Variability in measurement of abdominal aortic aneurysms

Frank A. Lederle, MD, Samuel E. Wilson, MD, Gary R. Johnson, MS, Donovan B. Reinke, MD, Fred N. Littooy, MD, Charles W. Acher, MD, Louis M. Messina, MD, David J. Ballard, MD, PhD, and **Howard J. Ansel, MD,** for the **Abdominal Aortic Aneurysm Detection and Management Veterans Administration Cooperative Study Group,*** *Minneapolis, Minn., Orange, Calif., West Haven, Conn., Hines, Ill., Madison, Wis., Ann Arbor, Mich., and Decatur, Ga.*

Purpose: The purpose of this study was to report interobserver and intraobserver variability of computed tomography (CT) measurements of abdominal aortic aneurysm (AAA) diameter and agreement between CT and ultrasonography observed in the course of a large, multicenter, randomized trial on the management of small AAAs.
Methods: CT measurements of AAA diameter from participating centers were compared with measurements made from the same scan by a central laboratory. Blinded central remeasurement of a randomly selected subset of these CT scans was used to assess intraobserver variability. Agreement between AAA measurements by CT and ultrasonography done within 30 days of each other was also assessed.
Results: For interobserver pairs of local and central CT measurements of AAA diameter (n = 806), the difference was 0.2 cm or less in 65% of pairs, but 17% differed by at least 0.5 cm. For intraobserver pairs of central CT remeasurements (n = 70), 90% differed by 0.2 cm or less, 70% were within 0.1 cm, and only one differed by 0.5 cm. Of 258 ultrasound-measured and central CT pairs, the difference was 0.2 cm or less in 44% and at least 0.5 cm in 33%. Ultrasound measurements were smaller than central CT measurements by an average of 0.27 cm ($p < 0.0001$). Local CT and ultrasound measurements showed a marked preference for recording by half centimeter.
Conclusions: A high degree of precision is possible in CT measurement of AAA diameter, but this precision may not be obtained in practice because of differences in measurement techniques. Differences between imaging modalities increase variability further. Variations in AAA measurement of 0.5 cm or more are not uncommon, and this should be taken into account in management decisions. Efforts to reduce variation in measurement are warranted and might include (1) seeking agreement between surgeons and radiologists on a precise definition of AAA diameter, (2) limiting the number of radiologists who measure AAAs, and (3) use of calipers and magnifying glass for CT measurements. (J VASC SURG 1995;21:945-52.)

The best predictor of rupture of an asymptomatic abdominal aortic aneurysm (AAA) is the maximum diameter of the aneurysm.[1] The decision to repair an AAA electively is therefore usually based primarily on this diameter. Ultrasonography and computed tomography (CT) are the principal modalities used to measure AAAs. Because a size difference of a few millimeters may determine whether a patient is

From the Department of Medicine and the Diagnostic Radiology Service, Veterans Affairs Medical Center, Minneapolis, the Department of Surgery, University of California, Irvine, the Cooperative Studies Program Coordinating Center, Veterans Affairs Medical Center, West Haven, the Department of Surgery, Veterans Affairs Medical Center, Hines, the Department of Surgery, Veterans Affairs Medical Center, Madison, the Department of Surgery, Veterans Affairs Medical Center, Ann Arbor, and Emory University Center for Clinical Evaluation Sciences, Decatur.

Supported by the Cooperative Studies Program of the Medical Research Service, Department of Veterans Affairs Central Office, Washington, D.C.

Reprint requests: Frank A. Lederle, MD, Veterans Affairs Medical Center, Department of Medicine (III-0), One Veterans Dr., Minneapolis MN 55417.

*See appendix.

