Intra- and inter-observer variability and accuracy in the determination of linear and angular measurements in computed tomography

An in vitro and in situ study of human mandibles

Edwin L. Christiansen, Joseph R. Thompson and Sigvard Kopp
Department of Endodontics, Section of TMJ and Orofacial Pain, School of Dentistry,
Loma Linda University, and Department of Radiation Sciences, Section of Neuroradiology,
Loma Linda University Medical Center, Loma Linda, California, USA, and Department of
Stomatognathic Physiology, School of Dentistry, University of Lund, Malmö, Sweden

Christiansen EL, Thompson JR, Kopp S. Intra- and inter-observer variability and accuracy in the determination of linear and angular measurements in computed tomography. An in vitro and in situ study of human mandibles. Acta Odontol Scand 1986;44:221–229. Oslo. ISSN 0001–6357.

The observer variability and accuracy of linear and angular computed tomography (CT) software measurements in the transaxial plane were investigated for the temporomandibular joint with the General Electric 8800 CT/N Scanner. A dried and measured human mandible was embedded in plastic and scanned in vitro. Sixteen observers participated in the study. The following measurements were tested: inter- and extra-condylar distances, transverse condylar dimension, condylar angulation, and the plastic base of the specimen. Three frozen cadaveric heads were similarly scanned and measured in situ. Intra- and inter-observer variabilities were lowest for the specimen base and highest for condylar angulation. Neuroradiologists had the lowest variability as a group, and radiology residents and paramedical personnel had the highest, but the differences were small. No significant difference was found between CT and macroscopic measurement of the mandible. In situ measurement by CT of condyles with structural changes in the transaxial plane was, however, subject to substantial error. It was concluded that transaxial linear measurements of the condylar processes free of significant structural changes had an error and an accuracy well within acceptable limits. The error for angular measurements was significantly greater than the error for linear measurements. ☐ *Condylar dimension; radiology; temporomandibular joint*

Edwin L. Christiansen, Department of Endodontics, Section of TMJ and Orofacial Pain, School of Dentistry, Loma Linda University, Loma Linda, CA 92350, USA

The need for enhanced radiographic imaging of the temporomandibular joint (TMJ) has led to the application of X-ray computed tomography (CT) in the study of TMJ disorders. Recent CT investigations of the TMJ have dealt with imaging of articular discs (1–9), various scan protocols (10, 11), three-dimensional imaging of the osseous joint components (12), and degenerative joint disease (13).

Questions have arisen as to the magnitude and importance of errors in linear and angular measurements of the osseous TMJ components when utilizing CT software programs. In the measurement of joint space or transverse condylar dimension, for example, how closely do CT measurements correspond to direct measurements of the joint components? What is the error of those individuals involved in the collection of patient-related and research data (inter-observer error)? And what is the error of an investigator collecting similar data on different occasions (intra-observer error)? We felt it important that these questions be addressed because they reflect on the credibility of clinical and research TMJ CT findings.

A review of the literature in English revealed no papers on the topic of observer variability or accuracy of CT-generated mandibular measurements. Nor were any papers found comparing in vitro and in situ CT measurements of the mandible or its component parts. The purpose of this investigation, therefore, was to study intra- and inter-observer variability and accuracy in the CT software determination of linear and angular TMJ measurements. Additionally, we attempted to determine what influence such variables as target factor, radiographic experience, the operator's CT experience, and learning effects might have on the interpretation of CT data.

Materials and methods

Phantom head

A dried and intact human mandible with a dental age of approximately 9 years was used in this study. There were no gross signs of abnormality in the body of the mandible or the condylar processes. Macroscopic linear measurements were made of the transverse condylar dimensions and the extra- and inter-condylar distances. These distances were measured in millimeters (mm), from the heights of contours of the respective condylar poles. These procedures were performed by five observers with an Etalon dial micrometer incremented to 0.02 mm, and the results were averaged. The extracondylar distance averaged 93.5 mm (SD = 0.05); the intercondylar distance, 66.1 mm (SD = 0.28); the right transverse condylar dimension, 14.0 mm (SD = 0.03); and the left transverse condylar dimension, 14.3 mm (SD = 0.09).

