Assessing Hepatomegaly:

Automated Volumetric Analysis of the Liver

Marius George Linguraru, DPhil, Jesse K. Sandberg, BS, Elizabeth C. Jones, MD, Nicholas Petrick, PhD, Ronald M. Summers, MD, PhD

Rationale and Objectives: The aims of this study were to define volumetric nomograms for identifying hepatomegaly and to retrospectively evaluate the performance of radiologists in assessing hepatomegaly.

Materials and Methods: Livers were automatically segmented from 148 abdominal contrast-enhanced computed tomographic scans: 77 normal livers and 71 cases of hepatomegaly (diagnosed by visual inspection and/or linear liver height by radiologists). Quantified liver volumes were compared to manual measurements using volume overlap and error. Liver volumes were normalized to body surface area, from which hepatomegaly nomograms were defined (H scores) by analyzing the distribution of liver sizes in the healthy population. H scores were validated against consensus reports. The performance of radiologists in diagnosing hepatomegaly was retrospectively evaluated.

Results: The automated segmentation of livers was robust, with volume overlap and error of 96.2% and 2.2%, respectively. There were no significant differences ($P > .10$) between manual and automated segmentation for either the normal or the hepatomegaly subgroup. The average volumes of normal and enlarged livers were 1.51 ± 0.25 and 2.32 ± 0.75 L, respectively. One-way analysis of variance found that body surface area ($P = .004$) and gender ($P = .02$), but not age, significantly affected normal liver volume. No significant effects were observed for two-way and three-way interactions among the three variables ($P > .18$). H-score cutoffs of 0.92 and 1.08 L/m^2 were used to define mild and massive hepatomegaly (95% confidence interval, ± 0.02 L/m^2). Using the H score as the reference standard, the sensitivity of radiologists in detecting all, mild, and massive hepatomegaly was 84.4%, 56.7%, and 100.0% at 90.1% specificity, respectively. Radiologists disagreed on 20.9% of the diagnosed cases ($n = 31$). The area under the receiver-operating characteristic curve of the H-score criterion for hepatomegaly detection was 0.98.

Conclusions: Nomograms for the identification and grading of hepatomegaly from automatic volumetric liver assessment normalized to body surface area (H scores) are introduced. H scores match well with clinical interpretations for hepatomegaly and may improve hepatomegaly detection compared with height measurements or visual inspection, commonly used in current clinical practice.

Key Words: Hepatomegaly; volumetric analysis; liver; segmentation; nomogram.

Hepatomegaly is an abnormal enlargement of liver size and is inherently defined by a volumetric change. Patient size has been reported to correlate with the volume of the liver (1–4). To date, there are no defined volumetric liver nomograms to detect hepatomegaly.

Hepatic size has been an important biomarker for assessing disorders (4–13) and surgical planning (8,11,14–18). Predictably, hepatic size estimates by physicians using palpation and percussion are approximate (19,20) and adequate for diagnosing only cases of massive hepatomegaly (21–24). Blendis et al (21) found that only half of enlarged livers detected by plain radiography were also identified by physical examination, while approximately half of normal livers were diagnosed as enlarged.

Radiologic imaging modalities such as magnetic resonance imaging, ultrasound, and computed tomographic (CT) imaging are better able to assess liver size (25–29). To detect hepatomegaly in radiologic scans, radiologists commonly focus on landmark-based visual evaluations of the hemidiaphragm, displacement of the stomach, the duodenum, hepatic flexure of the colon, the right kidney, and the lower costal cartilage (30). Nevertheless, the great diversity of normal liver shapes within the population can make landmark-based evaluation of liver volume unreliable (31). In the case of Riedel's lobe, a normal variant, the elongated right liver lobe extends past the lower costal cartilage (32,33), but these livers have normal volumes. Without easily accessible volume measurements, 3% to 31% of the population with this normal variation could potentially be misdiagnosed (32,34,35). Additionally, differentiating between mild and massive hepatomegaly (with the latter category of patients

Acad Radiol 2012; 19:588–598

From the Imaging Biomarkers and Computer-Aided Diagnosis Laboratory, Radiology and Imaging Sciences, Clinical Center, National Institutes of Health, Bethesda, Maryland (M.G.L., J.K.S., E.C.J., R.M.S.); The Sheikh Zayed Institute for Pediatric Surgical Innovation, Children's National Medical Center, 111 Michigan Avenue, NW, Washington, DC 20010 (M.G.L.); and the Center for Devices and Radiological Health, US Food and Drug Administration, Silver Spring, Maryland (N.P.). Received September 26, 2011; accepted January 28, 2012. This work was supported in part by the Intramural Research Programs of the National Institutes of Health, Clinical Center, and the US Food and Drug Administration. The mention of commercial products, their sources, or their use in connection with material reported herein is not to be construed as either actual or implied endorsements of such products by the US Department of Health and Human Services. **Address correspondence to:** M.G.L. e-mail: mlingura@cnmc.org

doi:10.1016/j.acra.2012.01.015

possibly benefiting from medical intervention) is unreliable without systematic quantitative measures of liver size (36,37).

Another popular method routinely used by clinicians is manually measuring the liver's height at the midhepatic line (MHL) from radiologic scans (7,38). The MHL is the half-point distance between the midpoint of the spine and the outermost point on the liver surface (right side) in axial planes. This two-dimensional measure does not fully characterize the liver's morphology, and livers with Riedel's lobe may exhibit large MHL heights. Although correlations between liver size and patient size have been reported (2–4,28,39,40), liver height measurements were not normalized by patient size in previous studies to detect hepatomegaly (7,38).

CT studies have shown that manual volumetric liver measurements are important for diagnoses (1,41). Manual segmentations suffer from two major drawbacks: they are time consuming because of the high and expensive human interaction, and if the same organ is segmented twice (by the same or different observers), the result will likely differ. Currently, most radiologists do not rely on liver volume or even height measurements for diagnosing hepatomegaly, because of the lack of robust and accessible methods adapted to the image-viewing software. Automated volumetric measurements of liver volume have the potential to assist clinicians in the systematic and accurate analysis of hepatic size.

In this report, nomograms are defined to identify and grade hepatomegaly from automated liver volumetry on CT imaging. The automated liver volumetric software was previously developed by our group (42) and allows the quantification of healthy and diseased livers. From readily available volumes of normal livers normalized to patient size (body surface area [BSA]), nomograms to detect hepatomegaly are defined to assist radiologists in routine clinical assessments. The nomograms are validated against consensus reports. Finally, we evaluate retrospectively the performance of radiologists to assess hepatomegaly.

