Of significance and relevance

Papageorgiou, Spyridon N

Of significance and relevance

Spyridon N. Papageorgiou

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

CONTACT

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

Words: 1069

Theoretical scenario

A private orthodontic practitioner wants to make his private practice more clinically efficient in terms of patient output, and therefore, financial profit. To this end he looks for ways to reduce the average treatment duration of patients undergoing a fixed appliance treatment, without compromising treatment outcomes.

The orthodontist runs a relevant electronic literature search in MEDLINE through PubMed (https://www.ncbi.nlm.nih.gov/pubmed/) limited to randomised clinical trials on treatment duration, as this study design is less prone bias (i.e. systematic error) than other designs. He finally ends up with four 2-group parallel randomised clinical trials (Trials A, B, C, and D) assessing accelerating measures to reduce overall treatment duration compared to conventional fixed-appliance treatment including novel bracket designs, adjunctive vibrational appliances, and corticotomy-assisted orthodontics.

The results of Trials A, B, C, and D are extracted in the provided Table 1 in terms of descriptive statistics for treatment duration in the experimental (fixed appliances plus proposed accelerating measure) – control group (plain fixed appliance) and statistical testing for differences between experimental - control groups with Student's t-tests for independent samples. The orthodontist uses data from Table 1 and judges that the measures from Trial B and Trial D can be used to considerably cut down mean treatment duration in a clinically relevant way.

Which of the following statements are correct, if any?

(A) Table 1 provides enough evidence to identify any clinically relevant benefits to be used in the way patients are treated.

(B) As Trial A reports statistically non-significant differences between groups, no relevant benefits are to be expected from the proposed accelerating measure.

(C) As Trial B reports statistically significant differences between groups, the accelerating measure can directly be adopted to enhance the efficiency of the orthodontist's clinical practice.

(D) As Trial C reports statistically non-significant differences between groups, no relevant benefits are to be expected from the proposed accelerating measure.

(E) As Trial D reports statistically significant differences between groups, the accelerating measure can directly be adopted to enhance the efficiency of the orthodontist's clinical practice.

Answers

Statement (A) is wrong. Table 1 reflects what is often reported in publications of clinical trials in orthodontics -- i.e. descriptive statistics for each group and inferential statistics for differences between the two groups. The former gives a broad overview of treatment duration in each group and the latter is simply a measure of the strength of evidence against the null hypothesis (of no difference between the experimental and control group). As such, it indicates that there *might* be an actual difference in treatment duration between the experimental and control group, but it does give us neither the magnitude of this estimated effect, nor our uncertainty around this estimate. Additionally, the 0.05 cut-off for P values used almost universally in biomedical literature was arbitrarily chosen by R. A. Fisher (1926) and is not based on any sound statistical reasoning. Generally, considerably more information about the results of clinical studies can be gained if we calculate the 95% Confidence Interval (CI) around an estimate (Greenland et al. 2016). This gives a range of values that would contain 95 times the true effect of the study, if the present study was conducted 100 times with the same sample.

Therefore, if we add the estimated differences in treatment duration between randomised groups and the corresponding 95% CIs to Table 1, we get Table 2. Although 95% CIs intervals are not perfect and are also often misinterpreted (Greenland et al. 2016), they are the preferred method of reporting the results of scientific studies over P values by journal editors and statisticians (Lang and Altman 2013).

Additionally, by giving a range of possible values for the trials' true effect, the 95% CIs enable quantification of the expected reduction in treatment duration based on evidence from existing trials. This in turn allows the judge how big is the reduction in treatment duration that can be gained with each accelerating measure. Based on an assumed average duration of fixed appliance treatment of about 20 months (Tsichlaki et al. 2016), we can depict the results of all four identified trials in a single graph with areas of varying magnitude in terms of 2-month (10%) changes (Figure).

With careful consideration of the Figure, it is clear that **statement (B) is correct**. Trial A reports a statistically non-significant (P>0.05) benefit of the experimental group in terms of treatment duration. This indicates that any observed differences between the experimental and the control group could be due to chance and therefore this trial cannot be used to improve clinical practice.