24/1/62997

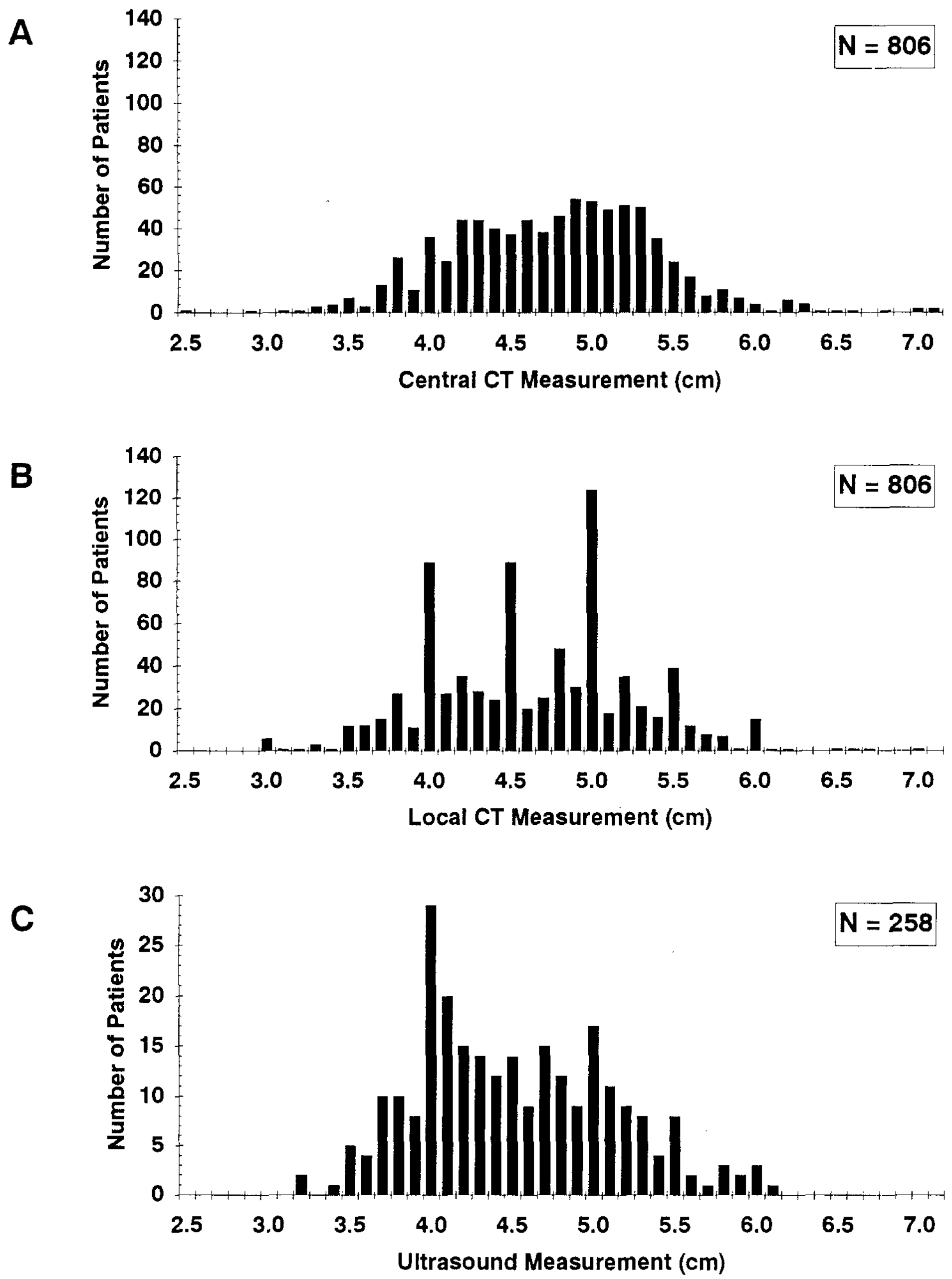

Fig. 1. Frequency distribution of measurements of maximum AAA diameter.

offered surgery, the precision of the measurement is of considerable importance.

Several recent studies have examined the variability of ultrasonography for measuring AAA diameter.[2-4] Comparison of ultrasound and CT measurements has been limited to a few small series,[3,5-7] and we know of no previous studies of the variability of CT measurements. In this article we report interobserver and intraobserver variability in CT measurements of AAA diameter and agreement between CT and ultrasound measurements, observed in the course of a large, multicenter, randomized trial on the management of small AAAs.

METHOD

The AAA Detection and Management (ADAM) Study is an ongoing Department of Veterans Affairs Cooperative Study that began in 1992. Details of the study design have been described previously.[8] Patients aged 50 to 79 years with AAAs 4.0 to 5.4 cm

Table I. Intraobserver and interobserver variability in CT measurement of AAA and agreement between CT and ultrasound measurements

	No. of pairs	*Mean difference (cm)*	*p Value**	*Limits of agreement (cm)†*
Central CT reading 1 vs central CT reading 2	70	0.003	NS	0.31, −0.30
Local CT reading vs central CT reading	806	−0.123	<0.0001	0.57, −0.81
Ultrasound vs central CT reading	258	−0.267	<0.0001	0.70, −1.24
Ultrasound vs local CT reading‡	258	−0.119	<0.0001	0.70, −0.94

