Noninvasive detection of hepatic lipidosis in dairy cows with calibrated ultrasonographic image analysis

A. Starke,*1,2 A. Haudum,*2 G. Weijers,†2 K. Herzog,* P. Wohlsein,‡ M. Beyerbach,§ C. L. de Korte,† J. M. Thijssen,† and J. Rehage*
*Clinic for Cattle, University of Veterinary Medicine, Bischofsholer Damm 15, D-30173, Hannover, Germany
†Clinical Physics Laboratory, Department of Pediatrics, Radboud University, Nijmegen Medical Center, PO Box 9101, 6500HB, Nijmegen, the Netherlands
‡Department of Pathology, University of Veterinary Medicine, Buenteweg 17, D-30559, Hannover, Germany
§Institute for Biometry, Epidemiology and Information Processing, University of Veterinary Medicine, Buenteweg 2, D-30559, Hannover, Germany

ABSTRACT

The aim was to test the accuracy of calibrated digital analysis of ultrasonographic hepatic images for diagnosing fatty liver in dairy cows. Digital analysis was performed by means of a novel method, computer-aided ultrasound diagnosis (CAUS), previously published by the authors. This method implies a set of pre- and postprocessing steps to normalize and correct the transcutaneous ultrasonographic images. Transcutaneous hepatic ultrasonography was performed before surgical correction on 151 German Holstein dairy cows (mean ± standard error of the means; body weight: 571 ± 7 kg; age: 4.9 ± 0.2 yr; DIM: 35 ± 5) with left-sided abomasal displacement. Concentration of triacylglycerol (TAG) was biochemically determined in liver samples collected via biopsy and values were considered the gold standard to which ultrasound estimates were compared. According to histopathologic examination of biopsies, none of the cows suffered from hepatic disorders other than hepatic lipidosis. Hepatic TAG concentrations ranged from 4.6 to 292.4 mg/g of liver fresh weight (FW). High correlations were found between the hepatic TAG and mean echo level (r = 0.59) and residual attenuation (ResAtt; r = 0.80) obtained in ultrasonographic imaging. High correlation existed between ResAtt and mean echo level (r = 0.76). The 151 studied cows were split randomly into a training set of 76 cows and a test set of 75 cows. Based on the data from the training set, ResAtt was statistically selected by means of stepwise multiple regression analysis for hepatic TAG prediction (R² = 0.69). Then, using the predicted TAG data of the test set, receiver operating characteristic analysis was performed to summarize the accuracy and predictive potential of the differentiation between various measured hepatic TAG values, based on TAG predicted from the regression formula. The area under the curve values of the receiver operating characteristic based on the regression equation were 0.94 (<50 vs. ≥50 mg of TAG/g of FW), 0.83 (<100 vs. ≥100 mg of TAG/g of FW), and 0.97 (<50 vs. ≥100 mg of TAG/g of FW). The CAUS methodology and software for digitally analyzing liver ultrasonographic images is considered feasible for noninvasive screening of fatty liver in dairy herd health programs. Using the single parameter linear regression equation might be ideal for practical applications.

Key words: digital ultrasonographic analysis, triacylglycerol, fatty liver, computer-aided ultrasound diagnosis

INTRODUCTION

The most common liver disorder in high-yielding dairy cows is the fatty infiltration of hepatic tissue during early lactation (Morrow, 1976; Drackley, 1999; Bobe et al., 2004). In hepatic tissue, accumulated fat is mainly triacylglycerol (TAG; Grummer, 1993; Drackley, 1999). Up to 65% of dairy cows are affected by a moderate [TAG 50–100 mg/g of liver fresh weight (FW)] or severe (TAG ≥100 mg/g of FW) fatty liver during early lactation (Jorritsma et al., 2001; Raoofi et al., 2001; Bobe et al., 2004). The economic relevance of fatty liver results from reduced milk yield and fertility and increased risks of other periparturient diseases and early culling in affected animals (Jorritsma et al., 2000; Bobe et al., 2004; Mulligan and Doherty, 2008).

The accurate assessment of hepatic fat content has been restricted up until now to biochemical (Atkinson et al., 1972; Hara and Radin, 1978; Bickhardt et al., 1988) or histological (Collins and Reid, 1980; Collins et al., 1985; Rehage et al., 1996) examination of liver biopsy specimens (Staufenbiel et al., 1993; Cebra et al., 1997; Kalaitzakis et al., 2007). For on-farm testing, liver biopsies are impractical because of the time needed for acquiring and analyzing them. There is also
a minor but always inherent danger of infection and hemorrhages because of the biopsy procedure (Smart and Northcote, 1985; Smith et al., 1997). According to studies in human medicine, it became evident that the biopsy procedure is painful, even when using local anesthesia, and pain may last for as long as 24 h (Eisenberg et al., 2003). A suitable noninvasive method might be transcutaneous ultrasonographic (US) imaging of the liver (Bobe et al., 1999).

The normal echo pattern of the liver is caused mainly by a series of alternate collagen-water interfaces. With increasing fat accumulation in hepatic tissue, an increased gray level (grl) in the US image is seen and the liver appears subjectively brighter (Joseph et al., 1979). However, the grl of US images are dependent not only on tissue structure but also on the actual characteristics and settings of the ultrasound scanner. For this reason, the equipment settings as well as the examination technique need to be standardized. Moreover, visual inspection of the ultrasound images by the examiner introduces a subjective factor into the assessment of image grl. Thus, the visual evaluation of the brightness of hepatic tissue in US images can give only a rough estimate of the liver fat content. Furthermore, the high inter- and intraobserver variability inherent in subjective methods prohibits a high level of reliability of the results (Strauss et al., 2007).

