

Errors and Discrepancies in Measurement of Tooth Size

W. STUART HUNTER and WILLIAM R. PRIEST

Department of Orthodontics, School of Dentistry, University of Michigan, Ann Arbor, Michigan

The purpose of this study was to evaluate the experimental errors and discrepancies involved in the measurement of tooth size using plaster casts. The terminology employed is taken principally from Beers,¹ who defines the terms used in this area as follows:

- A. *Error*. This word is used correctly with two different meanings:
 1. To denote the difference between a measured value and the "true" value. Except in a few trivial cases, the "true" value is unknown and the magnitude of the error is hypothetical.
 2. When a number such as (± 0.087) is given or implied, "error" refers to the estimated uncertainty in an experiment and is expressed in terms of such quantities as standard deviation, average deviation, probable error, or precision index.
- B. *Discrepancy*. This is the difference between two measured values of a quantity, such as the difference between two measured values of the same quantity obtained by two investigators. The word "error" is often used incorrectly to refer to such differences.
- C. *Experimental errors*. When a given measurement is repeated, the resulting values, in general, do not agree exactly. The differences between the resulting values are called "experimental errors." Also, they are sometimes called "erratic" or "accidental errors."
- D. *Systematic or constant errors*. If, on the other hand, all of the individual values are in error by the same amount, the errors are called *systematic* or *constant* errors.

In the measurement of dental casts, both experimental and systematic errors may be present. Sometimes they arise from the same source. For example, if an investigator uses dividers for the measurement of mesiodistal tooth diameter, the values are apt to be consistently larger than the "true" value because of the taper on the points of the dividers. The true points do not touch the greatest width of the tooth when placed on the tooth, but rather the sides of the points are in contact with the tooth. This error is included in the measure when read off on the scale, since it is the distance between the true points that is read off. The error introduced is thus a *systematic* error. When the tooth is remeasured by the same technique, the points of the dividers may not be placed in exactly the same position on the tooth, and the value read off on the scale will differ from the previous value. This is an *experimental* error.

It should be remembered that the terms "error" and "discrepancy" are used here in their technical sense and do not encompass what is commonly called "mistakes." A reading of 0.87 instead of 0.78 is a mistake, with which no systematic treatment of error could cope.

This investigation was supported by Research Grant D-224 from the N.I.D.R., U.S. Public Health Service.

Received for publication October 22, 1959.

When measuring dental casts, as in any investigation, it is obvious that errors and discrepancies will occur. If the data obtained are to be used for comparisons between various population and racial groups or between twins, it is important to know how much of the differences obtained are real differences and how much may be accounted for by errors and discrepancies. Selmer-Olsen² suggests that one is in danger of laying more stress on the numerical results than the accuracy of the measuring technique justifies. It follows, then, that the investigator must know what the numerical values of the errors and discrepancies in his methods are before any valid conclusions can be made.

A review of anthropologic and dental literature shows that little mention has been made of errors and discrepancies in the measurement of dental casts, with the notable exception of the Scandinavian investigators such as Selmer-Olsen,² Seipel,³ and Lundstrom.⁴ Recently, Moorrees *et al.*⁵ gave a detailed account of the discrepancies and errors in their study: "The errors of measurement by different investigators, both of mesiodistal crown diameter of individual permanent teeth and of the combined mesiodistal crown diameters of a series of such teeth ($P_{m_2} - P_{m_2}$), were calculated from a number of double determinations and expressed as standard deviations of the differences. For single permanent mandibular teeth of different tooth classes the error of measurement was 0.09 mm. ($n = 360$) and for the combined mesiodistal crown diameters of ten permanent mandibular ($P_{m_2} - P_{m_2}$) teeth it was 0.41 mm. ($n = 30$)."

In "Variation of Tooth Position," Seipel³ states: "The exactitude of the measurements depends on the safety of the chosen points, the precision of the measuring instrument, and on the way in which the investigator uses it" (p. 27).

PART I

The purpose of the first portion of this study was to evaluate the errors and discrepancies between two commonly used measuring instruments and between two investigators. The materials used consisted of the plaster casts of 24 subjects chosen randomly from those obtained in 1941 from the University of Michigan Elementary School Growth Study. Impressions were all taken with an alginate material and poured in model plaster.

METHOD

The first method of measurement involved the use of a pair of engineering dividers (with wheel-screw adjustment) and a millimeter scale especially engraved to permit readings in tenths of millimeters. The dividers were adjusted to the mesiodistal diameter of the tooth being measured, and the size was read off from the engraved millimeter scale by putting the points of the dividers on the scale. The other method involved the use of a regular sliding caliper (Helios style) with a Vernier scale as an integral part of the instrument. Both methods permit readings to the nearest 0.1 mm.

