Acta Radiologica

Adrenal lesions: variability in attenuation over time, between scanners, and between observers
Lilian Hammarstedt, Anne Thilander-Klang, Andreas Muth, Bo Wängberg, Anders Odén and Mikael Hellström

The online version of this article can be found at:

Published by:
$SAGE

http://www.sagepublications.com

On behalf of:

Nordic Society of Medical Radiology

Additional services and information for *Acta Radiologica* can be found at:

Email Alerts: http://acr.sagepub.com/cgi/alerts

Subscriptions: http://acr.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.com/journalsPermissions.nav

>> Version of Record - Aug 24, 2013

OnlineFirst Version of Record - Apr 30, 2013

What is This?

Adrenal lesions: variability in attenuation over time, between scanners, and between observers

Lilian Hammarstedt[1,2], Anne Thilander-Klang[3,4], Andreas Muth[5,6], Bo Wängberg[5,6], Anders Odén[7] and Mikael Hellström[1,2]

[1]Department of Radiology, Sahlgrenska University Hospital, Gothenburg; [2]Department of Radiology, Institute of Clinical Sciences, the Sahlgrenska Academy at the University of Gothenburg, Gothenburg; [3]Department of Medical Physics and Biomedical Engineering, Sahlgrenska University Hospital, Gothenburg; [4]Department of Radiation Physics, Institute of Clinical Sciences, the Sahlgrenska Academy at the University of Gothenburg, Gothenburg; [5]Department of Surgery, Sahlgrenska University Hospital, Gothenburg; [6]Department of Surgery, Institute of Clinical Sciences, the Sahlgrenska Academy at the University of Gothenburg, Gothenburg; [7]Department of Mathematical Sciences, Chalmers University of Technology, Gothenburg, Sweden
Correspondence to: Lilian Hammarstedt. Email: hammarstedt.lilian@gu.se

Abstract

Background: Measurements of attenuation (in Hounsfield units [HU]) and contrast wash-out are widely used to characterize adrenal lesions as benign or indeterminate/malignant at computed tomography (CT). Clinical experience suggests that such measurements of adrenal lesions may vary over time and between observers, making evaluation difficult.

Purpose: To investigate the change over time of adrenal lesion size, attenuation, and contrast wash-out at CT, to determine inter-observer variability, and to analyze other factors underlying the variability.

Material and Methods: In a cohort of patients, with or without malignant disease, undergoing CT, adrenal lesions were prospectively analyzed. Lesions with growth >20% or >5 mm over 6 months were excluded. Non-enhanced attenuation and contrast medium wash-out over 2-year follow-up were analyzed. An inter-observer analysis with five observers and a phantom study of eight different CT scanners were performed to assess measurement variability.

Results: Mean adrenal lesion non-enhanced attenuation values decreased by 0.5 HU/year during follow-up. Using 10 HU or 40% relative wash-out as threshold values for benign *versus* indeterminate lesions, 27 (20%) and 39 (29%) of 136 lesions, respectively, would be reclassified at some occasion during follow-up. In the observer analysis 37 of 40 lesions demonstrated agreement between all observers, using established threshold values. The phantom study showed an intra-scanner variability of 1–3 HU, but an inter-scanner variability of up to 8 HU for water.

Conclusion: The clinically widespread use of specific attenuation threshold values for characterizing adrenal lesions must be used with great caution, considering that multiple factors, related to patient, equipment, scanning technique, and observer influence the outcome.

Several studies have shown the ability of CT to distinguish between benign and malignant adrenal lesions by measuring the lesion attenuation in absolute Hounsfield unit (HU) values on native images or by the calculation of the time-dependent percentage wash-out of intravenously injected contrast material. Many benign adrenal adenomas are rich in intracytoplasmic lipid, which lowers the density and hence reduces the unenhanced CT number. Such benign adrenal lesions will typically have attenuation values ≤10 HU (1, 2), while some benign adenomas as well as most malignant adrenal lesions are lipid poor and therefore have higher attenuation values. If a lesion is lipid poor

the absolute or relative percentage wash-out of contrast material can be calculated using the CT numbers of the lesion on native and/or early (venous phase, 60–90 s) and late (10–15 min) contrast-enhanced series. Benign adenomas are characterized by a higher percentage wash-out in the late phase as compared to non-benign lesions, and cut-off values for differentiation between benign and non-benign lesions have been established (3–7). In daily practice, such cut-off values, e.g. 10 HU on non-enhanced images and 40–60% wash-out, are now widely used for clinical decision-making. However, clinical radiologists often experience difficulties in applying such strict cut-offs when measuring and interpreting the attenuation, as it may vary between CT slices of the same lesion, between scanners and sometimes over time in the same patient. Thus, several technical parameters have been shown to have an impact on the estimated attenuation of the tissue and the accuracy of CT numbers has been questioned (8–11). For example, if scanning protocols with similar parameters are used on different scanners on the same objects some diversity will still occur in the attenuation values (12).

As differentiation between a benign adenoma and a malignant lesion may have a profound impact on patient handling, it is of fundamental importance to understand, and minimize, the factors that contribute to the variability in characterizing adrenal lesions.

The aim of this clinical and experimental study was to investigate change over time of size, attenuation, and contrast wash-out characteristics of adrenal lesions, and to determine the inter-observer and inter-scanner variability.

Material and Methods

Variability of adrenal characterization over time

Patients

A prospective study on incidentally detected adrenal lesions was performed in the region of Västra Götaland including the northern part of the county of Halland (total population of 1.7 million inhabitants), Sweden, from October 2002 to April 2004. All adult patients ($\geq$18 years) with an adrenal lesion detected at CT, magnetic resonance imaging (MRI) or ultrasonography (US) were reported from the radiology departments and a 2-year clinical and radiological follow-up program was performed. A radiological re-evaluation of 3827 randomly selected CT cases from the original cohort was performed in order to confirm the detection rate, and patients with adrenal lesions not earlier reported to the study were thereby included in the follow-up. Detection and follow-up data have been presented earlier (13, 14).

In the present study, patients without as well as with extra-adrenal malignancy were included. In order to minimize the risk of including fast growing, malignant lesions that exhibit obvious interval change in size and attenuation, lesions had to exhibit size stability for at least 6 months at CT, using criteria adopted from RECIST, version 1.1 (15). Growth was defined as an increase in lesion diameter >20% (at least >5 mm) and decrease in size as a decrease in diameter >30%.