To avoid the limitations of subjective analysis of US images, quantitative methods were developed. Two quantitative approaches have been described: one is based on spectral analysis of the raw data (i.e., the radio frequency echograms; Thijssen, 2000; Lizzi et al., 2006) and the other approach is based on the analysis of brightness (B)-mode image statistics and texture characteristics (Knipp et al., 1997; Thijssen et al., 2002). Garra et al. (1987) found that analyzing radio frequency-derived B-mode data and spectral analysis of radio frequency data yields similar results. Thus, analyzing conventional B-mode images might be useful for improving clinical diagnosis.

Acorda et al. (1994) and Bobe et al. (2008) demonstrated that quantitative digital analysis of transcutaneously obtained US liver images improves accuracy of detection of bovine hepatic lipidosis. However, in both studies the various influences and artifacts of the equipment and of the investigated tissues on image characteristics were not considered. For instance, subcutaneous fat in the abdominal wall has a substantial effect on quality and characteristics of US images (Zamir et al., 2003; Thijssen et al., 2008). Recently, a computer program (computer-aided ultrasound diagnosis; CAUS) for quantitative analysis of B-mode images of the bovine liver considering careful calibration of the equipment and correction of images for artifacts was developed by the authors of this study. The feasibility of diagnosing fatty liver by means of CAUS was already shown in a relatively small group of dairy cows (n = 12; Thijssen et al., 2008). In the present paper, the approach of analyzing transcutaneous B-mode data by means of CAUS is applied (Thijssen et al., 2008) with some new aspects in the procedure to investigate the potential of the devised method for a noninvasive, practical means of estimating the liver fat content in dairy cows.

**MATERIAL AND METHODS**

**Animals**

The study was conducted under the guidelines of the Research Animal Act (research permit number 33.42502/06A372) of the Lower Saxony Federal State Office for Consumer Protection and Food Safety (Oldenburg, Germany).

In this study, from November 2006 to August 2007, 151 lactating German Holstein dairy cows (mean ± SEM; BW: 571 ± 7 kg; age: 4.9 ± 0.2 yr; DIM: 35 ± 5; DIM range: 2–350) were enrolled. All cows had left-sided abomasal displacement (LDA) and were referred to the Clinic for Cattle, University of Veterinary Medicine Hannover (Hannover, Germany) for surgical correction by right flank laparotomy and omentopexy according to Dirksen (1967). According to the results of the clinical examination (Stöber, 1990), in all studied animals no other apparent diseases apart from the LDA were found. For each animal, BW (weighing) and withers height (cattle measuring stick) were determined.

**Liver Biopsy**

A liver sample (approximately 1 g) was taken from the caudal lobe during right flank laparotomy for surgical correction of LDA. Laparotomy was performed on the standing, nonsedated animal under local anesthesia using paralumbar nerve blocks and infiltration of the incision line with about 160 mL of Isocaine 2% (Procainhydrochloride, Selectavet, Weyarn-Holzolling, Germany). One hour before surgery, all cows were given Finadyne RP (2.2 mg of flunixin meglumine/kg of BW; Essex Tierarznei, Munich, Germany) as an analgesic drug. Biopsies were stored at −85 °C until further analysis. A subsample of the biopsy was fixed in 10% formalin for histopathologic examination.

**Biochemical Analysis.** The TAG contents in liver tissue were determined according to Starke et al. (2010). Results are expressed as milligrams per gram of FW.

**Histopathologic Examination.** The formalin-fixed liver samples were embedded in paraffin according to
standard laboratory procedures. Sections (5 μm) were cut and stained with hematoxylin-eosin. In addition, frozen sections of formalin-fixed specimens were stained with sudan III for fat droplets. Histologic examination of sections was performed by light microscopy (BX60, Olympus, Tokyo, Japan; objective: magnification 4× to 40×) to evaluate the fat distribution (Mertens, 1992) and to exclude hepatic diseases other than hepatic lipodosis that might have influenced US image characteristics.

**Ultrasonography**

**Equipment.** The US imaging was performed with a Power Vision 6000 scanner (Toshiba Inc., Tokyo, Japan) using a convex array transducer of 2 to 7 MHz (PVM–375AT). To enable unambiguous quantitative analysis of the images, certain equipment settings were fixed throughout the entire study: transmission frequency, 4.2 MHz; transmit focus, 6 cm; image depth range, 11.8 cm; time gain compensation, all sliders in neutral position; overall gain, 80 decibel (dB); dynamic range (compression), 70 dB; postprocessing curve, 0. The images were digitally acquired on magneto-optical disk and transferred to a personal computer for further analysis.