The mandible used for in vitro examination was embedded in clear polymer plastic in preparation for scanning. The hardened plastic was cut to the approximate size and shape of a human skull, producing a phantom head. The sides and base of the phantom head were squared and paralleled. To mimic the effect of dental restorations, strips of an intraoral dental film lead shield (thickness, 0.06 mm) were cemented on opposite sides of the base. The CT images of the cortical margins of the left condyle were distinct, whereas those of the right condyle were more indistinct and could therefore presumably influence the measurements under study.

CT-imaging and measurement

Imaging was done with a General Electric 8800 CT/N Scanner with 1.5-mm-thick high-resolution axial sections sequenced every 1.0 mm. The prospective scan was targeted at 1.0 with a bone algorithm. The scan was ReViewed (General Electric Co.) for optimal bone detail and targeted at 1.2, 1.6, and 3.0 (Fig. 1). These target factors (TF) were chosen because they are used most frequently in skull, temporal bone, and TMJ CT studies. At a target factor of 3.0 it was necessary to ReView the right and left condyles individually because the images were too large to be simultaneously displayed on the video display terminal (VDT).

Computer measurements of the CT condylar images were made with the Exam and Dytest (General Electric Co.) software programs. Linear measurements are displayed in centimeters (cm) on the VDT to the nearest 0.01 cm. Angulations are displayed in degrees to the nearest 0.1°. Measurements in centimeters were converted to millimeters for convenience.

The level in the scan sequence selected for this study was determined by measuring the transverse condylar dimension for each axial section made in the scan sequence. This was done many times so that the level with the greatest extracondylar distance might be determined. This was taken to correspond to the level at which direct caliper measurements were made at the heights of contour.

A test protocol consisting of 32 specific, sequenced instructions along with a drawing of the condyles was used (Fig. 2). This protocol was designed in such a manner that responses to identical tasks did not appear on the same page, and the sequence of the tasks was rearranged in the second half of the test to minimize learning effects. The test consisted of eight measurements: extra- and inter-condylar distances, transverse condylar dimension, condylar angulation, and measurement of the width of the base of the phantom head at sections with and without the metal markers (Fig. 2). Each measurement was made twice at TF 1.0, 1.2, 1.6, and 3.0. Measurements at TF 3.0 were per-