All available patients with two or more dedicated adrenal CT examinations were analyzed ($n = 113$). Lesions radiologically characterized as a cyst or a hematoma were excluded from this analysis. Adrenal lesion attenuation on unenhanced series and relative percentage wash-out calculations were compared over time.

Among these 113 patients, 33 were randomly selected for a comparison of attenuation in the adrenal lesion, aorta and portal vein to investigate the influence of vascular enhancement on the attenuation of the adrenal lesions. Measurements in aorta were made at the level of the superior mesenteric artery exit and the portal vein was measured close to the superior mesenteric vein entry.

In addition, 57 patients with follow-up abdominal CT other than dedicated adrenal CT were included in an analysis regarding variations in lesion size and attenuation values on unenhanced CT over time. Thus, this analysis comprised a total of 170 patients.

The patient study was approved by the local Regional Ethical Review Board (s202–02).

CT scanners and protocol

During the inclusion and follow-up period, a variety of CT scanners were used in the centers of the study region. Initially, there were two non-spiral, 10 single-detector CT (SDCT), seven 4-slice, one 8-slice, and one 16-slice multidetector CT (MDCT) in the region. Throughout the study period until latest follow-up examinations in 2007, CT scanners were gradually replaced by 8-, 16-, and 64-slice MDCTs. All scanners were maintained and calibrated according to the manufactures' specifications by the manufacturers' service personnel.

The examinations at which the adrenal lesions were incidentally detected were performed according to routine protocols designed for the specific clinical situation. Follow-up examinations, on the other hand, were performed according to a predesigned dedicated adrenal CT protocol; an unenhanced scan followed by a contrast-enhanced scan 80 s after the start of injection and a delayed scan after 10 min. Single-detector CT scan parameters were: tube voltage, 120 kV; tube current, 250 mA; rotation time, 0.7–1.0 s; slice thickness, $\leq$3 mm; interval, 2.0 mm; pitch, 1:5; and reconstruction algorithm for abdominal imaging. A corresponding protocol was used with MDCT but usually with thinner scan slices reconstructed to 2–3 mm. Intravenous contrast medium used was of non-ionic low- or iso-osmolar type with a concentration of 300 mg I/mL and an injection rate of 3 mL/s.

All examinations from the different hospitals were sent to a common regional digital image archive which made it possible to view all images on the same digital radiological workstation (Centricity RA600; GE Healthcare, Chalfont St. Giles, UK) at the study center, using a high-resolution screen (Coronis 3 MP; Barco N.V., Kortrijk, Belgium) calibrated according to DICOM PS 3.14-2009 (16). All measurements were done using the same equipment.

Lesion measurements

For every adrenal lesion, comparison between CT examinations was made at the same occasion. Lesion size was
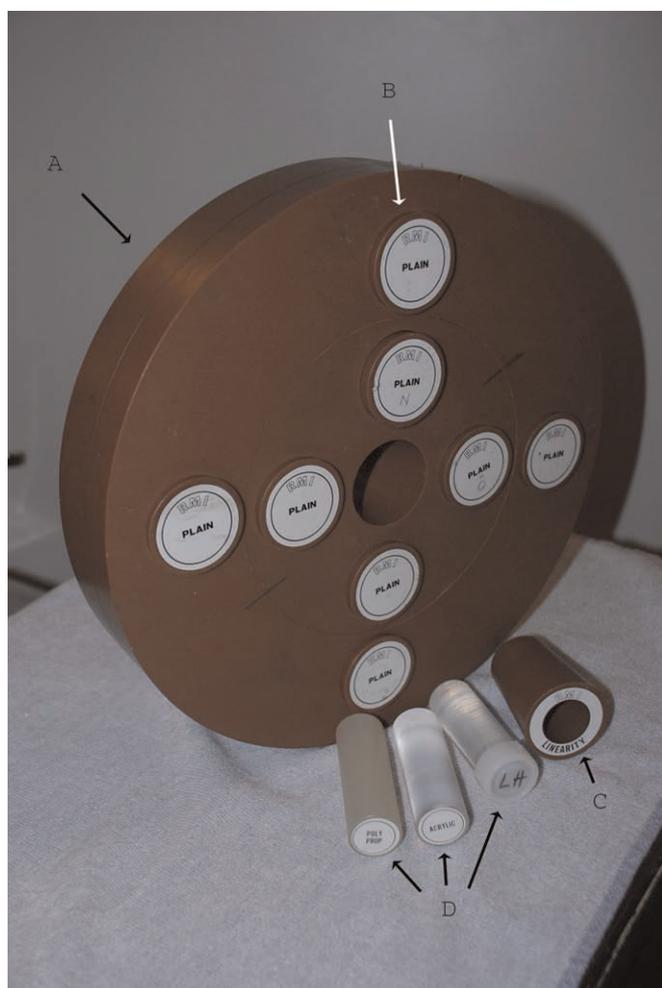
Fig. 1 (a) Body CT phantom used in the study (RMI, model 461A, snr 461A1077, Radiation Measurement Inc., Middleton, USA). (b) Conical tapered inserts of uniform water equivalent composition (Solid water®). (c) Centrally placed insert with possibility to contain test tubes. (d) Different test tubes containing tap water, acrylic, and polypropylene

defined as its largest diameter on transaxial images obtained in venous phase. Attenuation of the lesion was measured with an oval or round region of interest (ROI), as large as possible but avoiding lesion borders (at least covering half of the lesion area). Necrotic, cystic, hemorrhagic, and calcified areas were avoided if possible. A mean attenuation value was calculated from at least two measurements in each scanning phase (unenhanced, enhanced, and delayed phases). Relative wash-out was calculated with the aid of a computer software (OmniVis 4.0, GE Healthcare, Stockholm, Sweden) using the following equation for relative percentage wash-out (RPW):

$$RPW = \frac{Attenuation\ value_{enhanced} - Attenuation\ value_{delayed\ enhanced}}{Attenuation\ value_{enhanced}} \times 100$$

A threshold of ≤ 10 HU on unenhanced images and/or a RPW $>40\%$ was used to categorize lesions as benign; the other lesions were categorized as indeterminate.

Inter-observer analyses

Inter-observer variation in assessing size, attenuation, and contrast wash-out of adrenal lesions was determined by measurements performed by five observers independently. All were board-certified radiologists, specialized in abdominal imaging (experience ranging from 1–30 years). The observers were blinded to clinical history, radiological, and other patient information, and the patient identity was replaced by code numbers.

