Feature

Journal of Orthodontics

Journal of Orthodontics
2022, Vol. 49(3) 359–361
DOI: 10.1177/14653125221076142

On correlation coefficients and their interpretation

Spyridon N. Papageorgiou

Theoretical scenario

This article is based on a theoretical scenario of a clinical trial aiming to compare the dentoalveolar effects of two different methods for early correction of Class III malocclusion through maxillary protraction. For this trial, eligibility included (i) children of both sexes, (ii) in the late mixed or early permanent dentition, (iii) aged 9.0 to 13.0 years of age, (iv) having skeletal Class III malocclusion with maxillary deficiency (Wits appraisal of less than −1 mm), (v) having anterior crossbite or incisor edge-to-edge relationship, (vi) without previous orthodontic treatment, and (vi) without systemic disease or syndromes. A total of 80 children (32 girls, 48 boys) with an average age of 11.1 years (Standard Deviation [SD] 1.1 year) were ultimately included and randomised on a 1:1 basis to maxillary protraction with either a hybrid-protraction protocol (a hybrid expander anchored on 2 midpalatal miniscrews and 2 mandibular miniscrews positioned bilaterally distally to the permanent canines (Miranda et al. 2021); group 1) or a control group (group 2) with a conventional dentally anchored expander in the maxilla (Mandall et al. 2010) that has shown good results in the short- (Mandall et al. 2012) and long-term (Mandall et al. 2016).

Data are based off a recent publication (Miranda et al. 2021), but the sample has been doubled, and here only the outcome of length of the skeletal maxilla (Condylion (Co)-A point) from lateral cephalograms is analysed before and after treatment (Table 1).

In this piece the effect of patient age on the sagittal length of the maxilla is mostly discussed, and analysed statistically using Pearson's correlation with significance level set arbitrarily at 5%. The following results are given in Table 2 in terms of correlation coefficients, which can take any value ranging from -1 (perfect negative correlation), 0 (no correlation), to +1 (perfect positive correlation).

There are also several classifications of magnitude, like that of Evans (1996), which interprets:

- correlations <0.20 as very weak,
- correlations between 0.20-0.39 as weak,
- correlations 0.40-0.59 as moderate,
- correlations 0.60-0.79 as strong, and
- correlations >0.80 as very strong.

However, these cut-offs are set arbitrarily to refer to linear associations, which do not always exist. Therefore, such classifications should be used judiciously or avoided, and interpretation of correlation coefficients should be specific to the subject area. Nevertheless, higher absolute values and smaller associated P values are traditionally taken to imply a stronger departure from a null hypothesis of no correlation.

Which of the following statement are correct, if any?

(A) There is overall a moderate correlation between patient age and maxillary length.
(B) For each additional patient year, an increase of 0.38 mm in Co-A is expected.
(C) This moderate age-maxillary length correlation from (A) means that for each additional patient year a moderate increase in maxillary length is expected.
(D) The correlation between patient age and maxillary length is much larger in group 1 than in group 2, which means that for each additional patient year, a greater increase in maxillary length is expected in group 1 than in group 2.

Discussion

Starting with the basics, an overall correlation coefficient of 0.38 was observed between Co-A after treatment and baseline patient age. This according to conventional means can be taken as an indication of a moderate correlation between age and maxillary length. So (A) is true.

However, correlation coefficients are often misinterpreted by both laypersons and researchers in the biomedical field. Correlation coefficients give us an indication about

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, CH 8032, Switzerland.
Email: snpapage@gmail.com

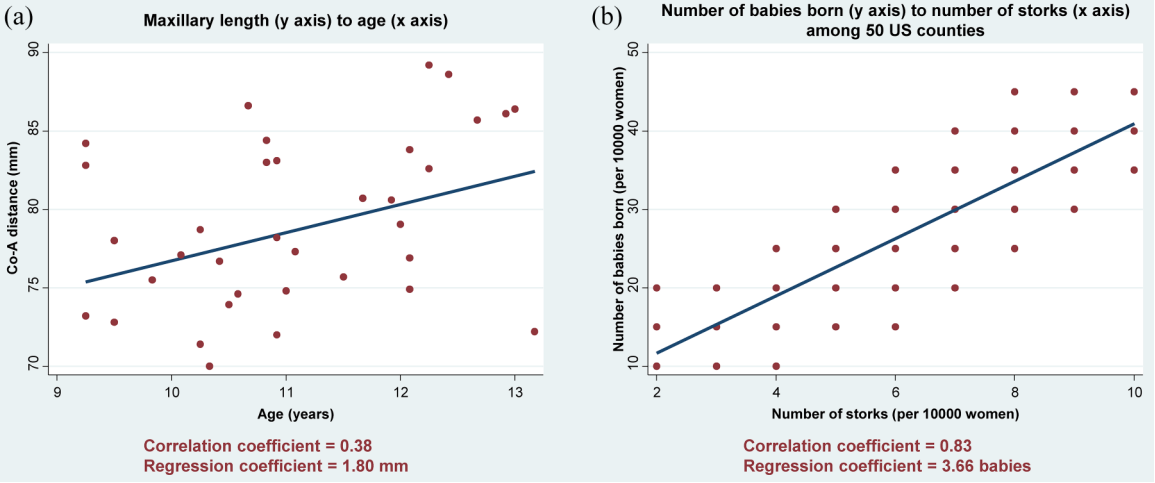
Table 1. Maxillary length measurements, given as means with SDs in parentheses.

Variable	Overall	Group 1	Group 2
Co-A before treatment (mm)	77.4 (5.0)	76.2 (5.3)	78.6 (4.5)
Co-A after treatment (mm)	78.7 (5.4)	76.9 (5.7)	80.5 (4.5)

Table 2. Correlation analyses run on the hypothetical trial's data between baseline patient age and Co-A distance after treatment.

