Scientific Section

Journal of
Orthodontics

Journal of Orthodontics
1–6
DOI: 10.1177/14653125221094333

journals.sagepub.com/home/joo
SAGE

Effect of prior knowledge about treatment on cephalometric measurements

Sabarinath Prasad[1,2], Gloria Denotti[3] and Mauro Farella[1]

Abstract

Objective: We hypothesised that prior knowledge of details for a growth modification treatment influences cephalometric measurements and results in a detectable bias.

Design: Observational study.

Setting: University teaching facility.

Methods: Six orthodontic residents assessed 48 lateral cephalograms taken before and after functional appliance treatment from 24 patients. The residents assessed six cephalometric measurements, (Cd-Pog, Cd-Me, Ar-Pog, Ar-Me, Go-Me, SNB) over three separate sessions, in either a random concealed order or as matched pairs with information about treatment and time disclosed.

Results: When information was disclosed, five out of the six cephalometric measurements were significantly higher that the corresponding cephalometric measurements taken randomly with undisclosed information. The bias was in the range of 1.6–3.2 mm for linear measurements and was 1.1° for SNB.

Conclusion: Disclosing treatment information does introduce systematic errors in cephalometric measurements. Cephalometric analysis in orthodontic clinical research should be carried out by assessors who are blinded to treatment details, to minimise risk of bias.

Keywords

cephalometry, bias, reproduciblity of results

Date received: 25 December 2021; revised: 27 March 2022; accepted: 29 March 2022

Introduction

Randomised controlled trials (RCTs) are considered the gold standard study design for clinical research, with systematic reviews (SRs) and meta-analyses (MAs) of RCTs being at the top of the hierarchy of evidence. Randomisation of interventions and blinding are important methodologic aspects of RCTs, which are undertaken to meet the assumptions of statistical testing and to minimise the risk of bias (Day and Altman, 2000). Assessors who are aware of treatment allocation or stage may introduce a measurement bias because they have an expectation about outcome, which in turn can result in an unintentional alteration of their assessment.

The effect of blinding in clinical research has now been well-established in medical literature (Hróbjartsson and Boutron, 2011; Poolman et al., 2007) and in oral health interventions, where non-blinded assessors tend to consistently overestimate treatment effect size (Saltaji et al., 2018). Reporting of blinding in orthodontic literature has been poor (Harrison, 2003) and remains suboptimal (Sandhu et al., 2015). Single-blinding can be found in 15% of orthodontic RCTs, while double-blinding or triple-blinding are rare (<3%) (Alharbi and Almuzian, 2019). Blinding in orthodontic clinical research is difficult primarily due to the

---

[1]Discipline of Orthodontics, Faculty of Dentistry, University of Otago, Dunedin, New Zealand
[2]Department of Orthodontics, Hamdan Bin Mohammed College of Dental Medicine (MBRU), Dubai, UAE
[3]Discipline of Orthodontics and Pediatric Dentistry, Department of Surgical Science, University of Cagliari, Cagliari, Italy

Corresponding author:
Sabarinath Prasad, Discipline of Orthodontics, Department of Oral Sciences, Faculty of Dentistry, University of Otago, PO Box 56, Dunedin, 9054, New Zealand.
Email: drsabarinath@gmail.com

visibility of appliances used in treatment unlike trials involving medications, where one can easily achieve blinding using placebos, but even when feasible, blinding is often not implemented in orthodontic clinical research (Abdulraheem and Bondemark, 2019).

Cephalometric measurements are widely used to assess the outcome of facial growth modification and other orthodontic treatments (Flores-Mir and Major, 2006; Tulloch et al., 1990). Nonetheless, they are prone to errors because anatomical structures may be difficult to identify on the radiographic images and are interpreted subjectively. This makes cephalometric measurements at a high risk of bias, especially when the assessors are not blinded to the treatment allocation and to the time radiographs were taken. However, to the best of our knowledge, no study has attempted to quantify biases associated with measurements on lateral cephalograms in relation to a growth modification treatment.

The aim of this study was to test the hypothesis that prior knowledge of treatment details for a growth modification treatment will influence cephalometric measurements and will result in a detectable systematic bias.