Patients with documented adrenal lesions, examined with a dedicated adrenal CT protocol at the study center during a 7-month period (from 1 May 2010 to 31 Dec 2010), were included in the analysis ($n = 40$). These patients were examined with 8- or 16-MDCT scanners routinely used at the study center (LightSpeed Ultra and LightSpeed Pro 16; GE Healthcare, Milwaukee, WI, USA).

The dedicated adrenal CT protocol included an unenhanced scan, a contrast-enhanced scan in the venous phase (70–80 s after the start of injection) and a late scan after 15 min. Helical 0.7 s per rotation scans were obtained with tube voltage 120 kV, tube current 440 mA, beam collimation 10 mm, slice thickness 2.5 mm with an interval of 2.0 mm, pitch 0.94 or 1.35, 50-cm scan field of view (SFOV) and display field of view (DFOV) of 36 cm and standard reconstruction algorithm.

The instructions to the observers were the same as those used in the original adrenal study, i.e. adrenal lesion size and attenuation were to be measured in the same way as described earlier. All measurements by all observers were performed on the same digital radiological workstation (Centricity RA600; GE Healthcare, Chalfont St. Giles, UK) using a high resolution screen (Coronis 3 MP; Barco N.V., Kortrijk, Belgium) calibrated according to DICOM part 14 (16) before the analysis. In case of bilateral adrenal lesions, the protocol specified which side to measure (right or left). If there were multiple adrenal lesions on one side, the observer should select the largest lesion for the analysis. The observers were asked to use the abdominal window settings by which they could appreciate the lesion best (close to window width 40 HU; level 400 HU). Based on the observer measurements, relative wash-out values for each lesion and each observer were later calculated by one radiologist (LH) according to the equations given above. As the dedicated adrenal CT protocol of the study center had changed after the adrenal study follow-up was completed a delayed phase of 15 min was used instead of 10 min. Thus a threshold of ≤ 10 HU on unenhanced images and/or a RPW $>50\%$ was used to categorize lesions as benign; the other lesions were categorized as indeterminate (5).

Phantom analysis

In order to analyze variation of CT attenuation in the same object between different scanner types, a body CT phantom (RMI, model 461A, snr 461A1077, Radiation Measurement Inc., Middleton, WI, USA) was used. The phantom is a uniform disc of Solid Water® phantom material (water equivalent), 6 cm thick and 33 cm in outer diameter (Fig. 1). It has nine conical tapered insert cavities of identical

Table 1 Convolution kernel, image slice thickness, and pitch factor used for the CT scanners in the phantom analysis

Scanner type	GE single	GE 4	GE 8	GE 16	GE 64	Toshiba 16	Siemens 16	Siemens 64
Kernel	Std	Std	Std	Std	Std	FC 12	B30s	B31s
Slice thickness	3.0	2.5	2.5	2.5	2.5	2.0	2.0	1.2
Pitch	1.5	1.5	1.5	1.35	1.375	1.438	1.5	1.5

sizes (4.5 cm in diameter centrally). In the central conical insert a test tube measuring 2.7 cm in diameter is placed, making it possible to scan different materials. Solid water inserts are used in the rest of the cavities. The center of the phantom was placed in the isocentre of the gantry. All images were analyzed in the same digital radiological workstation (Centricity RA600, GE Healthcare) as described above. Attenuation measurements of the central inserts were made in the three most central images. Measurements were made with a ROI size of approximately 200–250 mm^2, and a mean value was calculated. The ROI standard deviation displayed on the workstation was also noted.

The first test included two repeated scans of each different object in the central test tube (containing solid water insert, room temperature tap water, polypropylene, and acrylic inserts, respectively) in each scanner. The other test was to investigate possible differences in water attenuation using different temperature (cold, room temperature, and warm, respectively). In this test the GE 8-MDCT was used. The phantom was examined on one occasion at each CT machine. To investigate intra-scanner consistency over time the phantom was also examined three times during one day (at 07:00, 11:00, and 17:00) and two times >4 weeks apart with the GE 8-MDCT.

CT scanners and protocol in the phantom analysis

The CT scanners compared were one SDCT, and one each of 4-, 8-, 16-, and 64-MDCT from GE (HiSpeed FX/i, LightSpeed QX/i, LightSpeed Ultra, LightSpeed Pro 16, LightSpeed VCT, GE Healthcare, Milwaukee, WI, USA) and one 16-MDCT from Toshiba (Aquilion 16, Toshiba Medical Systems Corporation, Tokyo, Japan) and one each of 16- and 64-MDCT from Siemens (Somatom Sensation 16, Somatom Definition, Siemens Healthcare, Erlangen, Germany). The CT scanners are hereafter referred to as GE HiSpeed, GE 4, GE 8, GE 16, GE 64, Siemens 16, Siemens 64, and Toshiba 16.

Image acquisition and reconstruction parameters for the individual CT scanners were set to be as equal as possible. Helical scans with 1.0 s rotation time, tube voltage of 120 kV, tube current 440 mA, SFOV of 50 cm, and DFOV of 36 cm were used in all scanners. The rest of the reconstruction parameters are given in Table 1.

Statistical analyses

The analysis of the intrapatient variation with time of measurements of lesion size and attenuation in patients with adrenal lesions was performed in two steps. The first one comprised the calculation of the standard deviation (SD) of the difference between the baseline value and the value after approximately 1 year and the corresponding standard deviation after 2 years. Then an assumption for a special type of stochastic process, Wiener process, was applied to calculate the standard deviation for the difference for small steps. The variance when the difference in time is t units is $\sigma^2 \cdot t$, where σ^2 is the variance corresponding to 1 unit of time.

In the inter-observer analysis, for each patient five observers determined the value of the variables size, unenhanced HU, postcontrast venous phase HU, late postcontrast phase HU, and wash-out (%) was calculated. The mean and standard deviation (SD) was calculated for each patient. The SD reflected the variation between the observers. The square root of the sum of variances divided by the number of patients was calculated for each one of the variables. We also determined the mean over all patients as well as the correlation coefficient between mean and SD. The correlation between mean and SD was tested by Pitman's test (17), which is a non-parametric test of correlation. The regression function with SD as the dependent variable and mean as the independent one was determined in order to assess whether the coefficient of variation was an appropriate quantity to use.

The P values were determined by Pitman's test.

