Liver fat is reproducibly measured using computed tomography in the Framingham Heart Study

Elizabeth K. Speliotes[1,4], **Joseph M. Massaro**[5,6], **Udo Hoffmann**[2], **Meredith C. Foster**[5], **Dushyant N. Sahani**[2], **Joel N. Hirschhorn**[4,8,9], **Christopher J. O'Donnell**[3,5], and **Caroline S. Fox**[5,7]

[1] Department of Gastroenterology, Massachusetts General Hospital, Boston MA

[2] Department of Radiology, Massachusetts General Hospital, Boston MA

[3] Department of Cardiology, Massachusetts General Hospital, Boston MA

[4] Department of Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge MA

[5] NHLBI's the Framingham Heart Study, Framingham MA

[6] Department of Statistics, Boston University, Boston MA

[7] Departments of Medicine and Endocrinology, Brigham and Women's Hospital, Boston MA

[8] Departments of Endocrinology and Genetics, Children's Hospital, Boston MA

[9] Department of Genetics, Harvard Medical School

Abstract

Background & Aims—Fatty liver is the hepatic manifestation of obesity, but community-based assessment of fatty liver among unselected subjects is limited. We sought to determine the feasibility of and optimal protocol for quantifying fat content in liver in the Framingham Heart Study using multi-detector computed tomography (MDCT) scanning.

Methods—Participants (n=100, 49% women, mean age 59.4 years, mean BMI 27.8 kg/m^2) were drawn from the Framingham Heart Study Cohort. Two readers measured the attenuation of liver, spleen, paraspinal muscle, and an external standard from MDCT scans using multiple slices in chest and abdominal scans.

Results—The mean measurement variation was larger within a single axial CT slice than between multiple axial CT slices for liver and spleen whereas it was similar for paraspinal muscles. Measurement variation in liver, spleen, and paraspinal muscles was smaller in the abdomen than in the chest. Three vs. six measures of attenuation in liver and two vs. three measures in spleen gave reproducible measurements of tissue attenuation (intra-class correlation coefficient (ICCC) of 1 in the abdomen). Intra- and inter-reader reproducibility (ICCC) of the liver-to-spleen ratio was 0.98 and 0.99, of the liver-to-phantom ratio was 0.99 and 0.99, and of the liver-to-muscle ratio was 0.93 and 0.86, respectively.

Conclusions—One cross-sectional slice is adequate to capture the majority of variance of fat content in liver per individual. Abdominal as compared to chest scan measures of fat content in

Correspondence to: Elizabeth K. Speliotes MD PhD Department of Gastroenterology, Massachusetts General Hospital 15 Fruit St. Boston, MA 02114 (617)724-6113 espeliotes@partners.org. Caroline.S. Fox MD MPH Medical Officer, NHLBI's Framingham Heart Study 73 Mt Wayte Ave Suite #2 Framingham MA 01702 (508) 935-3447 foxca@nhlbi.nih.gov.

liver are more precise. Measurement of fat content in liver on MDCT scans is feasible and reproducible.

Keywords

Fatty Liver; reproducibility; CT scan; metabolic syndrome; measurement

Background and Aims

Concomitant with the increase in obesity worldwide is an increase in obesity-related illnesses including diabetes, hypertension, heart disease, and non-alcoholic fatty liver disease.[1,2] Fat deposition in the liver represents the hepatic manifestation of obesity. This deposition of fat in the liver can lead to liver inflammation and scarring and ultimately cirrhosis and liver failure.[2] Fatty liver is an independent correlate of metabolic syndrome and its components, including abdominal obesity, insulin resistance, and hypertriglyceridemia.[3] However, little is known about how fat in the liver relates to the development of metabolic syndrome and its components. By determining the best correlates of fatty liver in the community, we may begin to better understand the causes and consequences of fatty liver.