MATERIALS AND METHODS

Study Patients and CT Imaging

Data were acquired at the National Institutes of Health using contrast-enhanced CT scans at the portal venous phase without imaging or motion artifacts or large pathologies in the liver. This retrospective study was in compliance with the Health Insurance Portability and Accountability Act and was institutional review board approved, and the requirement for informed consent was waived.

From January 2001 to March 2009, 71 consecutive subjects with clinically diagnosed hepatomegaly (48 men [mean age, 43 years; age range, 18–76], 23 women [mean age, 41 years; age range, 19–66 years]) met the inclusion criteria shown in Appendix A. The criteria included adult patients with CT scans acquired with intravenous Isovue contrast agent (Bracco Diagnostics, Milan, Italy) in fully enhanced portal venous phase with available patient data. Additionally, the data should not have imaging or motion artifacts or missing slices in the liver. Diagnoses in the clinical reports were established by one of 11 radiologists without following any single criterion and including visual inspection and/or measurements of liver height. Cases were selected using search parameters in the radiology information system (Cerner Corporation, Kansas City, MO) and the clinical research information system (Eclipsys Corporation, Atlanta, GA). Radiologic reports included one of the following key terms: "hepatomegaly," "hepatosplenomegaly," "enlarged liver," or "liver enlarged" (see Appendix A for the study flowchart). Appendix B presents the clinical diagnoses of the hepatomegaly data.

The control population of normal livers was formed from consecutive kidney donors enrolled at the National Institutes of Health from January 2001 to August 2010. Liver function parameters were not available. Seventy-seven subjects with healthy livers (33 men [mean age, 43 years; age range, 17–76 years], 44 women [mean age, 44 years; age range, 18–72 years]) were selected.

The 148 cases (71 patients with hepatomegaly and 77 controls) were reevaluated by two experienced radiologists working by visual inspection and occasional linear measurements of the liver size in the cranial-caudal direction. Cases were presented in random order, and the two radiologists were blinded to the clinical reports (the first evaluation of data). The reassessment of data was used to create consensus reports between three radiologists: two who reevaluated the cases and one from the clinical report.

Additionally, 23 consecutive cases of partial hepatectomy fulfilled the selection criteria between January 2004 and March 2009 at the National Institutes of Health (15 men [mean age, 49 years; age range, 33–70 years], eight women [mean age, 54 years; age range, 28–71 years]). These cases were not included in the definition or evaluation of nomograms. Instead, they were used as an independent set to test the automated segmentation method.

Contrast-enhanced CT images were acquired in portal venous phase during a single breath using fixed delays (65–70 seconds depending on the scanner) or bolus tracking (43) after patients were administered 130 mL of Isovue-300. Data were collected on LightSpeed Ultra/QX/I (GE Healthcare, Milwaukee, WI), Brilliance 64 (Philips Medical Systems, Andover, MA), Definition (Siemens Healthcare, Erlangen, Germany), and Aquilion ONE (Toshiba Medical Systems, Tokyo, Japan) scanners at 100 to 240 mAs and 120 kVp. Image resolutions ranged between 0.52 and 0.93 mm in the axial view with slice thicknesses of 1 to 5 mm. Livers were manually segmented from 20 cases by two observers supervised by an experienced radiologist. The liver heights were manually measured at the MHL in all data, excepting the hepatectomy cases, by the two observers.

Last, 25 random pairs of supine and prone noncontrast CT data sets (slice thickness, 1 mm) from Walter Reed Army Medical Center in Washington, District of Columbia (14 men [mean age, 58 years; age range, 51–73 years], 11 women [mean age, 55 years; age range, 49–68 years]; data courtesy of J. Richard Choi, ScD, MD) were analyzed for intraobserver

intrapatient variability in MHL height measurements under the supervision of an experienced radiologist. Cases were acquired on LightSpeed or LightSpeed Ultra machines (GE Healthcare) (44) and declared exempt from institutional review board review by the National Institutes of Health's Office of Human Subjects Research.

Segmentation

The automated segmentation method quantifies liver volumes from contrast-enhanced CT imaging (42). The software was validated on normal, hepatomegaly, and partial hepatectomy cases and involved appearance, shape, and location statistics of livers. A liver atlas was aligned to the patients' CT data via nonlinear registration. The liver estimation was improved by geodesic active contour and adaptive convolution using patient-specific contrast enhancement. Last, shape and location statistics were used to optimize liver segmentation.

Liver volumes were computed automatically by summing voxel volumes in the rendered three-dimensional segmentations. Liver heights were estimated at the MHL by computer-aided diagnosis (CAD) along the sagittal plane (7).

Definition of Hepatomegaly and Nomograms

Liver volumes exhibiting hepatomegaly are outliers from the average liver size in the healthy population. To account for the relation between liver volume and patient size, volumes of livers were normalized by patients' BSA.

To date, there are no established nomograms for volumetric assessment of hepatomegaly. Nevertheless, standards have been defined for the detection, for example, of osteoporosis (45,46). According to the World Health Organization, osteopenia is diagnosed if the T score of bone mineral density is <1 standard deviation (SD) from the average of the healthy population. Osteoporosis is defined as <2.5 SDs from the average.

We defined the hepatomegaly score (H score) as the measure of liver volume normalized by patient's BSA. The average H score and its SD were computed from our healthy population according to the consensus reports. Only cases that were found normal by all radiologists were used in the computation of the nomograms. Following the approach used to determining osteoporosis, mild hepatomegaly was defined for H scores >1 SD from the average. Massive hepatomegaly was defined for H scores >2.5 SDs from the average.

Performance of Radiologists Relative to the Hepatomegaly Nomogram

The performance of radiologists to diagnose hepatomegaly was retrospectively evaluated using the H score as the reference standard. For comparison with previous nomograms on the basis of liver height (7,38), diagnostic performance using the liver MHL height with a 15.5-cm cutoff was compared to both the consensus reports and H-score reference standards.

TABLE 1. Average Automated Liver Volumes and Heights

Cases	Mean ± Standard Deviation
Normal volume (L) (n = 74)	1.52 ± 0.26
Normal volume/BSA (L/m^2) (n = 74)	0.81 ± 0.11
Normal MHL height (cm) (n = 74)	12.86 ± 2.21
Enlarged volume (L) (n = 43)	2.57 ± 0.76
Enlarged volume/BSA (L/m^2) (n = 43)	1.45 ± 0.40
Enlarged MHL height (cm) (n = 43)	19.32 ± 3.64
Hepatectomy volume (L) (n = 23)	1.56 ± 0.40

BSA, body surface area; MHL, midhepatic line.

Average volumes, volumes normalized to BSA, and heights at the MHL for normal and enlarged livers. Only average volumes are reported for partial hepatectomy cases, because of the frequent inability to calculate the MHL height and extraneous relationship to body size after surgery.