Results

Variability of adrenal characterization over time

From the total of 443 patient included for follow-up in the original adrenal study, 263 were excluded because of too short or no follow-up with CT ($n = 225$), adrenal hemorrhage ($n = 5$), or change in size according to RECIST criteria ($n = 33$). Thirty of these 33 patients had lesions that increased and three had lesions that decreased in size. Of the increasing lesions eight had otherwise benign characteristics, while 20 had characteristics suggesting non-benign etiology (heterogeneous, high attenuation at unenhanced series, and/or poor delineation).

Variability of adrenal lesion contrast wash-out over time

In 113 patients with 136 adrenal lesions two or more dedicated adrenal CT examinations with a follow-up time ≥6 months were performed. All these lesions were considered to be radiologically benign, taking all imaging data into account, but individual attenuation values varied over time. Data showing unenhanced HU and RPW over time are presented in Table 2. If only using ≤10 HU (unenhanced) as threshold value for benign *versus* indeterminate lesions, 27 lesions (25 patients) would have been reclassified at some occasion during follow-up. Twelve of these lesions had HU values suggesting benign etiology (≤10 HU) at first examination, but were indeterminate (>10 HU) at a later examination, while the reverse was

Table 2 Number of lesions with unenhanced CT attenuation value ≤10 HU and relative percentage wash-out (RPW) >40% (criteria for benign lesions) over time for the total of 136 lesions examined at least twice using dedicated adrenal CT protocol, and change in classification using these criteria for classifying lesions as benign *versus* indeterminate

Variable	Exam 1	Exam 2	Exam 3	At any of the three exams	Change in classification benign/indeterminate between examinations
HU ≤ 10	85 (62%)	88 (65%)	53 (74%)	100 (73%)	27 (20%)
RPW > 40%	105 (77%)	112 (82%)	60 (83%)	128 (94%)	39 (29%)
Total lesions	**136**	**136**	**72**	**136**	**136**

Numbers in parentheses denote percentages in relation to the total number of lesions

Table 3 The standard deviation (SD) of the difference in size (mm) and attenuation (HU) in patients with adrenal lesions and with a CT follow-up time of ≥12 months

Variables	Mean period (years)	Size (mm)		Attenuation (HU)	
		n	SD of the difference	*n*	SD of the difference
Baseline *vs.* 1 year	1.1	207	3.393	189	17.603
Baseline *vs.* 2 years	2.3	206	5.041	189	17.467

observed for 15 lesions. The mean difference in unenhanced attenuation value between the first and a later examination for these 27 lesions was 8.2 HU (range, 1–17 HU), while in the remaining lesions mean attenuation difference was 5.0 HU (range, 0–19 HU). Using RPW >40% as threshold value for benign *versus* indeterminate, 39 lesions in 35 patients would have been reclassified during follow-up (Table 2). The mean difference in RPW values between two examinations in these 39 lesions was 24% (range, 5–52%). Using either or both of these criteria (HU < 10; RPW >40%), 50 of a total of 113 patients (44%) would have been reclassified. In 13 of these 50 patients (26%) different CT scanners were used at follow-up.

In the 33 patients that were analyzed with comparison of the attenuation values in the aorta, portal vein, and adrenal lesions, there was a statistically significant correlation between contrast-enhanced attenuation in the adrenal lesion and the portal vein ($P < 0.001$).

Variability of adrenal lesion size and unenhanced attenuation over time

In total, 170 patients (100 women [59%]) with 207 lesions (mean diameter, 23.2 mm; range, 10–50 mm) had initial CT and follow-up CT examination performed after ≥1 year. Of these, 68% (140 lesions), had an attenuation value ≤10 HU at some occasion during follow-up.

The standard deviation (SD) of the difference of the intra-patient variation with time regarding lesion size and attenuation values (HU) is presented in Table 3. The result from the calculation of the Wiener process was used to simulate the variation for two pairs of individuals. The standard deviations (or variances) alone are not sufficient to give a good picture of the variation. The variation is therefore demonstrated also by the simulated curves of two imaginary individuals in the study (Fig. 2). Overall lesion size
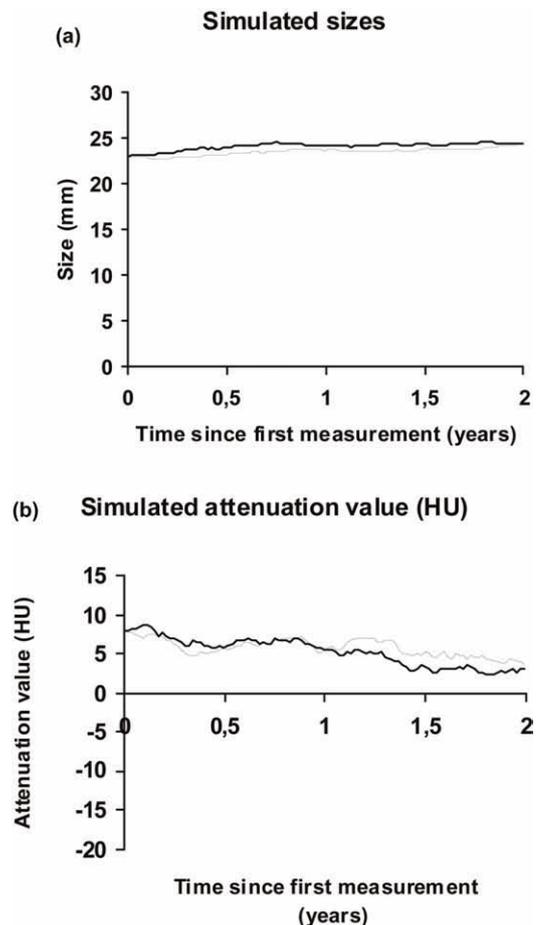

(a) Simulated sizes

(b) Simulated attenuation value (HU)

Fig. 2 Simulation of variation in size (a) and attenuation value (b) in two pairs of individuals with adrenal lesions using the "Wiener process assumption". The lesion size was assumed to be 23 mm and attenuation value assumed to be 8 HU at baseline for both individuals (using baseline mean values)

tended to increase slightly with time, 0.3 mm per year, and CT number decreased by 0.5 HU per year. In the simulations adjustments were performed for these changes.

Inter-observer analyses

Tables 4 and 5 show the variation between the five observers.

