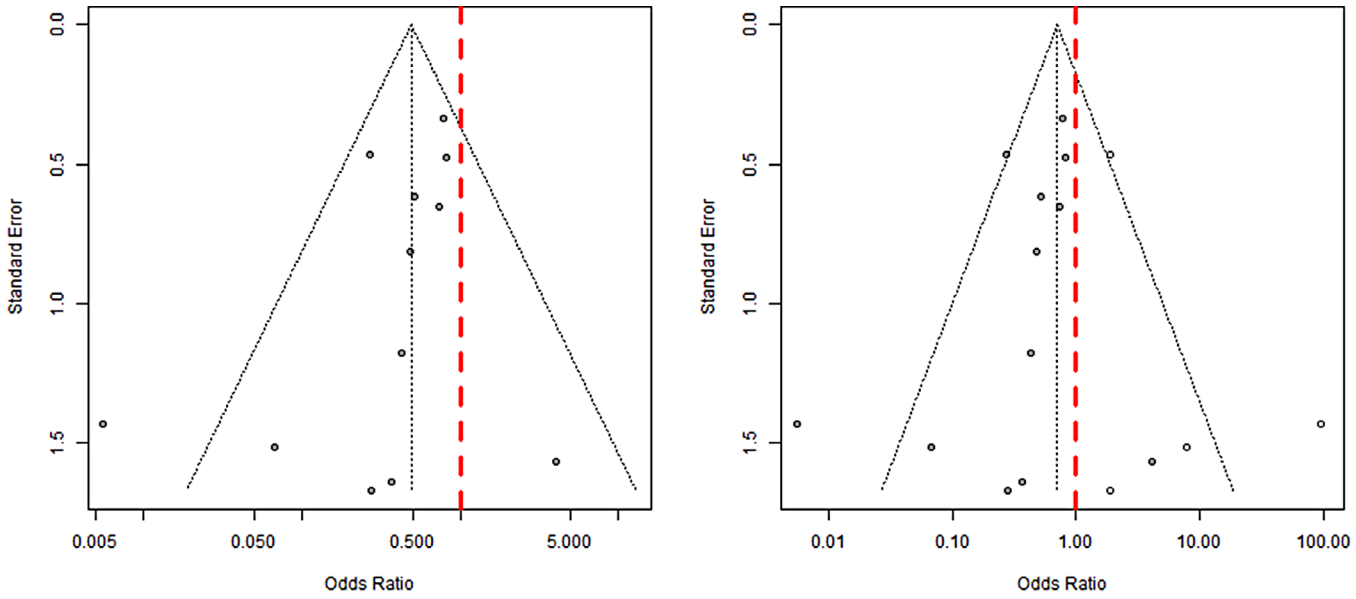

FIGURE 2 (Left) Funnel plot for meta-analysis 1. (Right) Funnel plot after applying trim-and-fill with four studies added to the right of the red vertical line (hollow circles).