TABLE 2. Correlations between Liver Volume and Height and Patient BSA and Age

Cases/Correlation Factor	Correlation (P value)
Normal volume (n = 74)/BSA	0.60 (<.001)
Normal volume (n = 74)/age	0.01 (.90)
Enlarged volume (n = 43)/BSA	0.41 (.005)
Enlarged volume (n = 43)/age	0.38 (.01)
Normal MHL height (n = 74)/BSA	−0.02 (.80)
Normal MHL height (n = 74)/age	−0.33 (.003)
Enlarged MHL height (n = 43)/BSA	0.08 (.50)
Enlarged MHL height (n = 43)/age	0.18 (.20)

BSA, body surface area; MHL, midhepatic line.

Correlation coefficients (and associated P values) are presented between volume and MHL height measurements and patient BSA and age for normal and enlarged livers.

Statistical Analysis

Retrospective power analysis was performed using a two-sided test with a .05 significance level and a binomial distribution to determine the power of our sample size (n = 148) to detect a significant difference between normal livers and hepatomegaly cases. The variation of the H score with the increase in the number of consecutive samples was analyzed. We also analyzed the receiver-operating characteristic (ROC) curve to identify the operating point with the highest accuracy of the H score, defined as the sum of sensitivity and specificity; the consensus reports were used as the reference standard for hepatomegaly.

Manual and automated volumetric segmentations were compared by volume overlap (twice the volume intersection over the union) and volume error (volume difference relative to the manual volume). Intraobserver and interobserver variability and error analysis for measuring liver height were performed following the Bland-Altman method (47). The Mann-Whitney U test was used to assess significance between interobserver and intraobserver and observer-CAD agreement.

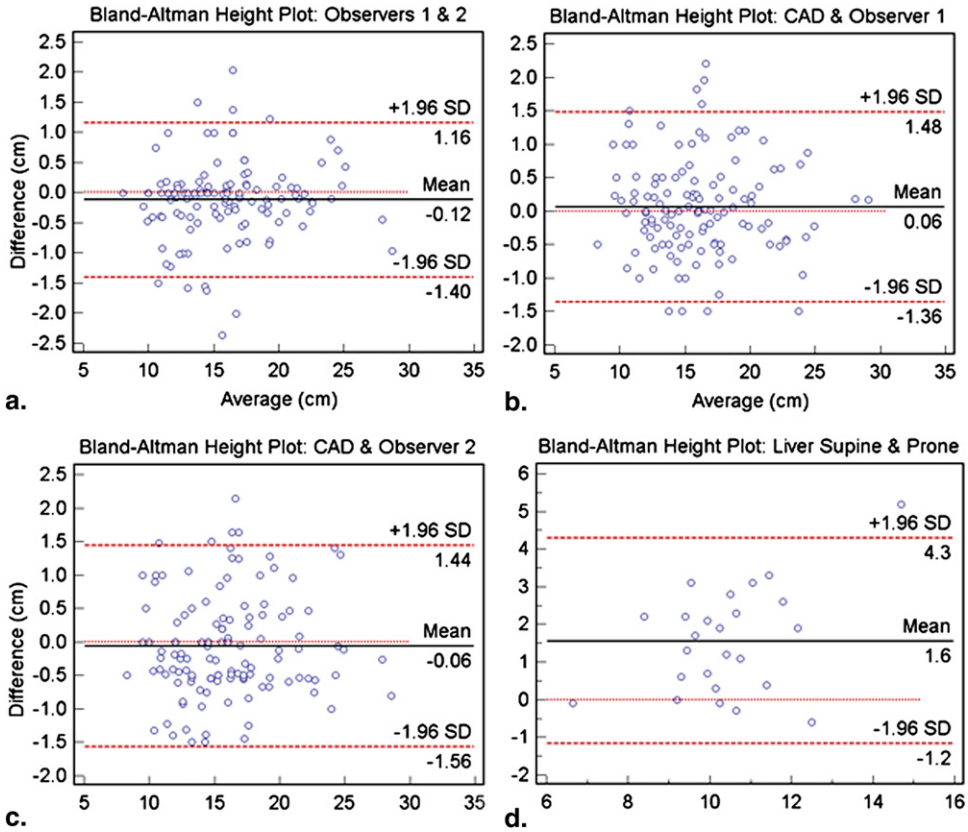


Figure 1. Bland-Altman agreement plots for liver height measurements at the midhepatic line between **(a)** two independent observers (n = 148), **(b)** computer-aided diagnosis (CAD) and observer 1 (n = 148), **(c)** CAD and observer 2 (n = 148), and **(d)** intraobserver supine and prone measurements (n = 25; note scale change on axes). The mean error is shown as the *solid line* and the 95% limits of agreement (±1.96 standard deviations [SDs]) as *dashed lines*.

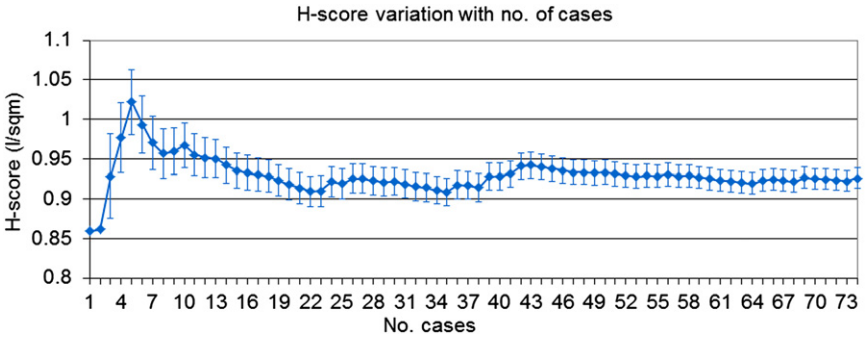


Figure 2. H-score (cutoff to detect hepatomegaly) variation with the number of consecutive normal samples used for its computation. The peak near the beginning of the graph is caused by an outlier with a large H score.

One-way, two-way, and three-way analyses of variance with full interaction were performed for combinations of patient BSA, age, and gender to determine the impact of these factors on normal liver volumes. Pearson's correlation coefficients were calculated between liver size and patient height, weight, BSA, and age for comparisons with the same metrics reported in the literature. Student's paired t tests were used to assess the significance of correlations after testing the normal distribution of data.

Fisher's exact test was used to assess the significance between the sensitivity of radiologists compared to the MHL height criterion to detect hepatomegaly with the H score as the reference standard. We assessed whether the sample size (n = 148) had 90% power to detect a 10% (0.08) change in sensitivity between the two criteria. Pearson's correlation between the H score and MHL height was also analyzed.