Liver biopsy is an invasive method of determining fat content in liver that can have multiple complications and thus is not practical for determining the presence of fatty liver in healthy individuals. Fortunately, fatty liver can be detected in clinically asymptomatic individuals in the community using computed tomography (CT). The attenuation of the liver on CT when controlling for the penetrance of the scan using an internal control can be correlated to the amount of fat in the liver.[4] Among 266 living liver donors, liver biopsy and concomitant CT revealed that a liver-to-spleen ratio (LSR: the CT Hounsfield Units (HU) of the liver divided by the CT HU of the spleen) of 1.1 was the most reliable way of non-invasively distinguishing the presence of 30% or more of steatosis with a sensitivity and specificity of 83% and 82%, respectively.[4]

To determine the best correlates of fatty liver we also need to measure the fat content of liver in a population-based sample of individuals that have been well-characterized genetically and epidemiologically. Over 3500 individuals from two generations (the Offspring and Third Generation) of the Framingham Heart Study have undergone abdominal multi-detector CTs. Whether measures of fat content in liver using multi-detector CT scans in a community-based sample is feasible and reproducible is not known. In this study we aim to determine the reproducibility of measuring fatty liver in the community-based Framingham Heart Study cohort.

Materials and Methods

Subjects

The Framingham Heart Study cohort is a community-based cohort of over 14,000 individuals across three generations who have been followed for the development of cardiovascular disease and its risk factors.[5] The cohort consists of three generations of participants. The original cohort includes 5209 participants who have had 28 exams to date starting in 1948; the offspring cohort includes 5124 participants who have completed seven exams to date starting in 1971; the third generation cohort includes 4094 participants who have had one exam to date starting in 2002. Over 3500 individuals from two generations (the Offspring and Third Generation) of the Framingham Heart Study have undergone abdominal multi-detector CTs.

1422 individuals from the offspring cohort have had CT scans conducted between 2002 and 2005. Of these, 100 individuals were randomly selected to have equal numbers of men and women as well and approximately equal numbers of participants in the age groups of 35–44, 45–54, 55–64, 65–74 and 74–84. Subjects were excluded from having CTs if they were female with definite or possible pregnancy, were >160kg (352lbs), or if they were <35 years of age (men) or 40 years of age (women). All subjects provided written consent and the study and protocol was approved by the institutional review boards of the Boston University Medical Center and Massachusetts General Hospital.

Multi-detector computed tomography *(MDCT) scan protocol*

Individuals were scanned using an eight-slice MDCT (LightSpeed Ultra, General Electric, Milwaukee, WI) in the supine position and amounted to a total effective radiation exposure of 2.7 mSv. Twenty-five contiguous five-mm thick slices (120 kVp, 400 mA, gantry rotation time 500 ms, table feed 3:1) were acquired covering 125 mm above the level of S1 and raw data were reconstructed using a 55 cm field of view. In the chest 48 continuous 2.5-mm thick slices (120 kVp, 320/400 mA [for < and > 220 pounds of body weight, respectively], gantry rotation time 500 ms, temporal resolution 330 ms) were acquired during a single breath hold and reconstructed using a 35 cm field of view.

A calibration control (phantom) (Image Analysis, Lexington, KY, US) with water equivalent compound (CT-Water™) and calcium hydroxyapatite at 0, 75, and 150 mg/cm^3 was placed under each subject. We used the 150 mg/cm^3 phantom to standardize all liver measurements as this phantom had the least percentage error in its measure (data not shown).

Quantitative Measures

Protocol Development—We measured the Hounsfield Units (HU) of the liver, spleen and paraspinal muscles as well as an external phantom control. In order to determine the optimal number of CT slices to interpret, we measured 2 separate areas over an area of 100 mm squared in the liver, intentionally avoiding blood vessels in the liver. We also measured two separate areas in the spleen and one area each in the paraspinal muscles, avoiding fat planes. We conducted the measures in two abdominal and two chest CT slices per individual, in a total of 10 individuals. In order to determine whether to use the chest or abdominal scans for fatty liver measurement, we measured 6 separate areas in liver, three in spleen and one each of the paraspinal muscles and determined whether the variation in these measurements was less for the chest or abdominal scans. In order to determine the most parsimonious number of measures necessary in the liver, spleen, and paraspinal muscles, we compared three versus six measured areas in the liver, and two versus three measures in the spleen.

Our final protocol used three measures of at least 100 mm squared in the liver, two in the spleen, one in the left and one in the right paraspinal muscles and one in an external phantom. Two independent observers (EKS and MCF) analyzed the same set of computed tomograms independent of each other and blinded to participant characteristics. One observer repeated reading the scans two weeks after the initial period of reading (EKS) to determine intra-reader correlations.

