HHS Public Access

Author manuscript

J Evid Based Dent Pract. Author manuscript; available in PMC 2021 March 01.

Published in final edited form as:

J Evid Based Dent Pract. 2020 March ; 20(1): 101403. doi:10.1016/j.jebdp.2020.101403.

Bayesian network meta-analysis of multiple outcomes in dental research

Menglu Liang[1], **Qinshu Lian**[1], **Georgios A. Kotsakis**[2], **Bryan S. Michalowicz**[3], **Mike T. John**[4], **Haitao Chu**[*,1]

[1]Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, MN USA

[2]Department of Periodontics, UTHealth San Antonio, San Antonio, TX, USA

[3]Department of Developmental and Surgical Sciences, University of Minnesota, Minneapolis, MN, USA

[4]Department of Diagnostic and Biological Sciences, University of Minnesota, Minneapolis, MN, USA

Abstract

Objectives—Dental research typically targets multiple outcomes. Interdental cleaning devices such as interdental brushes (IB), and waterjet devices (WJ) share a sizable portion of the medical device market. However, recommendations for device selection are limited by the conflicting evidence from multiple outcomes in available studies and the lack of an appropriate synthesis approach to summarize evidences taken from multiple outcomes. In particular, both pairwise meta-analyses and single-outcome network meta-analyses can give discordant results[1–5]. The purpose of this multi-outcome, Bayesian network meta-analysis is to introduce this innovative method to the dental research community using data from interdental cleaning device studies for illustrative purposes.

Methods—We reanalyzed a network meta-analysis of interproximal oral hygiene (IOH) methods in the reduction of clinical indices of inflammation, which included 22 trials assessing 10 IOH aids. We focused on the primary outcome of gingival inflammation, which was measured by two correlated outcome variables, the Gingival Index (GI) and Bleeding on probing (BOP).

Results—In our previous single-outcome analysis, we concluded that IB and WJ rank high for reducing gingival inflammation while TP and FL rank last. In this multi-outcome Bayesian network meta-analysis with equal weight on GI and BOP the surface under the cumulative ranking

[*]Correspondence: Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, MN 55455, chux0051@umn.edu.

Author Contributions

Ms. Liang, contributed to conception or design, drafted the manuscript; Drs. Chu, Lian, John, Michalowicz, and Kotsakis contributed to conception or design, critically revised the manuscript. All authors gave final approval and agree to be accountable for all aspects of the work.

curve (SUCRA) was 0.87 for WJ and 0.85 for IB. WJ and IB remained ranked as the two best devices across different sets of weightings for the GI and BOP.

Conclusion—In conclusion, multi-outcome Bayesian network meta-analysis naturally takes the correlations among multiple outcomes into account, which in turn can provide more comprehensive evidence.

Keywords

dental hygiene; meta-analysis; evidence-based dentistry/health care; Bayesian network meta-analysis; multiple outcomes

Introduction

The appropriate approach to synthesize research findings across studies has become a major challenge for scientists, especially when facing explosions in size and dimension of available evidences.[6]The growing interest in comparative effectiveness research and evidence-based medicine has led to a dramatic increase in the number of published systematic reviews and meta-analyses. Meta-analysis refers to a set of statistical methods to combine results from multiple independent studies to improve statistical power and reduce certain biases within individual studies,[7]and to contrast results from multiple studies to identify patterns and sources of disagreement.[8]The traditional pairwise meta-analysis compares two treatments, typically an intervention versus a control, to estimate treatment efficacy. For diseases with multiple treatment options available, a network meta-analysis synthesizes evidence from multiple treatments simultaneously.[9,10] For example, in 2015 the American Dental Association published “Evidence-Based Clinical Practice Guideline on the Nonsurgical Treatment of Chronic Periodontitis by Scaling and Root Planning with or without Adjuncts” and summarized the evidence with 72 meta-analyses.[11]A subsequent network meta-analysis (NMA)[12] summarized all evidence in one analysis, providing new insights into the treatment effects of nonsurgical periodontitis treatments and the existence of publication bias. However, the availability of multiple treatments and the consequent need for research synthesis of multiple treatment alternatives is only one challenge for research synthesis.

Synthesizing research findings to inform clinical decision making is complex because it not only involves multiple treatment comparisons, but often also involves contrasts across multiple outcomes. In a meta-analysis, typically one would select a primary outcome as the focus and supplement the results with multiple secondary outcomes. Those analyses are typically done separately. A multi-outcome meta-analysis can allow us to address multiple correlated outcomes simultaneously. Although it is possible to model correlation structures considering both outcomes and treatments under a random effects model setting, the estimations will become more complicated to accommodate numerous outcomes.[13–15]

Several methods related to NMA have been proposed, among which, Bayesian hierarchical network meta-analysis (BNMA) model is commonly used.[16]This model enables the estimation of rank probabilities of each treatment. A recent study has proposed a Bayesian hierarchical model to account for between-study heterogeneity in network meta-analysis of multiple treatments and multiple outcomes.[15]However, it is not well known in the dental

research community. The purpose of this paper is to introduce this innovative method into the dental research community to simultaneously consider multiple treatments and multiple outcomes. Specifically, we first describe the basic principles of a multi-outcome BNMA and then illustrate it via a re-analysis of data from a network meta-analysis of interproximal oral hygiene (IOH) methods in the reduction of clinical indices of inflammation.[4]

Method

The rationale behind a Bayesian hierarchical model with multiple treatments and multiple outcomes is straightforward. Suppose we have three treatments A, B, and C, as well as three outcomes O_1, O_2, and O_3. We are interested in estimating the most effective treatment as well as the differences in treatment effects of A versus B, B versus C, and A versus C. Ideally, one would integrate the treatment effects with respect to O_1, O_2, and O_3 simultaneously in the decision making. In addition, for the comparisons between treatment arms, denoted as δ_{AB}, δ_{BC} and δ_{AC}, one would integrate both direct evidence (i.e., from trials that make comparisons between the treatments of interest) and indirect evidence (i.e., using evidence from trials comparing A to C and B to C to make inferences about A versus B). Under the Bayesian framework, the expected mean value of A, B, and C with respect to each outcome O_1, O_2, and O_3 can be estimated from the posterior distribution.