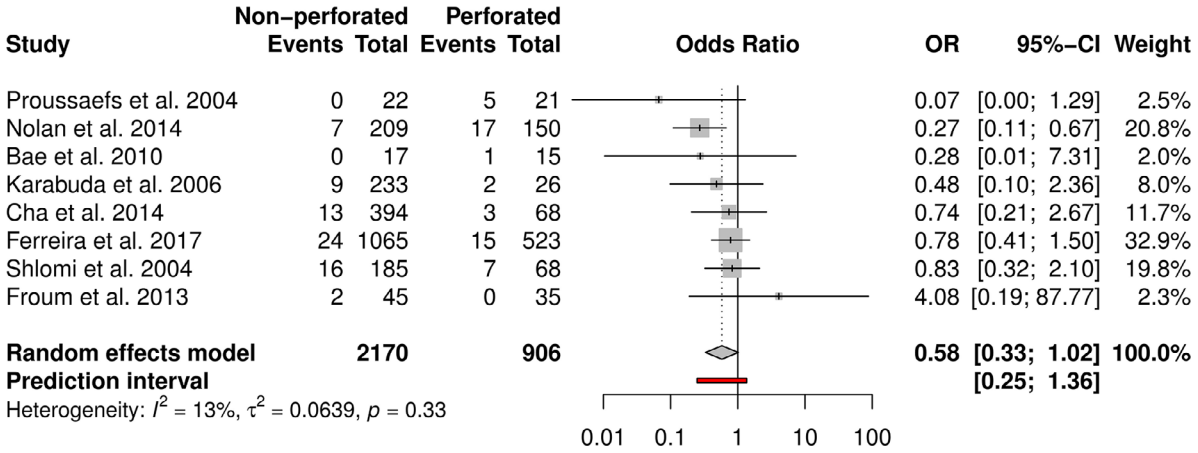

FIGURE 3 Meta-analysis 2, based on more homogeneous clinical procedures.

the pooled effect size. Again, we observed that the study of Hernandez-Alfaro et al. [15] has the biggest contribution to the heterogeneity. This study has the second largest sample size and has zero events in one arm. It should be noted that we are not suggesting that this study should be removed from the meta-analysis, but we are merely trying to examine in more detail the sources of heterogeneity.

Simulation meta-analyses 2–4

Meta-analysis 2 (Figure 3) (OR = 0.58, 95% CI = [0.33, 1.02], $p = 0.06$), based on more homogenous studies, included eight studies, and suggested on average 42% lower odds of failures for the non-perforated sites with a 95% PI for the OR of [0.25, 1.36], which crossed the line of no difference indicating up to 36% higher odds of failure in the non-perforated sites compared to the perforated sites. Table S2 reports the information used as background to support the rationale for deciding on the heterogeneity (more or less homogenous) of the primary studies included in meta-analysis 4. Meta-analyses 3 (Figure 4) (OR = 0.48, 95% CI = [0.13, 1.71], $p = 0.16$), and 4 (Figure 5) (OR = 0.83, 95% CI = [0.03, 23.61], $p = 0.89$), which were based on prospective and retrospective studies (Table S3), respectively, are not in agreement with meta-analysis 1 in terms of statistical significance. The corresponding 95% PIs for the OR are [0.08, 2.67] and [0.000, 7658.54], respectively. The 95% PI for meta-analysis 3 indicates extreme uncertainty, whereas the PI for meta-analysis 4 cannot be trusted, a finding encountered in studies with no or only very few events [42].

European Journal of Oral Sciences | WILEY

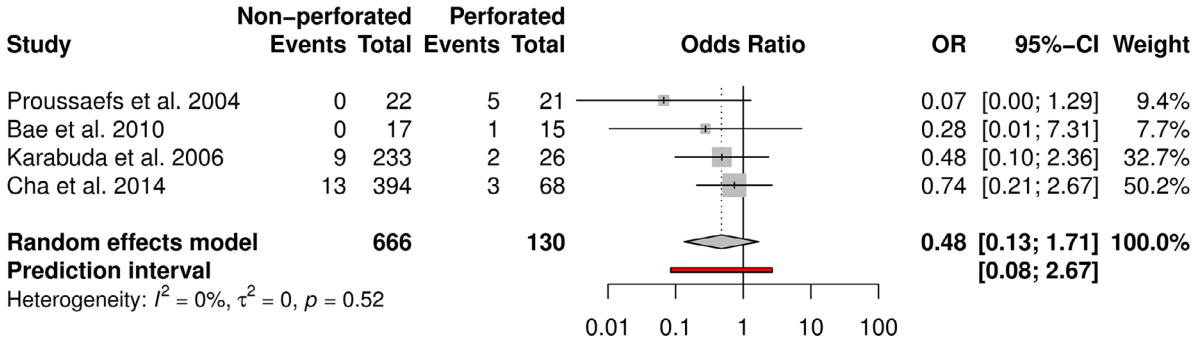

FIGURE 4 Meta-analysis 3, of prospective studies.