The coefficient of variation, which is the quotient between the standard deviation and the mean value, is a commonly used quantity to describe a measurement error. However, the ideal assumption for that quantity is that the regression

Table 4 Inter-observer analysis based on five observers with estimated standard deviation of measurement error $(\sigma SD^2/n)^{\frac{1}{2}}$, mean, range, correlation coefficient between mean and SD, and *P* value

Variable	$(\sigma SD^2/n)^{\frac{1}{2}}$	Mean	Range	R	P value
Size (mm)	2.38	23.2	8–50	0.04	0.7742
Attenuation, unenhanced (HU)	4.98	9.4	−17–56	0.46	0.0083*
Attenuation, enhanced (HU)	5.43	60.6	13–102	0.25	0.1148
Attenuation, late phase (HU)	4.38	26.4	−7–70	0.34	0.0471[†]
Relative wash-out (%)	8.22	55.5	−60–154	−0.35	0.0524

*Highly significant
[†]Significant

Table 5 Regression functions in the inter-observer analyses

Variable	SD
Size (mm)	$1.83 + 0.0061 \cdot mean$
Attenuation, unenhanced (HU)	$1.47 + 0.1367 \cdot mean$
Attenuation, enhanced (HU)	$1.88 + 0.0379 \cdot mean$
Attenuation, late phase (HU)	$1.05 + 0.0692 \cdot mean$
Relative wash-out (%)	$9.37 - 0.0655 \cdot mean$

function is $SD = 0 + c \cdot mean$. That assumption is not fulfilled for the studied variables (Table 5).

The estimated standard deviation of the measurement error $(\Sigma SD^2/n)^{\frac{1}{2}}$ can be used to elucidate the distribution of the difference between two observers when they both investigate the same, randomly chosen patient. Let X_1 and X_2 denote their measurement value and set σ (sigma) = $(\Sigma SD^2/n)^{\frac{1}{2}}$ then the probability that the difference is less than for example $0.5 \cdot \sigma$ is $P(|X_1 - X_2| \leq 0.5 \cdot \sigma) \approx \Phi(0.5 \cdot \sigma/\sqrt{2} \cdot \sigma) - \Phi(-0.5 \cdot \sigma/\sqrt{2} \cdot \sigma) = 0.28$.

If 0.5 times sigma is exchanged for 2.0, 1.0, 0.25, or 0.125 times sigma we will obtain the probabilities 0.84, 0.52, 0.14 and 0.07, respectively. If we consider the variable size, then $\sigma = 2.38$ (Table 4). Thus the probability that the two observers differ <2.38 mm is approximately 0.52.

The maximum difference in measurements of unenhanced attenuation value between the five observers ranged from 0 to 20 HU (mean, 4.6 HU). The larger range was seen in three lesions (15–20 HU) but in the rest (37 lesions) the range was 0–8 HU with a mean of 3.6 HU. In the late contrast-enhanced phase the difference was 0–23 HU (mean, 5.7 HU) but in 36 of the lesions the range was 0–9 HU. Of the 40 lesions analyzed, 31 lesions were categorized as benign by all the observers using the predefined threshold values; six were categorized as indeterminate by all observers while there was disagreement between the observers in three cases. In two of these three lesions 4/5 observers were in agreement and in the third case 3/5 were in agreement. These three lesions measured 10–15 mm, 29–34 mm, and 29–34 mm, respectively, while their unenhanced attenuation values measured 7–11 HU, 15–35 HU, and 56 HU (all observers), respectively, according to the five observers.

Phantom analysis

The variation in attenuation of solid and tap water inserts between scanners are given in Fig. 3 A and B. Fig. 4 shows the intra-scanner variation over time of measurement on the GE 8 scanner examined on four separate occasions.

The difference in HU between tap water of different temperature, and between tap water and saline, are also demonstrated in Fig. 4. The difference in mean attenuation value for tap water within each scanner was 2.7 HU (range, 1.8–5.1 HU) and between scanners a maximum difference of 0.3–8 HU was found. The stated manufacturer specifications for measurements on these scanners are to obtain a CT number of water within $\pm$ 3 HU or $\pm$ 4 HU. Fig. 5 demonstrates the inter-scanner variation of attenuation of inserts with polypropylene (a) and acrylic (b) test tubes. The inter-scanner difference between mean attenuation for polypropylene ranged from 0.4 to 9.9 HU and for acrylic insert from 0.4 to 17.7 HU.

Discussion

The rationale for using attenuation values for characterization of adrenal lesions has been supported by a large number of studies (3, 18, 19). However, the use of absolute CT attenuation values as a diagnostic tool has been questioned (8–12). First of all, it is obvious that the HU values of an adrenal lesion are associated with a considerable imprecision, reflected in the standard deviation of the measurement. The attenuation value is depending on several factors, e.g. CT scanner, scanning technique, the patient, and the observer.

Our follow-up data showed that there was a slight, but statistically insignificant increase over time of adrenal lesion size, and there was a slight reduction in mean HU values over time. This suggests a benign course, providing that stable size and increased lipid content (and thereby a lower HU value) are accepted as indicators of benign development. Nevertheless, repeated scanning of the same adrenal lesion may result in different risk stratifications. In our clinical material, 20% and 29% (totally 44%) of the patients would have their classification (benign *versus* indeterminate) altered at follow-up if applying 10 HU and 40% as thresholds for unenhanced and wash-out attenuation measurements, respectively. It should be remembered that patients with larger interval increase in size (>20% or >5 mm) of the adrenal lesions, suggesting malignancy, were not included in this analysis, as they are easy to identify and therefore do not constitute a clinical problem, especially if combined with other features of malignancy, e.g. inhomogeneous attenuation pattern, irregular shape, and poor delineation.

