Visualising the results of clinical trials to draw conclusions

Papageorgiou, Spyridon N

DOI: https://doi.org/10.1177/14653125221101911

Feature

Journal of Orthodontics

Visualising the results of clinical trials to draw conclusions

Journal of Orthodontics
2022, Vol. 49(2) 240–243
DOI: 10.1177/14653125221101911

Spyridon N. Papageorgiou

Theoretical scenario

In the current piece I will try to show how one can acquire an overview of available data on a specific research question by visualizing the totality of the underlying evidence base (that is the existing clinical studies) on a graph. Specifically, a forest plot will be used, which is the most common way to plot the results of individual studies included in a systematic review and their quantitative synthesis (meta-analysis).

The basis for this piece will be a recently published overview of systematic reviews on the effectiveness of surgical adjunctive procedures (corticotomy, micro-osteoperforation, and piezocision among others) in the acceleration of orthodontic tooth movement (Mheissen et al., 2021). In that paper, all systematic reviews of randomised and non-randomised clinical studies investigating the effectiveness of surgical adjunctive procedures in accelerating orthodontic tooth movement were included and their results were pooled among included studies through meta-analysis. The clinical studies' data provided in the paper was adapted for this piece with slight modifications and are put on forest plots for three separate comparisons (a) micro-osteoperforations versus control (no adjunctive procedure), (b) piezocision versus control, and (c) corticotomy versus control. In all instances the rate of canine retraction per month (mm/month) in the first month after the surgical insults is used.

Emphasis is given here on the results of the separate clinical studies included in each of the three comparisons and not in their meta-analytic summary provided at the bottom of each graph. The point of this exercise is to analyse the treatment effects provided by different studies in a hypothetical clinical scenario. It should not be assumed that (i) all pertinent studies that should have been identified and used in an evidence synthesis have been included, (ii) their data has been appropriately extracted and processed, and (iii) that the data given in this piece can be used robustly to facilitate clinical decision-making. Readers are encouraged to seek the published paper and critically appraise it.

The data on the three surgically assisted adjunct procedures is given as forest plots in Figures 1–3. Briefly, a forest plot provides the result of each used study in terms of a comparative effect measure and its (im)precision in a single row. The effect measure used here is the Mean Difference (MD), which is the absolute difference in canine retraction rate in the surgically assisted (experimental) group minus the control group, given in millimetres/month, and a negative sign indicates that greater canine retraction has been observed in the experimental group than the control group. The MDs are given in the forest plots by the green boxes for each study. The black horizontal whiskers connected to each green box depict the 95% Confidence Interval (CI) of the treatment effect estimate and give a measure of how uncertain we are of the average MD of that study given the data available in this study. Wide 95% CIs could be seen due to many reasons, including among others the analysed sample size (with large sample sizes usually giving narrower 95% CIs), the variability in the outcome measurement within the clinical study, the baseline probability that the event-of-interest will happen, and the analytical strategy used. Studies with their effect estimate (green box) on the left side of the forest plot indicate faster canine retraction with surgically assisted procedures, while studies on the right side indicate a slower retraction rate compared to the control group. The vertical black line in the middle of the forest plot is the "line of no effect" and here is placed at an MD of zero. A vague rule of thumb says that if the horizontal whiskers (95% CIs) of a single study cross this vertical line it means the results of that single study (the difference between experimental-control groups) are not statistically significant at the 5% level ($P \geqslant 0.05$). However, focussing overly on the "statistical significance" of any observed result is problematic (Amrhein et al., 2019) and disregards among other things the potentially clinically meaningful impact of a truly existing treatment effect. For this reason, the provided forest plots have been augmented with colour contours that show a potential scale of magnitude for the observed effects (Papageorgiou, 2014); lighter colours indicating smaller treatment effects that might be irrelevant for everyday practice and darker colours indicating treatment effects that might make a difference