**Examination.** The transcutaneous ultrasound examination was carried out on the nonsedated, standing animal before surgery of LDA. The skin was shaved, washed, degreased (alcohol, 70%), and covered with ultrasound contact gel (Waldeck GmbH & Co. KG, Muenster, Germany). The scanned area was on the right abdominal wall in the seventh to twelfth intercostal space (boundary: dorsal, Longissimus dorsi; ventral, elbow joint level; Braun, 1990). The liver was examined to establish the general quality of the images and screened for the presence of morphologic changes (e.g., abscess, tumor; Braun, 1990). None of the cows used in this study suffered from hepatic disorders other than hepatic lipodosis. The transducer was gently pressed against the skin to obtain a homogeneous echo level for the right lobe of the liver. Five independent images were acquired, avoiding any larger hepatic vessels being depicted in the images. Back fat thickness was measured with ultrasound (Power Vision 6000, Toshiba Inc.; linear transducer of 3–13 MHz, PLM – 805AT) according to Staufenbiel (1997). The mean grl of US images is depth-dependent. This phenomenon is related to the beam forming (diffraction and focusing) as well as to the attenuation within the imaged tissue. The depth dependence has to be calibrated by reference measurements. Because a tissue mimicking phantom does not contain skin and subcutaneous tissue layers, calibration was performed with the images obtained from a group of normal cows. The speckle size is tissue-specific and increases with increasing frequency characteristics (center frequency, bandwidth) of the received echoes. The speckle size is related to the effective number density of the scatterers within the tissue (Thijssen et al., 2007) as well as to the frequency characteristics (center frequency, bandwidth) of the received echoes. The attenuation within tissues is tissue-specific and increases with increasing frequency, the speckle size is also dependent on the analyzed tissue itself (Thijssen et al., 2007).

The mean grl of US images is depth-dependent. The methods are described in detail elsewhere (Thijssen et al., 2008). Some new aspects are introduced in this paper that are related to a procedure for automatic detection and removal of hepatic vessels from the images before the analysis. A brief summary of the various (pre)processing steps and methods of image analysis is given here; more details are presented in Appendix II.

Generally, the user is able to choose the setting of the grl look-up table (LUT) of the ultrasound equipment. The LUT decodes the stored echo levels into grl in the image (mostly called the postprocessing of the equipment). A logarithmic compression amplifier is present in the equipment and, therefore, a linear LUT preserves the logarithmic relation between image grl and relative echo level (expressed in dB). The authors proposed either to use a linear LUT or to retrospectively measure and linearize the LUT (Valckx et al., 2000; Thijssen et al., 2007). This latter approach is presented here.

A further step is to estimate the contrast sensitivity, called gamma of the ultrasound equipment (i.e., the number of image grl per dB of echo strength). The procedure for estimating the gamma by using a tissue mimicking phantom with calibrated contrast objects was described in Thijssen et al. (2007). In the equipment used for this study, the gamma was found to be 3.7 grl/dB.

An important aspect of the approach introduced by Valckx et al. (2000) is based on the observation by Oosterveld et al. (1985) that the lateral size of the speckles in US images of scattering and attenuating media increases linearly with depth. Moreover, in sector images, the speckles are curved because of the sector-wise imaging process. The so-called back-scan conversion (Valckx et al., 2000) of the sector image (Figure 1A) produces a rectangular image in which the image texture [i.e., lateral speckle size (LatSp) and shape] is practically independent of depth (Figure 1B). The speckle size is related to the effective number density of the scatterers within the tissue (Thijssen et al., 2007) as well as to the frequency characteristics (center frequency, bandwidth) of the received echoes. Because the attenuation within tissues is tissue-specific and increases with increasing frequency, the speckle size is also dependent on the analyzed tissue itself (Thijssen et al., 2007).

The mean echo level (MeanEcho) curves (in dB) versus depth of the transcutaneous images of the normal cows were subtracted from those obtained intraoperatively.
in a previous study (Thijssen et al., 2008). The mean dB difference was divided by the local overall thickness of the skin plus muscle and subcutaneous fat layer. This yielded the mean attenuation per centimeter of tissue (dB/cm) of these layers (Thijssen et al., 2008). This procedure is called automatic gain control (Figure 1C).

**Postprocessing (Individual Corrections Based on Image Characteristics).** In cattle, the hairy skin occasionally induces some shadowing in the images, even after careful shaving of the transducer contact zone. In this paper, the concept of lateral gain correction is used to automatically compensate for shadowing (Thijssen et al., 2008). This correction procedure was applied to the digitally acquired image data and no proprietary software was used for this processing step.

A final, innovative step in the postprocessing of the images in this study is the estimation of the residual attenuation (ResAtt) coefficient (Thijssen et al., 2008) of diseased (fatty) liver as compared with healthy liver from 12 control cows. The histology revealed in the control cows no hepatic pathology and fat infiltration and the TAG content in the liver was below 30 mg/g of FW. This coefficient is then used both for performing a secondary correction of the depth dependence of the MeanEcho and as a diagnostic feature. After correction, the MeanEcho is depth independent and can be calculated for the whole region of interest (ROI). Its standard deviation is then used to automatically detect and remove hepatic vessels from the images in an iterative procedure (Figure 1D).

**Presentation of Results**

The patient data, biopsy results, and transcutaneously obtained US parameters were statistically evaluated (ANOVA and correlations) for all of the 151 cows included in the study (Tables 1 and 2). The study

**Figure 1.** Brightness-mode transcutaneous ultrasonographic liver image (right abdominal wall in the eleventh intercostal space; Power Vision 6000 scanner, Toshiba Inc., Tokyo, Japan; convex array transducer of 4.2 MHz, PVM-375AT) from a 5-yr-old German Holstein dairy cow. A) Original transcutaneous bovine liver ultrasonographic image. B) Ultrasonographic image after back scan conversion. C) Ultrasonographic image after preprocessing. D) Ultrasonographic image after postprocessing and showing the removed hepatic vessels.
population (n = 151) was divided randomly into 2 comparable subpopulations: a training population (training set) and a test population (test set). For this purpose, the animals (n = 151) were listed in the order of their enrollment (study) number. The even numbers were allocated to the training set (n = 76) and the odd numbers to the test set (n = 75). No significant differences were found in any of the studied parameters when comparing the 2 groups (Table 3). Data from the animals of the training set (n = 76) were used to generate a (multiple) regression equation for the subsequent calculation of TAG from the US parameters. The regression formula was then applied to the data of the cows from the test set (n = 75) to verify the reliability of predicting TAG by receiver operating characteristic (ROC) analysis.