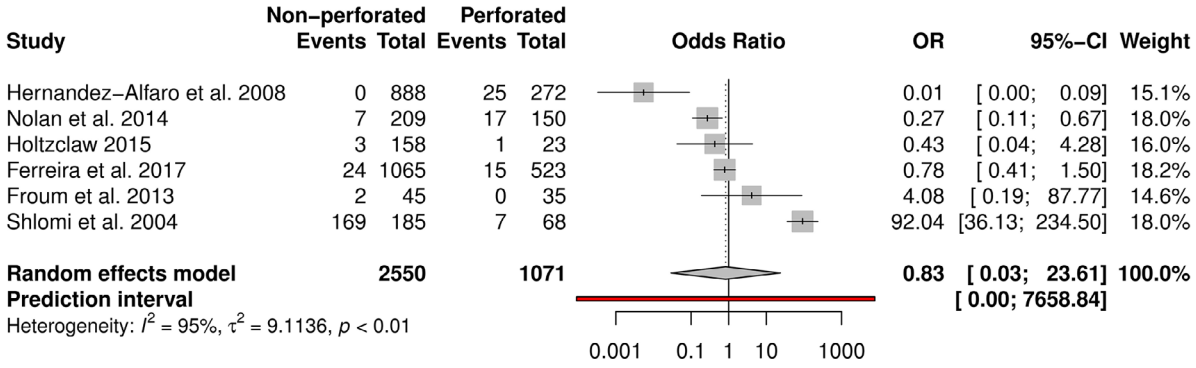

FIGURE 5 Meta-analysis 4, of retrospective studies.

Effect of clustering

There is empirical evidence that clustering effects are often ignored in oral health and elsewhere [43–45] and this also seemed to be the case for the included studies in the reproduced meta-analyses. On the x-axis in Figure 6 are the ICC values from 0 to 0.5, and on the y-axis is the standard error range. The line plots show the changes in the precision as a function of the ICC and the number of implants (line colour represents the number of implants). We can see that in the presence of four and five implants the results of meta-analysis become more imprecise (increased standard error) for some ICC values. While the resulting imprecision is not very predictable as indicated in the plot, it is nevertheless indicative of the fact that ignoring clustering can result in small p-values which may not be genuine. In an individual study, the effective sample size is expected to decrease up to a certain limit as the number of implants per patient and the ICC increase. The constraint is that effective sample size cannot be smaller than the number of patients when the ICC = 1. In meta-analysis, things are more complicated as shown by the fluctuations of the standard error across ICC values. In the random effects meta-analysis, we increase the width of the CIs for the individual studies by dividing the individual study sample sizes and the number of events by the design effect, and in some cases this resulted in a decrease of the between-study heterogeneity, leading to a decreased width of the 95% confidence interval [46]. This is a possible contributing factor to the explanation for the imprecision fluctuating instead of showing a monotonic increase.

Methodological challenges

Our findings emphasize the need for careful analysis of all biases and meta-analysis parameters that can interfere with the meta-analytic estimates. The original three meta-analyses included studies with low methodological quality. Most studies were retrospective and were designed to answer the study question without a control, that is, conducting a one-arm study. In fact, most studies were case series which analysed the implant survival in perforated and non-perforated membrane sinuses. We tried to contact authors of the primary studies included in the original meta-analyses to provide us with the individual patient data to calculate the clustering effect of the implants on the meta-analytic estimates. However, only one author responded, who was not able to provide the original dataset. Similar to the primary studies, the selected