Statistical evaluation—In order to capture the majority of variation in fatty liver we first determined whether there was more variation in fatty liver within or between axial CT slices. We calculated the mean difference and standard deviation (SD) in HU of two measures in the liver, two in the spleen and two in the paraspinal muscles between two different axial CT slices (inter-slice). We also calculated the mean difference and SD between two measures in the liver, spleen, and paraspinal muscles within a single axial CT

slice (intra-slice). In order to capture the majority of variation in the liver, from the inter-slice and intra-slice measurements we chose the one with the greatest variation. To determine whether measuring fat content in liver in chest versus abdominal scans was more precise, we determined the mean and SD of liver, spleen, and paraspinal muscle measures within individuals in individuals that had both a chest and abdominal scan and chose the one with the lower SD within individuals. To determine the inter-individual variation, the SD of the mean of the intra-individual measurement was determined. In order to determine the minimum number of measurements of the liver, spleen, and muscle necessary to precisely measure these organs, we calculated an intra-class correlation coefficient for one versus three, one versus six, and three versus six measures of attenuation in the liver, one versus two, one versus three, and two versus three measures in the spleen. The intra-class correlation coefficients were made by determining the mean square (MS) of the variance between individuals (MS between) minus the mean square of the variation within individuals (MS within) divided by MS between plus MS within for each variable noted. A value of 1 indicates perfect correlation and anything above 0.7 is excellent correlation. Liver-to-spleen (LSR), liver-to-muscle (LMR), and liver-to-phantom (LPR) ratios were calculated by taking the mean of the liver measures and dividing by the mean of the spleen measures, the mean of the muscle measures, or the single phantom measures, respectively, per individual. Comparisons of the LSR, LMR, and LPR were made within and between observers to determine the intra- and inter-reader class correlation coefficients. All analyses and plots of intra and inter-class correlation coefficients were generated in SAS.9.1.

Results

Among the 100 individuals available for analysis, 96 had CT slices through the liver that could be used to measure liver fat content and 94 had anthropomorphic data available; the mean age was 59.4 years and 49% were women (Table 1. The mean body mass index was 27.7 kg/m^2, the mean weight was 79.3 kg, the mean height was 168.8 cm, and the mean waist circumference was 98.7 cm (Table 1).

Optimal number of CT slices

The mean intra-slice variation was larger than the inter-slice variation for liver (6.7 versus 3.9 HU in the chest and 3.1 versus 1.8 HU in the abdomen) and spleen (12.2 versus 7 HU in the chest and 3.5 versus 3 HU in the abdomen) and was about the same for paraspinal muscles (7.5 versus 7.6 HU in the chest and 7.1 versus 7.7 HU in the abdomen) (Table 2). These results suggest that using one slice adequately captures the variance in measurement in these organs.

Fatty liver assessment on chest versus abdominal scans

Intra-individual variation in the liver, spleen, and muscle measurements were smaller in the abdominal scans than in the chest scans (Table 3). Specifically, the liver, spleen, and paraspinal muscle mean HU variation (SD) were 1.8, 1.3, and 2.2 HU in the abdomen and 5.4, 5.2 and 6.2 HU in the chest, respectively. Further, the variation (SD) in the liver measures between individuals was approximately equal between the two types of scans (12.0 versus 10.6 HU) whereas the variation for the spleen and muscle between individuals in the chest scans was greater (8.7 and 11.0 HU, respectively) in the chest than in the abdomen (2.5 and 7.4 HU, respectively) (Table 3). These findings suggest that measurement of all organs is more precise in the abdominal scans, where the intra-individual variation is smallest. Further, since the inter-individual variation of the spleen and paraspinal muscles is smaller in the abdominal measures than the chest measures, it suggests that the abdominal measures rather than the chest measures of these organs are a better control for the penetrance of the scans.