*Tests the hypothesis that the mean difference is not different from zero.
†The range within which 95% of the differences would be expected to occur, calculated as the mean difference ± 1.96 times the standard deviation of the differences.[10]
‡These readings were not made blinded to each other.

in diameter who are not at high surgical risk are randomized to immediate surgery or selective surgery (i.e., follow-up imaging at 6-month intervals, with surgery reserved for AAAs that enlarge to 5.5 cm, enlarge rapidly, or become symptomatic). CT measurement of AAA diameter is used to determine eligibility for randomization for all patients and to determine when surgical criteria have been met in the selective surgery group. To ensure consistent measurement throughout the 15 participating centers, all CT scans are read at a central CT laboratory. Most CT scans submitted for central reading have also been read by a radiologist at the participating center, called the "local reading." For quality-control purposes, a proportion of the CT scans read by the central laboratory are randomly selected for blinded central rereading.

In this report we examined interobserver variation in CT measurement of AAA diameter in all CT scans submitted for central reading (regardless of whether the patient was enrolled in the ADAM Study) for which a local reading was also available as of March 18, 1994. We also studied agreement between CT and ultrasound measurement in all ultrasound examinations done within 30 days before these CT scans. In addition, all central rereadings done on these CT scans were used to examine intraobserver variation in CT measurement.

Central CT readings were made blinded to local CT and ultrasound readings. Central rereadings were also blinded to the first central reading. Local CT or ultrasound readings may have been done with knowledge of each other but were done without knowledge of the central CT reading.

Local radiology services were asked to perform CT scans from the diaphragm to the symphysis pubis in 1 cm intervals with 1 cm slice thickness without contrast (unless otherwise indicated) and to measure the maximum external aortic diameter as described below. A copy of each scan was mailed to the central laboratory.

AAA size by CT measurement was defined as the maximum external diameter in any direction, a definition widely used in literature and practice. The diameter of regions of the aorta determined to be tortuous (by tracking aortic position in serial slices) was defined as the diameter perpendicular to the direction of tortuosity. This was done because aortic tortuosity can result in oblique slices on CT images and hence an overestimate of aortic diameter.[9] This aspect of our definition of AAA diameter is unique to the study (although most radiologists try to correct for tortuosity), but the definition was provided to local readers.

The central CT laboratory was staffed by two experienced CT radiologists, the principal reader (D.B.R.), who read 90% of the films, and a backup reader (H.J.A.). For the purposes of this article, these two radiologists are considered together as the "central reader." The central reader used calipers and a magnifying glass to measure the diameter of the aneurysm against the scale provided on the film. These scales are graded in 1 cm intervals, so considerable interpolation is required to determine the size in millimeters. Cursor measurements were not used by the central reader because they are technician dependent at the local site.

Description of the study technique and size definition were supplied to local CT readers on a request form to be used with each reading. No further effort was made to influence readings at the participating sites. Scans were read by the usual staff in their usual manner except as influenced by the study request form. Local measurements were done by staff radiologists or residents or technicians with staff review. An average of five staff radiologists per local site read abdominal CT scans, most of whom specialized in CT radiology. Local measurement techniques included cursor, calipers, and markings on a piece of paper.

The ADAM Study employed an ultrasonographer at each center to conduct the study screening program. This ultrasonographer was instructed to scan the aorta in both the anteroposterior and lateral planes and to report the maximum external AAA diameter in any direction. The ultrasound measurements reported below include both those done by the study ultrasonographer and those done in the usual way by the ultrasound service at the participating center, uninfluenced by study procedures or definitions. No special definition was used for measuring a tortuous aorta by ultrasonography because the probe can be tilted to measure true diameter. There were no central ultrasound measurements.

For each of the comparisons studied, the mean pair difference was calculated and the hypothesis that this value was not different from zero was tested with the paired *t* test. The limits of agreement, representing the range within which 95% of the differences would be expected to occur, were calculated as the mean difference ± 1.96 times the SD of the differences, according to the method of Bland and Altman.[10]

RESULTS

There were 806 interobserver pairs of local and central CT measurements, 70 intraobserver pairs of central CT remeasurements, and 258 CT and ultrasound pairs. Only five of the 806 CT scans demonstrated tortuosity in the widest region of the aneurysm by central CT reading. The distributions of measurements are shown in Fig. 1. The concentration of AAA diameters within a relatively narrow range reflects the 4.0 to 5.4 cm eligible range for the ADAM Study. A marked preference for recording by half centimeter can be seen in the local CT and ultrasound measurements. Forty-seven percent of the local CT measurements and 29% of the ultrasound measurements ended with 0.0 cm or 0.5 cm, significantly more than the 20% expected if all terminal digits were represented equally ($p < 10^{-8}$ for local CT; $p < 0.00015$ for ultrasonography). For local CT measurements, this figure was above 30% at 12 of the 15 participating centers. In comparison, 20.5% of central CT measurements ended with 0.0 cm or 0.5 cm ($p > 0.85$).