Multi-outcome Bayesian network meta-analysis

Assumptions and the specification of the model can be found in the supplementary materials. In addition, the assessment of selection bias, performance bias, detection bias, attrition bias, reporting bias, and other bias should also be evaluated based on the selected articles.[17]

There are couple of steps in the estimation of ranks for each treatment. The first step is to plot a network geometry to develop a coherent description of direct and indirect evidences presented in the meta-analysis. In clinical setting, the network [9] is often complex. In some cases, several treatment arms may have sparse connections in the network meta-analysis graph. Since uninformative evidence does not guarantee a reliable estimation, this can limit the choices of outcomes. In our example data, information regarding comparisons across treatment arms differ across outcomes and we can observe a rather sparse connection in the network graph for probing depth (PD) reduction (Appendix, eFigure1).

The next step is to build a random-effects model under the Bayesian framework. The random effects are separated into two independent sources, i.e. treatments and outcomes. When the outcome is continuous, the expected mean value is assumed to have a Gaussian distribution. Non-informative priors were used for the fixed effects throughout the analysis to minimize the impact of prior information on the results and a weakly-informative prior were assigned to the covariance matrix. Treatment effects can be estimated from the posterior samples taken from Gibbs sampling method.[18]

After the estimation for fixed effects in the model, by choosing a different set of weight values, we can estimate the summarized effects for each treatment arm. Then the mean effect of treatment A can be described as $\mu_A = \sum_i \omega_i(\mu_{AO_{i2}} - \mu_{AO_{i1}})$, where $\mu_{AO_{i1}}$ denotes the

baseline measure for treatment A given outcome O_i, μ_{AO_i2} denotes the endpoint measure for A given outcome O_i, and ω_i denotes the weight for outcome O_i with constrain of $\sum_i \omega_i = 1$. And the mean rank of a specific treatment A is based on the rank of μ_A among all the other available treatments. The treatment effect between treatment A and treatment B can be summarized as $\delta_{AB} = \mu_A - \mu_B$ where μ_A and μ_B denote treatment effects which are defined above for A and B respectively.

The last step is to estimate the rank of treatments. In each Markov Chain Monte Carlo (MCMC) run, each treatment of interest will be ranked based on its estimated mean value. Then, the probability of a treatment k ranking the best among K treatments can be estimated by the proportion of MCMC iterations in which the treatment k ranks first. To incorporate the uncertainty for ranking estimation, we can estimate the cumulative rank probabilities for each treatment using surface under the cumulative rank curve (SUCRA). The SUCRA index can be calculated as:

$$SUCRA_k = \frac{\sum_{r=1}^{K-1} cum_{k,r}}{K-1},$$

where k indicates a certain treatment arm, K indicates the total number of treatment arms, and r indicates the rank. SUCRA provided a single summary estimate of relative treatment efficacy by taking the average of cumulative rank probabilities of treatment k being ranked r-th best among j treatments. A treatment is nearly unequivocally the best when SUCRA index equals to 1, and the worst if the index equals to 0.

Interpretation for the estimates

In the presence of heterogeneity or inconsistency in the network, poor quality of the included studies, or only a small amount of data available, decisions on the best treatment could be controversial since the SUCRA scores for treatments of interest could be similar. Therefore, ranking statistics, i.e., the SUCRA values should be interpreted in the context of the actual treatment effect size.[19] In the model, choices of weights could be based on practical utility, for example, giving larger weight to clinically more important outcomes. Intuitively, choosing a different weight for different outcome makes sense in practice, while on the other hand, how to select weights is worthy of thorough discussion since it will affect the conclusion and interpretation of the results. One way to investigate the impact of assigning weights for different outcomes is to provide an intuitive graphic presentation using R Shiny App incorporating different weights. This flexibility in data visualization allows users to summarize evidence using their own weights.

A case study

Example data—We now illustrate how to apply our model in practice. We conducted a systematic review and network meta-analysis of the effectiveness of means of interproximal tooth cleaning regarding reduction in plaque accumulation, gingival inflammation and pocket depth reduction in a previous study.[4] In this illustration, we used the same published data set. The literature search strategy and quality assessment of included studies can be

found elsewhere.[4]Among those identified 22 clinical trials which meet the inclusion criteria, 18 interventions against toothbrushing controls were grouped into 10 intervention categories including flossing (FL), powered flossing (FL2), toothpicks (TP), toothpicks and intensive oral hygiene instructions (TO), water jet irrigation devices (WJ), interdental brushes (IB), gum massaging devices (MD), toothbrush only (Ctrl), powered, electric, sonic toothbrush (Powered Ctrl), and Powered Control and water jet (PW). Five outcomes were identified in these 22 trials: Gingival Index (GI), Bleeding on probing (BOP), plaque control record (PCR), plaque index (PI) and probing depth (PD).