Central incisors through first permanent molars in all four quadrants, or 24 teeth, were measured where possible. The changing dentitions eliminated some teeth from the study, as did poorly contoured restorations involving contact areas.

The light source was standardized for all measurements, with a desk-type fluorescent lamp and goose-neck tungsten lamp. By reading the instrument in the glare of

the light, the eye, instrument, and light source lay approximately in a straight line, thus reducing errors of parallax to a minimum.

When the scale gave a reading precisely between two 0.1 divisions (i.e., precisely between 7.6 and 7.7), a deck of cards, eight marked with a plus and eight marked with a minus, was consulted to decide whether to read the higher or the lower digit. After sixteen such selections, the deck was reshuffled.

The mesiodistal crown diameter was measured wherever possible, with the points of the dividers or calipers parallel to the long axis of the tooth and on the normal contact areas. The dividers or calipers were inserted from the buccal (labial), with the instrument held at a right angle to the long axis in most cases. However, in some instances the measurements were necessarily made with an insertion from the occlusal (incisal), with the instrument held in a plane parallel to the long axis of the tooth. In cases of linguoversion, a lingual insertion was used. When teeth were rotated, the normal contact areas were chosen. Since the widest portion of a tooth is usually

TABLE 1*
EIGHT MEASUREMENTS ON SAME TOOTH AND DIFFERENCE FROM
MEAN OF 8 MEASURES (5.15 MM.)

Measure	Method	Investigator	Actual Size	Difference from Mean	Investigator	Actual Size	Difference from Mean
First.....	{ Dividers S. Cal.	SH1	5.2	0.05	VL1	5.3	0.15
		SH1	5.0	-.15	VL1	5.1	-.05
Second.....	{ Dividers S. Cal.	SH2	5.2	.05	VL2	5.4	.25
		SH2	5.0	-.15	VL2	5.0	-.15

* All measures are in millimeters.

toward the buccal, it follows that the measurement from contact point to contact point is not necessarily the greatest mesiodistal diameter. However, the difference is slight.

To study differences between investigators and between first and second measures by the same investigator (as well as instrument differences), each tooth was measured by each investigator, using both dividers and sliding calipers. The procedure was then repeated; thus each tooth was measured eight times.

Since it was the error or variability, not the actual tooth size, in which we were interested for this study, the effect of tooth size was eliminated by calculating, for each tooth, the mean of the eight measurements and subtracting the mean from each measurement. The measurements for the right mandibular central incisor, case No. 371, are shown in Table 1.

These differences were accumulated for the 24 cases, and a $2 \times 2^*$ analysis of variance was performed for the first measurements and for the second measurements for the right mandibular central incisor. For this tooth, the sliding calipers (both investigators) gave a significantly smaller measurement than did the dividers for both first and second measurements. The difference between investigators (both instruments)

* The authors wish to express their thanks for the assistance given by Professor Paul Dwyer, Department of Mathematics, University of Michigan, in the design of the statistical treatment.

was significant for this tooth for the first measurement, but not for the second.

This type of analysis was repeated for all 24 tooth positions measured.

RESULTS

1. In all tooth positions, the dividers, on the average, gave a significantly larger measurement than did the sliding calipers.

2. The mean size of this difference (mean for all positions and both investigators) was 0.153 ± 0.026 mm. (S.E. of the difference).

3. Consideration of first and second measurements revealed that in 17 out of 24 positions the mean of the first measurement was larger than the mean of the second and contained more variation, but only 6 of these 17 were significantly larger. When the means of all the first measurements were pooled and compared with the means of all the second measurements, the reduction in variability was significant at the 0.01 level.

4. Means of differences between investigators were significant for measurements of all 4 cuspids, but not consistently for the other positions. It was felt that a sample of two investigators does not provide a sound basis for any conclusions in this matter.

5. Tests for differences in variances for the 6 anterior teeth as opposed to the 6 posterior teeth were significant, suggesting that a tooth-by-tooth analysis would be interesting.

CONCLUSIONS

1. Use of dividers gave, on the average, a significantly larger measurement of tooth size than did sliding calipers. This is probably due to the taper on the points of the dividers. The true points do not touch the greatest width of the tooth when placed on the tooth, but rather the sides of the points. This error is systematically included in the measure when read on the scale, since it is the distance between the true points that is read off. Therefore, there is both a theoretical and a practical advantage to the sliding caliper, since the measures which result when they are used are closer to the true value for the size of the tooth and are, in addition, more easily obtained.