Table I shows the agreement between the various measurements according to the parameters described above. On average, local CT measurements of AAA diameter were smaller than the central measurements, and this was the case at 13 of the 15 participating centers. The frequency distribution of the difference between local and central CT measurements is shown in Fig. 2, *B*. This difference was 0.2 cm or less in 65% of pairs (ranging across the 15 participating centers from 41% to 97%), but 17% differed by at least 0.5 cm. As expected, intraobserver variability for the central reader was considerably less (Fig. 2, *A*). The difference was 0.2 cm or less in 90% of the scans (70% were within 0.1 cm), and only 1% differed by at least 0.5 cm. On average, ultrasound measurements were smaller than CT measurements (Table I). The difference between the ultrasound and central CT measurements was 0.2 cm or less in 44% of pairs and at least 0.5 cm in 33% of pairs (Fig. 2, *C*). Agreement between ultrasound and local CT measurements was somewhat better (≤0.2 cm in 51% of pairs and ≥0.5 cm in 28% of pairs) but should be interpreted with caution because these measurements were not made blinded to each other.

Fig. 3 shows the difference between each pair of measures plotted against their mean, according to the method of Bland and Altman.[10] Variability in measurement was not affected substantially by AAA diameter within the range of AAA diameters studied.

DISCUSSION

Variability in AAA measurement has several components, including intraobserver and interobserver variability, differences in the definition of AAA diameter used by different readers, biologic and technical variability resulting from performing the same test on different occasions, and differences between tests such as CT and ultrasonography. Several of these components are addressed in this study.

We found that when measurements were done carefully with calipers and a magnifying glass and a defined protocol, intraobserver variability of measuring AAA diameter on the same CT scan was low. In most cases this difference was 1 mm or less, approaching the limit of accurate interpolation from the scale provided on the scan. Nevertheless, a few rereadings produced substantial differences. Anecdotally, these larger differences resulted from uncertainty as to where the iliac bifurcation began, uncertainty as to whether the aorta was tortuous, and probable interpolation or recording errors. CT measurement of tortuous AAAs is problematic, but tortuosity in the widest region of the aneurysm was rare in our series of mostly small AAAs.

