Journal of Orthodontics, Vol. 34, 2007, 234–242

SCIENTIFIC SECTION

Tooth-size discrepancy and Bolton's ratios: the reproducibility and speed of two methods of measurement

S. A. Othman and N. W. Harradine
Bristol Dental Hospital, Bristol, UK

Objective: To determine and compare the reproducibility and speed of two methods of performing Bolton's tooth-size analysis.

Design: Analysis of randomly selected clinical sample.

Setting: Bristol Dental Hospital, University of Bristol, United Kingdom.

Materials and methods: Pre-treatment study casts of 150 patients were selected randomly from 1100 consecutively treated Caucasian orthodontic patients. Bolton tooth-size discrepancies and ratios were measured using two methods; one method employed entirely manual measurement and the Odontorule slide rule, while the other employed digital callipers and the HATS analysis software. Twenty study casts were measured twice, a week apart with both methods. Another three investigators also measured 20 study casts twice with the HATS analysis.

Results: There were small or no systematic errors within or between these two methods. A very significant difference was evident for mean time measurements between the two methods (mean time for HATS was 3.5 minutes and for Odontorule was 8.9 minutes). There was relatively high error variance of both methods of measurement as a percentage of the total variance.

Conclusions: On-line electronic measurement was found to be more rapid than the manual method used. Both methods demonstrate relatively high random error and this has important consequences for the clinical use of Bolton's ratios.

Key words: Bolton's ratios, tooth-size discrepancy, reproducibility, methods of measurement

Received 18th August 2006; accepted 7th July 2007

Introduction

Many studies have investigated tooth-size discrepancy (TSD).[1–4] The best-known study of TSD in relation to treatment of malocclusion was published by Bolton[5] in 1958 in which he evaluated 55 cases with excellent occlusions. Bolton developed two ratios for estimating TSD by measuring the summed mesiodistal widths of the mandibular to the maxillary anterior teeth (anterior ratio) and the total width of all lower to upper teeth from first molar to first molar (overall ratio).

The traditional methods of measuring mesiodistal widths of teeth on dental casts have been either with needle-pointed dividers or a Boley gauge (Vernier callipers). Recent technological advances have allowed the introduction of digital callipers, which can be linked to computers for quick calculation of the anterior and posterior ratios. Alternatively, digitized or scanned images of study casts can be measured on-screen to assist with diagnosis and treatment planning. Proffit[6] stated that one advantage of digital measurement for space analysis is that the computer can quickly provide a subsequent tooth-size analysis.

Previous studies of Bolton's ratios have included very incomplete investigation or reporting of the reproducibility of their measurements.[7–9] Othman and Harradine[10] in their review paper on TSD concluded that reproducibility of measurement of TSD has been poorly investigated. Some well-known studies did not report the measurement error at all.[11] Crosby and Alexander[12] did not mention which statistical test was used or what values were tested for the measurement error or whether systematic error, random error or both types of error were evaluated. Araujo and Souki[13] found that there were no significant differences between two sets of measurements with digital callipers, but there was no measurement of random error. Also, their analysis was confined to Bolton's ratios, and did not include data on the required corrections in millimetres, which is the therapeutically relevant measurement. Conversely, Bernabé *et al.*[14] only evaluated random error not systematic error, and they too only examined the ratio

Address for correspondence: Mr N. W. T. Harradine, Department of Child Dental Health, University of Bristol Dental School, Lower Maudlin Street, Bristol BS1 2LY, UK.
Email: Nigel.harradine@bris.ac.uk

DOI 10.1179/146531207225022302

values but not the correction values in millimetres. If a measurement is to be used to determine a therapeutic intervention, then it is important to know the measurement errors in relation to the planned dimensions of that intervention. Houston[15] wrote that if any study using measurements is to be of value, it is imperative that such error analysis be undertaken and reported, and the very discipline of undertaking an error analysis should also improve the quality of results.

The aims and objectives of the present study were to determine and compare the reproducibility and speed of two methods of performing Bolton's tooth-size analysis.

The two methods were a manual measurement with the Odontorule (Dental Corporation of America, West Chester, PA, USA) and a computerized method – the Hamilton Arch Tooth System (HATS) (GAC International, Central Islip, NY, USA). The null hypothesis was that there would be no difference between the methods with regard to rapidity or reproducibility. The central hypothesis was that at least one method would be sufficiently rapid and sufficiently reproducible to be a robust and practicable method of measurement for an individual clinical case.