2. There was some reduction in variation in measures around the mean in each set of eight measures, as we go from first measure to second, suggesting that the measuring technique was perfected to some extent. This improvement was not large but suggests that a training period should be included in any study involving this type of measurement.

3. The training period provided in this study did standardize the techniques of both investigators, so that, aside from the reduction in variability seen in both, they measured with similar accuracy except for the cuspids.

4. Technique may depend upon tooth position, since investigators varied on the cuspid measure and variability as a whole was greater for the posterior teeth.

PART II

The purpose of the second part of these investigations was to compare variances for all teeth individually and to study the accidental and experimental errors between investigators and between two sets of replicate casts and measurements, using the same measuring instrument.

The material used for Part II consisted of replicate sets of plaster casts for the teeth of 23 individuals. The second set of casts were taken in each case, within an hour of the first set.

METHOD

In general, the method was identical with the method used in Part I, except that (1) only sliding calipers were used and (2) two sets of casts were available for each subject. Each cast was measured twice by the two investigators, giving again eight measures for each tooth.

Since the presence of fillings should have negligible effect on a study of discrepancies of measurement, the filled teeth were also measured, giving at least 18 determinations for each tooth position. When more than 18 determinations were possible, these were obtained, and the excess above 18 was deleted randomly down to 18.

Again, the mean of the 8 measures for each tooth was obtained, and the differences from this mean were used in the study of discrepancies.

The statistical treatment of the data was designed to provide a six-way analysis of variance, if the assumptions concerning the homogeneity of the variances were met.

The components of the analysis were maxilla versus mandible, left versus right sides, cast I versus cast II, tooth position, and interactions. The differences from the mean of the 8 measures are the variables being studied.

For tests of differences in variability of tooth position, right and left sides, and maxilla or mandible, an estimate was made for each tooth based on a pooled sum of squares of the eight groups of measurements (8 combinations of cast I, cast II, maxilla, mandible, investigator 1, and investigator 2) with 136 degrees of freedom (8×17).

RESULTS

The findings of this study are summarized in Tables 2, 3, 4, and 5. Table 2, which contains the results of testing the variances of the discrepancies for the left side against those of the right side shows that, although the discrepancies for the left side were more variable than the right side more frequently (10 out of 14 times), this difference was significant only for the maxillary central incisors and for the mandibular central and lateral incisors.

Table 3 summarizes the test of variances of discrepancies for maxilla versus those for the mandible and demonstrates that the variances of the discrepancies for the maxillary lateral incisors and first and second molars are significantly greater than those of their mandibular counterparts. On the other hand, the mandibular left cuspid discrepancy variability is significantly greater than that of the maxillary left cuspid. Essentially, this suggests that the maxillary lateral incisors and molars and mandibular cuspids are harder to measure than the corresponding teeth in the opposite arch. This is probably accounted for by the shape of these teeth.

Table 4 summarizes the results of Bartlett's tests (standard statistical test of within-cell homogeneity) on various groupings of differences within each quadrant. The overall Bartlett's test contraindicated a six-way analysis of variance (the variances were not sufficiently homogeneous). We may, however, arrange the variances of the discrepancies in a comparative fashion, as has been done in Table 4. Any tooth position within parentheses is homogeneous with respect to its variability to any other tooth position

TABLE 2
VARIANCE RATIO TESTS FOR WITHIN-CELL DIFFERENCES IN VARIANCE
USING POOLED ESTIMATES OF WITHIN-CELL-VARIANCE
(The Side Having the Greater Variability Is to the Left)

TOOTH	RIGHT VERSUS LEFT					
	Mandible			Maxilla		
		F	Prob.		F	Prob.
I ₁	L/R	1.58	0.01	L/R	1.99	0.001
I ₂	L/R	2.18	0.001	L/R	1.29
C.....	L/R	1.18	R/L	1.08
P ₁	L/R	1.18	L/R	1.35
P ₂	R/L	1.15	L/R	1.04
M ₁	L/R	1.36	R/L	1.13
M ₂	L/R	1.15	R/L	1.22

TABLE 3
VARIANCE RATIO TEST FOR WITHIN-CELL DIFFERENCE IN VARIANCE
USING POOLED ESTIMATES OF WITHIN-CELL VARIANCE
(The Jaw Having the Greater Variability Is to the Left.)