We found a slight decrease in CT attenuation of the adrenal lesions over time. It can be speculated if a change
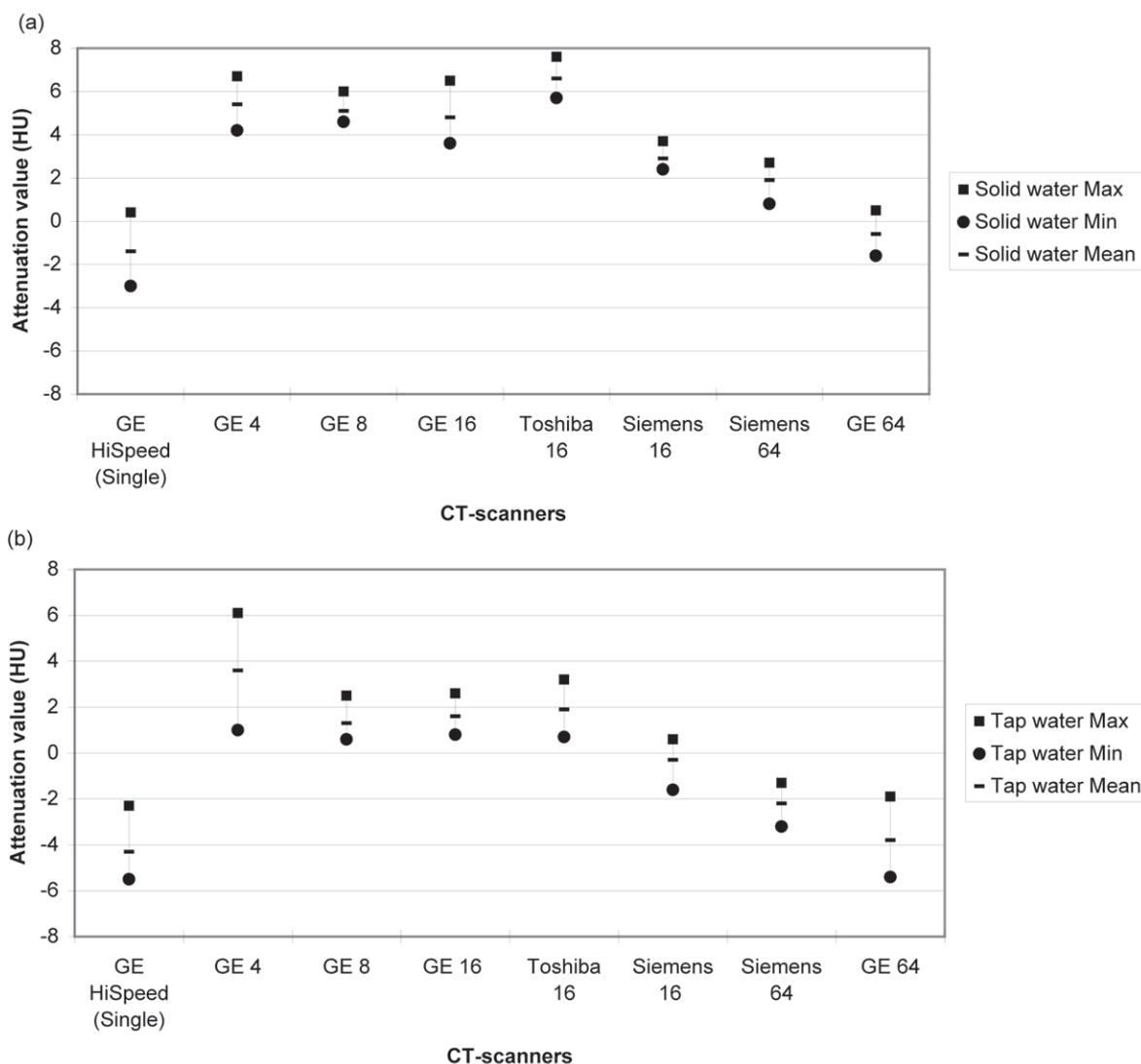
Fig. 3 Inter-scanner variability of attenuation measurements of the centrally placed room temperate solid water insert (a) and tap water insert (b) in the phantom

in individual lesion attenuation may be associated with a biological tissue alteration over time. This seems unlikely, as other characteristics, e.g. lesion size, changed very little during the limited follow-up period. The variability is more likely to be associated with technical or observer-related factors. It can also be speculated that variations in the circulation and distribution of contrast media (related to, for example, cardiac output and heart frequency) may have an impact on the attenuation of adrenal lesions. We could demonstrate that the contrast enhancement of the adrenal lesions correlated with the attenuation of the portal vein, while aortic attenuation did not. Further studies on the influence of contrast media concentration, rate of injection, and timing of post-injection scanning seems warranted to further elucidate its potential role for variations in attenuation.

One source of measurement variation is variation between observers. Our inter-observer study using clinical cases showed a moderate variation between abdominal radiologists with varying degrees of experience. Lesion size measurements appeared relatively stable between observers, while much larger variations between observers were noted regarding lesion attenuation measurements. Thus, there was a statistically significant difference between different observers regarding both unenhanced and late phase attenuation measurements. In 90% of the cases, however, the nominal differences were relatively small, ranging from 0 to 8 HU (mean, 3.5 HU) on unenhanced series, and 0–9 HU (mean, 4.3 HU) on late phase series, respectively. Nevertheless, there were cases with larger discrepancies, and even small variations around the cut-off value 10 HU, or any other cut-off, have an impact on clinical handling and follow-up. The relative percentage wash-out determinations, based on the attenuation measurements in the early and late contrast-enhanced phases, also showed notable inter-observer variability, but did not quite reach statistical significance. The calculated chances that two observers would differ by 2.4 mm in size estimates or by 5.0 HU in attenuation measurements were 50%. From a clinical viewpoint, these differences may still seem acceptable. Wider discrepancies could be expected in less homogenous groups of observers and in routine clinical
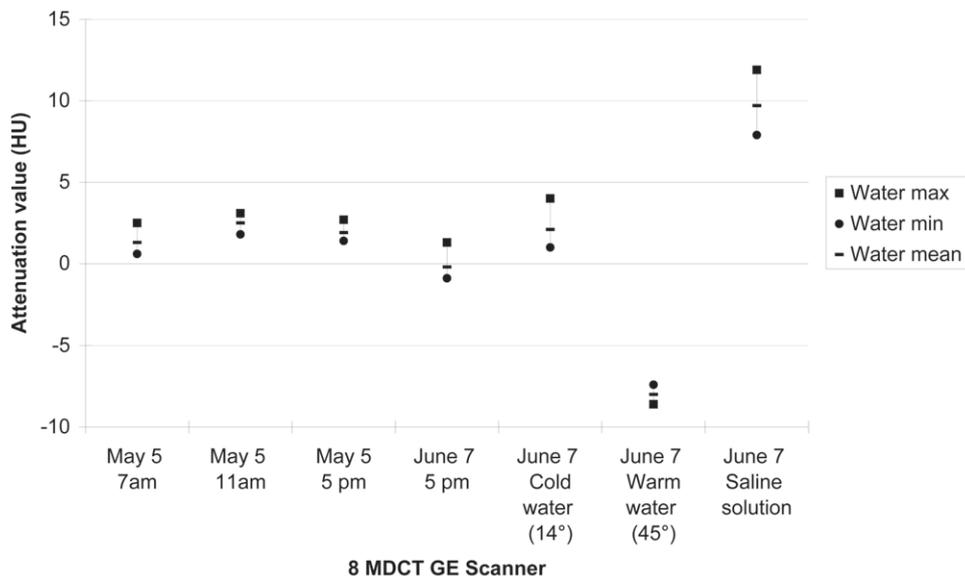
Fig. 4 Intra-scanner (8-MDCT GE scanner) variability over time of attenuation measurements at repeated CT examinations on different occasions (dates given), using room temperate tap water insert centrally in the phantom and cold tap water, warm tap water, and normal saline (0.9% solution) at room temperature in the inserts, respectively