Multi-outcome Bayesian network meta-analysis—The network geometry for the current study has been published elsewhere[4] and Figure 1 also gives the general picture of complexity of the network in this study. We only consider two correlated primary outcomes, i.e., BOP and GI, in the analysis illustrating the model. The main reason for selecting only two outcomes is the data set did not have sufficient information for the other three (Appendix, eFigure 1). Sparse data mixed with complex model also could lead to unstable estimation, including convergence issues. Bayesian network meta-analysis was conducted with non-informative priors to estimate the fixed effects throughout the analysis to minimize the impact of prior information on the posterior inference. The posterior distribution of mean effects will be estimated using MCMC algorithm. JAGS software (version 3.4.0) via the "rjags" package in R software was used to sample from the joint posterior distribution using MCMC methods. The posterior samples were drawn by Gibbs sampling algorithms.[18]The marginal distributions of the mean effects were summarized by the posterior medians and 95% credible intervals. Four chains of 200,000 MCMC samples were saved after 50,000 burn-in, and convergence was assessed using the Gelman-Rubin statistics and the trace plots.[20]

Results

The selection criteria, search strategy, data extraction, and quality assessment of the original single-outcome BNMA have been reported before.[4] In total, 22 clinical trials met the inclusion criteria, therefore were included in the current multiple-outcome BNMA. Four trials had low risk of bias, one had high risk of bias and the remaining trials had unclear bias. Details of the assessments can be found in our previous publication.[4]Similar grouping strategy was used to group the interventions in the included studies as we did before in single-outcome BNMA, and Figure 1 demonstrated the network plots for the primary outcome which was assessed by GI and BOP. The network plot illustrated that majority of direct comparisons were limited to one or two studies, and comparisons for BOP had the most information in our dataset, followed by GI.

Our previous single outcome BNMA concluded that FL did not yield a substantial reduction in GI. TP and TO showed the greatest BOP reduction in comparison to control, followed by WJ. IB had the largest effect size compared to control for reducing GI.[4]

Table 1 compared the ranks of each treatment from single- with multiple-outcome network meta-analyses. The single-outcome analysis failed to make a coherent conclusion for BOP and GI. The evidences from previous pairwise comparison study on this topic also presented

a large degree of heterogeneity.[1,2,21]The inconclusive messages were mainly attributed to the large number of interventions available and the small number of studies that compared the same intervention groups.

In contrast to previous single-outcome BNMA analysis, the multiple-outcome network meta-analyses, as shown in Table 1, synthesized evidence for different treatments and outcomes simultaneously to provide a single set of ranking. Figure 2 and Table 1 shows the treatment ranks based on varying weightings of the GI and BOP. Regardless of the outcome weightings, WJ and IB were consistently ranked as the two most effective treatments. Figure 2 contrasts the changes in rank of treatments by assigning more weight to one of the primary outcome. The ranking for WJ was relatively stable as the weighting for BOP increases (and GI decreases), while the rank for IB dropped with similar weighting changes (Figure 2, Panel A). Similar results were observed when using mean ranks (Figure 2, Panel B) as the statistical assessment. Figure 3 shows an example where consistent estimation of the best treatment was achieved using different weights in BOP and GI. From the bar plot presented in Figure 3A and 3B with different weights on BOP and GI, both IB and WJ were ranked as among the “best” treatment and were never ranked worst. In particular, IB and WJ were ranked among the top four treatments. TO and IB were most likely to be categorized the best among all interventions with a weight of 0.7 in BOP (49.78% and 30.48%, respectively; Figure 3, Panel A). However, TO and WJ were most likely to be categorized the best among all interventions when assigning a weight of 0.7 in GI (49.97% and 25.44%, respectively; Figure 3, Panel B). When BOP weighting increased, the rankings changed more for treatments with lower, initial ranks. As shown in Figure 3 and Table 1, TO also turns out to be likely identified as the worst treatment (Figure 3). When BOP and GI are equally weighted (=0.5), the rank of treatments are almost in between of the results taken from BOP and GI separately (Table 1). Given that information regarding TO only presents in BOP, in Table 1, we observed that although the rank for TO is the best for BOP, but it is only in the middle under equal weight condition.

Discussion

In the current study, we illustrated a multi-outcome BNMA with an analysis of oral hygiene interventions to reduce BOP and GI. In our previous single-outcome analysis, we concluded that IB and WJ ranked highest for reducing gingival inflammation.[4]In particular, TO and WJ were most likely to be categorized as the best in terms of BOP reduction, while IB was most likely the best for GI reduction.[4]In the current analysis, IB and WJ remained highly ranked among all the treatment arms in our multiple-outcome meta-analysis. The results in our current study based on summary statistics, i.e., SUCRA, are also consistent across different sets of weights for BOP and GI. Consequently, IB and WJ can be considered as the two best treatments. Although BOP and GI are correlated, they measure two phenomena. BOP occurs when the probe is inserted to the base of the sulcus or pocket. The GI is assessed by swiping the probe along the opening of the sulcus and, if done correctly, does not impinge on the base of the sulcus. It’s possible that some candidate treatments reach further below the gum line and better affect BOP than GI, while others may clean more superficially and affect mostly the GI. Therefore, we observed differences between Figure 3, Panel A and Figure 3, Panel B.

Although, the rank of the treatment in single-outcome analysis for BOP identified TO as the most likely best treatment (Table 1), when using more appropriate summary statistics like SUCRA in multiple-outcome analysis, TO was no longer identified as the best treatment across different set of weights for BOP and GI. Also, Appendix eTable 1 and Figure 2 demonstrates that rank for TO stays in the middle across different sets of weight score. This could be caused by the fact that TO is completely missing for GI, therefore the rank for TO drops in multi-outcome BMNA when the uncertainty in its ranking for GI is taken into consideration. Additionally, since summary results for each treatment comparison now incorporate indirect evidence from related treatment comparisons as well as dependence between outcomes. The posterior SUCRA score for TO is now relatively stable around 0.5 across different sets of weight, which implies the probability for TO be the best treatment is no better than random guess. In the current analysis, we did not observe the effects of TO in the raw data for GI; thus, we estimated the effects of TO in multiple-outcome BMNA for GI via borrowing information from TO's impact for BOP. And this is also an example of using both direct and indirect evidence in BNMA. The weight score for BOP and GI can be selected based on patient-oriented condition, such as the interdental anatomy and periodontal status etc., or interest of ad-hoc analysis. Shiny App would be a better choice for data visualization for these changes(https://mengluliang.shinyapps.io/ShinyMenglu/).