Materials and methods

Pre-treatment study casts of 150 patients were used in this study. The study model numbers of 1100 patients treated consecutively in a teaching hospital from 1999 to 2002 were obtained from the laboratory database and a computer-generated list of random numbers was used to select the sample from this consecutive series. If a case was discarded because it did not meet the selection criteria, the next consecutive eligible case was included. This sample therefore included a random selection of different malocclusions representative of an orthodontic treatment population. The Chairman of the Local Research Ethics Committee confirmed that ethical approval for measuring study casts was not required.

The following selection criteria were used:

- good quality pre-treatment models;
- a fully erupted permanent dentition from first molar to first molar;
- Caucasian ethnicity.

Rejection criteria included:

- gross restorations, build-ups, crowns, onlays, Class II amalgam or composite restoration that affect the tooth's mesiodistal diameter;
- congenitally missing teeth and impacted teeth.

The mesiodistal diameter tooth sizes were measured from first molar to first molar at the level of the contact points. Contact points were defined at the points on the proximal surfaces, as observed or estimated as those which should be touching when the teeth were perfectly aligned.

Method 1: this employed the Odontorule for analysis of the maxillary–mandibular tooth size relationship. This employs a rotating wheel, which is in-effect a circular slide rule, and was developed by Dr David C. Hamilton and Dr Charles W. Patton based upon studies by Dr Wayne A. Bolton as an aid to measurement which would be faster and more convenient than looking up tables of figures. The mesiodistal tooth sizes were measured in millimetres manually to the nearest 0.5 mm with Helios sliding callipers. The sum of the total maxillary and mandibular teeth (6–6) and sum of the anterior maxillary and mandibular teeth (3–3) were calculated using a calculator. The total and anterior ratios were determined by Bolton's formulas.[5] The amounts of correction in the maxillary and mandibular arches for the total ratio and anterior ratio are obtained on the rotating wheel. Each analysis was timed by a stopwatch from the first measurement to the final computation.

Method 2: this employed the HATS software which is available from GAC. All the study casts were measured to the nearest 0.01 mm with digital callipers (PRO-MAX Digital Callipers, Fred V. Fowler Co., Inc., Newton, MA, USA) connected to a computer. The HATS software calculates the Bolton's ratios and also recommends the tooth size correction in either arch to achieve Bolton's average ratio for an ideal occlusion. The entire procedure was timed from initial measurement to availability of the calculated results and the results were then printed. Neither the method of measurement nor the timing of the measurements was amenable to blinding of the assessor at recording.

Assessment of reproducibility

The principal intra-examiner reproducibility procedure consisted of the primary investigator (SAO) measuring 20 sets of study models randomly selected from the larger group of 150 patients. These patients were selected from the 150 by means of a computer-generated random number list. A sample of 20 sets of study models was deemed to be adequate and sufficiently representative in relation to the variance of the larger group and to the size of error regarded to be of clinical significance, which was judged to be half the size of TSD (1.5 mm) which Proffit[6] considered of potential clinical significance. All the teeth were measured twice for the two methods, with a week between the measurements. An inter-examiner calibration involved three additional examiners – an experienced orthodontist (NWTH) and two senior trainees (CD and SD) with four years of previous orthodontic training. These three examiners also measured the 20 study models twice using the HATS method only, to determine intra- and inter-examiner systematic error and to compare random errors for that method.

Statistical analysis

The distribution of data was evaluated for normality. For assessment of the systematic error, repeated measures analysis of variance (ANOVA) and the Greenhouse–Geisser approximation were used to test statistical significance. Greenhouse–Geisser is a standard method of dealing with sphericity with the assumption that each of the examiners was related to each other in the same way. The paired-sample *t*-test was used to evaluate the systematic error and the differences in timings for the two methods of measurement. Random error was calculated in terms of the standard deviation of the differences in replicate measurements as advocated by Houston.[15] The variance of the difference between two replicate measurements is double that of a single measurement, so the variance of the differences must be halved to give a correct estimate of the error for a single measurement. This measure was preferred to the root mean square error (as advocated by Dahlberg,[16] and which is still frequently employed), because it avoids the possibility of any systematic bias affecting the assessment of random error. Dahlberg's formula is only accurate if there is no systematic bias. The analysis in the present study included the percentage of the total sample variance that consists of error variance (the variance of replicate measurements), because Houston makes the crucial point that the potential effect of random error on interpretation of results can only be properly estimated in relation to the variance from all sources in a representative sample.

