OXFORD

Original article

Assessment of early exaggerated treatment effects in orthodontic interventions using cumulative meta-analysis

Jadbinder Seehra[1,], Daniel Stonehouse-Smith[1,] and Nikolaos Pandis[2]

[1]Department of Orthodontics, Faculty of Dentistry, Oral & Craniofacial Sciences, King's College London, UK, [2]Department of Orthodontics and Dentofacial Orthopedics, Dental School/Medical Faculty, University of Bern, Switzerland

Correspondence to: Jadbinder Seehra, Department of Orthodontics, Faculty of Dentistry, Oral & Craniofacial Sciences, King's College London, Floor 25, Guy's Hospital, Guy's and St Thomas NHS Foundation Trust, London SE1 9RT, UK. E-mail: jadbinderpal.seehra@kcl.ac.uk

Summary

Background: The reported initial strong treatment effects reported in early trials that are refuted in subsequent future studies assessing the same interventions have been attributed to novelty bias. The aim of this study was to determine whether there is any evidence of novelty bias in the reported treatment effects of orthodontics interventions.

Materials and methods: Relevant orthodontic systematic review (SRs) topics containing at least one meta-analysis on either binary or continuous outcomes with a minimum of three trials considered important areas in the field of orthodontic practice were identified. SR, meta-analysis, and primary study-level characteristics were extracted. Descriptive statistics were calculated at the SRs, meta-analysis, and at the individual study level. All SR and trial-level data were imported into the statistical software and all meta-analyses were replicated using the cumulative random-effects meta-analysis approach. Changes in the size and direction of the estimates between the first trial and the cumulative effect over time were recorded.

Results: Forty-seven meta-analyses were included. The total number of primary studies included within these meta-analyses was 408 ($N = 408$). Overall, the final effect size estimate decreased in 29 (61.7%, $N = 29/47$) cumulative meta-analyses whilst it increased in the remaining 18 (38.3%, $N = 18/47$). No association between the level of risk of bias and the cumulative absolute effect size was evident (OR 1.00; 95% CI: 0.98, 1.03; $P = 0.717$) after adjusting for year of the primary study ($P = 0.22$).

Conclusions: Clinicians should be wary of the results of trials reporting the effectiveness of new interventions as there is a possibility that the reported effect size will be often exaggerated.

Introduction

The conduct of orthodontic trials has increased substantially within the literature with just under 1100 trials published during a 10-year period (1). Despite clinical trials representing high-level evidence that can inform clinical practice, they are prone to several forms of bias which can distort the direction and the size of the effect estimates (2). The effect estimates indicate in a clinically relevant manner how effective an intervention is and together with the associated confidence intervals constitute a better way to communicate clinical relevance compared to testing for statistical significance (3). Biases can be exaggerated in the results of trials assessing the effectiveness of new interventions. The initial strong treatment effects reported in early trials can be refuted in subsequent future studies assessing the same interventions (4, 5–10). The term novelty bias has been introduced to account for differences between the reporting of initially exaggerated treatment effects which are then not supported in future studies (6,

11). Novelty bias has been proposed to occur as a result of other forms of bias such as selection, outcome reporting confirmation, and 'hot stuff' bias having a greater impact when the intervention is new (11).

The term cumulative meta-analysis depicts a meta-analytical approach where prospectively as the results of new relevant trials become available, previous meta-analysis pooled estimates are updated (12). This analysis describes statistically how the evidence base can evolve regarding a particular intervention (13). Previous cumulative meta-analyses have reported how initially favourable estimates have either decreased, increased, or undergone no change as the evidence-based has evolved through study replications (13). Cumulative meta-analyses can also be used as a tool to identify research waste (14), reinforce the need for the design of new studies to be informed by relevant systematic reviews (SRs) (15) and the results of new studies to be reported in the context of updated SRs of similar studies (16).