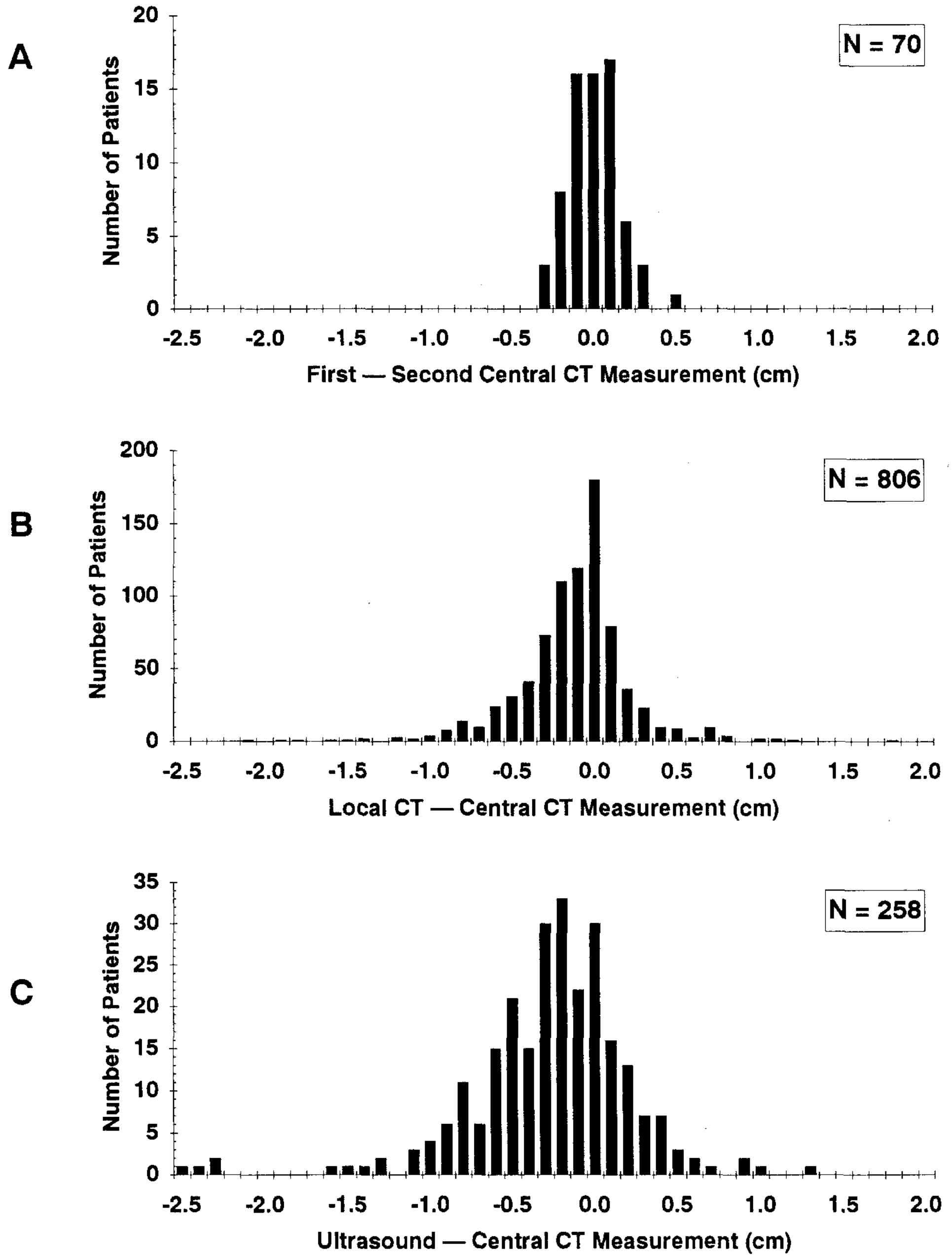

Fig. 2. Frequency distribution of difference between measurements of maximum AAA diameter. Each ultrasound study was done within 30 days of corresponding CT.

As expected, interobserver variability was greater than intraobserver variability. Most interobserver measurements differed by 2 mm or less, but one sixth differed by at least 5 mm. Several specific factors may have contributed to this increased variability. Despite the instructions provided, some local readers may have been using other definitions of maximum AAA diameter. Our definition, the maximum external diameter in any direction, appears to be the most widely used, but others have been proposed, including the maximum width of the shorter axis,[11] the average of the anteroposterior and lateral diameters,[12] and the maximum diameter of the lumen.[7]

Our data also suggest that some local measurements may have been done less meticulously than the central measurements. The distribution of local measurements shows rounding to the half centimeter, a phenomenon termed "terminal digit preference" that