work, where less favorable reading conditions prevail, including stress and time constraints. Siegel *et al.* (20) showed that agreement between observers in measuring attenuation is higher in simple than in complicated and small lesions. Hopper *et al.* (21) found similar results regarding size measurement on poorly defined and irregular lesions. As all observers in our inter-observer study analyzed the same cases using the same work station, it can be assumed that the inter-reader variability is due to individual selection of CT slice used for measurement, and individual selection of size and placement of ROI cursor.

Our phantom study shows that within a single CT scanner, repeated measurements of the same object at different times of the day and 1 month later appear relatively stable, resulting in a change in mean attenuation values by only 1.5 HU, while the range of individual measurements was 1–3 HU. The variation of attenuation values measured in the phantom analysis is thus in range of the CT manufacturers' specification. However, the temperature of the object had a great impact on the CT number as density of water changes with temperature. As intra- and inter-patient body temperature varies only slightly, this is of little significance when comparing HU measurements over time and between patients. However, it may be of importance when applying a calibration instrument/phantom, which ideally should be applied at a defined temperature, e.g. by using tap water at room temperature. As expected, the variation between scanners of different types and generations was greater than the intra-scanner variation, with mean attenuation for both tap and solid water ranging >8 HU between scanners. This has clear clinical implications when classifying and following adrenal lesions, and suggests that repeated examinations of the same adrenal lesion should preferably be performed with the same scanning protocol and at the same scanner providing that it has been properly calibrated.

As suggested by our results, a limitation of multicenter and follow-up studies, including our own (13, 14), is that CT examinations may not be uniform, when different scanners are used. A limitation of our study is that we used 10 min instead of 15 min delay to calculate the wash-out. This was due to the routine used at the time of the start of the study. The use of a 10-min delay has recently been questioned (22) and the proportion of lesions that were reclassified based on 10 min wash-out might possibly be larger than if we had used a 15-min delay. Still, the general attenuation variability shown in all enhancement phases would not be affected by this.

To summarize, in our clinical material we have shown that benign appearing adrenal lesions tend to increase only slightly in nominal size over a 2-year period, and nominal CT attenuation of the lesions tends to decrease slightly. However, when applying established, fixed cut-off values for unenhanced HU values (10 HU) and relative contrast wash-out (40%) for classification of adrenal lesions as benign or indeterminate, nearly half (44%) of the lesions would be reclassified if using the same criteria at a follow-up CT examination, even when using the same CT scanner. This does not necessarily indicate a systematic change in biological behavior of the adrenal lesions, as change in classification went both ways. Inter-observer variation in classifying lesions as benign or indeterminate was small, but may add to the variation at repeated examinations when using different observers. Some of the variation can be attributed to different hardware and software, as attenuation values also varied between CT scanners in the phantom analysis. Obviously, calibration of CT scanners is often suboptimal. Some authors suggest the use of a calibration instrument to adjust image attenuation values in CT scanning both in research and clinical practice, especially when the measurement value is in the range of 0–20 HU (11, 12). Our results support this but the calibration
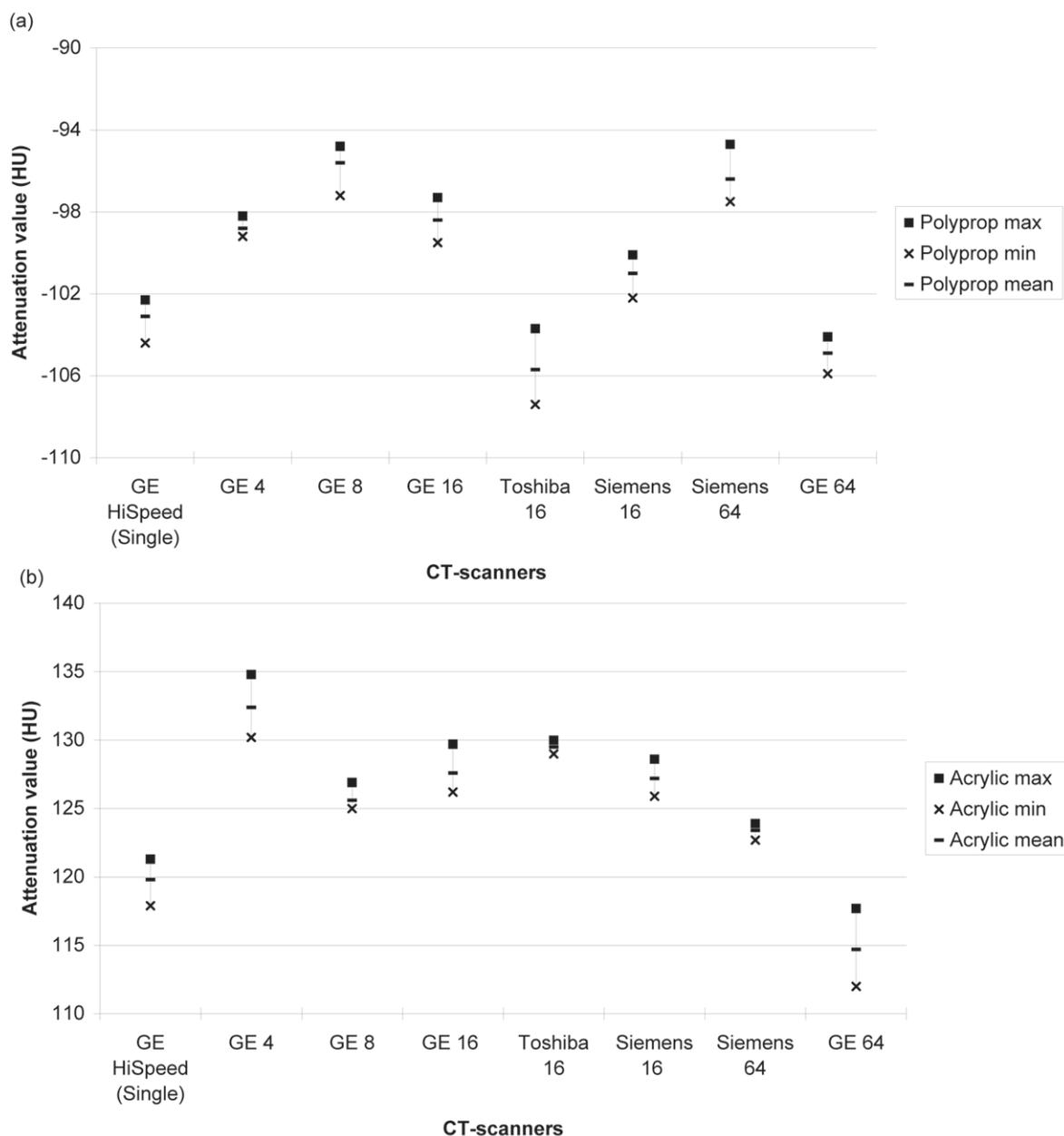
(a)