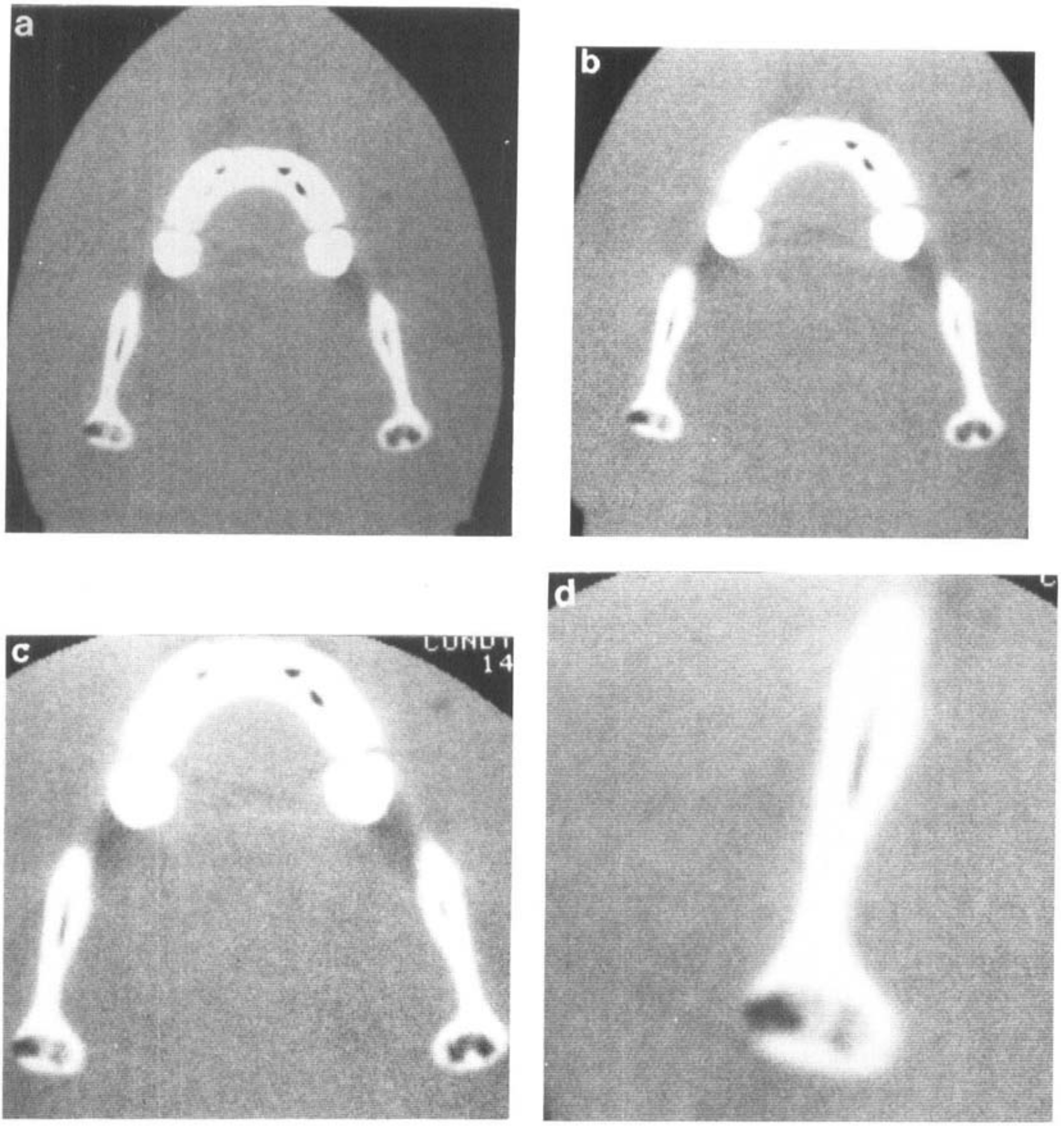

Fig. 1a. Axial section through the greatest transverse dimensions of the condylar processes at a target factor of 1.0. 1b. The same CT section as in 1a, targeted at 1.2. 1c. The same CT section as in 1a and b, targeted at 1.6. 1d. The right condyle from 1a–c, targeted at 3.0.

formed only for condylar transverse dimension and condylar angulation because both condyles could not be simultaneously displayed.

The participants in the study were divided into five groups in accordance with an arbitrarily presumed familiarity with CT images: group I, neuroradiologists ($n = 3$); group II, radiology residents ($n = 3$); group III, CT technologists ($n = 3$); group IV, dentists ($n = 4$); and group V, paramedical personnel ($n = 3$). These groups were studied to test whether experience with CT, with the TMJ, or with both, might affect the reproducibility and accuracy of the measurements.