Material and method

Lateral cephalograms used in the study were of individuals included in a large-scale orthodontic RCT that aimed to evaluate the efficacy of a functional appliance (Cioffi et al., 2008). Individuals were included in the RCT if they were in the late mixed / early permanent dentition, had at least half unit bilateral Class II molar and canine relationships, an initial overjet ⩾6mm, mild lower arch crowding, cervical vertebrae maturation stage of 3 or 4, and a retrognathic mandible. Patients with previous orthodontic treatment or craniofacial anomalies were excluded from the study. Participants were treated by a bite jumping functional appliance to correct the underlying Class II dental and skeletal malocclusion. Treatment time was in the range of 12.0–18.2 months.

The radiographic images consisted of scanned high resolution JPEG files, which were adjusted for optimal contrast and cropped to A4 size, while keeping the native resolution and magnification by software (Adobe Photoshop Element). Images were then printed on A4 (210 × 297 mm) size transparent plastic films using a business photocopy machine (Image Runner Advance C3500i III; Canon, Tokyo, Japan). The images were de-identified from patients' details and dates, and labelled using non-ordered sequences by student interns.

Postgraduate residents attending the postgraduate programme in Orthodontics at the University of Otago participated as assessors. All residents had received the same training in cephalometric analysis at the same centre, had at least a year of previous experience, and had successfully demonstrated adequate competence in tracing cephalometric radiographs in examinations conducted as a part of their initial orthodontic training. No additional steps for calibration of the assessors were undertaken.

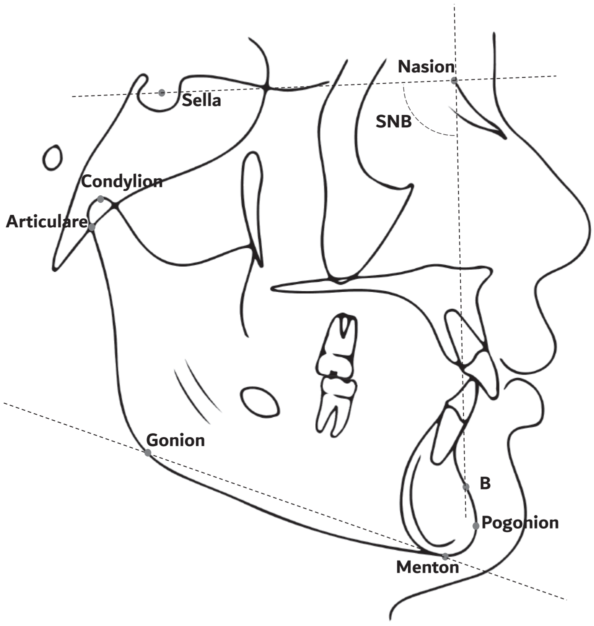
Each assessor took part in three separate sessions, which were separated by a two-week memory washout period. At each session, the six assessors were provided with a set of acetate tracing sheets and were requested to undertake six cephalometric measurements from 16 radiographs, which were taken from eight patients at T0 and T1. During the first session (S1), each assessor received a first block of 16 radiographs. The radiographs were presented to assessors using simple random sequences generated by on-line tools, with information about treatment and time (i.e. before and after status) pertaining to the case being withheld. During the second session (S2), each assessor received a second block of 16 radiographs, which were taken from eight different patients, and again given in a random order. This measurement session was used to estimate method errors of cephalometric variables measurements between different sessions and assessors.

During the third session (S3), a third block of 16 radiographs was presented to each assessor as matched pairs, with the before and after status disclosed. None of the residents involved in the tracing were aware that they were assessing lateral cephalometric radiographs of participants who underwent functional appliance therapy during the first session (S1). Information that study participants had received treatment by a functional appliance was disclosed to the assessors only during the third session (S3). At the end of the study, a total of 288 cephalometric tracings were obtained, with each assessor tracing 48 radiographs over three separate sessions.