Several factors differentiate our current method from traditional methods. First, the proposed approach synthesizes all studies in a complex evidence network by estimates the joint distribution of multiple outcomes for multiple treatment arms. It solves the problem of having inconsistent treatment ranking for different assessments that measure the same clinical outcome. Our model limits parameters to be estimated by modeling the relationships between treatments and outcomes simultaneously, yet independently. Secondly, since our random effects model adjusts baseline heterogeneity, it will be less biased than those estimations without taking this under control.[22–24] Furthermore, by using weighted outcomes, our model is more flexible in that it allows clinicians to make treatment recommendations based on individualized weightings of disease outcomes.

Conceptually, treatment effect, i.e., the changed health outcomes for a patient, is multidimensional. Whether a treatment works has typically multiple facets. The researcher who wants to synthesize all evidence has several options. First, the outcomes are differentiated in terms of their importance. In each individual study, a primary outcome and secondary outcome(s) may be selected to represent the most important health outcome aspects. In most cases, the primary outcome measure is the outcome that is considered the most important one. The role of secondary outcome(s) may serve as supplementary information to the primary outcome. However, different studies may select different primary endpoints. The primary endpoint in one study can be a secondary endpoint in another study. In some studies, co-primary endpoints may be used.[25–27] When decision is made via a meta-analytic, it requires a method which can incorporate evidence of all endpoints while having the flexibility to prioritize one or more endpoints of special interest. A second option, as proposed here, several outcomes are simultaneously included in the analysis. On one hand, the richness of the information would provide us with sources to synthesis evidence more globally; on the other hand, one could argue that correlations between those outcomes would make the estimation more complicated. Although couple of methods have been discussed

beforehand,[15,28–30] not a single universal solution has been recommended. Yet, ignoring the correlations between outcomes could lead to a biased estimate. A third, but conceptually different option is to capture the entire patient-perceived impact and to summarize it in one numeric score using the concept oral health-related quality of life. Although disease-oriented outcomes have several advantages such as having a clear clinical explanation and being relatively easily to quantify, they also have some drawbacks. For example, they may not reflect the whole picture in patient-perceived disease and treatment impact.[31] However, when capturing this impact with a questionnaire, one disadvantage is that all disease impacts are equally weighted in a typical analysis using questionnaires' sum scores. This does not agree with clinical intuition and this problem could be addressed with multiple-outcome BNMA analyzing the four dimensions of OHRQoL (Oral Function, Orofacial Pain, Orofacial Appearance, and Psychosocial Impact)[32,33] simultaneously. While summarizing these four dental patient-reported outcomes (dPROs)[34] is psychometrically justified, a multiple-outcome analysis could potentially provide new insights how the four components (Function, Pain/Discomfort), Pain, Psychosocial Impact) shape the patient's treatment experience. To comprehensively investigate the four-dimensional outcomes for all dental treatments across all oral conditions, multi-outcome BNMA is an essential tool for evidence-based dentistry[35,36] and value-based oral health care.[37]

In conclusion, regardless whether analyzing disease- or patient-oriented outcomes, the nature of the disease/intervention impact is multidimensional and a multi-outcome BNMA would be an appropriate tool to synthesize the evidence.

Our model possesses several desired features, however, as discussed in our supplementary materials, a couple of assumptions are made when conducting multiple outcome network meta-analysis. If these assumptions do not hold, the validity of the results will be questionable. Nevertheless, these assumptions are not different from those made by standard pair-wise meta-analysis, and both require high-quality RCT. In our case study, we only focus on the primary outcome which was assessed by two measurements while one can easily extend our model into more outcomes. Noted that including outcomes with limited information on treatment effects could lead to an inconclusive result, while excluding trials with sparse information tends to result in substantial changes in estimates. Therefore, it is essential to develop a strict systematic review protocol with careful considerations on exclusion and inclusion criteria to have robust and generalizable results from NMA.[38]

Conclusion

Clinical patient situations are complex and to address the patient's complex concerns several interventions are available that affect the patient in multiple ways. Compared to more traditional meta-analyses, the multi-outcome BNMA allows addressing these complexities better. First, multiple treatments options can be analyzed together; Second, multiple outcomes can be analyzed simultaneously with the consideration of personalized weights on outcomes. This makes the multi-outcome BNMA a more appropriate tool for clinical decisions making. Consequently, multi-outcome BNMA should become a major analysis tool for evidence-based dentistry and value-based oral health care.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

Drs. Chu, John and Kotsakis were supported by NIDCR grant R03 DE024750. Dr. Kotsakis has received an unrestricted research grant by Waterpik Inc. in the past. Ms. Liang and Drs. Lian, Michalowicz, John, and Chu report no potential conflicts of interest.