Intraobserver variability was studied specifically for two of the observers (nos. 4 and 5). Observer 4 is a neuroradiologist

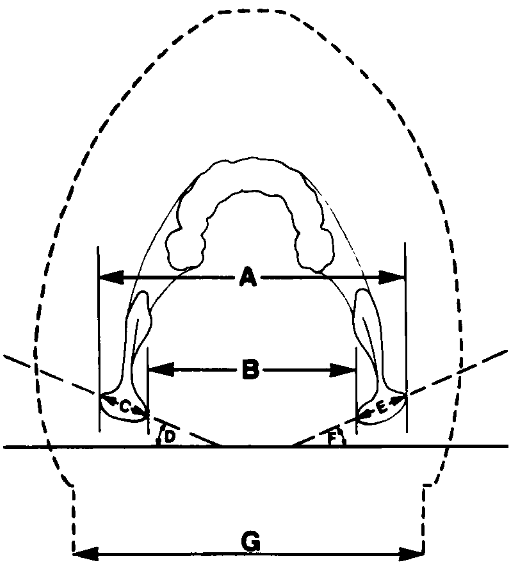

Fig. 2. Illustration used in the instruction booklet to demonstrate the configuration of the phantom head and the measurements to be made on the VDT with the software program. Extracondylar distance (A), intercondylar distance (B), transverse condylar distance (C, E), condylar angulation (D, F), and phantom head base (G).

specializing in pediatric neuroradiology with experience in TMJ CT. Observer 5 is a dentist and the director of a TMJ/orofacial pain clinic with experience in TMJ CT. These two observers performed the test five times at random intervals over a period of 6 months.

Fresh-frozen specimens

Three fresh-frozen human cadaveric heads were scanned with the protocol described above. A bone algorithm was selected, and the axial images were displayed with extended scale windowing (window 4000, level 80), and the maximum extracondylar, intercondylar, and transverse condylar distances were measured in succeeding sections at 1.0-mm intervals from superior to inferior. The specimens were thawed, and the TM joints were dissected. On removal of the mandibles from the skull base, single measurements were made of the extracondylar, intercondylar, and transverse condylar distances with the Etalon dial micrometer. The presence and severity of condylar structural changes were noted and were graded as mild, moderate, marked, or severe. Bilateral structural changes were found on the condyles of specimen 1 which were mild to moderate on the right and marked to severe on the left. No structural changes were noted on specimen 2. There were moderate structural changes on the right condyle and an unremarkable left condyle on specimen 3.

Statistics

The absolute observer error is expressed as the mean of absolute differences (di) between double determinations. The standard deviation of these differences (S di) was calculated as an estimation of the variance of the observer error. The relative error is expressed in percentage and was calculated by the formula: di × 100/mean of the variable. Variances of intra- and inter-observer differences were tested for equality by the F-test. Differences between intra- and inter-observer measurements and between CT measurements and macroscopic measurements were then tested with Student's t test (comparing the means of two small samples from normal populations). The levels of statistical significance (P values) are shown when $P < 0.05$, whereas $P > 0.05$ is considered not significant.

Results

Intraobserver variability

The lowest relative variability for observers 4 and 5 was found for the measurement of the base of the phantom head with or without metal markers (0.3%) and for the extracondylar distance (0.4–0.5%), and the highest was found for condylar angulation (12.2–16.0%). The absolute variability between duplicate measurements is shown in Fig. 3. A statistically significant difference was found between the two observers for measurement of right transverse condylar dimension. The variability was thereby lower for observer 5 (3.0%) than for observer 4 (5.7%; $P < 0.01$) (Fig. 3).