Definitions of the cephalometric landmarks were provided before the start of all sessions. As the patients had Class II skeletal relationship and were treated by functional appliances, measurements providing information on the size and position of the mandible were chosen in this study. Five linear (Cd-Pog, Cd-Me, Ar-Pog, Ar-Me and Go-Me) and one angular (SNB) cephalometric variables were assessed (Figure 1). All measurements were carried out under the same conditions and in the same clinical research environment. Tracing was done on cephalometric tracing paper (3M Unitek, USA) with a 0.5-mm HB mechanical pencil on illuminated fluorescent light boxes. All linear measurements were made to the nearest 0.5 mm and angular measurements to the nearest 0.5°. Cephalometric measurements were annotated on the tracing sheet and then entered in an Excel spreadsheet. No time constraints were placed for the measurement of the radiographs in any of the sessions.

A generalised linear model (GLS) was used for statistical analysis. Changes in cephalometric variables were used as response variables, while condition (i.e. disclosed vs. undisclosed) as factor. A random term was also entered in the model to identify the examiner in a mixed model analysis. Fixed effects were tested by Type III tests in SPSS version

Figure 1. Cephalometric landmarks and measurements used in the study. Ar, articulare; Cd, condylion; Go, gonion; Me, menton; N, nasion; Pog, pogonion; S, sella, Point B.



20.0 (IBM Corp., Chicago, IL, USA). Bland–Altman (BA) limits of agreements were carried out in order to assess inter-assessor agreement for cephalometric measurements.

Results

The BA plots for the linear and angular cephalometric measurements are given in Figure 2. The agreement between measurements taken from different assessors and across different sessions vary largely between cephalometric variables, being lowest for the linear measurements, which included Cd and Go as landmarks, and highest for Ar-Pog and SNB (Figure 2).

Comparisons of cephalometric measurements obtained under disclosed and undisclosed conditions, showed that the measurements were consistently higher when the assessors received the radiographs as matched pairs and were informed about treatment status and time. The bias was consistently positive for all measurements and was in the range of 1.5–3.0 mm for linear measurements and amounted to 1.4° for SNB. Differences were statistically significant ($P \leqslant 0.033$) for all cephalometric variables with the exception of Ar-Pog (Figure 3).

Discussion

This study showed that assessment of cephalometric measurements related to the size and position of the mandible are influenced by information regarding treatment delivered and timing of radiographs, resulting in a detectable bias.

When cephalometric radiographs were presented in a random concealed order, cephalometric measurements varied largely between different sessions (S1 and S2) and assessors. Landmarks Ar and/or Cd commonly represent the posterior endpoint limit of the mandible in linear cephalometric measurements of mandibular length. Cd-Pog, Cd-Me, Ar-Pog and Ar-Me were measurements used to represent mandibular length in this study. Cd is a difficult landmark to identify in closed-mouth lateral cephalograms, as it is often obscured by the superimposition of cranial base and middle cranial fossa structures. Measurements taken from Cd showed large variability in this study consistent with a previously published meta-analysis (Trpkova et al., 1997). Although close correlations between the Ar-Pog and Cd-Pog distances have been reported (Haas et al., 2001), evidence from this study suggest the contrary as linear measurements taken from Ar showed better repeatability than those from Cd. The Go to Me measurement also showed large variability as Go is a constructed landmark on a curvature and requires multiple steps to locate; making it more prone to measurement errors (Baumrind and Frantz, 1971). The good repeatability for SNB was expected as cephalometric landmarks S, N and Point B have all been found to be relatively easy to locate accurately (Trpkova et al., 1997).

Suspecting bias when outcome assessors are non-blinded in a trial is currently based on assumptions. Presently, the size and direction of outcome assessor bias (if any) is unknown in trials interpreting cephalometric data as no existing research has looked at the amount of bias when treatment information is known to outcome assessors. In this study, it was seen that despite having received the same training and having similar skills, systematic differences in the cephalometric measurements of the assessors were found when treatment information was disclosed and the radiographs were presented as matched pairs.

‘Treatment effect’ refers to the causal effect of a given treatment or intervention (in this study, growth modification) on an outcome variable of interest (e.g. size/position of the mandible). Oral health intervention studies report that non-blinded assessors tend to consistently overestimate treatment effect size (Saltaji et al., 2018). Statistically significant differences between blinded and unblinded assessors was seen in five out of six cephalometric measurements in this study, suggesting that bias in cephalometric measurements may result when outcome assessors are aware of the intervention received (in this case, a growth modification treatment by a functional appliance), leading to a detectable bias. When assessors were provided all treatment information, a greater risk of assessors erring on certain measurements to reflect the likely treatment outcome was found.