Clinic of Orthodontics and Pediatric Dentistry, Center of Dental Medicine, University of Zurich, Zurich, Switzerland

Corresponding author:
Spyridon N. Papageorgiou, Clinic of Orthodontics and Pediatric Dentistry, Center of Dental Medicine, University of Zurich, Plattenstrasse 11, Zurich, 8032, Switzerland.
Email: snpapage@gmail.com

Figure 1. Contour-enhanced forest plot depicting available clinical studies comparing adjunct micro-osteoperforations to no adjunct procedure (control group) with canine retraction rate (mm/month) within the first month. CI, confidence interval; MD, mean difference; RE, random-effects; SD, standard deviation.

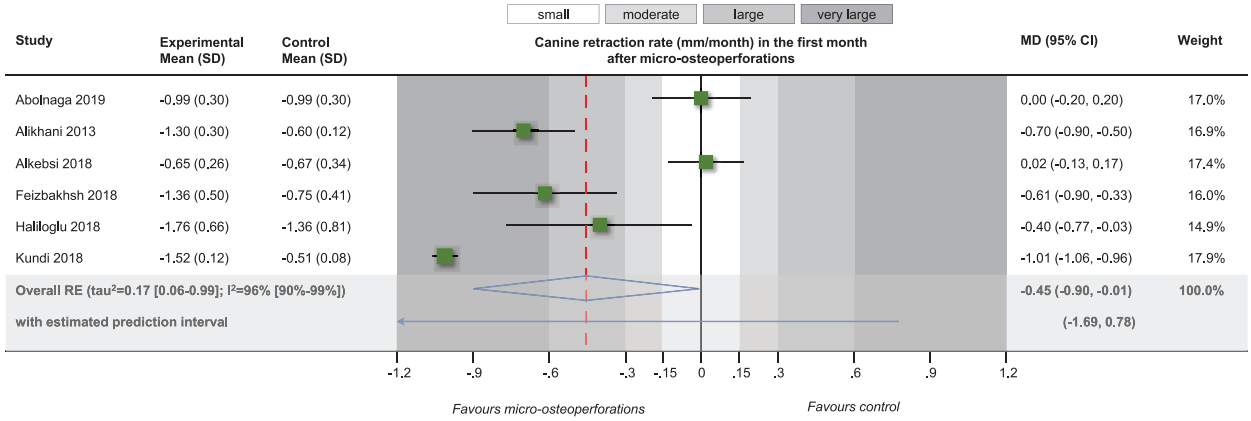
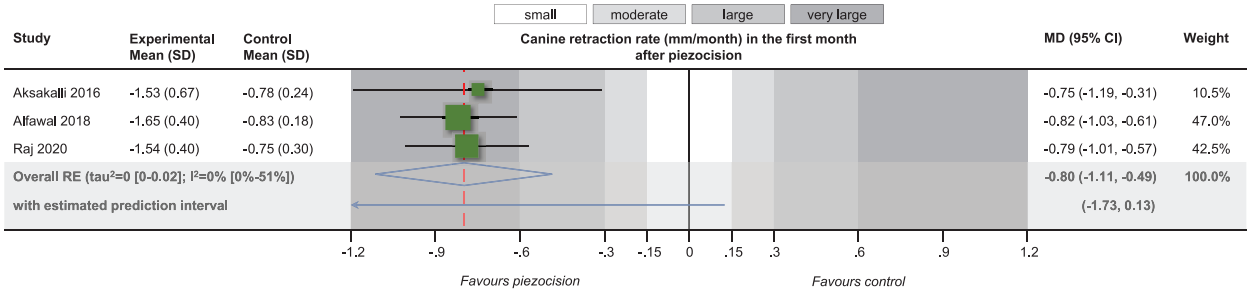


Figure 2. Contour-enhanced forest plot depicting available clinical studies comparing adjunct piezocision to no adjunct procedure (control group) with canine retraction rate (mm/month) within the first month. CI, confidence interval; MD, mean difference; RE, random-effects; SD, standard deviation.