TOOTH	MANDIBLE VERSUS MAXILLA					
	Left			Right		
		F	Prob.		F	Prob.
I ₁	Max/Mand	1.23	Mand/Max	1.02
I ₂	Mand/Max	1.03	Max/Mand	1.64	0.01
C.....	Mand/Max	1.73	0.01	Mand/Max	1.36
P ₁	Max/Mand	1.10	Mand/Max	1.04
P ₂	Max/Mand	1.14	Mand/Max	1.05
M ₁	Max/Mand	1.27	Max/Mand	1.95	0.001
M ₂	Max/Mand	1.33	Max/Mand	1.86	0.001

TABLE 4
TOOTH POSITIONS IN ORDER OF INCREASING VARIABILITY OF
DISCREPANCY, ACCORDING TO QUADRANT
(The Positions inside Parentheses Had Statistically
Homogeneous Variances.)

	Least Variability	Most Variability
Maxillary left.....	C, (I ₁ , I ₂ , P ₁ , P ₂),	M ₁ , M ₂
Maxillary right.....	(I ₁ , I ₂ , C), (P ₁ , P ₂),	M ₁ , M ₂
Mandibular left.....	(I ₁ , I ₂ , C, P ₁ , P ₂),	(M ₁ , M ₂)
Mandibular right.....	(I ₁ , I ₂), (C, P ₁),	(P ₂ , M ₁), M ₂

within the same parentheses. In general, the variability of the discrepancies increases from central incisor to second molar.

In lieu of the six-way analysis of variance, 28 three-way analyses of variance were performed (one for each tooth position). Each analysis was a test of the differences of means for investigator, cast, and measure and showed that only investigator differences were significant. The observed differences of means between investigators were calculated and are summarized according to position in Table 5. While 19 out of 28 are statistically significant at the 5 per cent level and 16 at the 1 per cent level, these differences are all (with the exception of the maxillary left second molar difference) less than 0.1 mm. in size. This does not seem to be a difference of practical significance. However, it does illustrate a point of some importance. For example, had one investigator measured the teeth of females and the other the teeth of males and the problem

TABLE 5
SUMMARY OF OBSERVED DIFFERENCES OF MEANS BETWEEN TWO
INVESTIGATORS ACCORDING TO TOOTH POSITION

	I1	I2	C	P1	P2	M1	M2	Quadrant Means
Maxillary left	0.074*	0.022	0.036*	0.049*	0.050*	0.043	0.104*	0.054
Maxillary right063*	.029*	.031*	.097†	.063*	— .100*	— .015	.024
Mandibular left068*	.074*	.011	.053*	.003	.028	.036	.039
Mandibular right	0.081*	0.028*	0.035†	0.025	0.072*	0.029	0.049†	0.046

* Significant at 1 per cent level.

† Significant at 5 per cent level.

been that of sex differences in mesiodistal tooth diameter, a statistically significant sex difference could have been obtained which very well might, in reality, have been an investigator discrepancy.

CONCLUSIONS

1. No significant trends are apparent in the analysis of the variability of discrepancies according to jaw (maxilla versus mandible) or side (right versus left). Any differences found may be termed "experimental errors."

2. As was suspected in Part I, the variability in discrepancies generally increases from central incisor to second molar, but these remain within the acceptable limits of experimental error.

3. Tests of mean differences between investigators, casts, and first and second measures were significant at the 0.1 level only for differences between investigators.

4. With one exception, the differences between investigators were less than 0.1 mm. in size. A difference of this magnitude may be of no practical significance. The total mean difference between investigators was 0.04 mm.

PART III

The purpose of the third part of the study was to compare measurements of tooth size made in the mouth with those made on plaster casts and to compare measurements

of tooth size made on soaped casts with those made on unsoaped casts. Twenty Sophomore dental students with complete permanent dentitions from central incisors through first molars provided the subjects for this investigation.

METHOD

Intraoral measurements were made on the teeth of the twenty subjects using the same sliding caliper, as in the previous studies. The same investigator then measured the teeth on the plaster casts. Both intraoral and cast measures were then repeated. Finally, the casts were soaped and measured again.

RESULTS

A preliminary comparison of the size of the differences between first and second measures on the casts and in the mouth showed that the intraoral differences were

TABLE 6
MEANS OF DIFFERENCES BETWEEN CAST AND INTRAORAL MEASURES WITH
STANDARD DEVIATIONS AND *t*-SCORES FOR 40 DOUBLE DETERMINA-
TIONS OF MESIODISTAL TOOTH DIAMETER