Statistics

All statistical analysis was done with the SAS system for Windows (version 9.1, SAS Institute Inc., Cary, NC). In each animal (n = 151) a set of 5 images was acquired. All of these images were analyzed and the mean and SD of the grl statistics, ResAtt, and texture parameters of the 5 images for each experimental condition of each animal was calculated. The data of the image with the largest deviation of the ResAtt coefficient from the mean of the 5 images per animal were deleted and all parameters were recalculated with the 4 remaining images. Biopsy-measured hepatic TAG \((\text{TAG}_{\text{meas}})\) was used as gold standard for evaluating US parameters.

Evaluation of the Entire Population. The patient population (n = 151) was divided into 4 TAG classes (Table 1), modified according to Bobe et al. (2004). The mean and standard error of the means of patient data and US parameters were calculated for each of the TAG classes separately (PROC MEANS of SAS). When the \(F\)-test in the GLM procedure of SAS was significant \((P < 0.05)\), the LSMEANS statement (PDIFF/TDIFF option of SAS) was used to test for significant group differences. The Pearson correlation coefficient (PROC CORR of SAS) was calculated among patient data, between patient data and CAUS parameters, and among CAUS parameters.

Evaluation of Training Set and Test Set. Group differences of studied parameters between the training and test set were tested for statistical significance by \(t\)-test (PROC TTEST of SAS). A multiple linear regression analysis (PROC REG of SAS) using the stepwise procedure (entry level: \(P < 0.05\)) was performed on the training set of 76 cows and the coefficient of determination was calculated to investigate the relation of the CAUS parameters with hepatic TAG content. The obtained regression equation was then used to predict hepatic TAG \((\text{TAG}_{\text{pred}})\) for each patient of the test set (n = 75) and, according to \(\text{TAG}_{\text{pred}}\), cows were assigned to one of the liver TAG classes. Using \(\text{TAG}_{\text{meas}}\) as gold standard, ROC curves (Greiner et al., 2000; Table 1. Division of total group of German Holstein dairy cows (n = 151) into subgroups of various hepatic triacylglycerol (TAG) ranges (modified from Bobe et al., 2004) and descriptive statistics of patient data and hepatic ultrasonography (mean)

<table>
<thead>
<tr>
<th>Item (^1)</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>Average</th>
<th>SEM</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hepatic fat content</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\text{TAG}^2) (mg/g of FW)</td>
<td>16.3</td>
<td>39.3</td>
<td>78.5</td>
<td>159.0</td>
<td>2.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patient data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (mo)</td>
<td>58.4</td>
<td>55.5</td>
<td>57.8</td>
<td>59.8</td>
<td>4.4</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>BW (kg)</td>
<td>547</td>
<td>536</td>
<td>581</td>
<td>583</td>
<td>14</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>DIM</td>
<td>103.0</td>
<td>24.6</td>
<td>16.6</td>
<td>14.7</td>
<td>8.5</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Withers height (cm)</td>
<td>144</td>
<td>145</td>
<td>144</td>
<td>146</td>
<td>1</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>Back fat thickness (mm)</td>
<td>13.3</td>
<td>14.7</td>
<td>18.4</td>
<td>20.9</td>
<td>1.0</td>
<td>NS</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Transcutaneous ultrasonography</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abdominal wall thickness (mm)</td>
<td>22.1</td>
<td>22.1</td>
<td>25.6</td>
<td>27.4</td>
<td>0.9</td>
<td>NS</td>
<td>0.003</td>
</tr>
<tr>
<td>MeanEcho (dB)</td>
<td>9.1</td>
<td>9.2</td>
<td>12.8</td>
<td>15.3</td>
<td>0.7</td>
<td>NS</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>SDEcho (dB)</td>
<td>3.58</td>
<td>3.29</td>
<td>3.26</td>
<td>3.12</td>
<td>0.19</td>
<td>NS</td>
<td>0.004</td>
</tr>
<tr>
<td>MeanEcho/SDEcho</td>
<td>2.52</td>
<td>2.75</td>
<td>3.97</td>
<td>4.96</td>
<td>0.22</td>
<td>NS</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Residual attenuation (dB/cm)</td>
<td>0.09</td>
<td>0.51</td>
<td>0.97</td>
<td>1.48</td>
<td>0.07</td>
<td>0.005</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Intercept (dB)</td>
<td>24.7</td>
<td>25.6</td>
<td>30.7</td>
<td>34.8</td>
<td>0.9</td>
<td>NS</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Axial speckle size (mm)</td>
<td>0.57</td>
<td>0.56</td>
<td>0.54</td>
<td>0.52</td>
<td>0.03</td>
<td>NS</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Lateral speckle size (mm)</td>
<td>2.59</td>
<td>2.51</td>
<td>2.43</td>
<td>2.47</td>
<td>0.03</td>
<td>NS</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

\(^1\)MeanEcho = mean echo level; SDEcho = standard deviation of the echo levels; MeanEcho/SDEcho = mean echo level over the standard deviation of echo levels; dB = decibel.