Optimal number of organ samplings needed

To determine the minimum number of samplings needed to precisely measure the liver and spleen we calculated the intra-class correlation coefficient for one versus three, one versus six, and three versus six measures in the liver and one versus two, one versus three, and two versus three measures in the spleen. The intra-class correlation coefficient for all of these measures was excellent on the abdominal slices and all were above 0.98 but the intra-class correlation coefficient was 1 only when three versus six measures in the liver and two versus threw measures in the spleen were compared (Table 4). Similar findings were observed when using measures from the chest scans where three versus six and two versus three measures in the liver and spleen were better than the other combinations (Table 4).

Intra-class correlations

The LSR, LMR, and LPR were calculated by taking the mean of the liver measures and dividing by the mean of the spleen or muscle measurements or the phantom measurement in order to correct for scan penetrance differences. The inter-and intra-class correlation coefficients for the LSR and LPR ratios were outstanding at >0.98 (Table 5 and Figure 2). The inter- and intra-class correlations for the LMRs were good at 0.86 and 0.93, respectively, but were less than those for the LSRs and LPRs (Table 5 and Figure 2).

Discussion

In this study our findings are four-fold. Measurement of fatty liver at just one slice level captures as much if not more variation compared to measuring multiple slices. We found that the variation in abdominal scans was less than in chest scans, suggesting that the abdominal scan may be more precise for measuring fatty liver. We also determined that three measures in the liver and two in the spleen were as good as more measures in these organs. Finally, we found that measurement of fatty liver in the abdomen is reproducible.

Measuring one slice per person per scan was sufficient to capture the variation in liver, spleen, and paraspinal muscle measures. Fat in the liver appears to be rather homogenously deposited. These findings are supported by the work of other investigators, who have noted that measuring fatty liver at one level is very representative of measures of fat in the entire liver.[6] A recent study of liver and spleen attenuation by CT in 439 individuals found that there was a slight statistically significant difference between slices 12 mm above the T12/L1 region of the spine, but the average magnitude of this difference was 1.5 HU in the liver and 2.0 HU in the spleen.[6] This difference between slices is smaller than the difference within a slice (1.8 HU in the liver and 3.4 HU in the spleen).[6] Similarly, we found that there was a 3.1 HU versus 1.8 HU difference in liver and 3.5HU vs. 3.0HU in spleen within versus between slices. We extend these findings to the paraspinal muscle, in which there were no major differences in the inter- versus intra-slice difference. In our study we also examined the variation in these organs in the chest which also showed that the intra-slice variation of the liver and the spleen was greater intra than inter-slice variation. These data support the concept that fat deposition in the liver is relatively homogeneous and most of the variation in the measurement of attenuation in these organs can be captured by measuring it at just one slice.

We also found that intra-individual variation in measuring liver intensity was less in the abdomen than in the chest, suggesting that the abdominal scan may be the more precise modality in which to measure fatty liver. The slices in the chest are half the size (2.5-mm versus 5-mm) of the abdomen, which may contribute in part to the increased variability of the chest measurements. Furthermore, the slices of the liver and spleen on the chest CTs are at the top of these organs near the diaphragm where respiratory change in their position may

cause more averaging of their true Hounsfield measurement with surrounding fat than in the abdomen where respiratory variation has less of an effect. To our knowledge this difference in the variance of measuring fatty liver on chest and abdominal scans has not been formally documented elsewhere but is an important point to note. Our data suggests that data taken from chest scans may have more variance in the measurement of fatty liver compared to abdominal scans. Whether this will impact relations between metabolic risk factors and genetic variants in relation to fatty liver is uncertain.

We also found that three measures in the liver and two in the spleen were the most parsimonious number of measures that can be used to reproducibly measure fatty liver. We have optimized a protocol which others may want to utilize to quantitate fatty liver on CT scans.

We also established high reproducibility for the LSR, whereas the LMR was not as reproducible a measure of fatty liver. In other studies, muscle has been substituted for the spleen when the spleen is not available or it is inaccurate to use the spleen as a standard.[7] Some of this variability in muscle measures stems from the fat planes within the paraspinal muscles that could not always be avoided during sampling. Further concern for using muscle as a standard for the liver measurement stems from the correlation between intra-muscular fat deposition and body mass index.[8] Overall adiposity will result in both a lower intra-muscular HU and a lower fatty liver HU, which may result in a falsely elevated LMR. Thus, the use of LMR as a "standard" for a fat measure such as fatty liver may require further evaluation.