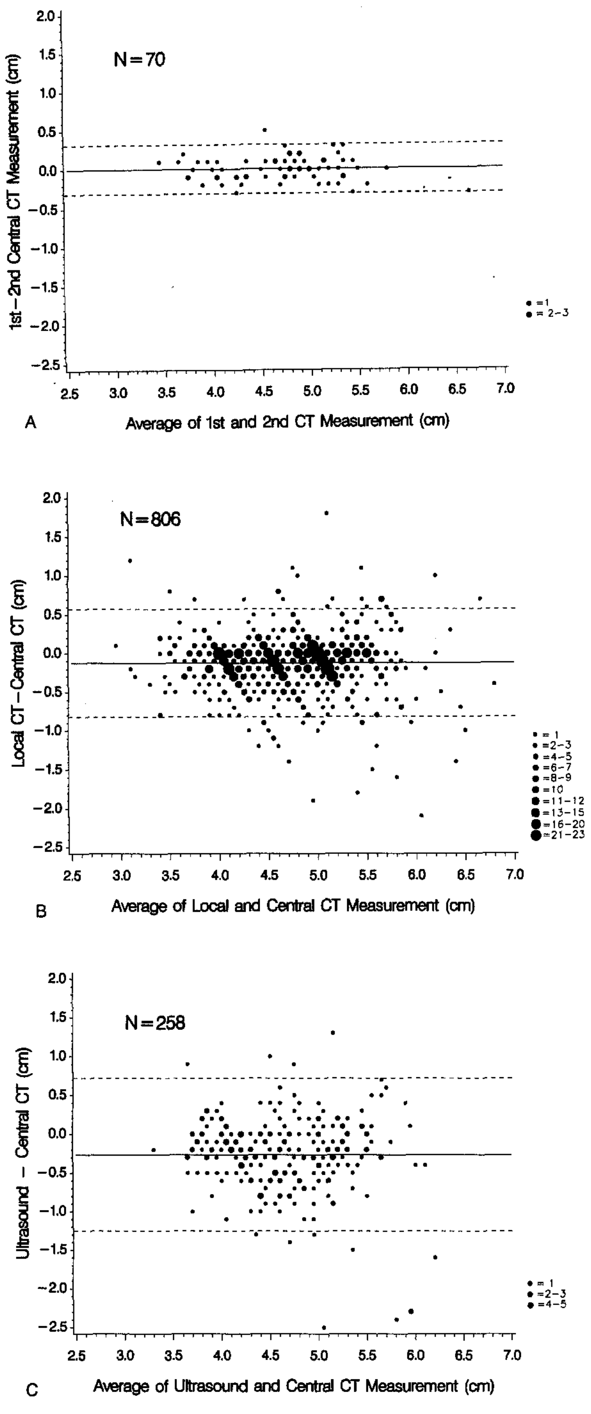

Fig. 3. Difference between measurements of maximum AAA diameter plotted against mean, shown with mean difference and limits of agreement.

is well described for other measurements.[13] Furthermore, the observation that the local measurements tended to be smaller than the central measurements could reflect a more rigorous search for the maximum diameter by the central reader.

The variability between CT and ultrasonography was greater still, again as expected. This comparison involves differences in reader, time of testing, and technology, so it is perhaps surprising that agreement was not worse. In our study, AAA diameter was generally smaller by ultrasonography than by CT, although in more than a fourth of cases the ultrasound measurement was larger. Ultrasound measurements showed better agreement with local CT measurements than with central CT measurements. Possible explanations for this finding include that (1) ultrasound and local CT measurements may have influenced each other because of their not being blinded, (2) the preference for recording by half centimeter, seen with both local CT and ultrasonography, could have enhanced agreement, and (3) a more rigorous search for the maximum AAA diameter by the central reader, as discussed above, could result in a difference relative to both local CT and ultrasonography.

The literature on variability of AAA measurement is limited. Our intraobserver results are similar to those reported for ultrasonography by Akkersdijk et al.[4] The four radiologists participating in that study observed each other for several months to standardize their technique, which may have allowed them to approximate the intraobserver level of agreement. Variability was determined separately for each dimension (e.g., anteroposterior), presumably reducing variability compared with our method of recording maximum diameter in any direction. Others have found intraobserver and interobserver variability of ultrasound measurement to be considerably worse,[2,3] more comparable to the interobserver variability we found with CT readings.

Previous studies have reported narrower limits of agreement between CT and ultrasound measurements of AAA diameter than we observed.[3,7] These studies consisted of small numbers of patients in a research setting, whereas our local CT and ultrasound measurements approximate conditions in clinical practice. The magnification of AAA diameter by ultrasonography compared with CT, reported by Ellis et al.,[3] was not seen in the majority of our patients.

Several points should be considered in applying our results to clinical practice. The intraobserver variability reported herein was attained in a research setting and may be difficult to duplicate in routine practice. Also, the intraobserver and interobserver variabilities reported in our study were obtained with the same CT scan for both measurements. Comparison of different CT scans, generating different images from different slices, would be expected to increase variability. In clinical practice, comparison is usually with another test obtained some months earlier. Our comparison of CT with ultrasonography done within 30 days more closely reflects clinical practice in this regard.

Our study is limited to measuring precision or variability of AAA measurement. An assessment of accuracy is problematic because no gold standard exists. Gomes et al.[5] used a caliper measurement at surgery to assess the accuracy of CT and found the two measures to be within 2 to 3 mm. However, Graeve et al.[6] found caliper measurements at surgery to be unreliable. The needle technique they preferred did not correlate well with CT. In both the literature and in practice, CT is often used as a gold standard.[3,7]

Our findings lend support to this strategy by demonstrating that a high degree of precision is possible with CT measurements of AAA diameter, but they also raise concern that this precision may not be obtained in routine practice or, therefore, in retrospective published reports. Variations in AAA measurement of 0.5 cm or more are not uncommon, and this should be taken into account in management decisions. Efforts to reduce variation in measurement are warranted and might include (1) seeking agreement between surgeons and radiologists on a precise definition of AAA diameter, (2) limiting the number of radiologists who measure AAA, and (3) the use of calipers and a magnifying glass for CT measurements.