\(^2\)Subgroups: A = no fat (<30 mg of TAG/g of fresh weight (FW); n = 32; 21%); B = mild fat (≥30 to 50 mg of TAG/g of FW; n = 13; 9%); C = moderate fat (≥50 to 100 mg of TAG/g of FW; n = 39; 26%); D = severe fat (≥100 mg of TAG/g of FW; n = 67; 44%).

\(^3\)No P-value statistic because subgroups classified after TAG.
Fawcett, 2006) were used to evaluate the accuracy of classifying patients by TAGpred. The true positive rate [sensitivity = true positive × (true positive + false negative)] was plotted versus the false positive rate [1 – specificity = false positive × (false positive + true negative)] by means of Sigmaplot 2001 (Systat Software Inc., Chicago, IL). The ROC curves were plotted for differentiation of hepatic TAGmeas (in mg/g of FW) <50 versus ≥50, <100 versus ≥100, and <50 versus ≥100 in steps of 10 mg/g of FW of TAGpred increments. As a measure of the quality of classification, the area under curve (AUC) of the ROC was estimated as a fraction of the total area (AUC = 1). This estimation was performed by counting the pixels of the area below the ROC and subsequently dividing this area by the total area of the graph. The overall sensitivity and the specificity of the classification were then estimated by using the cross section point of the ROC and the negative diagonal (i.e., the point indicating maximal sum of sensitivity and specificity) as a cut-off point. These latter statistical parameters were meant to enable a comparison with results from other studies, where an ROC can often not be estimated.

RESULTS AND DISCUSSION

Evaluation of the Entire Population

Liver Biopsies—Fat Concentration and Histopathology. In the present study, in hepatic biopsies of 151 cows with LDA a wide range for hepatic TAG concentration (4.6–292.4 mg/g of FW; Table 1) was found. Using the classification modified according to Bobe et al. (2004), about 21% of the studied cows showed no fatty liver (TAG <30 mg/g of FW; n = 32), about 9% showed mild (TAG ≥30–50 mg/g of FW; n = 13), about 26% showed moderate (TAG ≥50–100 mg/g of FW; n = 39), and about 44% showed severe fatty liver (TAG ≥100 mg/g of FW; n = 67; Table 1). This is a typical distribution pattern of different degrees of severity of hepatic lipidosis in LDA cows (Holtenius and Niskanen, 1985; Muylle et al., 1990; Rehage et al., 1996). However, it is also generally comparable to that seen in early-lactating dairy cows (Gerloff and Herdt, 1984; Herdt, 1988; Jorritsma et al., 2001).

The histopathologic examination of liver samples revealed a diffuse fat distribution in 127 patients. In accordance with Kalaitzakis et al. (2007), the lobular distribution of fat vacuoles ranged from few, small, scattered centrilobular fat droplets to panlobular fat accumulation. In 24 cows no visible fat vacuoles were found in liver samples. None of the cows participating in this study suffered from hepatic disorders other than hepatic lipidosis. The histopathologic examination was
performed to ensure that the evaluation of the US liver images was not biased by liver diseases other than fatty liver.

**Correlations of TAG to Patient Data.** During early lactation, a negative energy balance causes enhanced lipomobilization and an accompanying increased fat storage in the liver by way of a hepatic TAG increase (Grummer, 1993; Drackley, 1999). This conformed with our results in Table 1 showing considerable fat accumulation in hepatic tissue to be almost exclusively seen in cows in the first 50 DIM. In contrast, all cows above 90 DIM indicated hepatic TAG contents of less than 50 mg/g of FW. Accordingly, a negative correlation between hepatic TAG and DIM (range of DIM: 2 to 350; \( r = -0.37 \); Table 2) was found. Age, BW, and withers height had no significant effect on the liver TAG content in our patients (Tables 1, 2).

Predominately, cows in good body condition are prone to fat mobilization, resulting in fat accumulation in hepatic tissue (Staufenbiel et al., 1993; Drackley, 1999; Bobe et al., 2004). This observation was also confirmed in our study by a weak positive correlation between hepatic TAG and ultrasonographically measured back fat thickness (\( r = 0.07 \)–0.44) related to the US parameters (Table 2). Medium and high correlations (\( r = 0.35 \)–0.69) were found for back fat and abdominal wall thickness with most of the US parameters (Table 2). It is generally observed in literature that the thickness of the abdominal wall considerably affects the characteristics of US images (Zamir et al., 2003; Mottin et al., 2004; de Moura Almeida et al., 2008). However, in the present study the attenuation attributable to the abdominal wall was compensated for. The residual effect might be a distortion of the ultrasound propagation, resulting in changes in the speckle pattern and tissue attenuation.

**Correlation of TAG Content to US Parameters.** The correlations of US parameters with hepatic TAG content (\( r = 0.46 \); Table 2). This underlines the importance of considering the correction of effects of the abdominal wall on transcutaneously obtained US liver images, as done by the CAUS program (Thijssen et al., 2008).