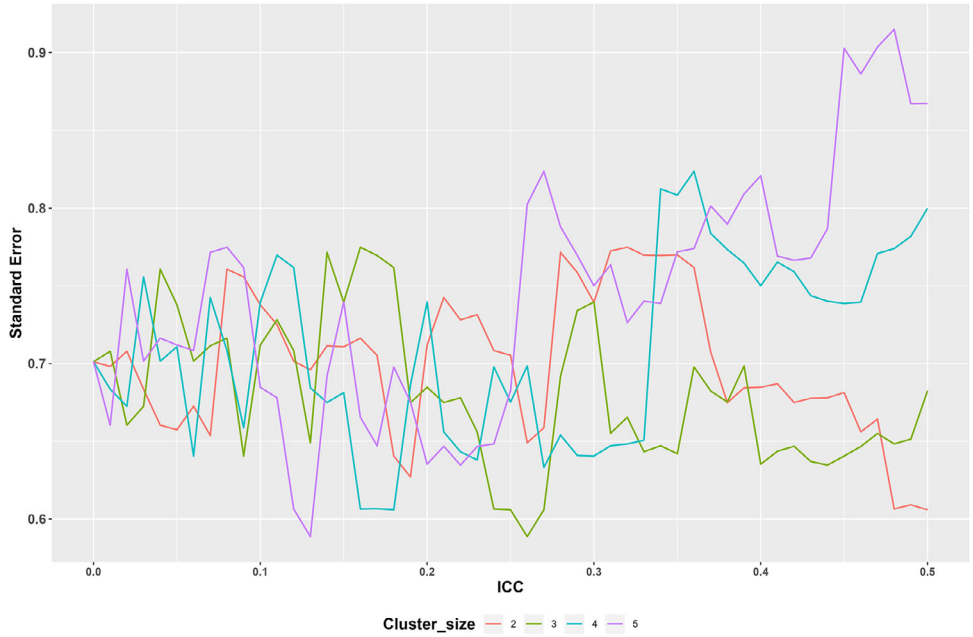

FIGURE 6 Line plots for the effect of the ICC and the number of implants on the standard error for the pooled estimate from meta-analysis 1. ICC, intra-cluster correlation coefficient.

systematic reviews were rated as low in methodological quality based on AMSTAR-2 scores, which further undermines the conclusions drawn from the meta-analysis.

Interpretation of the findings

The re-analysis of the available systematic reviews after considering clustering effects, publication bias, and clinical and methodological heterogeneity, provided variable results. Meta-analysis 1 included the greatest number of primary studies and provided the most precise estimate for the average effect. Meta-analysis 2, which included clinically homogeneous trials, showed, on average, higher odds of survival for non-perforated membrane sinuses; however, the range of the effects included estimates favouring also the perforated sites. Both meta-analyses involving prospective and retrospective studies point to inconclusive results. The results should be taken with caution due to potential inaccuracy in reporting the type of study on the part of the primary research authors [47, 48]. It has been established that making a distinction between retrospective and prospective designs can be challenging [49].

The confidence in the precision of the results can be further decreased if we analyse the data after considering publication bias and the potential clustering effects which were ignored, and by interpreting the PIs which in all cases show inconclusiveness. The use of the PI in addition to the CI in the reporting of meta-analytic findings involving three of more primary studies has been recommended [23]. Limitations of the CI for the effect size include the difficulty in retaining its coverage probability, the likelihood of underestimating the statistical error and generating overconfident results, as well as estimation problems when the number of included studies is small [50]. The use of the PI builds on the idea of applying the potential effect of treatment to a specific individual study setting that reflects reality, going beyond the average effect. In other words, the PI predicts the range for the true treatment effect in a future individual study originating in the same population mix of studies included in the meta-analyses at hand, by taking into account the existing heterogeneity [22].

The simulation study (comprising meta-analyses 2–4) showed that ignoring clustering effects can potentially change our conclusions and such practices should be avoided. However, individual studies should provide information about the cluster size and the ICC as suggested by the CONSORT extension for clustered randomized trials in order to accurately account for these correlations [51]. This information is equally useful for clustered designs beyond the realm of randomized controlled trials as those issues are similarly present. In the simulation study, not knowing the exact number of implants and ICC per study imposes some limitations and therefore we considered a large range of values to see the influence of those parameters on the statistical significance. However, the ICC and the number of implants do not seem to be the sole determinants of the expected precision and *p*-values.

This attempt to reproduce the conduct of a meta-analysis after considering several influential meta-analytical parameters is important as it increases the awareness of readers of scientific articles about the complexity of understanding and interpreting the results of meta-analyses.