RESULTS

The evaluation of the CAD tool (42) showed a volume overlap of 96.2% and a volume error of 2.2% between automatically and manually segmented livers. There was no statistically significant difference ($P > .10$) for either volume overlap or volume error between automated and manual segmentations on normal or abnormal cases.

The consensus reports agreed on 117 cases: 74 normal livers and 43 hepatomegaly cases. The radiologists disagreed on 31 cases (20.9%); three (3.8%) were normal and 28 (39.4%) were abnormal cases, according to the clinical reports. Table 1 presents the volumes, volumes normalized by BSA, and MHL heights for the automatically segmented normal, enlarged, and partial hepatectomy livers. There was no significant difference between the volumes of normal and partial hepatectomy livers ($P = .70$). Significant differences were found between normal and enlarged liver volumes and heights ($P < .001$).

Table 2 shows Pearson's correlation coefficients (R values) between liver measurements and patient BSA and age. The highest correlation ($R = 0.60$, $P < .001$) was noted between liver volumes and BSA in normal cases. Moderate correlations were also observed between the enlarged liver volumes and BSA and age of patients, while a negative moderate correlation was found between the liver MHL heights and ages of controls. No significant correlations existed between liver height and patient BSA.

One-way analysis of variance found that BSA ($P = .004$) and gender ($P = .01$), but not age ($P = .50$), significantly affected normal liver volume. No significant effects were observed for two-way and three-way interactions among the three variables ($P > .10$).

Bland-Altman agreement plots for height measurements between two observers and between each observer and CAD are shown in Figures 1a to 1c. Interobserver variability was 0.12 ± 1.28 cm, and the bias between the CAD method and each observer was 0.06 ± 1.5 cm at 95% limits of agreement. Significant correlations ($R = 0.98$, $P < .001$) were found between each observer's and CAD measurements, comparable to the interobserver measurement correlation ($R = 0.98$, $P < .001$). Outliers in Figure 1 corresponded to unusually shaped livers, which increased the variability of MHL height measurements.

Figure 1d presents Bland-Altman agreements for consecutive intraobserver MHL height measurements on pairs of supine and prone noncontrast CT scans of normal livers ($n = 25$). The variability in measurement was 1.6 ± 2.8 cm at 95% limits of agreement. These intraobserver errors were significantly larger ($P < .001$) than the errors between CAD and manual measurements on supine scans (Figs 1b and 1c).

An H-score cutoff of 0.92 L/m^2 was used to identify mild hepatomegaly (95% confidence interval, ± 0.02 L/m^2). The H-score cutoff for massive hepatomegaly was 1.08 L/m^2. Figure 2 shows the variation of the H score with the number of consecutive samples used to compute it. The H score became stable after approximately 50 cases. On ROC analysis, the operating point of the highest accuracy in the H score using the consensus reports as the reference standard was found at 0.93 L/m^2, which is within the 95% confidence interval. The sensitivity and specificity at this operating point were 97.6% and 86.4%, respectively. The area under the ROC curve of the H-score criterion for hepatomegaly detection was 0.98. The statistical power of our data set ($n = 148$) for the observed ΔH score = 0.81 was 99%.

TABLE 3. Retrospective Sensitivity and Specificity of Radiologists and Liver MHL Height to Detect Hepatomegaly

	Sensitivity/Specificity (%)		
Criterion ($n = 148$)	Mild Hepatomegaly	Massive Hepatomegaly	All Hepatomegaly
Radiologists	56.7/90.1	*100.0*/90.1	*84.4*/90.1
MHL height	40.0/88.7	85.1/88.7	67.5/88.7

MHL, midhepatic line.

The H score was used as the reference standard with cutoffs of 0.92 and 1.08 L/m^2 to identify mild and massive hepatomegaly. The radiologic reports were used to compute the sensitivity and specificity of the radiologists. A cutoff of 15.5 cm was used to identify hepatomegaly using liver MHL height. Significant differences ($P < .007$, shown in italics) were noted between the sensitivity of the radiologist and the sensitivity of the liver MHL height criterion for the detection of massive and all cases of hepatomegaly.

From H-score nomograms (the new reference standard), the performance of radiologists was retrospectively analyzed using the clinical reports. Table 3 presents the sensitivity and specificity of radiologists to detect hepatomegaly. The performance of diagnosis by liver MHL height is also shown. The statistical power of our data set ($n = 148$) to detect an effect size of 0.08 on the sensitivity between diagnosis criteria was 93%. There was no significant difference in the sensitivity of the detection of mild hepatomegaly between radiologists (56.7%, $P = .09$) and MHL height. The sensitivity of radiologists to detect massive hepatomegaly was significantly higher (100.0%, $P = .006$) than using MHL height. Radiologists also detected all cases of hepatomegaly with significantly higher sensitivity (84.4%, $P = .007$) than the MHL height criterion.

The linear regression model in Figure 3a showed a significant correlation between the H scores and MHL heights of normal cases in the consensus reports ($R = 0.34$, $P = .02$). A significant correlation was also found between the H scores and MHL heights of enlarged livers ($R = 0.58$, $P < .001$; Fig 3b). A 15.5-cm cutoff at MHL height (7,38) detected hepatomegaly with 83.7% sensitivity at 90.5% specificity using the consensus reports as the reference standard and 67.5% sensitivity at 88.7% specificity with the H score as the reference standard.

DISCUSSION

This retrospective study presents the evaluation of an automated method to segment livers from contrast-enhanced CT images (42). The automated technique was accurate and consistent with manual segmentations for liver volume and height for cohorts of normal and enlarged livers. Table 1 illustrates the average normal liver volumes from our data, which are compared to reports in the literature in Table 4 (48–52). Our average diseased liver volumes were generally larger than in the literature. The liver volumes of patients who underwent partial hepatectomy were comparable to those of normal controls. Although their average volumes were

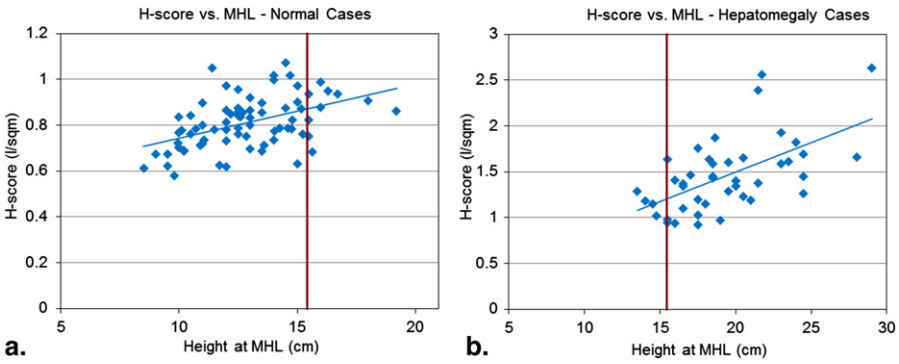


Figure 3. Correlations between the H score and liver midhepatic line (MHL) heights using a linear regression model: **(a)** normal controls and **(b)** patients with hepatomegaly (as defined by the H score). The *vertical red lines* show the cutoff for hepatomegaly as defined in the literature by an MHL height of 15.5 cm (7,38).