<table>
<thead>
<tr>
<th>Item</th>
<th>Training set</th>
<th>SEM</th>
<th>Test set</th>
<th>SEM</th>
<th>( \beta )-value, training set:test set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hepatic fat content</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TAG (mg/g of fresh weight)</td>
<td>96.4</td>
<td>8.1</td>
<td>98.9</td>
<td>7.3</td>
<td>NS</td>
</tr>
<tr>
<td>Patient data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (mo)</td>
<td>54.7</td>
<td>2.6</td>
<td>62.6</td>
<td>2.9</td>
<td>NS</td>
</tr>
<tr>
<td>BW (kg)</td>
<td>557</td>
<td>9</td>
<td>584</td>
<td>9</td>
<td>NS</td>
</tr>
<tr>
<td>DIM</td>
<td>33.4</td>
<td>6.8</td>
<td>36.0</td>
<td>7.9</td>
<td>NS</td>
</tr>
<tr>
<td>Withers height (cm)</td>
<td>144.2</td>
<td>0.5</td>
<td>145.5</td>
<td>0.4</td>
<td>NS</td>
</tr>
<tr>
<td>Back fat thickness (mm)</td>
<td>17.9</td>
<td>0.8</td>
<td>18.3</td>
<td>0.7</td>
<td>NS</td>
</tr>
<tr>
<td>Transcutaneous ultrasonography</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abdominal wall thickness (mm)</td>
<td>25.4</td>
<td>0.6</td>
<td>25.3</td>
<td>0.6</td>
<td>NS</td>
</tr>
<tr>
<td>MeanEcho (dB)</td>
<td>12.6</td>
<td>0.5</td>
<td>13.0</td>
<td>0.6</td>
<td>NS</td>
</tr>
<tr>
<td>SDEcho (dB)</td>
<td>3.26</td>
<td>0.04</td>
<td>3.28</td>
<td>0.04</td>
<td>NS</td>
</tr>
<tr>
<td>MeanEcho/SDEcho</td>
<td>3.96</td>
<td>0.18</td>
<td>4.03</td>
<td>0.19</td>
<td>NS</td>
</tr>
<tr>
<td>Residual attenuation (dB/cm)</td>
<td>0.97</td>
<td>0.08</td>
<td>0.97</td>
<td>0.08</td>
<td>NS</td>
</tr>
<tr>
<td>Intercept (dB)</td>
<td>30.6</td>
<td>0.7</td>
<td>31.0</td>
<td>0.8</td>
<td>NS</td>
</tr>
<tr>
<td>Axial speckle size (mm)</td>
<td>0.54</td>
<td>0.01</td>
<td>0.54</td>
<td>0.01</td>
<td>NS</td>
</tr>
<tr>
<td>Lateral speckle size (mm)</td>
<td>2.50</td>
<td>0.02</td>
<td>2.48</td>
<td>0.02</td>
<td>NS</td>
</tr>
</tbody>
</table>

\( ^{1}\text{ MeanEcho = mean echo level; SDEcho = standard deviation of the echo levels; MeanEcho/SDEcho = mean echo level over the standard deviation of echo levels; dB = decibel.} \)
of ResAtt of cows with TAG less than 30 mg/g of FW and those with TAG between 30 and 50 mg/g of FW (Table 1).

With increasing hepatic fat accumulation, the liver tissue also appears brighter in ultrasound images (Joseph et al., 1979). In this study, the MeanEcho increased significantly with higher hepatic TAG contents (Table 1). This was also shown by the significant positive correlation between the MeanEcho (r = 0.59; Table 2) and liver TAG. These results may be considered more conclusive than in literature because of the correction of the depth-dependent decrease in the mean grl in hepatic images by the computer program CAUS (Thijssen et al., 2008).

The correlation of hepatic TAG to axial speckle size (AxSp) was significantly negative (r = −0.60; Table 2). However, when comparing means of AxSp in cows of different hepatic fat classes (Table 1) the absolute differences were small. The negative correlation of the AxSp with TAG level implies an increase in the bandwidth of the received echoes because AxSp and bandwidth are inversely related (Thijssen, 2003). Because absorption of ultrasound only marginally influences the bandwidth and AxSp (Oosterveld et al., 1985), the negative correlations are explained by an enhancement of the backscattering level in fatty liver. Backscattering is higher the higher the frequency is, particularly if the scattering sites are small with respect to the wavelength. The increase may be quadratic with frequency. The fat droplets in the hepatocytes are between 1 and 10 μm (DiAugustine et al., 1973; Bradbury, 2006). These dimensions are small compared with the wavelength at 4.2 MHz in tissue (i.e., 360 μm). Therefore, the more droplets within a liver, the stronger the backscattered echoes are and the smaller the AxSp is.

The standard deviation of measured echo levels (SDEcho) and AxSp in ultrasound images correlated negatively with hepatic TAG (Table 2). From this outcome it is concluded that with increasing fat deposition in the liver the variability of echo levels in hepatic US images is substantially reduced and US images become more homogeneous and display a finer grain.

**Correlations Between US Parameters.** Most of the correlations between transcutaneous US parameters were statistically significant (Table 2). The ratio of the MeanEcho over the standard deviation of echo levels (MeanEcho/SDEcho) correlated highly with the MeanEcho because of inherent relatedness. Also, ResAtt showed high correlations (range: r = −0.58 to 0.85) with all but 1 of the other parameters. Interestingly, whereas the correlation with MeanEcho was positive (r = 0.76), the correlations with the SDEcho (r = −0.58) and with the AxSp (r = −0.67) and LatSp (r = −0.29) were negative after the correction of the depth-depen-

dent decrease in the mean grl in hepatic images by the computer program CAUS (Thijssen et al., 2008). This finding also indicates a brighter image on the one hand and a more homogeneous appearance and a finer grain of US images with increasing liver TAG on the other. These findings indicate a gradual shift of backscattering by small hepatic vessels toward a massive scattering by fat droplets, resulting echographically in a bright liver. It may be remarked that the SDEcho does not significantly correlate with the MeanEcho. This corresponds with the statistical character of logarithmically compressed echo data (Thijssen, 2003).