The extent of novelty bias within orthodontic trials or studies is unknown. Using the cumulative meta-analysis approach, the aim of the current investigation was, as the evidence from primary studies or trials accumulates, to examine whether there is any evidence of novelty bias in the reported treatment effects of orthodontics interventions.

Methods

Eligibility criteria

Relevant orthodontic SR topics which were considered important areas in the field of orthodontic practice were initially identified from published literature (17, 18). Further contemporary SR topics were also identified using an electronic database search. The final selected topics were agreed by consensus between two authors (JS and NP). To be included, the SR should include at least one meta-analysis on either binary or continuous outcomes containing a minimum of three trials. Furthermore, only English language and SR reporting interventional procedures involving human participants were included. Where multiple versions of the same SR existed, the latest version was selected.

Search of further SRs and primary studies

Both an electronic database search (Medline via PubMed) and search of Cochrane Library were undertaken using the following search terms: 'orthodontic' AND 'systematic review' OR 'meta-analysis' to identify further contemporary SR topics not covered in the published literature (17, 18). All titles and abstracts were initially screened by two authors (JS and NP). Full-text articles of abstracts meeting the inclusion criteria were retrieved and further assessed. For the selected SRs with a minimum of three studies, additional recent randomized clinical trials (primary studies) not included in the meta-analysis but with consistent aims and outcome measures to the SR were searched. Again, an electronic database search (Medline via PubMed) was undertaken using the following search terms and filters: 'randomized clinical trial' AND 'orthodontics' to identify further primary studies. All titles and abstracts were initially screened in duplicate (JS and NP). Any disagreements in the selection of the final SRs and additional trials were resolved by discussion among two authors (JS and NP).

Data extraction

Study characteristics were extracted from the forest plots, tables, and text of the final included SRs. A pre-piloting process of five SRs was undertaken to ensure consistency between authors (JS, DSS, and NP) regarding data extraction, interpretation of both data variables, and forest plots independently. Once 100% agreement had been achieved, all study characteristics were then extracted by a single author (DSS) and entered into a pre-piloted Microsoft Excel® (Microsoft, Redmond, Washington, USA) data collection sheet. A second author (JS) independently reviewed the collected data. Any disagreements were resolved by discussion.

At the SR level, the following information was extracted: number of authors, continent of corresponding author, year of publication, and PROSPERO registration at the meta-analysis level, the following were collected: number of primary studies included, type of effect measure (continuous or binary), and estimates with 95% CIs. The following information from the SRs at the primary study level were also extracted: year of publication and risk of bias assessment, number of patients, number of arms, type of outcome (binary or continuous), sample size, means, and standard deviation or number of events, where applicable, per treatment group, estimates, 95% CIs and *P*-values <0.05 of the contributing studies. For all newly identified trials, a risk of bias assessment was undertaken using the Cochrane risk-of-bias tool for randomized trials (19). This was conducted independently by two authors (JS and DSS). Any disagreements were resolved by discussion. When more than one meta-analysis was present, the meta-analysis directly related to the main outcome of the study was selected. When two or more meta-analyses were related to the main outcome, we selected the meta-analysis with the greatest number of primary studies included. All SR and trial-level data were imported to the statistical software and all meta-analyses were replicated using the cumulative meta-analysis approach.

Statistical analysis

Descriptive statistics were calculated at the SR, meta-analysis, and at the individual study level. All SR and trial-level data were imported to the statistical software and all meta-analyses were replicated using the cumulative random-effects meta-analysis approach. Changes in the estimates between the first study and the cumulative pooled estimated were represented in barplots. For the binary outcomes, the estimates were log-transformed before plotting in order to improve visualization. Ordinal logistic regression was used to assess the absolute cumulative effect adjusted for year of the trial on the probability of belonging to high, unclear, or low risk of bias. All analyses were conducted using Stata 16.1 (StataCorp, College Station, Texas, USA).