TABLE 4. Review of Average Volumes for Normal and Diseased Livers

Study	Patient Livers	Sex	Number of Cases	Mean Volume (L)	Modality
Andersen et al (1)	Normal	Male	7	1.60	CT
Kwo et al (49)	Normal	Male	10	1.54 ± 0.08	CT
Sandrasegaran et al (41)	Normal	Male	8	1.67	CT
Andersen et al (1)	Normal	Female	16	1.34	CT
Kwo et al (49)	Normal	Female	10	1.48 ± 0.06	CT
Sandrasegaran et al (41)	Normal	Female	11	1.52	CT
Farraher et al (50)	Normal	Male/female	18	1.66 ± 0.35	MRI
Henderson et al (51)	Normal	Male/female	11	1.45 ± 0.17	CT
Kardel et al (48)	Normal	Male/female	20	1.61 ± 0.19	Ultrasound
Mazonakis et al (30)	Normal	Male/female	27	1.47 ± 0.23	MRI
Stapakis et al (52)	Normal	Male/female	22	1.32 ± 0.42	CT
Farraher et al (50)	Diseased	Male/female	9	1.99 ± 0.52	MRI
Henderson et al (51)	Diseased	Male/female	12	1.64 ± 0.7	CT
Mazonakis et al (30)	Diseased	Male/female	11	1.94 ± 0.19	MRI
Van Thiel et al (13)	Diseased	Male/female	99	1.78 ± 0.09	CT

CT, computed tomography; MRI, magnetic resonance imaging.
Table 1 reports the average volumes of normal and diseased livers in our study.

similar to those of the normal controls, we preferred to compute hepatomegaly nomograms from data from patients who did not undergo liver surgery.

It is generally agreed that liver volume and patient BSA and weight are correlated (Table 5 [53–56]). This observation, and the statistical analysis in this paper, supports the normalization of the liver volume by BSA in the definition of the H score. A significant correlation between normal liver volume and age has been demonstrated (3,53–55); this was confirmed in Andersen et al (1) and in our study (Table 2). Reports also do not agree on the correlations between liver volume and patient height (1,2,28,39,53) or liver volume and patient age (3,53–55). The sensitivity and specificity of the H score to detect hepatomegaly using consensus reports as the reference standard were 97.6% and 86.4%, respectively, with an area under the ROC curve of 0.98.

Despite an abundance of studies analyzing normal and diseased liver volumes (Table 4), we are not aware of another effort to define nomograms to classify hepatomegaly using a volumetric measurement such as the H score. More commonly, hepatomegaly is classified using liver MHL height. Gosink and Leymaster (7) found that 93% of normal livers had MHL heights <13 cm, while 75% of patients with MHL heights >15.5 cm had hepatomegaly. Similarly, Rosenfield and Schneider (38) defined an upper limit of 15.5-cm MHL

TABLE 5. Correlations between Liver Volume and Patient Data

Study	Body Surface Area	Weight	Height	Age
This study	Yes	Yes	Yes	Yes
Andersen et al (1)	–	Yes	No	No
Boyd (54)	–	–	–	Yes
Calloway et al (55)	–	–	–	Yes
Hauksen et al (39)	Yes	Yes	Yes	–
Kardel et al (48)	–	Yes	–	–
Kratzer et al (53)	–	Yes	Yes	Yes
Leung et al (2)	Yes	Yes	No	–
McNeal et al (56)	–	Yes	–	–
Raeth et al (28)	Yes	Yes	Yes	–
Urata et al (40)	Yes	–	–	–
Wynne et al (3)	Yes	–	–	Yes
Zoli et al (4)	Yes	Yes	–	–

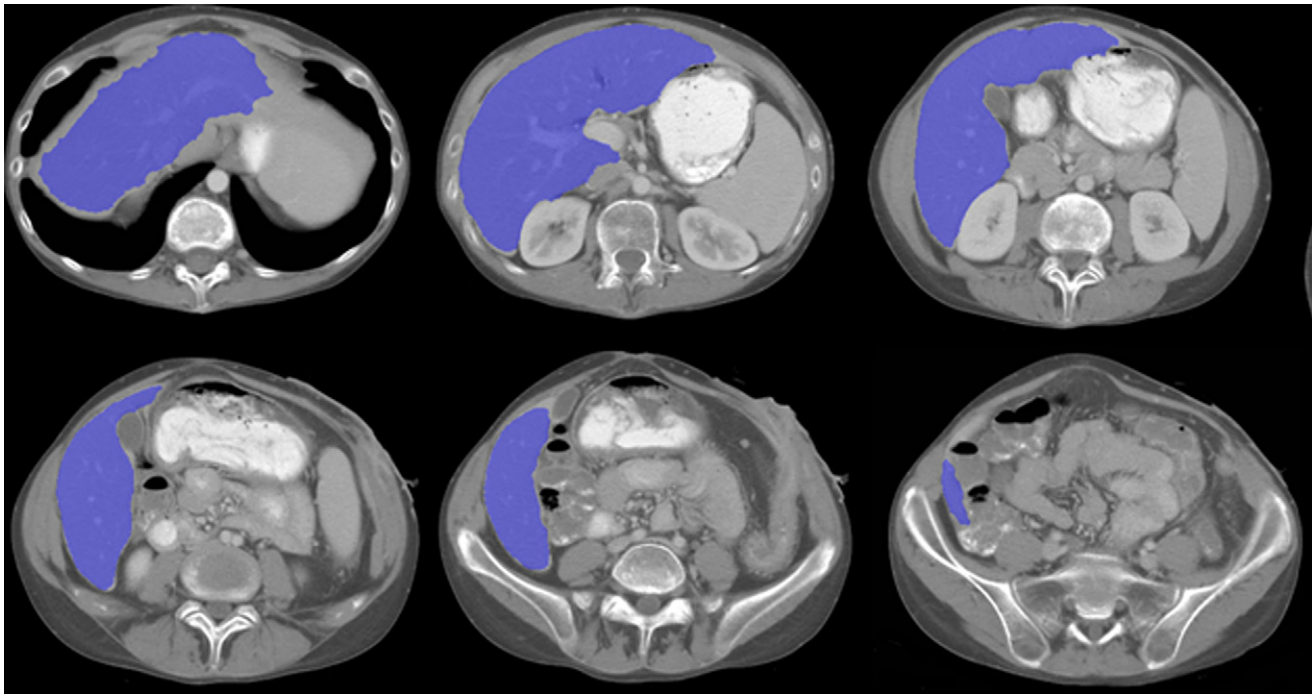


Figure 4. Image data of a patient with Riedel's lobe with abnormal liver midhepatic line height of 17.6 cm and normal H score of 0.88 L/m^2. The automatically segmented liver is overlaid in *blue* over the computed tomographic image and shown in *axial slices*.

height for normal livers. Niederau and Sonnenberg (57) defined the upper normal limit of liver height at the midclavicular line to be about 12.7 cm. The standards in the studies of Gosink and Leymaster and Rosenfield and Schneider were based on 36 and 33 cases, respectively. Our study found that a 15.5-cm cutoff in MHL height detected hepatomegaly with 67.5% sensitivity at 88.7% specificity using the H score as the reference standard.