Figure 2. Bland–Altman plots showing a set of duplicate cephalometric measurements taken by different examiners, over a two-week period, and with undisclosed information about treatment.

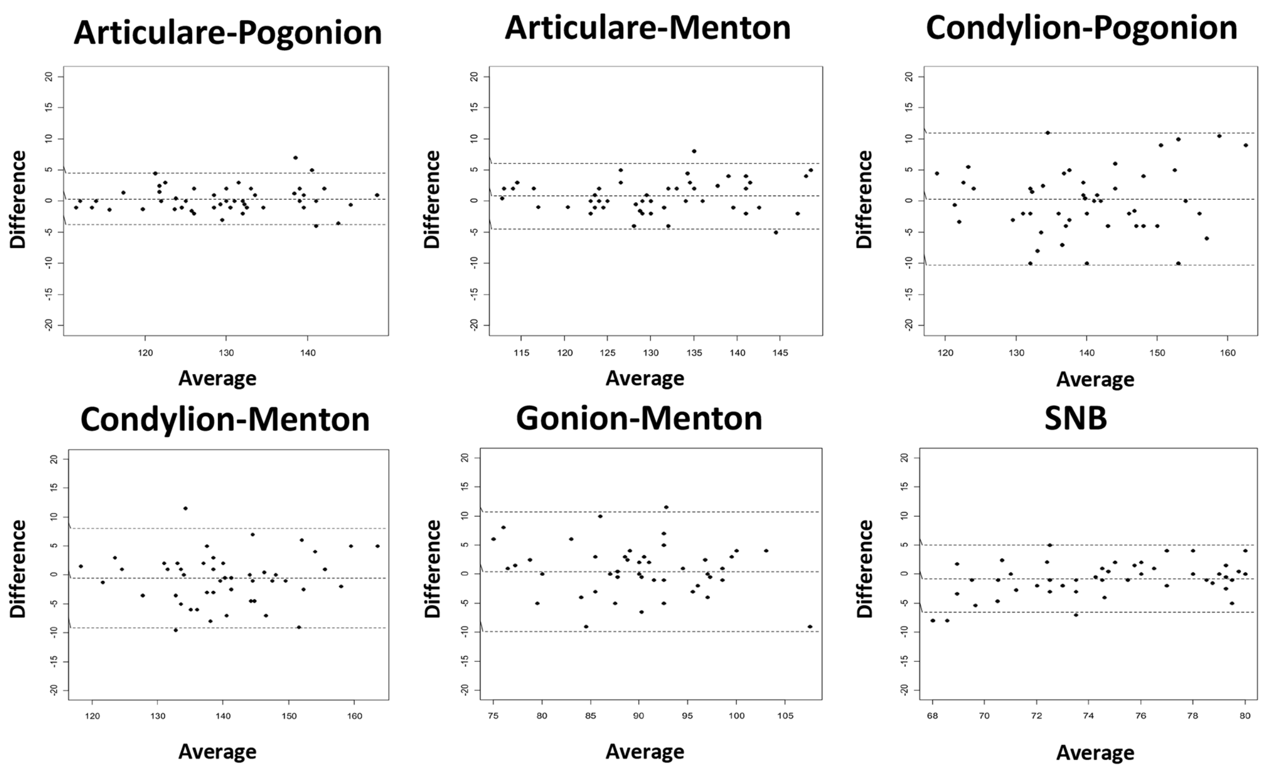
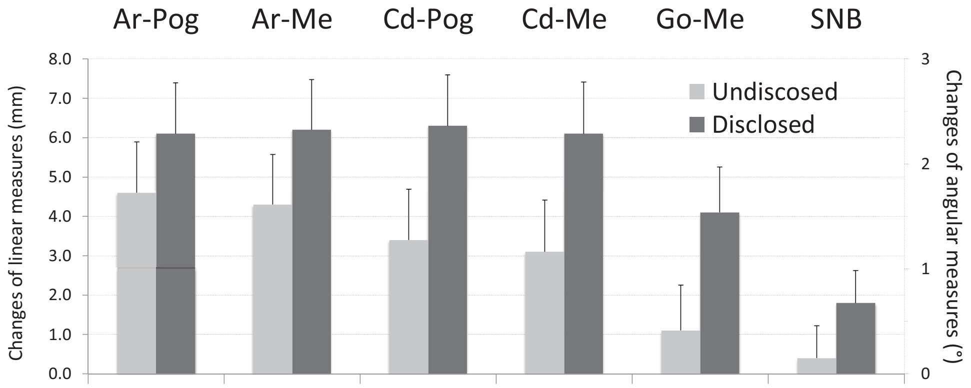