**Evaluation of Training Set and Test Set**

**Multiple Regression of US Parameters Versus TAG.** After splitting the total number of the 151 studied cows, the patient data and US parameters of the training set (n = 76) and the test set (n = 75) of cows did not differ significantly (Table 3).

Using the data of cows of the training set (n = 76), by means of step-wise multiple regression analysis using all US parameters, only ResAtt was selected by the software for the regression equation for predicting the hepatic TAG content. The equation of the linear regression for ultrasonographically predicted hepatic TAG content of the training set was

\[ \text{TAG}_{\text{pred}} = 17.2 + 81.8 \times \text{ResAtt}. \]

The correlations between US parameters (Table 2) may be considered indicative of the parameter selection to be expected in the step-wise approach. Parameters showing rather high correlations with ResAtt were not selected because they supplied little additional information to the regression. It became evident that the ResAtt coefficient was the most relevant parameter and, because of the high correlations, MeanEcho, SD of MeanEcho, and MeanEcho/SDEcho were excluded. The speckle sizes in axial and lateral directions, although changing with the TAG level, were excluded as well. This may be interpreted as a minor change in the granular appearance of the echograms as previously mentioned.

In Figure 2 the linear regression with 99% confidence interval between hepatic \(\text{TAG}_{\text{meas}}\) (gold standard) and the \(\text{TAG}_{\text{pred}}\) derived with the before-mentioned regression equation for the test set is depicted. The calculated coefficient of determination of 0.58 shows the high quality of ultrasonography in estimating the hepatic fat content. This indicates that in cows, ResAtt is the dominant diagnostic parameter when using mean grl versus depth analysis of liver US images in detecting hepatic lipidosis, as is the case in humans (Davies et
al., 1991; Gaitini et al., 2004; Szebeni et al., 2006). About 75% of predicted hepatic TAG values were found within a range of ±40 mg/g of FW of measured TAG, although in some cases the predicted TAG deviated substantially from measured hepatic TAG values. On average, by means of digital analysis of hepatic US images, $\text{TAG}_{\text{pred}}$ was slightly over- or underestimated by about 10 to 25 mg/g of FW when the hepatic $\text{TAG}_{\text{meas}}$ content was below or above a threshold of 90 mg/g of FW, respectively (Figure 2). At this stage we have no explanation for this finding. However, results indicate that there is still room to improve the digital analysis of hepatic US images.

Because it was reported that accumulated fat in the liver is homogeneously distributed over the entire organ (Gaál and Husveth, 1983), it appears to be justified to take the TAG content of small hepatic biopsies as representative of the whole liver and therefore as a gold standard. However, recently Mohamed et al. (2004) reported for the first time that an inhomogeneous fat distribution in hepatic tissue may occur in some cows. Thus, such an inhomogeneous hepatic fat distribution could also explain the deviation of predicted hepatic TAG values from measured values in this study.

**Diagnostic Accuracy.** Receiver operating characteristic analysis (Figure 3) was used to evaluate the diagnostic and predictive power of the digital analysis of hepatic US images for discriminating between mild (<50 mg/g of FW), moderate (50–100 mg/g of FW), or severe (>100 mg/g of FW; Bobe et al., 2004) hepatic lipidosis. The TAG content found in the liver biopsies served as gold standard. For each cow in the test set ($n = 75$) the $\text{TAG}_{\text{pred}}$ was calculated by means of the regression function obtained from the training set ($n = 76$) and, according to the results of calculated $\text{TAG}_{\text{pred}}$, cows were classified into TAG classes.

For discriminating mild and moderate from severe fatty liver as described in the Statistics section (threshold liver $\text{TAG}_{\text{meas}}$ 100 mg/g of FW), about 83% of the 75 cows of the test set were correctly classified (AUC: 0.83; Figure 3; Table 3). The optimal cut-off value (intercept point of negative diagonal with ROC curve) was
about 109 mg of TAG\textsubscript{pred}/g of FW, with a sensitivity and specificity of 0.79 and 0.80, respectively (Table 4). Better results were found for discrimination of mild from moderate and severe fatty liver (threshold liver TAG\textsubscript{meas} 50 mg/g of FW): 94% of cows in the test set were correctly classified (Figure 3; Table 3). The optimal cut-off value was about 72 mg of TAG\textsubscript{pred}/g of FW (sensitivity and specificity: 0.83 and 0.86, respectively; Table 3). The ROC analysis revealed almost optimal results for discrimination of mild (<50 mg of TAG\textsubscript{meas}/g of FW) from severe (>100 mg of TAG\textsubscript{meas}/g of FW) fatty liver (AUC: 0.97; Table 4). In this model, sensitivity and specificity were 0.90 and 0.88, respectively, at an optimal cut-off value of about 79 mg of TAG\textsubscript{pred}/g of FW. As already mentioned above, the obtained cut-off values show also that the digital analysis of liver US images by CAUS slightly over- and underestimated measured TAG contents in the low and the high TAG range, re-

**Figure 3.** Receiver operating characteristic (ROC) curves of the classification of cows in test set (n = 75) in hepatic triacylglycerol (TAG) groups [TAG <50 vs. ≥50 mg/g of fresh weight (FW), TAG <100 vs. ≥100 mg/g of FW, and TAG <50 vs. ≥100 mg/g of FW] by means of ultrasonographic prediction of TAG. Gold standard = biochemically measured TAG in liver biopsies. Ultrasonographic prediction of TAG using linear regression equation from training set (n = 76) = 17.2 + 81.8 × residual attenuation. Cut-off point = cross-section of ROC with negative diagonal. AUC = area under curve.
The current study indicated that digital analysis of B-mode ultrasonograms after correction for abdominal wall effects and removal of artifacts by means of CAUS has high potential to grade the degree of hepatic TAG infiltration and to estimate liver TAG content in cows. This implies that practical application in further prospective studies is warranted. Furthermore, the described methods illustrate the feasibility of screening cows for the occurrence of fatty liver for dairy herd health management. For the authors, the use of cows for the evaluation of the diagnostic accuracy of CAUS was a model study and they suggest that the technique of CAUS can also be used successfully for US diagnosis of nonalcoholic fatty liver in humans because the phenomena of bright liver and simultaneously enhanced echo level attenuation are also observed in human medical literature on echographic diagnosis of fatty liver.