Statement (C) however, is wrong. Trial B reports a (statistically significant) benefit of the experimental group in terms of treatment duration compared to the control group that, according to the 95% CI, might lie between 17%-29% of the average duration. This is a reduction in treatment duration that could potentially be clinically relevant to both the orthodontist and the patient—especially if this benefit can be attained with an easy-to-implement, cheap, and safe measure, like an alternative bracket design.

Trial C reports a statistically non-significant (P>0.05) benefit of the experimental group in terms of treatment duration. This means that any observed differences between the experimental and the control group could be due to chance and therefore this trial cannot be used to improve clinical practice. Additionally, the very wide 95% CI (ranging from 59% reduction to 6% increase) together with the very small sample of included patients (just 20 patients; 10 in each group) indicates that trial C is probably not adequately powered to precisely investigate the effect of the proposed accelerating measure. So based on current evidence, no clinical benefit can be expected by the measure proposed in trial C—**statement (D) is correct**, although future adequately-powered trials might change this.

Finally, **statement (E) is wrong**. Even though trial D reports a statistically significant (P<0.05) benefit for the experimental group compared to the control group, this reduction lies according to the 95% CI between 0% and 10%. This might have little clinical relevance to either the orthodontist or the patient, especially if this benefit comes with the cost of additional treatment costs (as is the case for vibrational adjuncts) or with a potentially increased risk of adverse effects (as is the case for surgically-assisted orthodontics). Such small benefits might be desirable in some cases, but issues like cost-effectiveness and safety need to be discussed with the patient and agreed upon prior to treatment.

References

Fisher RA. 1926. The arrangement of field experiments. *J Minist Agric Gt Britain* 33:503–513.

Greenland S, Senn SJ, Rothman KJ, Carlin JB, Poole C, Goodman SN, Altman DG. 2016. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *Eur J Epidemiol*. 31:337–350.

Lang T, Altman D. 2013. Basic statistical reporting for articles published in clinical medical journals: the SAMPL Guidelines. In: *Smart P, Maisonneuve H, Polderman A (eds). Science Editors' Handbook, European Association of Science Editors*.

Tsichlaki A, Chin SY, Pandis N, Fleming PS. 2016. How long does treatment with fixed orthodontic appliances last? A systematic review. *Am J Orthod Dentofacial Orthop*. 149:308–318.

Table 1. Results of identified randomized trials on the effectiveness of various measures aimed to reduce treatment duration. All trials measured treatment duration from bond to de-bond in months.

Trial	Trial sample	Experimental Mean (SD)	Control Mean (SD)		P value	Statistically significant
A	80	18.7 (4.9)	20.1 (4.6)		0.173	No
B	160	15.7 (4.1)	20.3 (3.9)		<0.001	**
C	20	15.4 (8.2)	20.7 (7.5)		0.114	No
D	240	18.8 (3.7)	19.9 (3.7)		0.016	*

SD, standard deviation.

* P < 0.05

** P < 0.001

Table 2. Results of identified randomized trials on the effectiveness of various measures aimed to reduce treatment duration. All trials measured treatment duration from bond to de-bond in months.

Trial	Trial sample	Experimental Mean (SD)	Control Mean (SD)		P value	Statistically significant		Difference in means	95% CI
A	80	18.7 (4.9)	20.1 (4.6)		0.173	No		-1.4	-3.4 to 0.6
B	160	15.7 (4.1)	20.3 (3.9)		<0.001	**		-4.6	-5.8 to -3.4
C	20	15.4 (8.2)	20.7 (7.5)		0.114	No		-5.3	-11.9 to 1.3
D	240	18.8 (3.7)	19.9 (3.7)		0.016	*		-1.2	-2.1 to -0.2

SD, standard deviation; CI, confidence interval.

* P < 0.05

** P < 0.001

Figure Legends

Figure. Graph depicting the calculated difference in treatment duration (red boxes) between the means of experimental and control group in each identified 2-group parallel randomised clinical trial. Results are supplemented with the 95% Confidence Intervals (CI; red horizontal lines) around the difference in means. Graph is divided in gray areas of varying effect magnitude in 2-month intervals (10% of the average treatment duration).