CONCLUSIONS

This case study has highlighted the influence of the various assumptions adopted in the meta-analysis on the interpretation of estimates reported in quantitative systematic reviews. Readers need to carefully consider the impact of meta-analytic parameters such as heterogeneity of the primary studies, publication bias, and clustering effects which can affect the size and direction of the reported estimates. In addition, authors, reviewers, and editors should collaborate with methodologists (including biostatisticians) as the issues discussed here are all well known but do require advanced knowledge, which is usually beyond the skills of clinicians.

AUTHOR CONTRIBUTIONS

Conceptualization: Clovis Mariano Faggion Jr; Nikolaos Pandis; **Methodology**: Clovis Mariano Faggion Jr; Nikolaos Pandis; Michail Tsagris; **Formal analysis**: Clovis Mariano Faggion Jr; Momen A. Atieh; Michail Tsagris; Jadbinder Seehra; Nikolaos Pandis; **Investigation**: Clovis Mariano Faggion Jr; Momen A. Atieh; **Data Curation**: Clovis Mariano Faggion Jr; Momen A. Atieh; **Writing—original draft preparation**: Clovis Mariano Faggion Jr; **Writing—review and editing**: Clovis Mariano Faggion Jr; Momen A. Atieh; Michail Tsagris; Jadbinder Seehra; Nikolaos Pandis; **Visualization**: Clovis Mariano Faggion Jr; Nikolaos Pandis; **Supervision**: Nikolaos Pandis; **Project administration**: Clovis Mariano Faggion Jr; **Funding acquisition**: This work was self-funded.

ACKNOWLEDGMENT

No external funding, apart from the support of the authors' institution, was provided for this study.

Open access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

ORCID

Clovis Mariano Faggion Jr https://orcid.org/0000-0001-8323-909X
Momen A. Atieh https://orcid.org/0000-0003-4019-9491
Michail Tsagris https://orcid.org/0000-0002-2049-3063
Jadbinder Seehra https://orcid.org/0000-0002-3243-1580
Nikolaos Pandis https://orcid.org/0000-0003-0258-468X

REFERENCES

1. Faggion CM Jr. Is the evidence supporting dental procedures strong? A survey of Cochrane systematic reviews in oral health. J Evid-Based Dent Pract. 2012;12:131-134.e14.e14.
2. Pandis N, Fleming PS, Worthington H, Salanti G. The quality of the evidence according to GRADE is predominantly low or very low in oral health systematic reviews. PloS One. 2015;10:e0131644. https://doi.org/10.1371/journal.pone.0131644
3. Seehra J, Bertl K, Faggion CM Jr, Pandis N. The certainty of the evidence in oral health has not improved according to GRADE: a meta-epidemiological study. J Clin Epidemiol. 2021;142:29–37.
4. Koletsi D, Fleming PS, Michelaki I, Pandis N. Heterogeneity in Cochrane and non-Cochrane meta-analyses in orthodontics. J Dent. 2018;74:90–94.