Results

Forty-seven meta-analyses were included in this study (Supplementary Table 1). At the SR level, the most frequent year of publication were 2017 (19.1%) and 2018 (19.1%), with the corresponding author based in Europe (42.3%) and most SRs not being registered with PROSPERO (59.6%). The median number of SR authors was 5 (IQR 4–6). The title of each meta-analysis is provided in Supplementary Table 1. Out of the 47 meta-analyses, 36 and 11 reported continuous and binary outcomes, respectively (Table 1). Twelve additional primary studies or trials (Supplementary Table 1) were identified and hence the total number of primary studies included within the meta-analyses was 408 (*N* = 408). Overall, the final effect size estimate decreased in 29 (61.7%, *N* = 29/47) cumulative meta-analyses whilst it increased in the remaining 18 (38.3%, *N* = 18/47) (Supplementary Table 2). For continuous outcomes, in 24 cumulative meta-analyses, the final effect size estimate decreased (*N* = 24/36) (Figure 1).

Table 1. Descriptive statistics for the included systematic reviews.

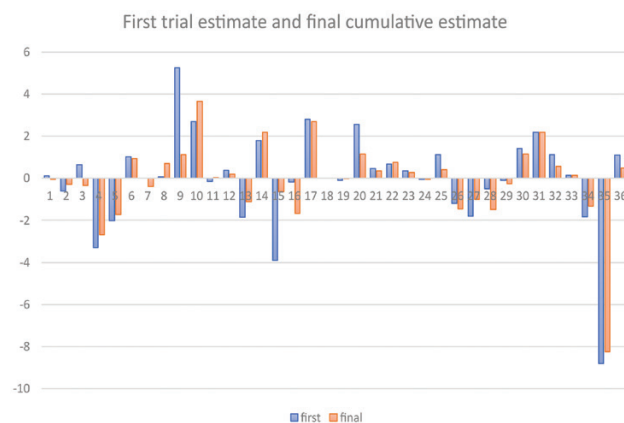| Characteristic	N (%)
Year of publication	
2010	1 (2.1)
2012	2 (4.3)
2013	2 (4.3)
2014	7 (14.9)
2015	4 (8.5)
2016	6 (12.8)
2017	9 (19.1)
2018	9 (19.1)
2019	6 (12.8)
2020	1 (2.1)
Continent of corresponding author	
Europe	20 (42.5)
Americas	10 (21.3)
Asia or other	17 (36.2)
Registration	
Yes	19 (40.4)
No	28 (59.6)
Total	47 (100.0)

Figure 1. Barplot for the estimate of the first study (blue) and the cumulative estimate (orange) for meta-analyses with continuous outcomes.
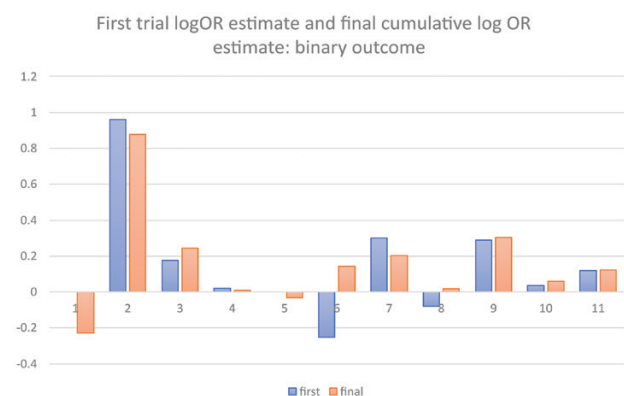

Figure 2. Barplot for the estimate of the first study (blue) and the cumulative estimate (orange) for meta-analyses with binary outcomes. Estimates have been log-transformed.

Conversely, in analyses reporting binary outcomes, the final effect size estimate decreased in five cumulative meta-analyses (N = 5/11) (Figure 2). No association between the level of risk of bias and the cumulative absolute effect size was evident (OR 1.00; 95% CI: 0.98, 1.03; *P* = 0.717) after adjusting for year of the primary study (likelihood ratio test *P*-value = 0.22).