Metric	Overall (80 patients)	Group 1 (40 patients)	Group 2 (40 patients)
Pearson's correlation coefficient	0.38	0.60	0.41
P value	<0.001	<0.001	<0.001

Figure 1. Scatter plots of data (a) from the theoretical maxillary protraction trial and (b) from the fictional data on storks and babies.



the strength of an association between two variables – that is, how well do the data fit (or are close) to a hypothetical best-fitting straight line between the two variables. Correlation coefficients do not quantify the amount one variable changes as the other increases. Or here in our example the correlation coefficient of 0.38 does not tell us how much does Co-A changes for each one-year increase in patient age. To reach this conclusion, one would need to do a linear (least squares) regression analysis of patient age on Co-A length. Its output, as one might see in an orthodontic journal, would typically include an unstandardised regression coefficient of 1.80 mm, a 95% Confidence Interval (CI) of 0.81 to 2.79 mm, and a p value being <0.001. This could then be interpreted as follows: for each additional year of patient age, one might expect on average a 1.80 mm increase in Co-A length (and which increment might actually lie somewhere between 0.81 and 2.79 mm). If we wanted to see this graphically, this increase of 1.80 mm would be actually the slope of the 'least squares' blue line in Figure 1a. Based on this, statement (B) is wrong.

We can compare the differences between correlation coefficients and coefficients from regression analysis using an empirical example from Neyman (1952), which is often used in statistical lessons as an example of spurious correlation. This example is based on data gathered for a single year from 50 US counties regarding the number of babies born and the number of storks (both divided by the number of existing women to adjust for population size) and aims to verify the theory that storks bring babies. As can be seen in Figure 1b, the data from Neyman are gathered much more closely to the best-fitting blue line, which fits to the correlation coefficient of 0.83. Additionally, each additional stork is associated with 3.66 babies being born (both per 10000 women; p value<0.001).

If one wanted now to assess the magnitude of a regression coefficient, and whether this corresponds to a clinical effect

of small, moderate, large, or very large magnitude, considerable differences exist. This decision entails both a large degree of subjectivity and critical thinking since (a) a large effect for one clinician might be considered moderate or small for another, (b) different risks for adverse effects and costs might be expected for different treatment options, and (c) these decisions are subject-specific. In an effort to (over-) simplify this, one might use the SD of the response variable as a gauge (here the baseline Co-A) and denote a small effect as one being up to 0.5 SD, a moderate being 0.5-1.0 SD, and a large being 1.0-2.0 SDs (with the abovementioned misinterpretation risks). Using this approach, a Co-A increase of 1.80 mm per patient year falls is less than 2.50 mm (half SD) and therefore the effect of age on maxillary length would probably be considered small (and statement C is wrong).

Coming to the final statement of comparing the effect of patient age on the Co-A distance in group 1 and group 2 now. One might indeed find that the correlation coefficient between these two variables is larger in group 1 (0.60) compared to group 2 (0.41), but this does not necessarily mean that each additional patient year is associated with different increase in Co-A length, only that the data fit to the best-fit line better for this group. In order to quantify this association, again, a linear regression model would be needed, which takes into account patient age, treatment-grouping, and a formal association term between these two. Indeed, the results of this model do not indicate that patient age affects differently the Co-A distance (interaction-term p value=0.30)—which is of course what we might expect on biological grounds. Statement D is wrong.

Possible explanations for this difference in correlations might lie with other differences, such a potential small difference in patient age between the two groups in this fictional example (age in group 1: 11.4±1.2 years; age in group 2: 10.8±1.0 years; independent-samples t-test P=0.01), which might happen also in real clinical trials purely due to chance (Roberts and Torgerson, 1999).

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

Anne Mandall N, Cousley R, DiBiase A, Dyer F, Littlewood S, Mattick R, Nute S, Doherty B, Stivaros N, McDowall R, Shargill I, Ahmad A, Walsh T and Worthington H (2012) Is early Class III protraction facemask treatment effective? A multicentre, randomized, controlled trial: 3-year follow-up. *Journal of Orthodontics* 39: 176–185.

Evans JD (1996) *Straightforward Statistics for the Behavioral Sciences*. Brooks/Cole Publishing; Pacific Grove, Calif.

Mandall N, Cousley R, DiBiase A, Dyer F, Littlewood S, Mattick R, Nute SJ, Doherty B, Stivaros N, McDowall R, Shargill I and Worthington HV (2016) Early class III protraction facemask treatment reduces the need for orthognathic surgery: a multi-centre, two-arm parallel randomized, controlled trial. *Journal of Orthodontics* 43: 164–175.

Mandall N, DiBiase A, Littlewood S, Nute S, Stivaros N, McDowall R, Shargill I, Worthington H, Cousley R, Dyer F, Mattick R and Doherty B (2010) Is early Class III protraction facemask treatment effective? A multicentre, randomized, controlled trial: 15-month follow-up. *Journal of Orthodontics* 37: 149–161.

Miranda F, Cunha Bastos JCD, Magno Dos Santos A, Janson G, Pereira Lauris JR and Garib D (2021) Dentoskeletal comparison of miniscrew-anchored maxillary protraction with hybrid and conventional hyrax expanders: A randomized clinical trial. *American Journal of Orthodontics & Dentofacial Orthopedics* 160: 774–783.

Roberts C and Torgerson DJ (1999) Understanding controlled trials: baseline imbalance in randomised controlled trials. *British Medical Journal* 319: 185.

Neyman J (1952) *Lectures and Conferences on Mathematical Statistics and Probability*, 2nd edn, pp. 143-154. Washington DC: US Department of Agriculture.