Results

By inspection, all the data were demonstrated to come from a normally distributed population and parametric tests were therefore used.

Systematic error

The systematic error (bias) of the principal investigator (SAO) for the Odontorule is detailed in Table 1a. There was a small but statistically significant difference between the two means for the upper total correction, lower total correction and the time. Table 1b shows the same data for the HATS method where there was no significant difference found between the two sets of measurements, however the P value for time was close to statistical significance ($P=0.058$) and again showed a small reduction in time with the second set of measurements. Table 1c contains the analysis of systematic differences between four examiners using the HATS method. Statistically significant differences were found between the four examiners for the three total arch measurements, but not for the anterior arch measurements or the time. Results also reflected the results for SAO in Table 1b, in that there was no within-operator systematic bias for the HATS method for any of the operators. The mean figures in Table 1c are the averages of the duplicated measurements.

Random error

Random errors are given in Table 2 for Odontorule and Table 3 for HATS for observer SAO. The error variance is a high percentage of the total variance for all measures and for both methods of measurement. The total variance for the sample of 150 was, by chance, much higher than the total variance for the sample of 20 used for the duplicate measurements. A sample size of 20 was judged to be completely sufficient for all analytical purposes except for the comparison of error variance to total variance. The total variance of the randomly chosen 20 subjects was unpredictably larger by chance than for the sample of 150, so the variance of the full 150 sample was the better choice for the comparison of error variance to total variance. Extending the duplication of measurements to a number greater than 20 would only stand a small chance of increasing the reliability of all other values in the reproducibility analysis. Table 4(a–d) therefore contains the random error analyses for all four observers for the HATS method, using this more representative complete sample variance for comparison with the error variance. The percentages of error variance were correspondingly lower than in Table 3, but still high.

Systematic error (bias) between the two methods

Table 5 shows that the mean differences for tooth-size discrepancy are very small and not statistically significant, but the speed of the measurement and analysis for the two methods was very different, the HATS method being much faster.

Discussion

Systematic error

Table 1(a,b) lists the systematic error results of the two methods of measurement performed on the same casts twice by the same investigator. The HATS results showed no significant differences between the means although the mean reduction in measurement time of 0.21 minutes or 12 seconds approached significance. For the Odontorule, the mean reduction in measurement time of 42 seconds for the second measurement was statistically significant. Both these results suggest that a process of familiarization was still occurring during this part of the study, in spite of fairly extensive use of both methods by the principal investigator prior to the study. This familiarization factor is of potential significance for the occasional user in a clinical setting.

The Odontorule results also show statistically significant differences in the mean total correction values. The mean differences are approximately 0.6 mm and are therefore small in terms of clinical significance – Proffit[6] felt that a discrepancy of <1.5 mm is rarely of significance. The same values when measured with HATS show a similar trend but to a smaller and non-significant extent. Table 1c shows statistically significant

Table 1 (a) Intra-examiner reproducibility (systematic error): Odontorule. Observer SAO (n=20).

				95% Confidence intervals		
Descriptive	Mean time 1	Mean time 2	Mean difference	Lower	Upper	*P* value
Total Bolton ratio (%)	91.23	91.90	–0.68	–1.13	–0.22	0.054
Upper total correction (mm)	–0.29	0.45	0.72	–1.26	–0.18	0.007*
Lower total correction (mm)	0.07	–0.49	0.56	0.04	1.07	0.012*
Anterior Bolton ratio (%)	78.07	78.80	–0.73	–1.40	–0.17	0.088
Upper anterior correction (mm)	0.80	0.94	–0.14	–0.53	0.25	0.216
Lower anterior correction (mm)	–0.60	–0.77	0.17	–0.23	0.54	0.270
Time (minutes)	9.53	8.82	0.70	0.34	1.06	0.001*

*$P<0.05$.

Table 1 (b) Intra-examiner reproducibility (systematic error): HATS. Observer SAO (n=20).