Discussion

The findings based on the change between the initial and the cumulative meta-analysis pooled estimates have shown evidence of exaggeration of initial treatment effects or reporting of novel agent effects (novelty bias) in orthodontic interventions. Overall, in 61.7% (N = 29/47) cumulative meta-analysis, the final treatment effect size estimate decreased compared to the reported first study treatment effect size estimate (Figure 3A and 3B). Similar results have been reported elsewhere. In the investigation of highly cited studies, in 16% of studies, the effects of the first trial were stronger than those of subsequent studies (20). This observation is supported by the findings of a cumulative network meta-analysis undertaken regarding the effectiveness of anti-depressants. Despite reported initial stronger effects of newly approved drugs, the effect estimate tended to decrease and stabilize over time with the addition of new evidence (21). Conversely, in 38.3% of cumulative meta-analysis, an opposite effect was observed with the final absolute treatment effect size estimate increased compared to the first reported absolute treatment effect size estimate (Figure 3B). This trend has been previously reported but tends to be fewer and weaker than the finding of the final absolute effect size decreasing (13). Within the medical literature, evidence of novelty bias has been reported (4–8). Using a multiple treatment meta-analysis model, the presence of novelty bias has been reported to result in an exaggeration of cancer treatment effect size estimate by 6% (6). Additionally in the assessment of randomized clinical trials of lipid-lowering anti-glaucoma drugs, the effectiveness of three out of four of the interventions decreased over time (5). In summary, based on the meta-analyses of clinical trials, the presence of novelty bias can result in an intervention appearing to be between 2% and 27% better when the treatment is novel (11).

Inflated newly reported effects compared to the true effect size may be the result of a combination of biases present in the study methodology including the conduct and reporting of a trial or study (22). Regarding study design, observational studies tend to report exaggerated effects which are often contradicted by the results of randomized clinical trials (23). More importantly, if replication of study design can be achieved, the results of such studies are more likely to contradict the reported initial stronger results over time (20). However, this has to be mitigated by the finding that reproducibility of research study design is both sub-optimal (24) and undertaken infrequently (25). The decrease in absolute treatment effect size estimate appears to be independent of the level of risk of bias within the primary study. It could be assumed that novelty bias exists as a result of the poor-quality studies. However, assumptions made regarding trial quality based on reported trial aspects which may deviate from what was actually done in each trial is not straightforward. Additionally, the relationship between trial quality and novelty bias has not been established (6). It should be remembered that if subsequent trials or studies contradict the findings of initial studies it does not mean that these studies were wrong in the first place and that later studies were larger or employed a controlled methodology (20). Indeed, increasing the sample size does not necessarily result in a change in the direction of the results but can increase precision. The latter explains the finding in the current study where a decreased effect from the cumulative estimate can be statistically significant whereas in the initial trial a larger effect is not significant. Alternative reasons for the overserved differences could be the differences in eligibility criteria or the use of additional interventions (26).