(b)

Fig. 5 Inter-scanner variability of attenuation measurements of centrally placed polypropylene inserts (a) and acrylic inserts (b) in the phantom

instrument has to be accurate and easy to use, e.g. a syringe containing room temperature water, as measurements of water are temperature sensitive.

In conclusion, the clinically widespread use of specific CT numbers as thresholds for characterizing adrenal lesions must be used with great caution, considering that multiple factors, related to patient, equipment, scanning technique, and observer influence the outcome. This emphasizes that not only native HU and wash-out values, but also other criteria, such as growth rate, lesion homogeneity, and demarcation need to be taken into account in classifying adrenal lesions. Ideally, follow-up examinations of adrenal lesions should be performed at one and the same, well-calibrated CT scanner and images should be read by the same reader, in order to minimize measurement error.

Conflict of interest: None.

REFERENCES

1 Korobkin M, Giordano TJ, Brodeur FJ, *et al.* Adrenal adenomas: relationship between histologic lipid and CT and MR findings. *Radiology* 1996;**200**:743–7

2 Boland GW, Lee MJ, Gazelle GS, *et al.* Characterization of adrenal masses using unenhanced CT: an analysis of the CT literature. *Am J Roentgenol* 1998;**171**:201–4

3 Boland GW, Blake MA, Hahn PF, *et al.* Incidental adrenal lesions: principles, techniques, and algorithms for imaging characterization. *Radiology* 2008;**249**:756–75

4 Szolar DH, Kammerhuber FH. Adrenal adenomas and nonadenomas: assessment of washout at delayed contrast-enhanced CT. *Radiology* 1998;**207**:369–75

5 Korobkin M, Brodeur FJ, Francis IR, *et al.* CT time-attenuation washout curves of adrenal adenomas and nonadenomas. *Am J Roentgenol* 1998;**170**:747–52

6 Pena CS, Boland GW, Hahn PF, *et al.* Characterization of indeterminate (lipid-poor) adrenal masses: use of washout characteristics at contrast-enhanced CT. *Radiology* 2000;**217**:798–802

7 Kebapci M, Kaya T, Gurbuz E, *et al.* Differentiation of adrenal adenomas (lipid rich and lipid poor) from nonadenomas by use of washout characteristics on delayed enhanced CT. *Abdom Imaging* 2003;**28**:709–15

8 Baxter BS, Sorenson JA. Factors affecting the measurement of size and CT number in computed tomography. *Invest Radiol* 1981;**16**:337–41

9 Levi C, Gray JE, McCullough EC, *et al.* The unreliability of CT numbers as absolute values. *Am J Roentgenol* 1982;**139**:443–7

10 Stadler A, Schima W, Prager G, *et al.* CT density measurements for characterization of adrenal tumors ex vivo: variability among three CT scanners. *Am J Roentgenol* 2004;**182**:671–5

11 Nelson JC, Kronmal RA, Carr JJ, *et al.* Measuring coronary calcium on CT images adjusted for attenuation differences. *Radiology* 2005;**235**:403–14

12 Birnbaum BA, Hindman N, Lee J, *et al.* Multi-detector row CT attenuation measurements: assessment of intra- and interscanner variability with an anthropomorphic body CT phantom. *Radiology* 2007;**242**:109–19

13 Hammarstedt L, Muth A, Wangberg B, *et al.* Adrenal lesion frequency: A prospective, cross-sectional CT study in a defined region, including systematic re-evaluation. *Acta Radiol* 2010;**51**:1149–56

14 Muth A, Hammarstedt L, Hellstrom M, *et al.* Cohort study of patients with adrenal lesions discovered incidentally. *Br J Surg* 2011;**98**:1383–91

15 Eisenhauer EA, Therasse P, Bogaerts J, *et al.* New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer* 2009;**45**:228–47

16 National Electrical Manufacturers Association. *Digital imaging and communications in medicine (DICOM) part 14: grayscale standard display functions. PS 3.14-2009.* Rosslyn, VA: NEMA, 2009

17 Good P. *Permutation Tests.* New York, NY: Springer Verlag, 2000

18 NIH state-of-the-science statement on management of the clinically inapparent adrenal mass ("incidentaloma"). *NIH Consens State Sci Statements* 2002;**19**:1–25

19 Al-Hawary MM, Francis IR, Korobkin M. Non-invasive evaluation of the incidentally detected indeterminate adrenal mass. *Best Pract Res Clin Endocrinol Metab* 2005;**19**:277–92

20 Siegel CL, Fisher AJ, Bennett HF. Interobserver variability in determining enhancement of renal masses on helical CT. *Am J Roentgenol* 1999;**172**:1207–12

21 Hopper KD, Kasales CJ, Van Slyke MA, *et al.* Analysis of interobserver and intraobserver variability in CT tumor measurements. *Am J Roentgenol* 1996;**167**:851–4

22 Sangwaiya MJ, Boland GW, Cronin CG, *et al.* Incidental adrenal lesions: accuracy of characterization with contrast-enhanced washout multidetector CT–10-minute delayed imaging protocol revisited in a large patient cohort. *Radiology* 2010;**256**:504–10