	LEFT				RIGHT			
	Mean Diff. (Mm.)	S.D.	<i>t</i>	Prob.	Mean Diff. (Mm.)	S.D.	<i>t</i>	Prob.
<i>Maxilla:</i>								
I ₁	0.08	0.38	1.40	0.04	0.20	1.27
I ₂	-.02	.14	0.89	-.02	.35	0.27
C.....	.08	.16	3.35	0.01	.10	.35	1.71
P ₁09	.26	2.13	.05	.08	.26	2.06	0.05
P ₂16	.29	3.55	.01	.18	.25	4.71	.01
M ₁37	.32	7.55	.01	.31	.34	5.92	.01
<i>Mandible:</i>								
I ₁	-.02	.12	1.300	.39
I ₂	-.07	.18	2.39	.05	-.04	.21	1.21
C.....	.08	.19	2.74	.01	.04	.17	1.61
P ₁04	.25	1.1004	.14	2.13	.05
P ₂10	.24	2.54	.05	.03	.22	0.90
M ₁	0.13	0.33	2.45	0.05	0.23	0.28	5.16	0.01

clearly larger and more variable in the maxilla. In the mandible, however, the size of the mean difference between first and second measures was the same intraorally as on the casts, although the variability of the former was very slightly greater.

The cast measures were consistently larger than the intraoral measures, although in the mandible this constituted a smaller amount than in the maxilla. This difference is approximately 0.1 mm., on the average, with the posterior teeth showing the largest differences, especially in the maxilla.

The second measures were somewhat less variable than the first but showed larger mean differences posteriorly than did the first. Since the variances of first and second measures in the cast versus intraoral comparisons were closely similar, the first and second measures were combined.

Table 6 shows these pooled findings. Note the rather large differences to be found in posterior measures. The average mandibular difference is less than 0.1 mm., and the average maxillary difference is 0.12 mm.

The study of soaped versus non-soaped casts showed slightly larger measures for tooth size on soaped casts. While 10 of the 24 teeth studied were significantly larger, the actual size of the difference was less than 0.08 mm. in all cases, and significance was achieved by small variances rather than by large actual differences.

CONCLUSIONS

1. It is clearly more difficult to measure teeth in the mouth than on plaster casts. However, the differences which resulted from this difficulty were not important for anterior teeth. On the other hand, for second bicuspid and molars the cast measurements were systematically 0.1 mm. larger than the intraoral measurements. This was probably due to the difficulty encountered in establishing the widest mesiodistal diameter, particularly in the maxilla. There would seem, then, to be considerable advantage in making measurements of tooth size on plaster casts rather than in the mouth.

2. The comparison between soaped and non-soaped casts suggests that soaping systematically increases the size of the measurement, probably simply by the addition of a slight film. However, the size of this increase is insignificant as far as individual teeth are concerned.

SUMMARY

This investigation of experimental errors and discrepancies involved in the measurement of the mesiodistal tooth diameter on plaster casts included nine comparisons. These consisted of consideration of differences between first and second measurements, differences between measures made by two investigators, differences between measures made with dividers and with sliding calipers, differences between two sets of replicate casts, differences in measuring maxilla versus mandible and left versus right, differences between tooth types, differences between measures made on casts and in the mouth, and differences between measures made on soaped and unsoaped casts. The mean differences and their variabilities were studied.

Of the nine comparisons, four showed only experimental errors. These were comparisons between two investigators, between two sets of replicate casts, between maxillary and mandibular measures, and between left and right sides.

Certain teeth, such as maxillary molars and lateral incisors and mandibular cuspids, were found to be more difficult to measure for morphologic reasons.

To some extent, the more an investigator measures, the more accurately he does so. Also, the differences between measurements made by two investigators can usually be shown to be statistically significant, although these differences are of small magnitude.

The remaining three comparisons showed systematic differences: measurements made with dividers were approximately 0.1 mm. larger, on the average, than measurements made with sliding calipers; measurements made intraorally of maxillary teeth averaged almost 0.1 mm. smaller than those made on plaster casts of the same teeth; and measurements made of soaped casts were systematically larger, though very slightly so, than measures made on unsoaped casts.

REFERENCES

1. BEERS, YARDLEY. *Introduction to the Theory of Error*. Cambridge, Mass.: Addison-Wesley Publishing Co., Inc., 1953.
2. SELMER-OLSEN, REIDER. *An Odontometrical Study on the Norwegian Lapps*. Oslo: I Kommissjon Hos Jacob Dybwad, 1949.
3. SEIPEL, C. M. Variation of Tooth Position, *Svensk Tandlakare Tidskr.*, suppl. 39, 1946.
4. LUNDSTROM, ANDERS. *Tooth Size and Occlusion in Twins*. Basle: S. Karger, 1948.
5. MOORREES, C. F. A., THOMSEN, S. O., JENSEN, E., and KAI-JEN YEN, P. Mesio-distant Crown Diameters of the Deciduous and Permanent Teeth in Individuals, *Am. J. Dent. Res.*, **36**:40, 1957.