Short-term effects of removable functional appliances
on maxillary growth (SNA) in the treatment of Class II malocclusions
compared to untreated controls
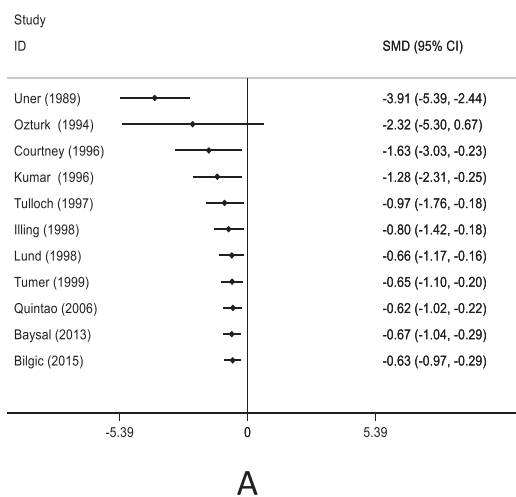
Improvement in skeletal relationship (ANB) in Class III patients
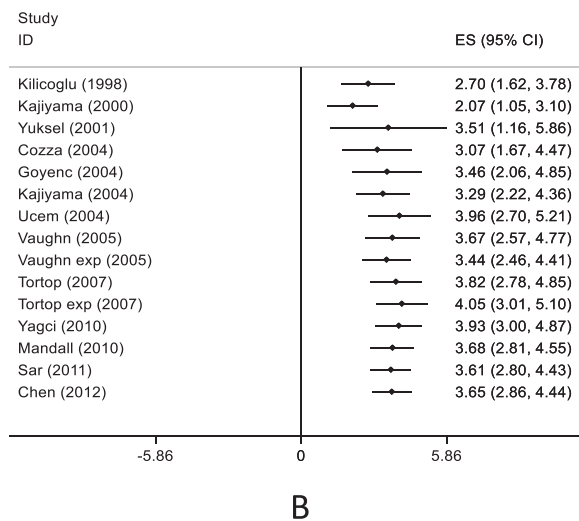treated with Facemask versus untreated controls

Study ID	SMD (95% CI)
Uner (1989)	-3.91 (-5.39, -2.44)
Ozturk (1994)	-2.32 (-5.30, 0.67)
Courtney (1996)	-1.63 (-3.03, -0.23)
Kumar (1996)	-1.28 (-2.31, -0.25)
Tulloch (1997)	-0.97 (-1.76, -0.18)
Illing (1998)	-0.80 (-1.42, -0.18)
Lund (1998)	-0.66 (-1.17, -0.16)
Tumer (1999)	-0.65 (-1.10, -0.20)
Quintao (2006)	-0.62 (-1.02, -0.22)
Baysal (2013)	-0.67 (-1.04, -0.29)
Bilgic (2015)	-0.63 (-0.97, -0.29)

-5.39   0   5.39

A

Study ID	ES (95% CI)
Kilicoglu (1998)	2.70 (1.62, 3.78)
Kajiyama (2000)	2.07 (1.05, 3.10)
Yuksel (2001)	3.51 (1.16, 5.86)
Cozza (2004)	3.07 (1.67, 4.47)
Goyenc (2004)	3.46 (2.06, 4.85)
Kajiyama (2004)	3.29 (2.22, 4.36)
Ucem (2004)	3.96 (2.70, 5.21)
Vaughn (2005)	3.67 (2.57, 4.77)
Vaughn exp (2005)	3.44 (2.46, 4.41)
Tortop (2007)	3.82 (2.78, 4.85)
Tortop exp (2007)	4.05 (3.01, 5.10)
Yagci (2010)	3.93 (3.00, 4.87)
Mandall (2010)	3.68 (2.81, 4.55)
Sar (2011)	3.61 (2.80, 4.43)
Chen (2012)	3.65 (2.86, 4.44)

-5.86   0   5.86

B

Figure 3. Compared to the initial effect size estimate, decrease in final absolute treatment effect size estimate (A) and increase in final absolute treatment effect size estimate (B). Although the result is still statistically significant, its clinical relevance is less evident.

To reduce novelty bias effects, recommendations have been suggested (11). Investigators of new or novel interventions should be encouraged to state if the reported observed effect could be the result of novelty bias. In addition, when reporting the results of novel interventions analytical methods that correct for the anticipated inflation should be used, strict protocols for analyses can be employed and complete and transparent reporting of all results should be pre-requisite (22). Replication of studies should be encouraged (22). In this event, investigators who test the same hypotheses in future studies and report less favourable results should try to identify and explain factors that may account for the differences between the studies (11). Clinicians can also have a role in identifying potential novelty bias when interpreting the results of new interventions and should exhibit caution regarding newly discovered effect sizes (22). Furthermore, adherence to reporting guidelines by investigators should be encouraged to promote transparency in the conduct and reporting of studies (27).