Figure 3. Change in linear and angular cephalometric measurements before and after treatment under undisclosed and disclosed conditions. The two sessions were separated by a two-week period. Under undisclosed conditions, the radiographs were given in a random and concealed order, while under disclosed conditions, the radiographs were given as matched pairs and the assessors were informed about treatment delivered (i.e. a functional appliance) and stage (i.e. before or after treatment). The height of the bars represents mean values while the error bars represent the upper 95% confidence limit.



Results from a meta-analysis (Cacciatore et al., 2019) report a 2.9-mm mandibular length (Cd-Gn distance) increase after growth modification treatment at the end of growth (95% confidence interval [CI] = 0.5–5.3) and 3.2 mm at 18 years and above (95% CI = 1.3–5.1). Angular improvement of the mandibular projection (SNB angle) reported was also significant at 18 years of age (mean difference = 0.7°, 95% CI = 0.0–1.3). Interestingly, the magnitude of the biases found in this study may be equal to or even larger than the cephalometric changes that have been previously reported for growth modification appliances.

Knowledge of intervention received by study participants has likely influenced the assessment of outcome. This emphasises that when there is previous knowledge of the effect of an intervention and when the intervention is known, unconscious expectations of the assessor may lead to biased measurement of treatment outcomes. Personal a priori expectations of researchers may obscure their objectivity and influence their analytical reasoning. This may in turn lead to unconscious manipulation of data in the form of differential rounding up/down, reassessing atypical observations and discarding inconvenient data. For example, in the past, unconsciously biased craniometric measurements of skull volume have led to flawed conclusions relating racial differences in skull volume to intellectual faculties (Gould, 2016). The bias identified in this study may account to explain why the evaluation of skeletal effects of growth modification appliances still remains contentious after more than a century of clinical research (Cacciatore et al., 2019; D'Antò et al., 2015). Although this study involved measurement of cephalometric radiographs of participants who had growth modification treatment, it is reasonable to expect that similar systematic biases may exist in cephalometric measurements involving other orthodontic treatment modalities. Even if blinding of participants and personnel during orthodontic trials is not feasible, blinding of outcome assessors in most instances is, and needs to be practised to reduce risk of bias. Increased automation in RCTs (Soboczenski et al., 2019) and the emergence of artificial intelligence based cephalometric measurements (Yu et al., 2020) may also help reduce bias in future.

This study has some limitations. First, the study was carried out in a convenience sample and no prior sample size calculation was performed. In addition, being an observational study, any inferences on causation needs to be avoided. Second, it may be argued that as the number of assessors and tracing occasions increases, it may produce a larger magnitude of error. However, this has been found to be not true (Trpkova et al., 1997). Third, even though all the assessors received the same training in cephalometric analysis, being postgraduate students they were relatively inexperienced and this may have also possibly contributing to the large variability of some of the measurements. Next, although no time limits were imposed on the assessors, it is possible that some assessors may have rushed the analyses. Lastly, since cephalometric measurements in this study were manual, there is the possibility of fatigue arising and possibly affecting measurements (Polat-Ozsoy et al., 2009).

Conclusion

Disclosing treatment information to assessors does introduce the risk of systematic measurement errors in cephalometric radiographs, underlining the necessity of blinding of outcome assessors to produce reliable data. It is recommended that cephalometric analysis in orthodontic trials should be carried out by assessors who are blinded to treatment details to minimise risk of bias.

Acknowledgements

We wish to thank Mickey Zhou and John Jo for help with data collection. We wish to thank Roberto Martina from the University of Naples Federico II for providing the lateral cephalometric radiographs used in the study.