REFERENCES

1. Nevitt MP, Ballard DJ, Hallett JW. Prognosis of abdominal aortic aneurysms: a population-based study. N Engl J Med 1989;321:1009-14.
2. Yucel EK, Fillmore DJ, Knox TA, Waltman AC. Sonographic measurement of abdominal aortic diameter: interobserver variability. J Ultrasound Med 1991;10:681-3.
3. Ellis M, Powell JT, Greenhalgh RM. Limitations of ultrasonography in surveillance of small abdominal aortic aneurysms. Br J Surg 1991;78:614-6.
4. Akkersdijk GJM, Puylaert JBCM, Coerkamp EG, De Vries AC. Accuracy of ultrasonographic measurement of infrarenal abdominal aortic aneurysm. Br J Surg 1994;81:376.
5. Gomes MN, Schellinger D, Hufnagel CA. Abdominal aortic aneurysms: diagnostic review and new technique. Ann Thorac Surg 1979;27:479-88.
6. Graeve AH, Carpenter CM, Wicks JD, Edwards WS. Discordance in the sizing of abdominal aortic aneurysm and its significance. Am J Surg 1982;144:627-34.
7. Grimshaw GM, Docker MF. Accurate screening for abdominal aortic aneurysm. Clin Phys Physiol Meas 1992;13:135-8.
8. Lederle FA, Wilson SE, Johnson GR, et al. Design of the Abdominal Aortic Aneurysm Detection and Management (ADAM) Study. J VASC SURG 1994;20:296-303.
9. Ellis M, Powell JT, Place J, et al. The limitations of ultrasound in surveillance of small abdominal aortic aneurysms. In: Greenhalgh RM, Mannick JA, Powell JT, eds. The cause and management of aneurysms. London: WB Saunders, 1990: 117-21.
10. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1:307-10.
11. Hirose Y, Hamada S, Takamiya M, Imakita S, Naito H, Nishimura T. Aortic aneurysms: growth rates measured with CT. Radiology 1992;185:249-52.
12. Zollner N, Zoller WG, Spengel F, Weigold B, Schewe CK. The spontaneous course of small abdominal aortic aneurysms: aneurysmal growth rates and life expectancy. Klin Wochenschr 1991;69:633-9.
13. Wen SW, Kramer MS, Hoey J, Hanley JA, Usher RH. Terminal digit preference, random error, and bias in routine clinical measurement of blood pressure. J Clin Epidemiol 1993;46:1187-93.

Submitted Oct. 13, 1994; accepted Dec. 27, 1994.

APPENDIX

Study Co-Chairmen: Frank A. Lederle, MD, VA Medical Center, Minneapolis, Minn., Samuel E. Wilson, MD, University of California Irvine Medical Center, Orange, Calif.

Project Coordinator, Co-Chairman's Office: Jean Larson, VA Medical Center, Minneapolis, Minn.

Central CT Scan Laboratory: Donovan B. Reinke, MD, Howard J. Ansel, MD, Diagnostic Radiology Service, VA Medical Center, Minneapolis, Minn.

Investigators at Department of Veterans Affairs (VA) Medical Centers in the following cities: *Ann Arbor, Mich.:* Louis M. Messina, MD, Mary Lingg, RN, Latonya Trohallis; *Cleveland, Ohio:* Linda Graham, MD, Evelynn Wingard, RN, MS, John Lindesmith, RN,* Julie Francosky, RT; *Miami, Fla.:* Anselmo A. Nunez, MD, Christine Estep, RN, James Taylor, RT; *Denver, Colo.:* William C. Krupski, MD, Pamela Strecker, RN, Linda Schoening, CCVT; *Hines, Ill.:* Fred N. Littooy, MD, John Maggio, PhD, Wendy Cote, RVT; *Little Rock, Ark.:* Gary W. Barone, MD, Bernard W. Thompson, MD,* Brenda Kackley, RN, Rob Little, RT, Terry Wood, BSRT*; *Long Beach, Calif.:* Ian Gordon, MD, Cheryl Kohl, RN, Rebecca Complin, RVT, Manju Akkinepali*; *Madison, Wis.:* Charles W. Acher, MD, Judy Archibald, RN, Traci Jo McGuire, RTR, Jennifer Fitzsimons, BS*; *Minneapolis, Minn.:* Edmund P. Chute, MD, Catherine Proebstle, RN, Michael Richardson, RT; *Pittsburgh, Pa.:* Michel S. Makaroun, MD, Tricia DeBoo, RN, Jean Fuhs, RT; *Portland, Ore.:* Gregory L. Moneta, MD, Roland Jemerson, RN, Robert McCartney, RVT; *Richmond, Va.:* Raymond G. Makhoul, MD, Laura