Where a minimum of three studies/trials were included in the meta-analysis, an attempt was made by the authors to identity further primary studies which were consistent with the aims and outcome measures of the SR. The aim of this approach was to reduce potential uncertainty of the results by increasing the number of primary studies and range of publication dates. These studies were searched independently by two authors to reduce potential selection bias. The decision to select orthodontic SR topics which were considered important areas in the field of orthodontic practice was based on discussion between two authors (JS and NP). Although this may have led to potential selection bias, this was minimized during the selection process as both authors referred to literature documenting topics in orthodontics where meta-analyses have been commonly undertaken (17, 18). Meta-analyses published in only English were only included which can impact the generalizability of the results.

Conclusions

To our knowledge, this is the first study in the field of orthodontics to address the topic of novelty bias using a cumulative meta-analysis approach. In approximately 60% meta-analyses, the final effect size

estimate of the treatment intervention decreased as the evidence from primary studies evolved. There was no association between the level of risk of bias and the cumulative absolute effect size. Clinicians should be wary of the results of trials reporting the effectiveness of new interventions as there is a possibility that the reported effect size will be often exaggerated.

Supplementary material

Supplementary material is available at the *European Journal of Orthodontics* online.

Supplementary Table 1 Titles of included systematic reviews and additional primary studies or trials.

Supplementary Table 2 Outcomes, comparison groups, number of studies included, period in years, risk of bias, effect measure and 95% CIs for first study and cumulative estimate, statistical significance, and change in direction of the effect size between initial and cumulative estimate for the included meta-analyses.

Funding

Conflict of interest

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

Data availability

The data underlying this article are available in the article and in its online supplementary material.

References

1. Papageorgiou, S.N. and Eliades, T. (2019) Evidence-based orthodontics: too many systematic reviews, too few trials. *Journal of Orthodontics*, 46, 9–12.

2. Savovic, J., Turner, R.M., Mawdsley, D., Jones, H.E., Beynon, R., Higgins, J.P.T. and Sterne, J.A.C. (2018) Association between risk-of-bias assessments and results of randomized trials in Cochrane reviews: the ROBES Meta-Epidemiologic Study. *American Journal of Epidemiology*, 187, 1113–1122.

3. Altman, D.G., Machin, D., Bryant, T.N. and Gardner, M.J, eds. (2000) *Statistics with Confidence*. 2nd edn. BMJ Books, London, UK.

4. Barbui, C., Cipriani, A., Brambilla, P. and Hotopf, M. (2004) "Wish bias" in antidepressant drug trials? *Journal of Clinical Psychopharmacology*, 24, 26–130.

5. Gehr, B.T., Weiss, C. and Porzsolt, F. (2006) The fading of reported effectiveness. A meta-analysis of randomised controlled trials. *BMC Medical Research Methodology*, 6, 25.

6. Salanti, G., Dias, S., Welton, N.J., Ades, A.E., Golfinopoulos, V., Kyrgiou, M., Mauri, D. and Ioannidis, J.P. (2010) Evaluating novel agent effects in multiple-treatments meta-regression. *Statistics in Medicine*, 29, 2369–2383.

7. Cipriani, A., et al. (2018) Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. *The Lancet*, 391, 1357–1366.

8. Tinè, F., Attanasio, M., Muggeo, V.M.R. and Crainiceanu, C.M. (2017) Evidence of bias in randomized clinical trials of hepatitis c interferon therapies. *Clinical Trials*, 14, 483–488.

9. Cappelleri, J.C., Ioannidis, J.P., Schmid, C.H., de Ferranti, S.D., Aubert, M., Chalmers, T.C. and Lau J. (1996) Large trials vs meta-analysis of smaller trials: how do their results compare? *JAMA*, 276, 1332–1338.