Declaration of conflicting interests

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

Funding

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

ORCID iD

Sabarinath Prasad https://orcid.org/0000-0001-6509-2055

References

Abdulraheem S and Bondemark L (2019) The reporting of blinding in orthodontic randomized controlled trials: where do we stand? *European Journal of Orthodontics* 41: 54-58.

Alharbi F and Almuzian M (2019) The quality of reporting RCT abstracts in four major orthodontics journals for the period 2012-2017. *Journal of Orthodontics* 46: 225–234.

Baumrind S and Frantz RC (1971) The reliability of head film measurements. 1. Landmark identification. *American Journal of Orthodontics* 60: 111–127.

Cacciatore G, Ugolini A, Sforza C, Gbinigie O and Plüddemann A (2019) Long-term effects of functional appliances in treated versus untreated patients with Class II malocclusion: A systematic review and meta-analysis. *PLoS One* 14: e0221624.

Cioffi I, Martina R, Michelotti A, Chiodini P, Tagliaferri R and Farella M (2008) Web-based randomised controlled trials in orthodontics. *Evidence-based Dentistry* 9: 118–120.

D'Antò V, Bucci R, Franchi L, Rongo R, Michelotti A and Martina R (2015) Class II functional orthopaedic treatment: a systematic review of systematic reviews. *Journal of Oral Rehabilitation* 42: 624–642.

Day SJ and Altman DG (2000) Statistics notes: blinding in clinical trials and other studies. *BMJ* 321: 504.

Flores-Mir C and Major PW (2006) A systematic review of cephalometric facial soft tissue changes with the Activator and Bionator appliances in Class II division 1 subjects. *European Journal of Orthodontics* 28: 586–593.

Gould SJ (2016) *The mismeasure of man*. Brantford, Ontario: W. Ross MacDonald School Research Services Library.

Haas DW, Martinez DF, Eckert GJ and Diers NR (2001) Measurements of mandibular length: a comparison of articulare vs condylion. *Angle Orthodontist* 71: 210–215.

Harrison JE (2003) Clinical trials in orthodontics II: assessment of the quality of reporting of clinical trials published in three orthodontic journals between 1989 and 1998. *Journal of Orthodontics* 30: 309–315; discussion 297–298.

Hróbjartsson A and Boutron I (2011) Blinding in randomized clinical trials: imposed impartiality. *Clinical Pharmacology and Therapeutics* 90: 732–736.

Polat-Ozsoy O, Gokcelik A and Toygar Memikoglu TU (2009) Differences in cephalometric measurements: a comparison of digital versus hand-tracing methods. *European Journal of Orthodontics* 31: 254–259.

Poolman RW, Struijs PA, Krips R, Sierevelt IN, Marti RK, Farrokhyar F, et al. (2007) Reporting of outcomes in orthopaedic randomized trials: does blinding of outcome assessors matter? *Journal of Bone and Joint Surgery* 89-A: 550–558.

Saltaji H, Armijo-Olivo S, Cummings GG, Amin M, Da Costa BR and Flores-Mir C (2018) Influence of blinding on treatment effect size estimate in randomized controlled trials of oral health interventions. *BMC Medical Research Methodology* 18: 42.

Sandhu SS, Sandhu J and Kaur H (2015) Reporting quality of randomized controlled trials in orthodontics—what affects it and did it improve over the last 10 years? *European Journal of Orthodontics* 37: 356–366.

Soboczenski F, Trikalinos TA, Kuiper J, Bias RG, Wallace BC and Marshall IJ (2019) Machine learning to help researchers evaluate biases in clinical trials: a prospective, randomized user study. *BMC Medical Informatics and Decision Making* 19: 96.

Trpkova B, Major P, Prasad N and Nebbe B (1997) Cephalometric landmarks identification and reproducibility: a meta analysis. *American Journal of Orthodontics and Dentofacial Orthopedics* 112: 165–170.

Tulloch JF, Medland W and Tuncay OC (1990) Methods used to evaluate growth modification in Class II malocclusion. *American Journal of Orthodontics and Dentofacial Orthopedics* 98: 340–347.

Yu HJ, Cho SR, Kim MJ, Kim WH, Kim JW and Choi J (2020) Automated skeletal classification with lateral cephalometry based on artificial intelligence. *Journal of Dental Research* 99: 249–256.