Four patients clinically misclassified with hepatomegaly had confirmed Riedel's lobe. Figure 4 exemplifies the segmentation of a liver with Riedel's lobe with an abnormally high liver MHL height but a normal H score.

Radiologists use experience and visual landmarks to diagnose the liver. Unsurprisingly, Table 3 shows that radiologists detected hepatomegaly with significantly higher sensitivity than the liver MHL height criterion. The H score, a systematic quantitative approach to diagnosing hepatomegaly, should be preferable to visual inspection for consistency and generalization. A uniform quantitative assessment of hepatomegaly has additional advantages; the criterion can be subdivided into categories, such as mild and massive hepatomegaly, which is less reliable using visual inspection. This finer categorization has the potential to customize treatments and monitor stages of diseases, such as cirrhosis or drug-related hepatitis. Selected patients with massive hepatomegaly can also benefit from operative intervention (36,37,58).

The large intraobserver errors noted for consecutive liver height measurements on supine and prone CT image pairs (Fig 1d) indicate that MHL height measurement is sensitive to patient position in the scan and unreliable for detecting hepatomegaly. However, Leung et al (2) also found substantial diurnal variations in liver volume, with a minimum value between 12 AM and 2 PM (17% fall average; range, 9%–31%). Our study did not include time and position variables.

This study had certain limitations. The distribution of the patient population with hepatomegaly may not be representative. Nevertheless, patients with hepatomegaly were not admitted to the National Institutes of Health on the basis of hepatomegaly diagnosis, which was a secondary radiologic finding. The definition of the H score would also benefit from a larger data sample. Finally, the reader variability in the clinical determination of hepatomegaly is potentially very large, as radiologists disagreed in 20.9% ($n = 31$) of cases, but representative for the clinical environment of a medium-sized to large hospital.

We proposed using liver size nomograms for detecting hepatomegaly. Liver size is inherently a volume and correlated to patient BSA. CAD can offer robust and reproducible liver volume measurements in a fully automated manner to support routine radiologic image analysis. In this light, the H score defined from liver volume normalized to patient BSA may be useful as a systematic indication of hepatomegaly.

With the introduction of volumetric measurements in routine radiologic assessment, whether organ or tumor volumes, nomograms have the potential to offer systematic and reproducible tools for diagnosis. Whether supported or not by CAD, the introduction of liver volume nomograms promises to reduce variability and error in the radiologic interpretation of abdominal data. Our study found 20.9% disagreement between radiologists in diagnosing hepatomegaly. If liver

CAD were to be adopted in the clinical work flow in combination with volumetric nomograms, the diagnosis of hepatomegaly could become a seamless automated process. Finally, differentiating between cases of mild and massive hepatomegaly is unreliable without methodical quantitative measures of liver size.

Automatic volumetric liver assessment normalized to BSA may improve the identification of hepatomegaly detection compared with height measurements or visual inspection, commonly used in clinical practice. Using the H score, a systematic quantitative measure of liver size, as the reference standard, radiologists detected all, mild, and massive cases of hepatomegaly with 84.4%, 56.7%, and 100.0% sensitivity, respectively, at 90.1% specificity.

ACKNOWLEDGMENTS

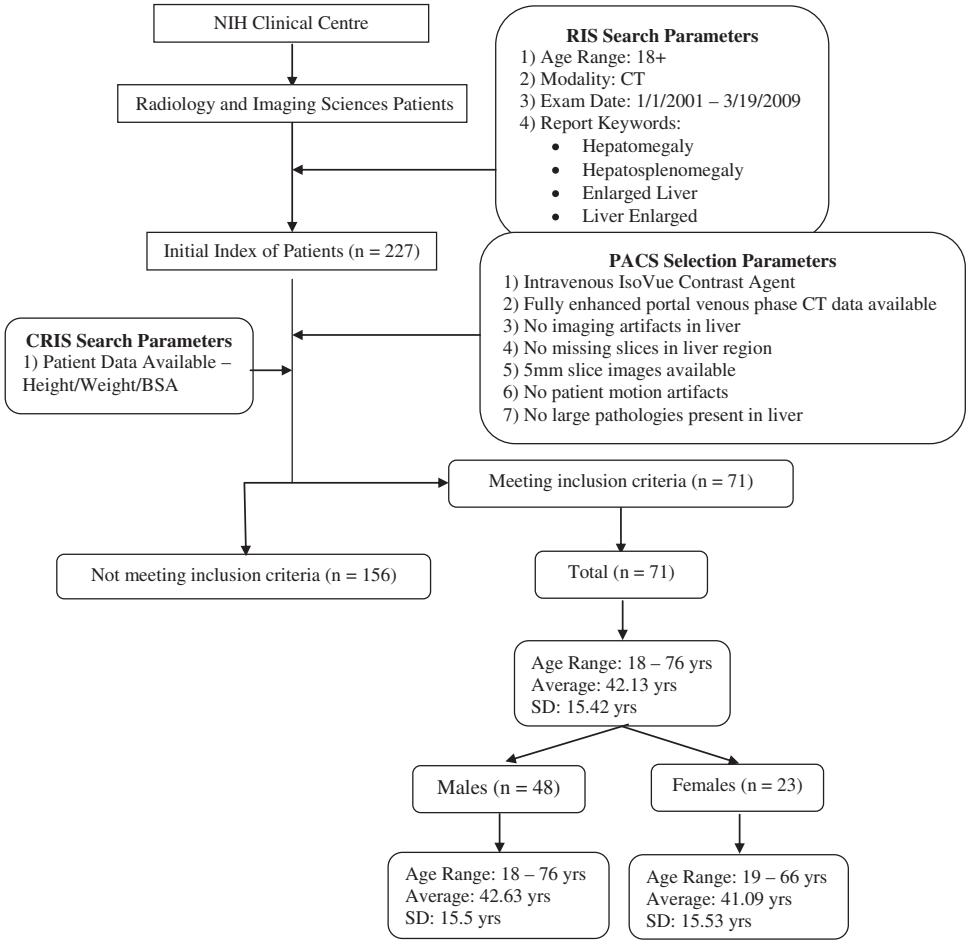
We would like to thank Zhixi Li, Furhawn Shah, and Visal Desai for assistance with data analysis and Andrew J. Dwyer, MD, for critical review.