Fortunately, in our study we had an external phantom that can be used to control for the penetrance of the scans when measuring fatty liver. This was particularly useful in our study, where we have a phantom in all scans but measurable spleens in only half of our scans. Therefore, by using the phantom to standardize the fatty liver Hounsfield unit measure, we are able to make use of the majority of our CT scans. Furthermore, the reproducibility of the LPR is outstanding and indeed may be slightly better than the LSR. This is not entirely surprising, since the phantom is a fixed and calibrated standard whereas the spleen may have biological variation in its measure.

Strengths and limitations

Strengths of the current study include having a well-characterized cohort of individuals with a wealth of clinical covariates and an external phantom that can be used to control for scan penetrance. Limitations of the current study include that the participants are predominantly of European ancestry. The predominant European structure of the study reflects the population of Framingham in 1948 which is when the cohort began. Therefore, the results of this study may not apply to individuals of other ethnicities. Further, we are missing spleen measurements in the half of the scans Due to the targeted abdominal CT protocol that was done primarily to capture calcification of the abdominal aorta. Fortunately, the use of an external phantom allows us to control for scan penetrance in all scans. Finally, there is no established best way of quantifying fat in the liver. Liver biopsy is limited in that it samples only a part of the liver and is still an invasive procedure with risks to the patient that cannot be ethically justified in healthy individuals. Imaging alternatives such as ultrasound, computed tomography, and magnetic resonance spectroscopy have been used and correlated with liver biopsy fat content and serve as non invasive alternatives to measuring fat content in liver.

In a population-based sample unselected for adiposity-related traits or fatty liver, quantification of fatty liver by MDCT was feasible and reproducible within and between readers. We have also optimized a protocol for fatty liver measurement.

Acknowledgments

Literature Cited

[1]. Procopiou M, Philippe J. The metabolic syndrome and type 2 diabetes: epidemiological figures and country specificities. Cerebrovasc Dis 2005;20(Suppl 1):2–8. [PubMed: 16276079]

[2]. Harrison SA, Neuschwander-Tetri BA. Nonalcoholic fatty liver disease and nonalcoholic steatohepatitis. Clin Liver Dis 2004;8:861–79. ix. [PubMed: 15464659]

[3]. Zelber-Sagi S, Nitzan-Kaluski D, Halpern Z, Oren R. Prevalence of primary non-alcoholic fatty liver disease in a population-based study and its association with biochemical and anthropometric measures. Liver Int 2006;26:856–63. [PubMed: 16911469]

[4]. Iwasaki M, Takada Y, Hayashi M, et al. Noninvasive evaluation of graft steatosis in living donor liver transplantation. Transplantation 2004;78:1501–5. [PubMed: 15599315]

[5]. Dawber, T. The Framingham Study: The Epidemiology of Atherosclerotic Disease. Harvard University Press; Cambridge, MA: 1980.

[6]. Davidson LE, Kuk JL, Church TS, Ross R. Protocol for measurement of liver fat by computed tomography. J Appl Physiol 2006;100:864–8. [PubMed: 16293702]

[7]. Panicek DM, Giess CS, Schwartz LH. Qualitative assessment of liver for fatty infiltration on contrast-enhanced CT: is muscle a better standard of reference than spleen? J Comput Assist Tomogr 1997;21:699–705. [PubMed: 9294555]

[8]. Tarnopolsky MA, Rennie CD, Robertshaw HA, Fedak-Tarnopolsky SN, Devries MC, Hamadeh MJ. Influence of endurance exercise training and sex on intramyocellular lipid and mitochondrial ultrastructure, substrate use, and mitochondrial enzyme activity. American journal of physiology 2007;292:R1271–8. [PubMed: 17095651]
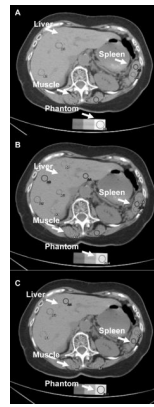
Figure 1. Measurement of Fatty Liver
A. Two measures (circles) in the liver, two in the spleen, two in the paraspinal muscles were taken. B. Six measures (circles) from the liver, three from the spleen, two from the paraspinal muscles, and one from the phantom were taken. C. Three measures (circles) from the liver, three from the spleen, two from the paraspinal muscles, and one from the phantom were taken.
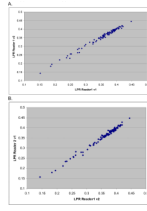
Figure 2. Inter and Intra reader LPR measurements
LPR measures within one reader (A) and between two different readers (B) are plotted. The intra-class correlation coefficient for both plots is 0.99.