Reference

1. Noorlin I, Watts TLP. A comparison of the efficacy and ease of use of dental floss and interproximal brushes in a randomised split mouth trial incorporating an assessment of subgingival plaque. Oral Health Prev Dent. 2007;5(1):13–18. [PubMed: 17366756]
2. Rosema NAM, Hennequin-Hoenderdos NL, Berchier CE, Slot DE, Lyle DM, van der Weijden GA. The effect of different interdental cleaning devices on gingival bleeding. J Int Acad Periodontol. 2011;13(1):2–10. [PubMed: 21387981]
3. Yost KG, Mallatt ME, Liebman J. Interproximal gingivitis and plaque reduction by four interdental products. J Clin Dent. 2006;17(3):79–83. [PubMed: 17022370]
4. Kotsakis GA, Lian Q, Ioannou AL, Michalowicz BS, John MT, Chu H. A network meta-analysis of interproximal oral hygiene methods in the reduction of clinical indices of inflammation. J Periodontol. 2018;89(5):558–570. doi:10.1002/JPER.17-0368 [PubMed: 29520910]
5. Husseini A, Slot DE, Van der Weijden GA. The efficacy of oral irrigation in addition to a toothbrush on plaque and the clinical parameters of periodontal inflammation: a systematic review. Int J Dent Hyg. 2008;6(4):304–314. doi:10.1111/j.1601-5037.2008.00343.x [PubMed: 19138181]
6. Challenges Densen P. and opportunities facing medical education. Trans Am Clin Climatol Assoc. 2011;122:48–58. [PubMed: 21686208]
7. Lin L, Chu H. Quantifying publication bias in meta-analysis. Biometrics. 2018;74(3):785–794. doi:10.1111/biom.12817 [PubMed: 29141096]
8. Hoaglin DC, Hawkins N, Jansen JP, et al. Conducting indirect-treatment-comparison and network-meta-analysis studies: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 2. Value Health. 2011;14(4):429–437. doi:10.1016/j.jval.2011.01.011 [PubMed: 21669367]
9. Mills EJ, Ioannidis JPA, Thorlund K, Schunemann HJ, Puhan MA, Guyatt GH. How to use an article reporting a multiple treatment comparison meta-analysis. JAMA. 2012;308(12):1246–1253. doi:10.1001/2012.jama.11228 [PubMed: 23011714]
10. Mills EJ, Thorlund K, Ioannidis JPA. Demystifying trial networks and network meta-analysis. BMJ. 2013;346:f2914. doi:10.1136/bmj.f2914 [PubMed: 23674332]
11. Smiley CJ, Tracy SL, Abt E, et al. Evidence-based clinical practice guideline on the nonsurgical treatment of chronic periodontitis by means of scaling and root planing with or without adjuncts. J Am Dent Assoc. 2015;146(7):525–535. doi:10.1016/j.adaj.2015.01.026 [PubMed: 26113100]
12. John MT, Michalowicz BS, Kotsakis GA, Chu H. Network meta-analysis of studies included in the Clinical Practice Guideline on the nonsurgical treatment of chronic periodontitis. J Clin Periodontol. 2017;44(6):603–611. doi:10.1111/jcpe.12726 [PubMed: 28370333]
13. Abrams KR, Gillies CL, Lambert PC. Meta-analysis of heterogeneously reported trials assessing change from baseline. Stat Med. 2005;24(24):3823–3844. doi:10.1002/sim.2423 [PubMed: 16320285]
14. Gajic-Veljanoski O, Cheung AM, Bayoumi AM, Tomlinson G. A tutorial on Bayesian bivariate meta-analysis of mixed binary-continuous outcomes with missing treatment effects. Stat Med. 2016;35(12):2092–2108. doi:10.1002/sim.6791 [PubMed: 26553369]
15. Hong H, Chu H, Zhang J, Carlin BP. A Bayesian missing data framework for generalized multiple outcome mixed treatment comparisons. Res Synth Methods. 2016;7(1):6–22. doi:10.1002/jrsm.1153 [PubMed: 26536149]

16. Zhang J, Carlin BP, Neaton JD, et al. Network meta-analysis of randomized clinical trials: reporting the proper summaries. Clin Trials. 2014;11(2):246–262. doi:10.1177/1740774513498322 [PubMed: 24096635]
17. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med. 2009;6(7):e1000097. doi:10.1371/journal.pmed.1000097 [PubMed: 19621072]
18. Geman S, Geman D. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. IEEE Trans Pattern Anal Mach Intell. 1984;6(6):721–741. [PubMed: 22499653]
19. Rouse B, Cipriani A, Shi Q, Coleman AL, Dickersin K, Li T. Network Meta-analysis for Clinical Practice Guidelines: A Case Study on First-Line Medical Therapies for Primary Open-Angle Glaucoma. Ann Intern Med. 2016;164(10):674–682. doi:10.7326/M15-2367 [PubMed: 27088551]
20. Gelman A, Rubin DB. Inference from Iterative Simulation Using Multiple Sequences. Stat Sci. 1992;7(4):457–472. doi:10.1214/ss/1177011136
21. Rasines G The use of interdental brushes along with toothbrushing removes most plaque. Evid Based Dent. 2009;10(3):74. doi:10.1038/sj.ebd.6400666 [PubMed: 19820738]
22. White IR, Barrett JK, Jackson D, Higgins JPT. Consistency and inconsistency in network meta-analysis: model estimation using multivariate meta-regression. Res Synth Methods. 2012;3(2):111–125. doi:10.1002/jrsm.1045 [PubMed: 26062085]
23. Higgins JPT, Jackson D, Barrett JK, Lu G, Ades AE, White IR. Consistency and inconsistency in network meta-analysis: concepts and models for multi-arm studies. Res Synth Methods. 2012;3(2):98–110. doi:10.1002/jrsm.1044 [PubMed: 26062084]
24. Dias S, Welton NJ, Sutton AJ, Ades AE. Evidence synthesis for decision making 5: the baseline natural history model. Med Decis Making. 2013;33(5):657–670. doi:10.1177/0272989X13485155 [PubMed: 23804509]
25. Li QH. Evaluating co-primary endpoints collectively in clinical trials. Biom J. 2009;51(1):137–145. doi:10.1002/bimj.200710497 [PubMed: 19219905]
26. Tonetti M, Palmer R. Clinical research in implant dentistry: study design, reporting and outcome measurements: consensus report of Working Group 2 of the VIII European Workshop on Periodontology. J Clin Periodontol. 2012;39(s12):73–80. doi:10.1111/j.1600-051X.2011.01843.x
27. McClure JB, Blasi PR, Cook A, et al. Oral health 4 life: Design and methods of a semi-pragmatic randomized trial to promote oral health care and smoking abstinence among tobacco quitline callers. Contemp Clin Trials. 2017;57:90–97. doi:10.1016/j.cct.2017.04.003 [PubMed: 28412230]
28. Dias S, Ades AE. Absolute or relative effects? Arm-based synthesis of trial data. Res Synth Methods. 2016;7(1):23–28. doi:10.1002/jrsm.1184 [PubMed: 26461457]
29. Zhao H, Hodges JS, Ma H, Jiang Q, Carlin BP. Hierarchical Bayesian approaches for detecting inconsistency in network meta-analysis. Stat Med. 2016;35(20):3524–3536. doi:10.1002/sim.6938 [PubMed: 27037506]
30. Zhang J, Chu H, Hong H, Virnig BA, Carlin BP. Bayesian hierarchical models for network meta-analysis incorporating nonignorable missingness. Stat Methods Med Res. 2017;26(5):2227–2243. doi:10.1177/0962280215596185 [PubMed: 26220535]
31. Chiappelli F Evidence-Based Dentistry: Two Decades and Beyond. J Evid Based Dent Pract. 2019;19(1):7–16. doi:10.1016/j.jebdp.2018.05.001 [PubMed: 30926103]
32. John MT, Reissmann DR, Feuerstahler L, et al. Exploratory factor analysis of the Oral Health Impact Profile. J Oral Rehabil. 2014;41(9):635–643. doi:10.1111/joor.12192 [PubMed: 24909881]
33. John MT, Feuerstahler L, Waller N, et al. Confirmatory factor analysis of the Oral Health Impact Profile. J Oral Rehabil. 2014;41(9):644–652. doi:10.1111/joor.12191 [PubMed: 24909797]
34. John MT. Health Outcomes Reported by Dental Patients. J Evid Based Dent Pract. 2018;18(4):332–335. doi:10.1016/j.jebdp.2018.09.001 [PubMed: 30514446]
35. Hua F Increasing the Value of Orthodontic Research Through the Use of Dental Patient-Reported Outcomes. J Evid Based Dent Pract. 2019;19(2):99–105. doi:10.1016/j.jebdp.2019.04.005 [PubMed: 31326063]
36. Reissmann DR. Dental Patient-Reported Outcome Measures Are Essential for Evidence-Based Prosthetic Dentistry. J Evid Based Dent Pract. 2019;19(1):1–6. doi:10.1016/j.jebdp.2019.01.003 [PubMed: 30926098]