in everyday clinical practice. It is important to stress that setting arbitrary cut-off values for the various magnitude categories (here the half, one, and two standard deviations of the response variable in the control group) is an inherently challenging process and different persons might have conflicting notions for what is a small, moderate, large, or very large effect. Furthermore, the particularities of each clinical scenario and trade-offs in terms of potential risks or costs should also be considered. The fluidity of these discriminatory cut-offs notwithstanding, assessing the clinical relevance of a treatment effect is crucial for evidence-based decision-making, as will be seen in a while.

Which of the following statements is correct, if any?

(A) Based on existing clinical studies on micro-osteoperforations (Figure 1), we can expect micro-osteoperforations to consistently lead to increased canine retraction.

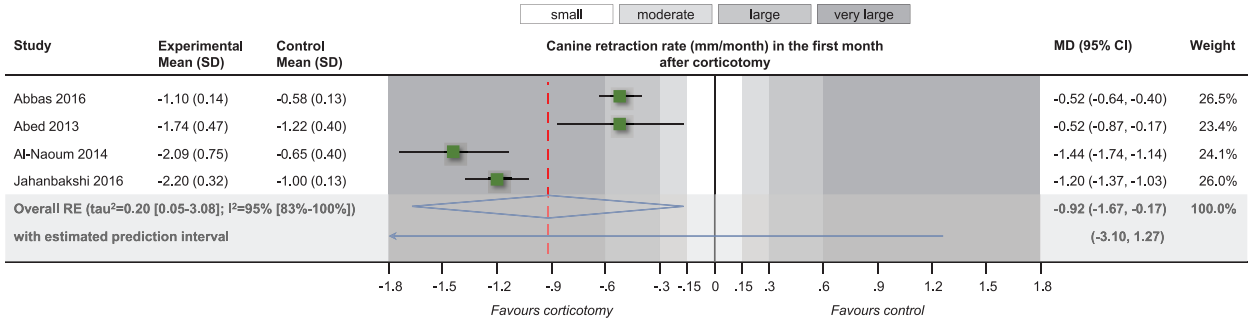
(B) Based on existing clinical studies on micro-osteoperforations (Figure 1), the clinical response to micro-osteoperforation is highly heterogeneous. This influences only our confidence in the magnitude of the micro-osteoperforations' benefit (increased canine retraction), which ranges from moderate to very large.

(C) Based on existing clinical studies on piezocision (Figure 2), piezocision can be expected to be beneficial in terms of a large to probably very large acceleration in canine retraction.

(D) Based on existing clinical studies on corticotomy (Figure 3), the clinical response to corticotomy is very heterogenous; one can expect to have either no benefit from it or have a large benefit in terms of expedited canine retraction.

(E) This observed heterogeneity in treatment response across studies from Figure 3 can influence our certainty both about the magnitude of the treatment effects (how large is the clinical benefit) and about

Figure 3. Contour-enhanced forest plot depicting available clinical studies comparing adjunct corticotomy to no adjunct procedure (control group) with canine retraction rate (mm/month) within the first month. CI, confidence interval; MD, mean difference; RE, random-effects; SD, standard deviation.



whether corticotomy is effective in accelerating canine retraction or not (if the result is statistically significant).

Discussion

The first statement indicates that all available clinical studies agree in terms of the direction of the effect of micro-osteoperforations ("consistently lead to increased canine retraction"). For this to be true, two conditions need to be met. For one, all studies should be on the same side of the forest plot (the left one that indicates micro-osteoperforations are beneficial), which means that all estimated MD should have negative values. Additionally, no horizontal whiskers (95% CIs) of any single study should cross the vertical line of no effect, since that would indicate that the findings of the clinical study are compatible with a scenario of no difference between surgically assisted and control groups ($P > 0.05$). It becomes very quickly obvious that both these prerequisites are not met, since (a) the Abolnaga 2019 and the Alkebsi 2018 studies do not have negative MDs and (b) those same two studies have 95% CIs containing zero. Therefore, statement (A) is wrong.