				95% Confidence intervals		
Descriptive	Mean time 1	Mean time 2	Mean difference	Lower	Upper	*P* value
Total Bolton ratio (%)	91.18	91.58	−0.40	−1.15	0.03	0.135
Upper total correction (mm)	−0.40	0.29	−0.43	−1.00	0.14	0.134
Lower total correction (mm)	0.13	−0.26	0.40	−0.13	0.91	0.135
Anterior Bolton ratio (%)	78.15	78.30	−0.15	−0.79	0.50	0.639
Upper anterior correction (mm)	0.54	0.66	−0.12	−0.50	0.30	0.389
Lower anterior correction (mm)	−0.41	−0.51	0.10	−0.21	0.41	0.393
Time (minutes)	3.97	3.46	0.21	−0.00	0.43	0.058

*$P<0.05$.

Table 1 (c) Intra-examiner reproducibility (systematic error): HATS (n=20).

Descriptive	Mean (SAO)	Mean (NWTH)	Mean (CD)	Mean (SD)	*P* value for difference between people
Total Bolton ratio (%)	91.38	91.21	91.15	91.71	0.014*
Upper total correction (mm)	0.07	−0.10	−0.16	0.43	0.014*
Lower total correction (mm)	−0.07	0.09	0.15	−0.40	0.011*
Anterior Bolton ratio (%)	78.2	78.04	78.01	77.99	0.760
Upper anterior correction (mm)	0.60	0.42	0.48	0.46	0.783
Lower anterior correction (mm)	−0.46	−0.34	−0.37	−0.36	0.735
Time (minutes)	3.56	3.75	4.73	2.99	0.077

*$P<0.05$.

Table 2 Random error: Odontorule. Observer SAO (n=20).

	Total ratio	Total correction Upper	Total correction Lower	Anterior ratio	Anterior correction Upper	Anterior correction Lower	Time
Total variance	1.82	2.06	1.64	2.21	0.87	0.68	0.68
Mean difference	−0.68	−0.72	0.56	−0.73	−0.14	0.17	0.70
SD of differences	0.98	1.16	1.11	0.92	0.84	0.83	0.77
Variance/2	0.48	0.67	0.61	0.42	0.36	0.34	0.29
Error variance (%)	26.23	32.42	37.54	9.28	40.68	0.32	43.57
Correlation coefficients (r)	0.746	0.665	0.619	0.658	0.605	0.545	0.508

systematic differences in the mean measurements obtained by the four operators. These mean differences were very small (less than 0.5% and less than 0.5 mm) and were again confined to the total arch measurements. Nevertheless, the existence of any systematic error suggests that considerable familiarity with these techniques is required before there is stability of point identification and that occasional use of this analysis is not appropriate. Inter-operator errors were not analysed for the manual Odontorule method because it had already become apparent that the substantial additional time required for this method, with no evidence of greater reproducibility, made this a method which could not be recommended for clinical use.

Random error

The results for the main examiner (SHO) for the two methods are shown in Tables 2 and 3, and they are similar for both methods. The standard deviation of replicate measurements is of the order of 1 mm for correction and 1% for ratios. These standard deviations are significant, being more than half the size of TSD (1.5 mm) which Proffit[6] considered of significance. It is also important to place the error variance in the context of the total variance of the sample. Midtgård *et al.*[17] suggested that the error variance should not exceed 3% of the total variance, and if it exceeded 10% the applied method of measuring is probably inappropriate. The results from the full sample of 150 revealed a substantially larger total variance and the expert statistical advice was that the larger sample is a more valid indicator of total variance in the orthodontic population. The results using this total variance are in Table 4 and show that for the main examiner (SAO, Table 4a), this estimate of the percentage random error was much closer to the 10% recommended by Midtgård *et al.*[17] but remains higher than is desirable for a robust measurement method.

The inter-examiner reproducibility was then assessed to see whether this percentage of error variance was particular to the main examiner. It can be seen in Table 4(b–d) that for all examiners the error variance was a higher percentage of the total variance than advocated by Midtgård *et al.*[17] and by Houston.[15] There were some differences between the examiners, but none of these differences in random error was statistically significant as assessed by Greenhouse–Geisser. Two investigators (CD and SD) were much less familiar with the HATS method, but their random error was similar to the main investigator (SAO) who was significantly more familiar with this process. NWTH was familiar with the method and over a longer period of time and had lower random error values, which were within the recommended 10%, but higher than the ideal 3%. The results suggest that experience may reduce random error

Table 3 Random error: Hats. Observer SAO (n=20).