37. Listl S Value-Based Oral Health Care: Moving Forward With Dental Patient-Reported Outcomes. J Evid Based Dent Pract. 2019;19(3):255–259. doi:10.1016/j.jebdp.2019.101344 [PubMed: 31732101]

38. Zhang J, Yuan Y, Chu H. The Impact of Excluding Trials from Network Meta-Analyses - An Empirical Study. PLoS One. 2016;11(12):e0165889. doi:10.1371/journal.pone.0165889 [PubMed: 27926924]

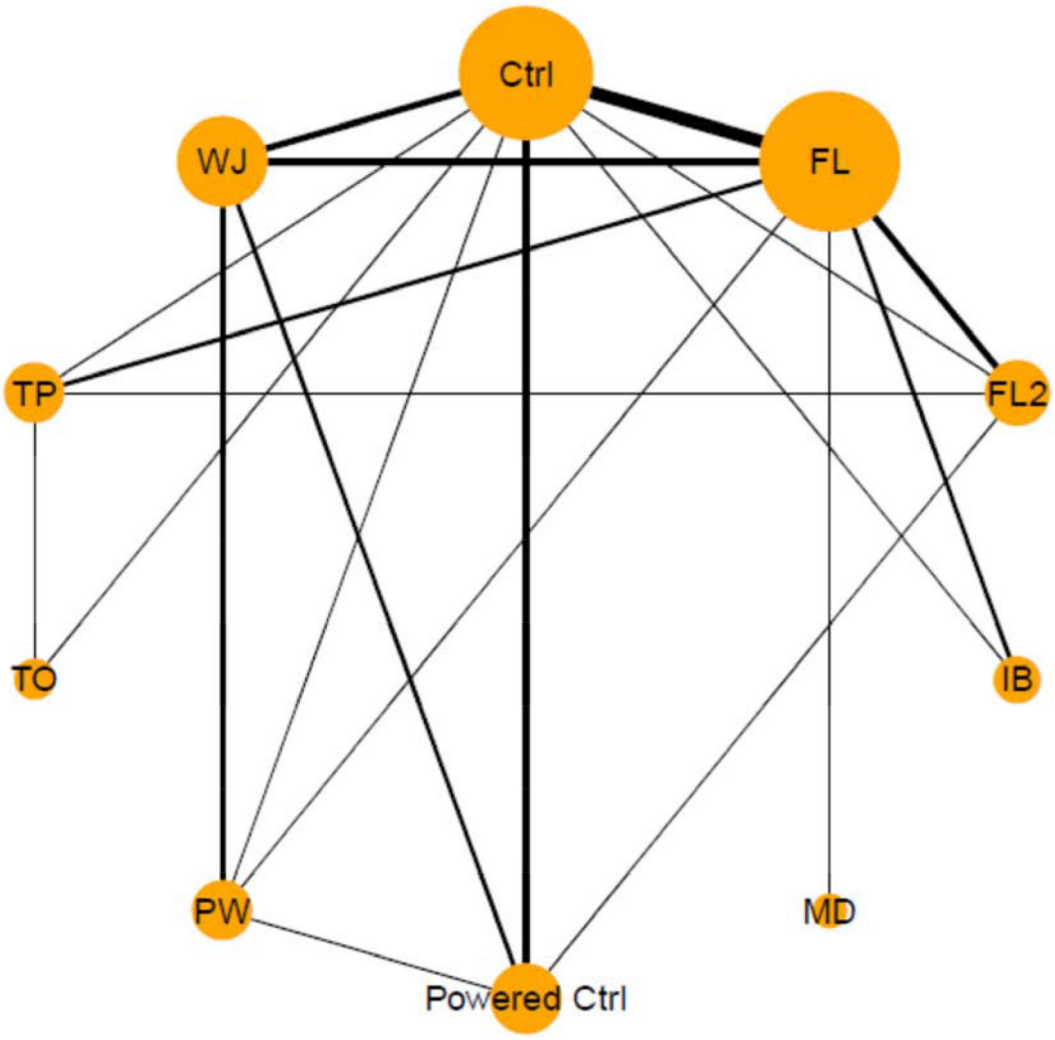

Figure 1.