It is also not that difficult to spot that the response to osteo-perforations as depicted from available studies in Figure 1 is not homogenous. Evidence indicates that one might see after micro-osteoperforation a variety of effects ranging from no clinical benefit (studies on the white region) to a very large benefit (as in the Alikhani 2013 and the Kundi 2018 studies). This observed heterogeneity across studies can influence our certainty about the magnitude of micro-osteoperforations' benefit, meaning we are uncertain how big that benefit is (since we have studies on almost all the colour contours), and statement (B) seems correct. This is however not true, since there are also studies indicating that no benefit exists (the Abolnaga 2019 and the Alkebsi 2018 studies). This would lead one to conclude that the heterogeneity across existing studies can also influence our certainty about whether micro-osteoperforations work or not, depending on which studies we look at. (B) is false.

Figure 2 depicts the results of studies on piezocision. Here we can see that we have three studies that are relatively consistent with their average effects (MDs) all falling into the same magnitude category (that of a 'very large' effect). This relatively low heterogeneity is somewhat accentuated by the uncertainty around the observed results (the studies' 95% CIs), which for two of the three studies extend also to the area of a 'large' effect, while the Alfawal 2018 study is contained within a single contour. Therefore, there is some uncertainty about the true magnitude of the piezocision's clinical benefits. These most probably would be expected by most readers to be 'very large' and had all studies included larger patient samples (all three of them included < 40 patients each) then their estimates might be more precise (or their 95% CIs might be narrower). In any case however, statement (C) seems to be correct.

Statement (D) indicates that the results of studies on corticotomy are very heterogeneous. This is easy to believe, since we see that there is a gap between the expected values of the studies' treatment effects (not all their 95% CIs overlap). If we consider the 95% CIs of each separate study, we see that we might expect treatment effects of 'very large', 'large' or even 'moderate' magnitude, which means that statement (D) is wrong (as it also includes the possibility of no clinical benefit).

Finally, statement (E) implies that the observed heterogeneity across studies might lead us to doubt the actual magnitude of clinical benefits, as well as our certainty of whether corticotomy works or not. This is however not true. All available studies seem to agree on the fact that corticotomy truly has a (statistically significant) benefit in terms of accelerating canine retraction. What they don't agree on is whether this effect is moderate, large, or very large. These are two distinct conclusions drawn from the available studies and might indeed be used differently by a clinician when weighing the relative pros and contras of a surgically assisted procedure and whether this should be adopted or not.

This method of visually plotting the results of all available clinical studies on a research question has been shown to be helpful in intuitively interpreting their results from a clinician's point of view. This could be used ideally within the framework of a systematic review and the provided forest plots but could also be used independently—either by drawing these plots or mentally doing so while assessing the evidence base in order to facilitate clinical decision-making. Finally, this piece has also hinted at the possibilities to assess a study's imprecision around its observed effects and notions of heterogeneity / inconsistency across-studies, which will be further discussed in upcoming pieces.

Declaration of conflicting interests

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

Funding

The author received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Spyridon N. Papageorgiou https://orcid.org/0000-0003-1968-3326

References

Mheissen S, Khan H, Alsafadi AS and Almuzian M (2021) The effectiveness of surgical adjunctive procedures in the acceleration of orthodontic tooth movement: A systematic review of systematic reviews and meta-analysis. *Journal of Orthodontics* 48: 156–171.

Amrhein V, Greenland S and McShane B (2019) Scientists rise up against statistical significance. *Nature* 567: 305–307.

Papageorgiou SN (2014) Meta-analysis for orthodontists: Part II--Is all that glitters gold? *Journal of Orthodontics* 41: 327–336.