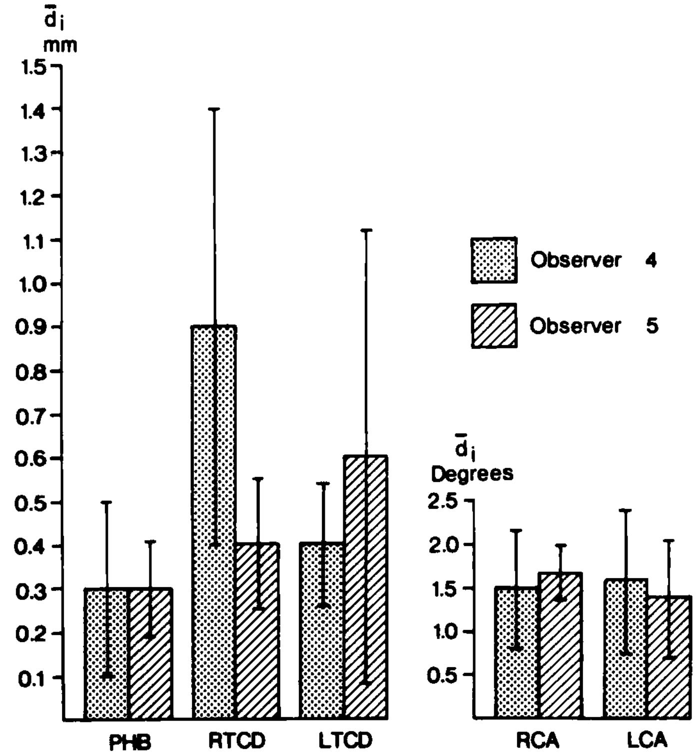

Fig. 3. Absolute intraobserver variability for observers 4 and 5. Differences (di) between measurements are calculated on the basis of three double (one at each target factor) determinations taken on five different occasions ($n = 15$). The variables are phantom head base (PHB), right and left transverse condylar dimensions (RTCD and LTCD), and right and left condylar angulation (RCA and LCA). The mean value of the absolute differences between the double determinations (di) is shown, with ±1 standard deviation of the absolute differences. Measurements in millimeters and degrees.

The variability of measurement for the right (indistinct cortical margin) and left (distinct cortical margin) transverse condylar dimension was statistically different for observer 4 ($P < 0.01$).

Intraobserver variability differed slightly with target factor, and the variability of measurements was lowest at TF 1.6 for extracondylar distance, intercondylar distance, right transverse condylar dimension, condylar angulation, and phantom head base distance. The difference between TF 1.2 and 1.6 was statistically significant for intercondylar distance ($P < 0.05$).

Interobserver variability

The lowest relative variability was found for measurement of the base of the phantom head with or without metal markers (0.2–0.3%) and extracondylar distance (0.7%) and the highest for condylar angulation (15.8–20.3%). The absolute variability is shown in Fig. 4. The lowest interobserver variability was found at TF 1.6 and 3.0. There was a statistically significant difference between TF 1.2 and 1.6 for measurement of intercondylar distance ($P < 0.001$) and between TF 1.2 and 3.0 for left condylar angulation ($P < 0.01$).

The lowest relative variability of measurement for a group was found for the neuroradiologists for five of the eight variables (ECD, 0.5%; RTCD, 2.7%; LTCD, 2.0%; RCA, 8.8%; and PHB, 0.2%). CT radiology technicians had the lowest variability for two (ICD, 0.9%; and LCA, 15.3%), whereas the group of dentists had the lowest variability for one (PHB, 0.1%) and shared the lowest variability for measurement of ECD (0.5%). There was a statistically significant difference in variability between the groups for measurement of intercondylar distance ($P < 0.05$), for which the group of CT technicians showed the lowest variability and the radiology residents the highest (Fig. 5). The highest variability was found for the group of radiology residents for three of the variables

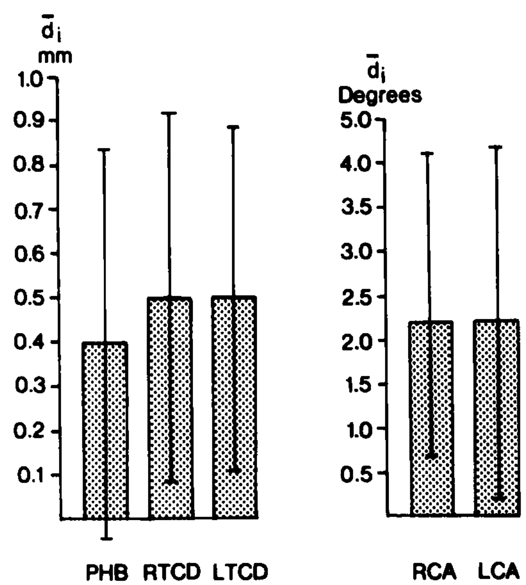

Fig. 4. Absolute interobserver variability of 16 observers on the basis of differences (di) between 3 or 4 double determinations (one at each target factor). The variables are phantom head base (PHB), right and left transverse condylar dimension (RTCD and LTCD), and right and left condylar angulation (RCA and LCA). The mean of the absolute differences (di) between the double determinations is shown, with ±1 standard deviation of the absolute differences. Measurements in millimeters and degrees.