REFERENCES

1. Andersen V, Sonne J, Sletting S, et al. The volume of the liver in patients correlates to body weight and alcohol consumption. Alcohol Alcohol 2000; 35:531–532.
2. Leung NW, Farrant P, Peters TJ. Liver volume measurement by ultrasonography in normal subjects and alcoholic patients. J Hepatol 1986; 2: 157–164.
3. Wynne HA, Cope LH, Mutch E, et al. The effect of age upon liver volume and apparent liver blood flow in healthy man. Hepatology 1989; 9: 297–301.
4. Zoli M, Cordiani MR, Marchesini G, et al. Prognostic indicators in compensated cirrhosis. Am J Gastroenterol 1991; 86:1508–1513.
5. Ellert J, Kreel L. The role of computed tomography in the initial staging and subsequent management of the lymphomas. J Comput Assist Tomogr 1980; 4:368–391.
6. Elstein D, Hadas-Halpern I, Azuri Y, et al. Accuracy of ultrasonography in assessing spleen and liver size in patients with Gaucher disease: comparison to computed tomographic measurements. J Ultrasound Med 1997; 16:209–211.
7. Gosink BB, Leymaster CE. Ultrasonic determination of hepatomegaly. J Clin Ultrasound 1981; 9:37–44.
8. Kawasaki S, Makuuchi M, Matsunami H, et al. Preoperative measurement of segmental liver volume of donors for living related liver transplantation. Hepatology 1993; 18:1115–1120.
9. Okamoto E, Kyo A, Yamanaka N, et al. Prediction of the safe limits of hepatectomy by combined volumetric and functional measurements in patients with impaired hepatic function. Surgery 1984; 95:586–592.
10. Sekiyama K, Yoshiba M, Inoue K, et al. Prognostic value of hepatic volumetry in fulminant hepatic failure. Dig Dis Sci 1994; 39:240–244.
11. Soyer P, Roche A, Elias D, et al. Hepatic metastases from colorectal cancer: influence of hepatic volumetric analysis on surgical decision making. Radiology 1992; 184:695–697.
12. Tsushima Y, Endo K. Spleen enlargement in patients with nonalcoholic fatty liver: correlation between degree of fatty infiltration in liver and size of spleen. Dig Dis Sci 2000; 45:196–200.
13. Van Thiel DH, Hagler NG, Schade RR, et al. In vivo hepatic volume determination using sonography and computed tomography. Validation and a comparison of the two techniques. Gastroenterology 1985; 88: 1812–1817.
14. Broering DC, Sterneck M, Rogiers X. Living donor liver transplantation. J Hepatol 2003; 38(suppl):S119–S135.
15. Chen YS, Cheng YF, De Villa VH, et al. Evaluation of living liver donors. Transplantation 2003; 75(suppl):S16–S19.
16. Hermoye L, Laamari-Azjal I, Cao Z, et al. Liver segmentation in living liver transplant donors: comparison of semiautomatic and manual methods. Radiology 2005; 234:171–178.
17. Kasai H, Makuuchi M, Kawasaki S, et al. Intraoperative color Doppler ultrasonography for partial-liver transplantation from the living donor in pediatric patients. Transplantation 1992; 54:173–175.
18. Sugarbaker PH. Surgical decision making for large bowel cancer metastatic to the liver. Radiology 1990; 174:621–626.
19. Delp MH, Manning RT. Examination of the abdomen and gastrointestinal and hepatobiliary disorders. In: Major's physical diagnosis. 9th ed. Philadelphia, PA: Lippincott Williams & Wilkins, 1981.
20. Sherlock S. Disease of the liver and biliary system, Vol 1. Boston: Blackwell Scientific, 1981.
21. Blendis LM, McNeilly WJ, Sheppard L, et al. Observer variation in the clinical and radiological assessment of hepatosplenomegaly. Br Med J 1970; 1:727–730.
22. Castell DO, O'Brien DK, Myench H, et al. Estimation of liver size by percussion in normal individuals. Ann Intern Med 1969; 70:1183–1189.
23. Meyhoff HH, Roder O, Andersen B. Palpatory estimation of liver size. Within- and between-observer variation. Acta Chir Scand 1979; 145:479–481.
24. Riemenschneider PA, Whalen JP. Relative accuracy of estimation of enlargement of liver and spleen by radiologic and clinical methods. Am J Roentgenol Radium Ther Nucl Med 1965; 94:462–468.
25. Cheng YF, Chen CL, Huang TL, et al. Single imaging modality evaluation of living donors in liver transplantation: magnetic resonance imaging. Transplantation 2001; 72:1527–1533.
26. Schroeder T, Nadalin S, Stattaus J, et al. Potential living liver donors: evaluation with an all-in-one protocol with multi-detector row CT. Radiology 2002; 224:586–591.
27. Heymsfield SB, Fulenwider T, Nordlinger B, et al. Accurate measurement of liver, kidney, and spleen volume and mass by computerized axial tomography. Ann Intern Med 1979; 90:185–187.
28. Raeth U, Johnson PJ, Williams R. Ultrasound determination of liver size and assessment of patients with malignant liver disease. Liver 1984; 4:287–293.
29. Rasmussen SN. Liver volume determination by ultrasonic scanning. Br J Radiol 1972; 45:579–585.
30. Mazonakis M, Damilakis J, Maris T, et al. Comparison of two volumetric techniques for estimating liver volume using magnetic resonance imaging. J Magn Reson Imaging 2002; 15:557–563.
31. McAfee JG, Ause RG, Wagner HN, Jr. Diagnostic value of scintillation scanning of the liver. Arch Intern Med 1965; 116:95–110.
32. Gillard JH, Patel MC, Abrahams PH, et al. Riedel's lobe of the liver: fact or fiction? Clin Anat 1998; 11:47–49.
33. Kudo M. Riedel's lobe of the liver and its clinical implication. Int Med 2000; 39:87–88.
34. Baum S, Locko RC, Davignon MB. Functional-anatomy and radionuclide imaging—Riedels lobe of the liver. Anat Clin 1982; 4:121–123.
35. Sham R, Sain A, Silver L. Hypertrophic Riedel's lobe of the liver. Clin Nucl Med 1978; 3:79–81.
36. Aussilhou B, Douflé G, Hubert C, et al. Extended liver resection for polycystic liver disease can challenge lver transplantation. Ann Surg 2010; 252:735–743.
37. Schnelldorfer T, Torres VE, Zakaria S, et al. Polycystic liver disease. Ann Surg 2009; 250:112–118.
38. Rosenfield AT, Schneider PB. Rapid evaluation of hepatic size on radioisotope scan. J Nucl Med 1974; 15:237–240.
39. Hausken T, Leotta DF, Helton S, et al. Estimation of the human liver volume and configuration using three-dimensional ultrasonography: effect of a high-caloric liquid meal. Ultrasound Med Biol 1998; 24:1357–1367.
40. Urata K, Kawasaki S, Matsunami H, et al. Calculation of child and adult standard liver volume for liver transplantation. Hepatology 1995; 21: 1317–1321.
41. Sandrasegaran K, Kwo PW, DiGirolamo D, et al. Measurement of liver volume using spiral CT and the curved line and cubic spline algorithms: reproducibility and interobserver variation. Abdom Imaging 1999; 24: 61–65.
42. Linguraru MG, Sandberg JK, Li Z, et al. Automated segmentation and quantification of liver and spleen from CT images using normalized probabilistic atlases and enhancement estimation. Med Phys 2010; 37: 771–783.
43. Goshima S, Kanematsu M, Nishibori H, et al. Multi-detector row CT of the kidney: optimizing scan delays for bolus tracking techniques of arterial, corticomedullary, and nephrographic phases. Eur J Radiol 2007; 63: 420–426.