Bartnicki, RN, Linda Tyree Joyner, RT; *San Diego, Calif.:* Robert J. Hye, MD, Gerry Cali, RN, Terri Scala, RN,* Roger Hull, RCVT; *Tampa, Fla.:* Dennis Bandyk, MD, Dolores Bou-Eid, RN, Maureen Jackson, RT; *Milwaukee, Wis.:* Julie Freischlag, MD, Christa Kallio, RN, Susan Framberg, RN, Sandra Towne, RN,* Cleveland Weller, RT.

Cooperative Studies Program Coordinating Center (CSPCC), VA Medical Center, West Haven, Conn.: *Study Biostatistician:* Gary R. Johnson, MS; *Statistical Assistants:* Rae Bartozzi, Ray Kilstrom, MBA,* Kathy Riester, BA*; *Computer Programmers:* Kathy Newvine, BS, Robert Goodwin, MS*; *Chief, CSPCC:* Dorothea Collins, MS; *Administrative Officer:* Peggy Antonelli; *Forms Design:* Mary Smith; *Travel Clerk:* Darrell Burns; *Data Entry:* Lillie Franklin, Pattie Collins, Stella Marcinauskis, Bonita Hunter.

Planning Committee: Frank A. Lederle, MD (chairman), Minneapolis, Minn.; Samuel E. Wilson, MD, Long Beach, Calif.; Gary R. Johnson, MS, West Haven, Conn.; Louis Messina, MD, Ann Arbor, Mich.; Charles W. Acher, MD, Madison, Wis.; Fred N. Littooy, MD, Hines, Ill.; Jack L. Cronenwett, MD, Hanover, N.H.; David J. Ballard, MD, PhD, Decatur, Ga.; David B. Matchar, MD, Durham, N.C.; Christopher R. B. Meritt, MD, New Orleans, La.

Executive Committee: Frank A. Lederle, MD (chairman), Minneapolis, Minn.; Samuel E. Wilson, MD, Long Beach, Calif.; Gary R. Johnson, MS, West Haven, Conn.; Louis M. Messina, MD, Ann Arbor, Mich.; Charles W. Acher, MD, Madison, Wis.; Fred N. Littooy, MD, Hines, Ill.; David J. Ballard, MD, PhD, Decatur, Ga.; C. William Cole, MD (consultant), Ottawa, Ontario, Canada.

Surgery Committee: Samuel E. Wilson, MD (chairman), Long Beach, Calif.; Charles W. Acher, MD, Madison, Wis.; Fred N. Littooy, MD, Hines, Ill.

Outcomes Committee: Martha D. McDaniel, MD, White River Junction, Vt.; Geoffrey C. Lamb, MD, Milwaukee, Wis.; Roy M. Fujitani, MD, Lackland AFB, Texas.

Data Monitoring Board: G. Patrick Clagett, MD (co-chairman), University of Texas Southwestern Medical Center, Dallas, Texas; John D. Corson, MB, ChB (co-chairman), University of Iowa, Iowa City, Iowa; William C. Cushman, MD, VA Medical Center, Memphis, Tenn.; C. Seth Landefeld, MD, University Hospital of Cleveland, Cleveland, Ohio; Theodore G. Karrison, PhD, University of Chicago, Chicago, Ill.; John P. Matts, PhD, University of Minnesota, Minneapolis, Minn.; Dorothea Collins, MS (ex-officio), CSPCC, VA Medical Center, West Haven, Conn.

West Haven CSPCC Human Rights Committee: Karina Courtmanche, APRN (chairperson), Diane Marshall, Esquire, Jack H. Evans, Esquire, Willis Pritchett, RPh, James Solomon, MD, William Field, MD, Robert Marks, MD, James C. Niederman, MD, Barbara A. Kathe, PhD.

Cooperative Studies Program Administration (VA Central Office): Daniel Deykin, MD (chief), Janet Gold (administrative officer), Ping C. Huang, PhD (staff assistant).

*Past members of the study group.