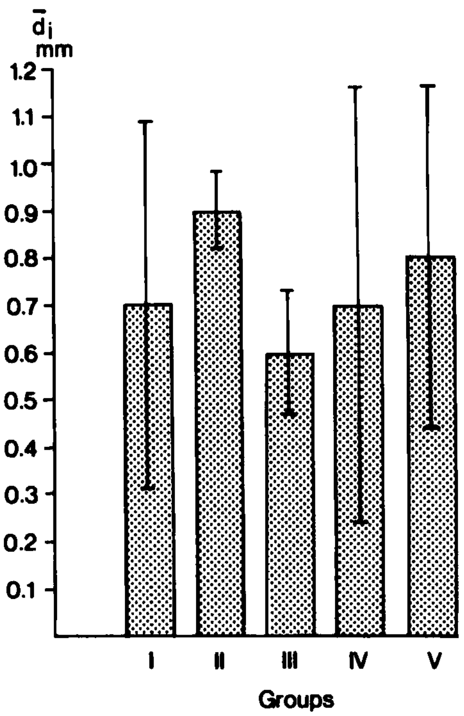

Fig. 5. Interobserver variability for CT measurement of intercondylar distance (ICD) in five groups of observers with different experience with CT and temporomandibular joint anatomy/pathology. Neuroradiologists (group I; $n = 3$), radiology residents (group II; $n = 3$), CT technicians (group III: $n = 3$), dentists (group IV; $n = 4$), and paramedical personnel (group V; $n = 3$). The mean of the absolute differences (di) of each group is shown based on three double determinations (one at each target factor) of each observer, with ±1 standard deviation of the absolute differences. Measurements in millimeters and degrees.

(ICD, 1.4%; LTCD, 4.2%; and PHB, 0.5%), for paramedical personnel for two (ECD, 1.1%; and RCA, 28.8%), and for CT technicians and dentists for one (RTCD, 5.7%; and LCA, 25.0%, respectively). No difference was found in interobserver variability between measurement of the right and left condylar transverse dimension or between right and left condylar angulation.

Accuracy

The average value of the extracondylar distance measured by CT by the 16 observers was greater than the corresponding value from the macroscopic measurement, and the difference was on an average 0.8 mm (SD = 0.94), whereas the intercondylar distance was on an average 0.4 mm shorter (SD = 0.58). The value of CT measurement was greater than that of macroscopic measurement for both the right and left transverse condylar dimension, and the average difference was 0.4 mm (SD = 0.74 and 0.79, respectively). The difference between CT and macroscopic measurement was not significantly different from zero for any of the variables.

No statistically significant differences in observer accuracy of the CT measurements could be found between the different target factors. No statistically significant differences in observer accuracy were found between groups with different observer experience. There was no difference with regard to accuracy of CT measurement between the left and right condylar transverse dimension—the distinct and indistinct cortical margins.

Fresh-frozen specimens

The accuracy of linear measurements by CT utilizing three fresh-frozen specimens is shown in Table 1. The joint components with structural changes introduced substantial error in CT linear measurements.

Discussion

The size of the picture element (pixel) in the VDT grid is determined by the target (magnification) factor and varies inversely with targeting. The amount of information contained in a pixel at a target factor of 3.0 is less than that contained in a pixel at a target factor of 1.0. Pixel size at the target factors used in this study with the General Electric 8800 was as follows: TF 1.0 pixel = 0.80 mm, TF 1.2 pixel = 0.6667 mm, TF 1.6 pixel = 0.50 mm, and TF 3.0 pixel = 0.2667 mm.