Network plots of the studies assessing gingival inflammation

Nodes represent the interventions, and edges connecting two nodes indicate that the direct evidences of the corresponding intervention (flossing (FL), powered flossing (FL2), toothpicks (TP), toothpicks and intensive oral hygiene instructions (TO), water jet irrigation devices (WJ), interdental brushes (IB), gum massaging devices (MD), toothbrush only (Ctrl), powered, electric, sonic toothbrush (Powered Ctrl), and Powered Control and water jet (PW)) comparisons. The node size is proportional to the number of studies that include the corresponding intervention. The thickness of the edge is proportional to the number of studies that directly compare the corresponding pair of interventions.

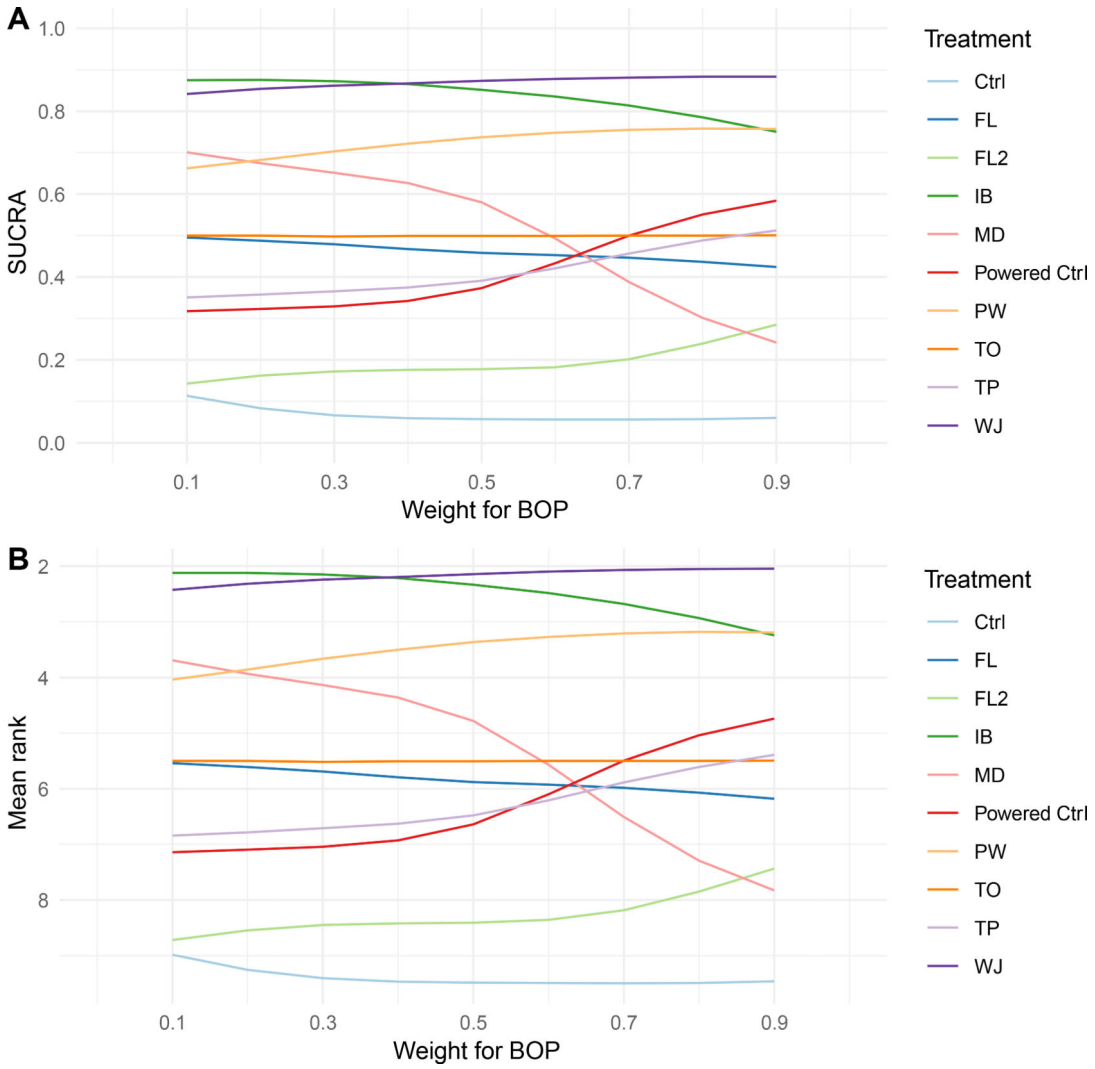

Figure 2.

Changes in rank of treatments across different combinations of weight scores

Cumulative rank probabilities for each treatment was estimated using surface under the cumulative rank curve (SUCRA). SUCRA provided a single summary estimate of relative treatment efficacy by taking the average of cumulative rank probabilities of treatment k being ranked r-th best among j treatments. A particular treatment is best almost surely when SUCRA index equals to 1, and the worst if equals to 0.