	Total ratio	Total correction Upper	Total correction Lower	Anterior ratio	Anterior correction Upper	Anterior correction Lower	Time
Total variance	1.61	1.71	1.42	2.50	0.90	0.53	0.15
Mean difference	−0.40	−0.43	0.40	−0.15	−0.12	0.09	0.21
SD of differences	1.16	1.22	1.12	1.37	0.86	0.66	0.47
Variance/2	0.67	0.75	0.63	0.94	0.37	0.22	0.11
Error variance (%)	41.58	43.85	43.96	37.88	40.93	41.25	74.85
Correlation coefficients (r)	0.505	0.561	0.560	0.633	0.609	0.609	0.254

Table 4 (a) Random error: Hats. Observer SAO. Using complete sample variance (n=150) for error variance %.

	Total ratio	Total correction		Anterior ratio	Anterior correction	
		Upper	Lower		Upper	Lower
Complete sample variance	3.42	3.55	2.96	5.25	1.80	1.07
Mean difference	−0.40	−0.43	0.40	−0.15	−0.12	0.09
SD of differences	1.16	1.22	1.12	1.37	0.86	0.66
Variance/2	0.67	0.75	0.63	0.94	0.37	0.22
Error variance (%)	19.5	21.1	21.28	17.90	20.55	20.56

Table 4 (b) Random error: Hats. Observer NWTH. Using complete sample variance (n=150) for error variance %.

	Total ratio	Total correction		Anterior ratio	Anterior correction	
		Upper	Lower		Upper	Lower
Complete sample variance	3.42	3.55	2.96	5.25	1.80	1.07
Mean difference	0.22	0.23	−0.21	0.17	0.11	−0.02
SD of differences	0.71	0.72	0.65	0.85	0.50	0.43
Variance/2	0.26	0.26	0.21	0.36	0.13	0.09
Error variance (%)	7.60	7.32	7.09	6.85	7.22	8.40
Correlation coefficients (r)	0.838	0.852	0.853	0.851	0.852	0.814

Table 4 (c) Random error: Hats. Observer CD. Using complete sample variance (n=150) for error variance %.

	Total ratio	Total correction		Anterior ratio	Anterior correction	
		Upper	Lower		Upper	Lower
Complete sample variance	3.42	3.55	2.96	5.25	1.80	1.07
Mean difference	−0.25	−0.25	0.23	−0.17	−0.09	0.07
SD of differences	0.73	0.77	0.72	1.15	0.72	0.56
Variance/2	0.26	0.31	0.26	0.66	0.26	0.16
Error variance (%)	7.60	9.73	9.78	12.57	14.44	14.95
Correlation coefficients (r)	0.786	0.771	0.772	0.737	0.722	0.723

Table 4 (d) Random error: Hats. Observer SD. Using complete sample variance (n=150) for error variance %.

	Total ratio	Total correction		Anterior ratio	Anterior correction	
		Upper	Lower		Upper	Lower
Complete sample variance	3.42	3.55	2.96	5.25	1.80	1.07
Mean difference	0.14	0.14	−0.14	0.17	0.10	−0.08
SD of differences	0.89	0.91	0.83	1.39	0.83	0.64
Variance/2	0.40	0.42	0.35	0.98	0.34	0.20
Error variance (%)	11.69	11.83	11.82	18.67	18.87	18.69
Correlation coefficients (r)	0.823	0.826	0.828	0.683	0.689	0.689

to a worthwhile extent, but not to a level where confidence can be placed in a single measurement. Statistical analysis of this suggestion is complicated by the difficulty in quantifying with validity the relevant experience in a group of four operators.

These results are clearly important in relation to the assessment of a single patient. Great caution should be exercised before instituting an intervention on the basis of one measurement of the Bolton discrepancy. Confidence in the calculation of discrepancy is particularly important if the resulting intervention is reduction of tooth width by interdental stripping or extraction. To reduce the random error for both methods of measurement explored in this study, a clinician is strongly advised to measure the same study models three or four times and then average the values obtained before committing to any active intervention.