Table 1

Characteristics of the participants.

Variable	Units	N	Mean	SD	Min	25th Pctl	50th Pctl	75th Pctl	Max
AGE	(yrs.)	94	59	13	37	48	60	71	83
WEIGHT	(kg)	94	79.3	16.4	48.5	67.1	79.6	90.7	137.0
HEIGHT	(cm)	94	168.8	9.2	150.5	160.0	168.3	175.9	190.5
BMI	(kg/m2)	94	27.7	4.6	18.2	24.8	27.3	30.4	41.6
WAIST	(cm)	94	98.7	12.5	71.8	90.2	99.1	106.0	132.1

Abbreviations: N= number of individuals; SD- standard deviation of mean; Min- minimum value; Pctl- percentile; Max- maximum value; BMI- body mass index; Waist- waist circumference.

Table 2

Intra-slice variation as compared to inter-slice variation

		Liver		Spleen		Muscle	
		Mean Δ (HU)	SD of mean Δ (HU)	Mean Δ (HU)	SD of mean Δ (HU)	Mean Δ (HU)	SD of mean Δ (HU)
Within slice *	Chest	6.7	3.8	12.2	11.6	7.5	4.9
Between slices **	Chest	3.9	3.3	7	3.6	7.6	4.9
Within slice *	Abdomen	3.1	2.3	3.5	1.6	7.1	4.8
Between slices **	Abdomen	1.8	1.2	3	1.6	7.7	4.9

*
The Hounsfield units from two measures in the liver (left and right lobe), spleen or paraspinal muscles within one slice were subtracted from each other to calculate the within one slice differences (one for each of two separate slices) and the two measures averaged to get the mean delta and SD for the within slice categories.

**
The Hounsfield units from two measures in the liver, spleen, or paraspinal muscles were subtracted from parallel measures in an adjacent slice to obtain the between slices differences. These two between slices measures were averaged to obtain the mean delta and SD for the between slices categories.

Table 3

Measurement variation in the abdomen as compared to the chest.

Scan	Tissue	N	Mean (HU)	Mean SD within an individual (HU)	SD of mean values between individuals (HU)
Chest	Liver	89	61.5	5.4	12.0
Chest	Spleen	70	45.3	5.2	8.7
Chest	Muscle	89	48.7	6.2	11.0
Abdomen	Liver	96	64.6	1.8	10.6
Abdomen	Spleen	55	55.6	1.3	2.5
Abdomen	Muscle	96	52.1	2.2	7.4

Six measures in the liver, three in the spleen, and two in the paraspinal muscles were measured in N individuals. The mean of these measures per individual were calculated and its SD presented above as the SD within an individual. The mean of these means is shown above as the mean and its SD as the SD between individuals.

Table 4

The intra class correlation of different numbers of measures in the liver and spleen.

	N	Liver (LPR)			N	Spleen (LSR)		
		1 vs. 3	1 vs. 6	3 vs. 6		1 vs. 2	1 vs. 3	2 vs. 3
ICCC chest	89	0.86	0.83	0.98	70	0.93	0.79	0.9
ICCC abdomen	91	0.99	0.98	1	55	0.99	0.99	1

Different numbers of measures in the liver as noted in table were used with one phantom measurement to create LPRs. Six measures from the liver were used with different numbers of spleen measures as noted in the table to create LSRs. LPR- liver/phantom ratio; LSR- liver/spleen ratio.

Table 5

Intra and inter observer intra class correlation coefficients (ICCCs).

	N	Inter-observer ICCC	Intra-observer ICCC
LSR	96	.99	.98
LPR	96	.99	.99
LMR	53	.86	.93

The intra- and inter-class correlation coefficient were calculated for LSR, LPR, and LMR in N individuals using three liver, two spleen and one phantom measures from abdominal scans and shown above.