ACKNOWLEDGMENTS

The authors thank the H. Wilhelm Schaumann Stiftung (Hamburg, Germany) for their generous grant and Frances C. Sherwood-Brock (Hannover, Germany) for revising the English manuscript.

REFERENCES


APPENDIX I: DEFINITION OF TERMS AND UNITS

**Mean echo level (MeanEcho):** Echo level (dB) averaged over ROI of each image and averaged over 4 images, selected out of 5, for each patient. Echo level is expressed in dB relative to the MeanEcho of a tissue mimicking phantom obtained with the same equipment and settings.

**Standard deviation of echo levels (SDEcho):** Standard deviation of echo levels (dB) in ROI of each image and averaged over 4 selected images per patient.

**Mean echo level over standard deviation of echo levels (MeanEcho/SDEcho):** Ratio of echo level mean over standard deviation of each image, and averaged over 4 selected images. This ratio is generally termed signal to noise ratio.

**Residual attenuation coefficient (ResAtt):** Slope of linear fit through average echo level versus depth curve (dB/cm) for ROI in each image after removal of hepatic vessels and correction with MeanEcho versus depth curve from 12 normal livers of cows and averaged over 4 images per animal.

**Intercept:** Cross section with vertical axis of linear fit through echo level versus depth curve for each image after correction with MeanEcho versus depth curve from 12 normal animals, and averaged over 4 images per animal. Expressed in dB.

**Axial speckle size (AxSp):** Average size (mm) in axial (depth) direction of speckles within ROI for each image and averaged over 4 images per animal. This size is defined as the full width at half maximum of the axial autocovariance function of the ROI (Oosterveld et al., 1985).

**Lateral speckle size (LatSp):** Average size (mm) in lateral direction of speckles within ROI for each image and averaged over 4 images per animal. This size is defined as the full width at half maximum of the lateral autocovariance function of the ROI (Oosterveld et al., 1985). For further details see Thijssen et al. (2008).

APPENDIX II: SCHEME OF PROCESSING OF TRANSCUTANEOUS US LIVER IMAGES

**Preprocessing**

1. Back-scan conversion and interpolation, yielding rectangular images. The image data are upsampled by a factor of 2, yielding a square pixel size of 0.1764 mm.

2. Linearizing the grl LUT: The first step in the preprocessing of all the images is to linearize the LUT and apply the already estimated gamma (3.7 grl/dB) to transfer the image grl to dB relative to the phantom images.

3. Automatic gain control: The next step is the application of the automatic gain control derived from the control group of normal cows.

4. Correction of abdominal wall attenuation: In each of the images to be analyzed, a contour is interactively drawn at the fat–liver boundary and all the vertical scan lines of the images are corrected one by one for the skin/fat layer attenuation using the estimated attenuation coefficient value (4.64 dB/cm).

5. Automatic selection of a ROI. The full width of the image is used, minus 25 image lines (i.e., rectangular columns of 25 pixels wide) at both sides, thus omitting the inadequate parts of images. The installed depth range is from 5 to 10 cm in the image frame. This means that the ROI starts at a depth always greater than the thickest subcutaneous layers present in the population of cows in this study.

**Postprocessing**

6. Estimation of and correction for ResAtt within each image. This is a first guesstimate, still disturbed by hepatic vessels, and it is used to start removing the depth dependence of echo levels.

7. Estimation of mean and SDEcho within ROI.

8. Automatic detection of hepatic vessels (Figure 1D). This processing is based on echo level thresholding of small image windows (4 × 16...
pixels; corresponding to 0.71 × 2.82 mm, axial × lateral size). A window is marked as “vessel” if more than 70% of pixels within the window exceed one of the thresholds [high or low score (i.e., mean lesser or greater than 1.5 SD, respectively)]. The window shape is then changed to a larger size, the axes of which are set at 12 and 32 pixels length (2.11 × 5.65 mm), thus still being adapted to the elliptic form of speckles. Marked windows are excluded from analysis (i.e., logical OR-operation applied to windows).

9. Removal of first attenuation correction. This step is a preparation for a second assessment of the ResAtt.

10. Estimation of and correction for ResAtt. This step is performed without taking into account the subregions that are detected as being hepatic vessels.

11. Estimation of lateral gain correction profile and correction, without taking the hepatic vessel regions into account.

12. Estimation of mean and standard deviation.

13. Automatic detection of hepatic vessels. The former vessel mask of the image is removed and a window detection procedure is performed for a second time.

14. Final estimation of ResAtt and correction. This value of the ResAtt is also used as a diagnostic parameter.

15. Final estimation of mean and SDEcho. The ratio of the mean and standard deviation (i.e., the signal to noise ratio) is taken as a diagnostic parameter as well.

16. Estimation of axial and LatSp. This calculation is done with a 2-dimensional algorithm, and the full width at half maximum in axial and lateral directions is determined.