Error in relation to tooth irregularity

Locating contact points on a crowded dentition is difficult. The sample in this present study consisted of a variety of malocclusions with a range of crowding. Shellhart *et al.*[8] found that every investigator made at least one error in measurement that was greater than a clinically significant value for the tooth-size excess when measuring Bolton discrepancies on crowded dentitions (at least 3 mm of crowding) with a Boley gauge and needle-point dividers. It would be possible to take study casts in the middle of treatment for analysis once alignment had been achieved and in very crowded dentitions, this is advisable.

Comparison of errors between methods of measurement

The present study indicated that there were no differences between the two methods of measurement. This result is in agreement with that reported by Tomassetti *et al.*[7] who compared the HATS system and Vernier callipers. Their correlation coefficient between the HATS and the Vernier callipers was $r=0.825$. However, there was no separate test for random error with either method in their study.

Shellhart *et al.*[8] studied the reliability of Bolton's tooth-size analysis when applied to crowded dentitions using needle-pointed dividers and the Boley gauge. For 14 of the 16 measures, there was no statistically significant difference. Random error was estimated by correlation coefficients. These varied very greatly from a reasonable correlation of $r=0.79$ to a very low figure of $r=-0.15$ for intra-investigator errors. Intraclass coefficients for measurements made by four investigators ranged from 0.80 to 0.29. Many of these values are therefore very much lower than would be considered desirable for a good method of measurement. The authors agreed with this view and stated that, 'If a clinician's repeatability of the Bolton analysis is average, calculations of tooth-size discrepancy should be viewed as ± 2.2 mm.' This recommended confidence level is very large in relation to a clinically significant TSD and their conclusion begs the question as to what method should, in their view, actually be used to decide on therapeutic intervention.

Zilberman *et al.*[9] also reported that measurement with digital callipers on plaster models showed better reproducibility than measurements on virtual computerized models (OrthoCAD). Importantly, the repeated measures of the total tooth-size widths were evaluated, but not Bolton's ratios or the discrepancies. They found both random and systematic errors were very small and clinically insignificant. The error of the sum of the tooth widths is, however, likely to be much smaller than the error in the calculated Bolton's ratios or correction in millimetres, because the sum of tooth widths is a much larger absolute figure. Direct comparison of the reproducibility of Zilberman *et al.*[9] with the present study is therefore not reliable. As has been mentioned, some well-known studies did not report the measurement error at all[11] or very inadequately.[12–14]

Table 5 Comparison of mean results for the Odontorule and HATS methods ($n=20$). Observer SAO. Paired t-test.

Descriptive	Mean Odontorule	Mean HATS	Mean difference	95% Confidence intervals		P value
				Lower	Upper	
Total Bolton ratio (%)	91.23	91.18	−0.05	−0.56	0.45	0.825
Upper total correction (mm)	−0.29	−0.40	0.14	−0.41	0.68	0.603
Lower total correction (mm)	0.07	0.13	0.05	−0.46	0.56	0.825
Anterior Bolton ratio (%)	78.07	78.15	0.07	−0.59	0.84	0.774
Upper anterior correction (mm)	0.80	0.54	−0.26	−0.64	0.12	0.168
Lower anterior correction (mm)	−0.60	−0.41	0.18	−0.17	0.61	0.341
Time (minutes)	8.93	3.49	−5.45	−5.88	−5.01	<0.001

Differences in timing

The present study noted that the HATS method was quicker than the Odontorule. This average difference of 5.45 minutes is significant both statistically and clinically. The comparative times for the four examiners (Table 1c) show a range of mean times for the HATS method which was much smaller (1.75 minutes) than the difference between the two methods. These results are similar to those of Tomassetti *et al*.[7] This part of the study strongly supported the abandonment of manual methods of measurement and calculation in favour of direct electronic entry of measurements into a software package. The two methods gave essentially the same mean values (Table 5) but the HATS method was much faster, potentially less prone to blunder or fatigue problems and was of comparable or better reproducibility. Both methods are technically easy to use. The Odontorule costs approximately £30 and the HATS software approximately £250. Electronic callipers are used with both methods and cost approximately £80.

Strengths and weaknesses of this study

This study benefits from a sample which is very representative of clinical orthodontic practice. The statistical analysis importantly measures both random error and systematic error and relates that error directly to the dimensional changes of clinical significance. This enables important clinical conclusions, namely that random error is high, so repeat measures are needed on an individual case and that infrequent use of this method is not appropriate. The study shows that the manual method of measurement no longer has a clinical role, being much slower and with the same reproducibility.