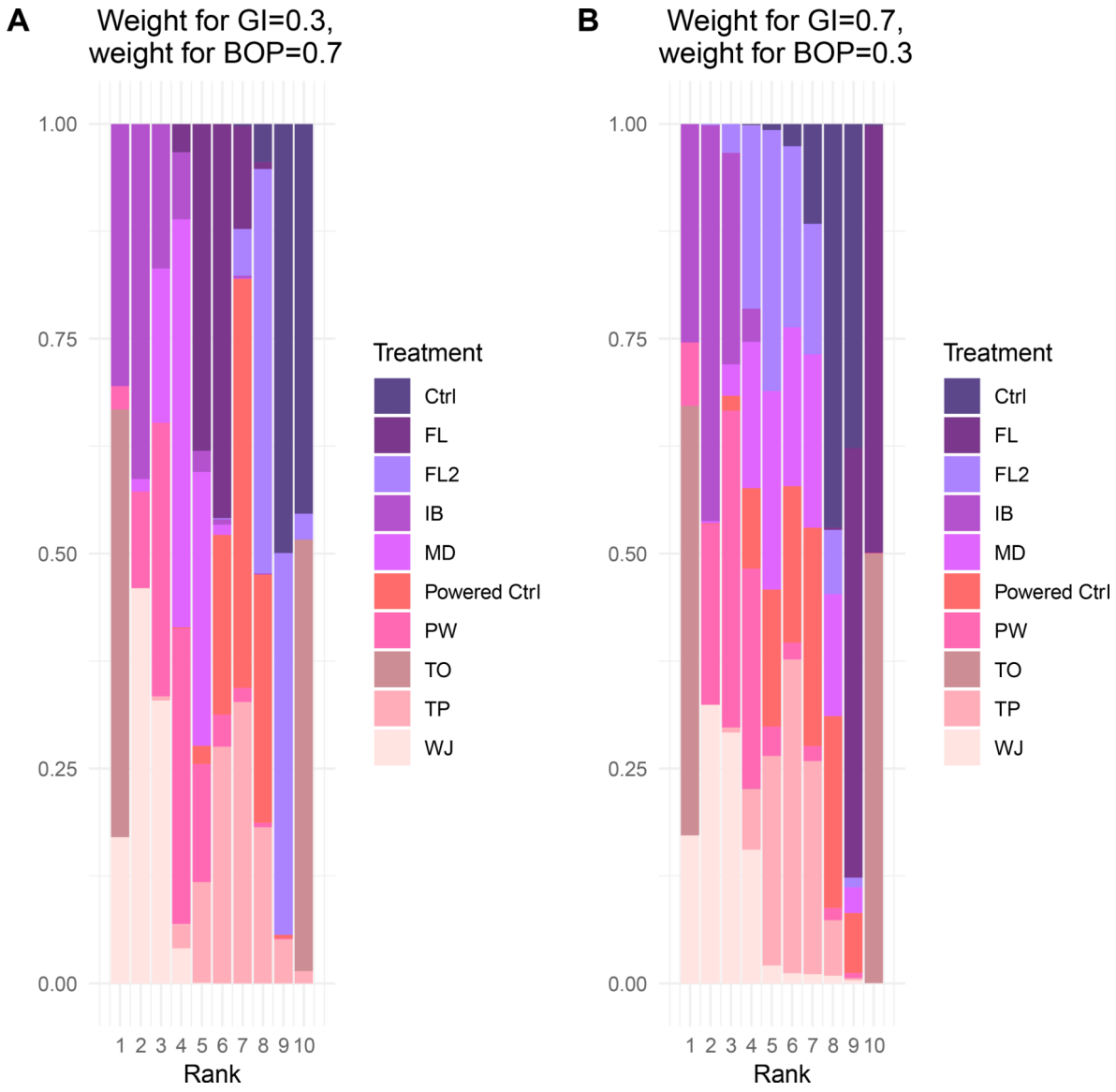

Figure 3.

Bar-plot of ranks based on weighted BOP reduction and weighted GI reduction. BOP and GI were weighted differently in Panel A and Panel B, all the 10 treatment were evaluated by the probability to be ranked as best to worst based on the estimated posterior mean effects.

Table 1

Comparison between the ranks of each treatment from single- and multiple-outcome network meta-analyses

	Single-outcome NMA[1]						Multi-outcome NMA		
	BOP			GI[5]			Weight1[2]*BOP + Weight2[3]*GI		
Treatment	**Mean rank**	**SUCRA[4]**	**SUCRA_rank**	**Mean rank**	**SUCRA[4]**	**SUCRA rank**	**Mean Rank**	**SUCRA[3]**	**SUCRA rank**
TO	1.70	0.92	1				5.50	0.50	5
WJ	2.31	0.85	2	1.95	0.88	2	2.14	0.87	1
PW	3.48	0.72	3	3.71	0.66	4	3.36	0.74	3
IB	3.83	0.69	4	1.73	0.91	1	2.33	0.85	2
TP	5.52	0.50	5	6.42	0.32	7	6.48	0.39	7
Powered Ctrl	5.72	0.48	6	6.36	0.33	6	6.64	0.37	8
FL	6.58	0.38	7	5.22	0.47	5	5.88	0.46	6
FL2	7.34	0.30	8	7.62	0.17	8	8.41	0.18	9
MD	8.63	0.15	9	3.04	0.75	3	4.78	0.58	4
Ctrl	9.89	0.01	10	8.96	0.01	9	9.49	0.06	10

[1] results for single-outcome NMA were replicated from previous published paper (Kotsakis et al. 2018).

[2] weight1 stands for weight for BOP, weight1 equals 0.5 in this table.

[3] weight2 stands for weight for GI, weight2 equals 0.5 in this table.

[4] SUCRA: Surface under the cumulative ranking curve. The larger the SUCRA value for the treatment k, the higher its rank among the available treatment options. SUCRA would be 1 when a treatment is certain to be the best and 0 when a treatment is certain to be the worst.

[5] in single-outcome analysis, data for GI only has information containing 9 treatments.