44. Pickhardt PJ, Choi JR, Hwang I, et al. Computed tomographic virtual colonoscopy to screen for colorectal neoplasia in asymptomatic adults. N Engl J Med 2003; 349:2191–2200.
45. Link TM, Koppers BB, Licht T, et al. In vitro and in vivo spiral CT to determine bone mineral density: initial experience in patients at risk for osteoporosis. Radiology 2004; 231:805–811.
46. Engelke K, Adams JE, Armbrecht G, et al. Clinical use of quantitative computed tomography and peripheral quantitative computed tomography in the management of osteoporosis in adults: the 2007 ISCD official positions. J Clin Densitom Assess Skel Health 2008; 11:123–162.
47. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986; 1: 307–310.
48. Kardel T, Holm HH, Rasmussen SN, et al. Ultrasonic determination of liver and spleen volumes. Scand J Clin Lab Invest 1971; 27:123–128.
49. Kwo PY, Ramchandani VA, O'Connor S, et al. Gender differences in alcohol metabolism: relationship to liver volume and effect of adjusting for body mass. Gastroenterology 1998; 115:1552–1557.
50. Farraher SW, Jara H, Chang KJ, et al. Liver and spleen volumetry with quantitative MR imaging and dual-space clustering segmentation. Radiology 2005; 237:322–328.
51. Henderson JM, Heymsfield SB, Horowitz J, et al. Measurement of liver and spleen volume by computed tomography. Assessment of reproducibility and changes found following a selective distal splenorenal shunt. Radiology 1981; 141:525–527.
52. Stapakis J, Stamm E, Townsend R, et al. Liver volume assessment by conventional vs. helical CT. Abdom Imaging 1995; 20:209–210.
53. Kratzer W, Fritz V, Mason RA, et al. Factors affecting liver size: a sonographic survey of 2080 subjects. J Ultrasound Med 2003; 22:1155–1161.
54. Boyd E. Normal variability in weight of the adult human liver and spleen. Arch Pathol 1933; 16:350–372.
55. Calloway NO, Foley CF, Lagerbloom P. Uncertainties in geriatric data. II. Organ size. J Am Geriatr Soc 1965; 13:20–28.
56. McNeal GR, Maynard WH, Branch RA, et al. Liver volume measurements and three-dimensional display from MR images. Radiology 1988; 169: 851–854.
57. Niederau C, Sonnenberg A. Liver size evaluated by ultrasound: ROC curves for hepatitis and alcoholism. Radiology 1984; 153:503–505.
58. Kornasiewicz O, Dudek K, Bugajski M, et al. Choice of transplantation techniques and indications for liver transplantation in polycystic liver disease in patients with no signs of end-stage liver disease. Transplantation Proceedings 2008; 40:1536–1538.

APPENDIX A. STARD FLOWCHART FOR HEPATOMEGALY STUDY



BSA, body surface area; CRIS, clinical research information system; CT, computed tomography; NIH, National Institutes of Health; PACS, picture archiving and communication system; RIS, radiology information system; SD, standard deviation; STARD, Standards for the Reporting of Diagnostic Accuracy Studies.

APPENDIX B. CLINICAL DIAGNOSES OF THE HEPATOMEGALY DATA

Liver Disorder	Number of Cases
Acute myeloid leukemia	2
Adult T-cell leukemia/lymphoma	1
Alveolar soft part sarcoma	1
Aplastic anemia	2
Appendiceal cancer	1
Autoimmune lymphoproliferative syndrome type 12	1
B-cell chronic lymphocytic leukemia	1
Chronic granulomatous disease	7
Chronic myelogenous (or myeloid) leukemia	2
Colon cancer with liver metastases	1
Cutaneous T-cell lymphoma	1
Dermatomyositis lipodystrophy	1
Desmoplastic small round cell tumor	1
D-MAC/myelodysplastic syndromes	1
Follicular lymphoma	1
Hepatitis C	1
Hepatosplenic T-cell lymphoma	1
Hereditary papillary renal carcinoma type I	1
Histoplasmosis (Darling's disease)	1
HIV	1
HIV with MAC	2
HIV/hepatic steatosis	1
HIV/KS/multicentric Castleman's disease	1
HIV/primary effusion lymphoma/multicentric Castleman's disease/KS	1
HIV with Burkitt's non-Hodgkin's lymphoma	1
HIV/lymphoma	1
Hypertriglyceridemia/pancreatitis	1
Hypertension	1

(*continued on next page*)

(*continued*)

Liver Disorder	Number of Cases
KS	2
Large granular lymphocytic leukemia	2
Lymphoma	3
Mantle cell lymphoma	1
Mastocytosis	2
Melanoma	4
Mesothelioma	1
Multicentric Castleman's disease	3
MAC	1
Mycobacterium avium intracellulare	1
No diagnosis	3
Nontuberculous mycobacteria	1
Ocular melanoma with liver metastases	3
Peritoneal mesothelioma	1
Prostate carcinoma	2
Pure red cell aplasia	1
Systemic lupus erythematosus	1
Thalassemia	1
Total	71

HIV, human immunodeficiency virus; KS, Kaposi's sarcoma; MAC, mycobacterium avium complex.