Such a study could be improved by increasing the number of operators and by measuring on more than two occasions. This could provide data on how much experience with this method is required to eliminate any tendency to drift in point identification and therefore statistical bias. It might also confirm whether, and by how much, greater experience also reduces random errors.

Conclusions

- The HATS and Odontorule methods had similar reproducibility, but the HATS method was significantly quicker.
- Single estimations of TSD should be treated with great caution and replicate measurements are advised if active clinical intervention is planned.
- There is evidence to suggest that significant experience with measuring TSD has a worthwhile effect in reducing error. Analysis on an occasional basis is not advisable.

Contributors

Siti Adibah Othman was responsible for material and data collection, conducting the experiment, data analysis and writing of the article. Mr Nigel Harradine was responsible for study design, drafting of the article, contributed to the writing of the article, data analysis and interpretation, critical revision and final approval of the published version. Mr Nigel Harradine is the guarantor.

Acknowledgements

The authors would like to express their thanks to Christian Day and Scott Deacon for their involvement in the assessment of inter-examiner reproducibility and to Mrs Rosemary Greenwood for her thorough and expert guidance throughout the statistical analysis.

References

1. Ballard ML. Asymmetry in tooth size: a factor in the etiology, diagnosis and treatment of malocclusion. *Angle Orthod* 1944; **14:** 67–71.
2. Neff CW. Tailored occlusion with the anterior coefficient. *Am J Orthod* 1949; **35:** 309–14.
3. Steadman SR. The relation of upper anterior teeth to lower anterior teeth as present on plaster models of a group of acceptable occlusions. *Angle Orthod* 1952; **22:** 91–97.
4. Lundstrom A. Intermaxillary tooth width ratio and tooth alignment and occlusion. *Acta Odontol Scand* 1954; **12:** 265–92.
5. Bolton WA. Disharmony in tooth size and its relation to the analysis and treatment of malocclusion. *Angle Orthod* 1958; **28:** 113–30.
6. Proffit WR. *Contemporary Orthodontics*, 3rd edn. St Louis: Mosby, 2000, 170.
7. Tomassetti JJ, Taloumis LJ, Denny JM, Fischer JR Jr. A comparison of 3 computerized Bolton tooth-size analyses with a commonly used method. *Angle Orthod* 2001; **71:** 351–57.
8. Shellhart WC, Lange DW, Kluemper GT, Hicks EP Kaplan AL. Reliability of the Bolton tooth size analysis when applied to crowded dentitions. *Angle Orthod* 1995; **65:** 327–34.
9. Zilberman O, Huggare JA, Parikakis KA. Evaluation of the validity of tooth size and arch width measurements using conventional and three-dimensional virtual orthodontic models. *Angle Orthod* 2003; **73:** 301–6.

10. Othman SA, Harradine NW. Tooth-size discrepancy and Bolton's ratios: a literature review. *J Orthod* 2006; **33:** 45–51.
11. Freeman JE, Maskeroni AJ, Lorton L. Frequency of Bolton tooth-size discrepancies among orthodontic patients. *Am J Orthod Dentofacial Orthop* 1996; **110:** 24–27.
12. Crosby DR, Alexander CG. The occurrence of tooth size discrepancies among different malocclusion groups. *Am J Orthod Dentofacial Orthop* 1989; **95:** 457–61.
13. Araujo E, Souki M. Bolton anterior tooth size discrepancies among different malocclusion groups. *Angle Orthod.* 2003; **73:** 307–13.
14. Bernabé E, Major PW, Flores-Mir C. Tooth-width ratio discrepancies in a sample of Peruvian adolescents. *Am J Orthod Dentofacial Orthop* 2004; **125:** 361–65.
15. Houston WJ. The analysis of errors in orthodontic measurements. *Am J Orthod* 1983; **83:** 382–90.
16. Dahlberg G. *Statistical Methods for Medical and Biological Students.* New York: Interscience publications, 1940.
17. Midtgård J, Björk G, Linder-Aronson S. Reproducibility of cephalometric landmarks and errors of measurements of cephalometric cranial distances. *Angle Orthod* 1974; **44:** 56–61.