5. Deeks JJ, Higgins JP, Altman DG. Analysing data and undertaking meta-analyses. In: Cochrane handbook for systematic reviews of interventions [Internet]. John Wiley & Sons; 2008 [cited 2021 Jun 26]. p. 243–296. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470712184.ch9
6. Rücker G, Schwarzer G, Carpenter JR, Schumacher M. Undue reliance on I2 in assessing heterogeneity may mislead. BMC Med Res Methodol. 2008;8:79. https://doi.org/10.1186/1471-2288-8-79
7. von Hippel PT. The heterogeneity statistic I(2) can be biased in small meta-analyses. BMC Med Res Methodol. 2015;15:35. https://doi.org/10.1186/s12874-015-0024-z
8. Montori VM, Smieja M, Guyatt GH. Publication bias: a brief review for clinicians. Mayo Clin Proc. 2000;75:1284–1288.
9. Hayes RJ, Bennett S. Simple sample size calculation for cluster-randomized trials. Int J Epidemiol. 1999;28:319–326.
10. Altman DG, Bland JM. Statistics notes. Units of analysis. BMJ. 1997;314:1874.
11. Chia KS. "Significant-itis"–an obsession with the P-value. Scand J Work Environ Health. 1997;23:152–154.
12. Savitz DA. Is statistical significance testing useful in interpreting data? Reprod Toxicol Elmsford N. 1993;7:95–100.
13. Kerry SM, Bland JM. The intracluster correlation coefficient in cluster randomisation. BMJ. 1998;316:1455.
14. Hirsch JM, Ericsson I. Maxillary sinus augmentation using mandibular bone grafts and simultaneous installation of implants. A surgical technique. Clin Oral Implants Res. 1991;2:91–96.
15. Hernández-Alfaro F, Torradeflot MM, Marti C. Prevalence and management of Schneiderian membrane perforations during sinus-lift procedures. Clin Oral Implants Res. 2008;19:91–98.
16. Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ. 2017 21;358:j4008. https://doi.org/10.1136/bmj.j4008
17. Cochrane Linked Data. PICO ontology [Internet]. [cited 2021 Oct 16]. Accessed 21 Nov 2023. Available from: https://linkeddata.cochrane.org/pico-ontology
18. Gagnier JJ, Moher D, Boon H, Beyene J, Bombardier C. Investigating clinical heterogeneity in systematic reviews: a methodologic review of guidance in the literature. BMC Med Res Methodol. 2012;12:111. https://doi.org/10.1186/1471-2288-12-111
19. Parker LA, Saez NG, Porta M, Hernández-Aguado I, Lumbreras B. The impact of including different study designs in meta-analyses of diagnostic accuracy studies. Eur J Epidemiol. 2013;28:713–720.
20. West SL, Gartlehner G, Mansfield AJ, Poole C, Tant E, Lenfestey N, et al. Comparative effectiveness review methods: clinical heterogeneity [Internet]. Rockville (MD): Agency for Healthcare Research and Quality (US); 2010 [cited 2021 Jun 27]. (AHRQ Methods for Effective Health Care). Available from: http://www.ncbi.nlm.nih.gov/books/NBK53310/
21. Shea BJ, Bouter LM, Peterson J, Boers M, Andersson N, Ortiz Z, et al. External validation of a measurement tool to assess systematic