Table 1. Accuracy in the determination of linear measurements in computed tomography (CT) of three fresh-frozen mandibles (nos. 1, 2, and 3). CT measurements (CTM) and direct measurements (DM) in millimeters of the transverse condylar dimensions of the right and left condyles (RTCD, LTCD) and of the extracondylar distance (ECD) and the intercondylar distance (ICD) were made. The ratio of direct measurements to CT measurements (DM/CTM) and the difference in millimeters (di) is shown. Specimen 1, mild structural changes on the right condyle, marked to severe on the left; specimen 3, moderate changes on the right side, normal on the left

	CTM	DM	DM/CTM	di
Mandible 1				
RTCD	18.9	18.8	0.99	0.1
LTCD	11.7	14.7	1.26	−3.0
ECD	113.8	111.1	0.98	2.7
ICD	81.3	79.0	0.97	2.3
Mandible 2				
RTCD	19.5	19.2	0.99	0.3
LTCD	19.5	19.7	1.01	−0.2
ECD	121.4	121.3	1.00	0.1
ICD	83.4	83.4	1.00	0.0
Mandible 3				
RTCD	17.4	18.2	1.05	−0.8
LTCD	17.7	18.2	1.05	−0.5
ECD	116.0	116.7	1.01	−0.7
ICD	83.3	83.9	1.01	−0.6

Cursor movements on the VDT are not linear but incremental, related to pixel size. The maximum error in a cursor measurement on a vertical or horizontal scale would be equal to 1 pixel, whereas the maximum error in any pixel would be the diagonal of the pixel, and would be the square root of 2 multiplied by the pixel size (personal communication, Technical Staff, General Electric Company).

Because of the relationship between target factor and pixel size, testing for accuracy by varying the target factor is identical to testing in accordance with pixel size. Hence, the minimal linear instrumental measurement error that can be expected at TTF 1.0 would be plus or minus 0.80 mm, whereas the minimal error at TF 3.0 would be plus or minus approximately 0.25 mm. Interestingly, these figures compare very closely with the average intraobserver error (0.4–0.9 mm), interobserver error (0.5–0.8 mm), and accuracy (0.4–0.8 mm) of the CT linear measurements of the phantom head.

Observer variability was greatest for measurement of condylar angulation and lowest for the linear measurements, as anticipated. Both the intra- and inter-observer variability were lowest in the measurement of the phantom head base, as might also be expected since the visualization of the air/plastic interface could be made more readily than that of the bone/plastic interface. The metal markers on the phantom head base did not influence the observer variability consistently in this study.

Both intra- and inter-observer variability of the linear measurements were low compared with those usually obtained in clinical studies. In a study in which Kopp & Wenneberg (14) looked at the relative intraobserver and interobserver error in measuring jaw mobility, they found errors between 2% and 50%, depending on the jaw movement being evaluated.

Observers 4 and 5 made measurements on 5 different days to investigate the intraobserver variability during a 6-month period and to ascertain whether learning took place. There was no evidence of a learning effect in this study, since no improvement was noted in their results from one test to another.

Reproducibility of measurement seems to be little affected by experience, since the differences between groups were generally numerically small. The neuroradiologists had the lowest variability overall, whereas the radiology residents and paramedical personnel had the highest. However, there was no difference in observer variability of practical significance between the groups. The neuroradiologists, having the greatest experience with CT images, were expected to produce the most consistent measurements, which in fact they did. However, both the intra- and inter-observer errors in the linear measurements were small. It seems reasonable to conclude that paramedical personnel can make such measurements as accurately and reproducibly as do neuroradiologists, provided the variables are well defined.

The extensive testing of observers 4 and 5 was done to determine whether their additional experience with TMJ CT would result in more accurate and consistent responses. Such a difference was not substantiated by the results of this study.