10. LeLorier, J., Grégoire, G., Benhaddad, A., Lapierre, J. and Derderian, F. (1997) Discrepancies between meta-analyses and subsequent large randomized, controlled trials. *The New England Journal of Medicine*, 337, 536–542.

11. Catalogue of Bias Collaboration, Persaud, N. and Heneghan, C. Novelty bias. In: *Catalogue of Bias*. https://catalogofbias.org/biases/novelty-bias/ (7 January 2021, date last accessed).

12. Lau, J., Antman, E.M., Jimenez-Silva, J., Kupelnick, B, Mosteller, F. and Chalmers, T.C. (1992) Cumulative meta-analysis of therapeutic trials for myocardial infarction. *The New England Journal of Medicine*, 327, 248–254.

13. Clarke, M., Brice, A. and Chalmers, I. (2014) Accumulating research: a systematic account of how cumulative meta-analyses would have provided knowledge, improved health, reduced harm and saved resources. *PLoS One*, 9, e102670.

14. Chalmers, I., Bracken, M.B., Djulbegovic, B., Garattini, S., Grant, J., Gülmezoglu, A.M., Howells, D.W., Ioannidis, J.P. and Oliver, S. (2014) How to increase value and reduce waste when research priorities are set. *The Lancet*, 383, 156–165.

15. Jones, A.P., Conroy, E., Williamson, P.R., Clarke, M. and Gamble, C. (2013) The use of systematic reviews in the planning, design and conduct of randomised trials: a retrospective cohort of NIHR HTA funded trials. *BMC Medical Research Methodology*, 13, 50.

16. Clarke, M., Hopewell, S. and Chalmers, I. (2010) Clinical trials should begin and end with systematic reviews of relevant evidence: 12 years and waiting. *The Lancet*, 376, 20–21.

17. Summaries of Selected Systematic Reviews (2018) *Evidence-Based Orthodontics*. John Wiley & Sons, Ltd., pp. 83–91. https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119289999.fmatter2 (1 September 2020, date last accessed).

18. Papadopoulos, M.A. (2010) Meta-analyses and orthodontic evidence-based clinical practice in the 21 century. *The Open Dentistry Journal*, 4, 92–123.

19. Sterne, J.A.C., et al. (2019) RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ (Clinical Research Ed.)*, 366, l4898.

20. Ioannidis J.P. (2005) Contradicted and initially stronger effects in highly cited clinical research. *JAMA*, 294, 218–228.

21. Luo, Y., Ostinelli, E.G., Sahker, E., Chaimani, A., Kataoka, Y., Ogawa, Y., Cipriani, A., Salanti, G. and Furukawa, T.A. (2021) Antidepressant prescriptions have not fully reflected evolving evidence from cumulative network meta-analyses and guideline recommendations. *Journal of Clinical Epidemiology*, 133, 14–23.

22. Ioannidis, J.P. (2008) Why most discovered true associations are inflated. *Epidemiology*, 19, 640–648.

23. Tatsioni, A., Bonitsis, N.G. and Ioannidis, J.P. (2007) Persistence of contradicted claims in the literature. *JAMA*, 298, 2517–2526.

24. Ioannidis J.P. (2011) An epidemic of false claims. Competition and conflicts of interest distort too many medical findings. *Scientific American*, 304, 16.

25. Open Science Collaboration (2015) Psychology. Estimating the reproducibility of psychological science. *Science (New York, N.Y.)*, 349, aac4716.

26. Lau, J., Ioannidis, J.P. and Schmid, C.H. (1998) Summing up evidence: one answer is not always enough. *The Lancet*, 351, 123–127.

27. Sarkis-Onofre, R., Cenci, M.S., Demarco, F.F., Lynch, C.D., Fleming, P.S., Pereira-Cenci, T. and Moher, D. (2015) Use of guidelines to improve the quality and transparency of reporting oral health research. *Journal of Dentistry*, 43, 397–404.