reviews (AMSTAR). PloS One. 2007;2:e1350. https://doi.org/10.1371/journal.pone.0001350
22. Riley RD, Higgins JPT, Deeks JJ. Interpretation of random effects meta-analyses. BMJ. 2011;342:d549. https://doi.org/10.1136/bmj.d549
23. IntHout 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
24. Spineli LM, Pandis N. Prediction interval in random-effects meta-analysis. Am J Orthod Dentofacial Orthop. 2020;157:586–588.
25. Knapp G, Hartung J. Improved tests for a random effects meta-regression with a single covariate. Stat Med. 2003;22:2693–2710.
26. IntHout J, Ioannidis JP, Borm GF. The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. BMC Med Res Methodol. 2014;14:25. https://doi.org/10.1186/1471-2288-14-25
27. Wiksten A, Rücker G, Schwarzer G. Hartung-Knapp method is not always conservative compared with fixed-effect meta-analysis. Stat Med. 2016;35:2503–2515.
28. Langan D, Higgins JPT, Jackson D, Bowden J, Veroniki AA, Kontopantelis E, et al. A comparison of heterogeneity variance estimators in simulated random-effects meta-analyses. Res Synth Methods. 2019;10:83–98.
29. Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M, et al. Chapter 23: Including variants on randomized trials. In: Higgins JPT, Eldridge S, Li T. (editors). Cochrane handbook for systematic reviews of interventions version 6.4 (updated August 2023). Cochrane, 2023. Available from: www.training.cochrane.org/handbook
30. Atieh MA, Alsabeeha NH, Tawse-Smith A, Faggion CM Jr, Duncan WJ. Piezoelectric surgery vs rotary instruments for lateral maxillary sinus floor elevation: a systematic review and meta-analysis of intra- and postoperative complications. Int J Oral Maxillofac Implants. 2015;30:1262–1271.
31. Stacchi C, Troiano G, Berton F, Lombardi T, Rapani A, Englaro A, et al. Piezoelectric bone surgery for lateral sinus floor elevation compared with conventional rotary instruments: a systematic review, meta-analysis and trial sequential analysis. Int J Oral Implantol Berl Ger. 2020;13:109–121.
32. Al-Dajani M. Incidence, risk factors, and complications of Schneiderian membrane perforation in sinus lift surgery: a meta-analysis. Implant Dent. 2016;25:409–415.
33. Al-Moraissi E, Elsharkawy A, Abotaleb B, Alkebsi K, Al-Motwakel H. Does intraoperative perforation of Schneiderian membrane during sinus lift surgery causes an increased the risk of implants failure?: a systematic review and meta regression analysis. Clin Implant Dent Relat Res. 2018;20:882–889.
34. Kim JS, Choi SM, Yoon JH, Lee EJ, Yoon J, Kwon SH, et al. What affects postoperative sinusitis and implant failure after dental implant: a meta-analysis. Otolaryngol Head Neck Surg. 2019;160:974–984.
35. Page MJ, Sterne JAC, Higgins JPT, Egger M. Investigating and dealing with publication bias and other reporting biases in meta-analyses of health research: a review. Res Synth Methods. 2021;12:248–259.
36. Peters JL, Sutton AJ, Jones DR, Abrams KR, Rushton L. Comparison of two methods to detect publication bias in meta-analysis. JAMA. 2006;295:676–680.
37. Duval S, Tweedie R. Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. Biometrics. 2000;56:455–463.
38. Hoaglin DC. Misunderstandings about Q and "Cochran's Q test" in meta-analysis. Stat Med. 2016;35:485–495.
39. Viechtbauer W, Cheung MWL. Outlier and influence diagnostics for meta-analysis. Res Synth Methods. 2010;1:112–125.
40. Olkin I, Dahabreh IJ, Trikalinos TA. GOSH—a graphical display of study heterogeneity. Res Synth Methods. 2012;3:214–223.
41. Baujat B, Mahé C, Pignon JP, Hill C. A graphical method for exploring heterogeneity in meta-analyses: application to a meta-analysis of 65 trials. Stat Med. 2002;21:2641–2652.
42. Efthimiou O. Practical guide to the meta-analysis of rare events. Evid Based Ment Health. 2018;21:72–76.
43. Campbell MK, Grimshaw JM. Cluster randomised trials: time for improvement. BMJ. 1998;317:1171–1172.
44. Fleming PS, Koletsi D, Polychronopoulou A, Eliades T, Pandis N. Are clustering effects accounted for in statistical analysis in leading dental specialty journals? J Dent. 2013;41:265–270.
45. Polychronopoulou A, Pandis N, Eliades T. Appropriateness of reporting statistical results in orthodontics: the dominance of P values over confidence intervals. Eur J Orthod. 2011;33:22–25.
46. Borenstein M, editor. Introduction to meta-analysis. Chichester, U.K: John Wiley & Sons; 2009. p. 421.
47. Esene I, Ngu J, Zoghby M, Soraloglu I, Sikod A, Kotb A, et al. Case series and descriptive cohort studies in neurosurgery: the confusion and solution. Childs Nerv Syst. 2014;30:1321–1332.
48. Grimes D. "Case-control" confusion: mislabeled reports in obstetrics and gynecology journals. Obstet Gynecol. 2009;114:1284–1286.
49. Dekkers OM, Groenwold RHH. Study design: what's in a name? Eur J Endocrinol. 2020;183:E11–E13.
50. Veroniki AA, Jackson D, Bender R, Kuss O, Langan D, Higgins JPT, et al. Methods to calculate uncertainty in the estimated overall effect size from a random-effects meta-analysis. Res Synth Methods. 2019;10:23–43.
51. Campbell MK, Elbourne DR, Altman DG, CONSORT group. CONSORT statement: extension to cluster randomised trials. BMJ. 2004;328:702–708.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Faggion CM Jr, Atieh MA, Tsagris M, Seehra J, Pandis N. A case study evaluating the effect of clustering, publication bias, and heterogeneity on the meta-analysis estimates in implant dentistry. Eur J Oral Sci. 2023;e12962. https://doi.org/10.1111/eos.12962