The difference in the radiographic integrity of the cortical margins of the right and left condyle of the phantom head mandible did not affect observer variability consistently. Observer variability was generally lowest at the two highest target factors, although the differences, with a few exceptions, were small. Neither target factor nor observer experience affected the accuracy of measurement to any significant degree.

No significant difference was found between CT and macroscopic measurements of the normal mandibles either in vitro or in situ; however, the in situ CT measurements of condyles with structural changes were subject to substantial error.

Because of the technical difficulties involved in the determination of condylar angulation on the phantom head and the fresh-frozen specimens, these measurements were omitted, and, accordingly, no estimation of the accuracy of angular measurement was made.

It can be concluded that linear measurements by CT have an observer error and accuracy within acceptable limits whether they are done in vitro or in situ on normal temporomandibular joint components. Linear measurements by CT of the osseous components of the TMJ can be expected to be accurate to within 0.4–0.8 mm. Measurements of diseased joints in the transaxial plane may produce inaccurate results.

References

1. Thompson JR, Christiansen EL, Hasso AN, Hinshaw DB Jr. Dislocation of the temporomandibular joint disk demonstrated by CT. AJNR 1984;5:115–6.
2. Thompson JR, Christiansen EL, Hasso AN, Hinshaw DB Jr. Temporomandibular joints: high resolution computed tomographic evaluation. Radiology 1984;150:105–10.
3. Thompson JR, Christiansen EL, Sauser DD, Hasso AN, Hinshaw DB Jr. Contrast arthrography versus computed tomography for the diagnosis of dislocation of the temporomandibular joint meniscus. AJNR 1984;5:747–50.
4. Christiansen EL, Thompson JR. Anteriorly displaced temporomandibular joint articular disc; a case diagnosed by computed tomography. Oral Surg 1984; 58: 355–7.
5. Christiansen EL, Thompson JR, Kopp S. Temporomandibular joint research: a status report. J Cal Dent Assoc 1984;12:155–7.
6. Katzberg RW, Dolwick MF, Keith DA, Helms CA, Guralnick WC. New observations with routine and CT-assisted arthrography in suspected internal derangement of the temporomandibular joint disc. Oral Surg 1981;51:569–74.
7. Helms CA, Moorish RB, Kircos LT, Katzberg RW, Dolwick MF. Computed tomography of the meniscus of the temporomandibular joint: preliminary observations. Radiology 1982;145:719–22.
8. Helms CA, Vogler JB, Morrish RB. Diagnosis by computed tomography of temporomandibular joint meniscus displacement. J Prosthet Dent 1984; 51(4):544–7.
9. Helms CA, Katzberg RW, Moorish R, Dolwick MF. Computerized tomography of the temporomandibular joint. J Oral Maxillofac Surg 1983;41:512–7.
10. Manzione JV, Seltzer SE, Katzberg RW, Hammerschlag SB, Chiango BF. Direct sagittal computed tomography of the temporomandibular joint. AJNR 1982:3;677–9.
11. Manzione JV, Katzberg RW, Brodsky GL, Seltzer SE, Mellins HZ. Internal derangements of the tem-

poromandibular joint: diagnosis by direct sagittal computed tomography. Radiology 1984;150:111–5.
12. Roberts D, Pettigrew J, Udupa J, Ram C. Three-dimensional imaging and display of the temporomandibular joint. Oral Surg 1984;58:461–74.
13. Christiansen EL, Thompson JR, Kopp S, Hasso AN, Hinshaw DB Jr. Radiographic signs of temporomandibular joint disease: an investigation utilizing x-ray computed tomography. J Dento-Maxillo-Fac Radiol 1985;14(2):83–91.
14. Kopp S, Wenneberg B. Intra- and interobserver variability in the assessment of signs of disorder in the stomatognatic system. Swed Dent J 1983;7:239–46.

Received for publication 14 February 1986