

Reliability and Accuracy of Clinical Assessment and Digital Image Analysis for Steatosis Evaluation in Discarded Human Livers

Hamid Abudhaise ^{a,*}

hamid.abudhaise.11@ucl.ac.uk

Tu Vinh Luong ^b

Jennifer Watkins ^b

Barry J. Fuller ^a

Brian R. Davidson ^a

^a UCL Division of Surgery and Interventional Sciences, Royal Free Hospital, London, United Kingdom

^b The Department of Cellular Pathology, Royal Free London NHS Foundation Trust, London, United Kingdom

* Address correspondence to Hamid Abudhaise, UCL Division of Surgery and Interventional Sciences, Royal Free Hospital, London, United Kingdom. Tel: +447809477667.

This work should be attributed to UCL Division of Surgery and Interventional Sciences, Royal Free Hospital, London, United Kingdom.

Funding: This work was supported by a DoH-funded NIHR i4i collaboration (Grant No. HTD 556) with Organ Recovery Systems entitled "Improved utilization of marginal liver grafts by machine perfusion." Sponsors funded the consumable materials including histology processing costs. Sponsors were not involved in designing the study, collecting, analyzing or interpreting the data, or writing the manuscript.

Abstract

Background

Accurate assessment of steatosis in procured livers is crucial to reduce the poor outcome associated with high-grade steatosis and to optimize the utilization of donor grafts. Clinical examination and digital image analysis (DIA) have been used for steatosis evaluation, but the validity of these methods is debated. This study aimed to compare these methods with standard histology for assessment of steatosis severity in human livers and to evaluate a revised classification system for automated fat measurement.

Methods

Clinical assessment of liver steatosis at time of retrieval and automated measurement were compared with standard histology in paraffinized and hematoxylin and eosin-stained slides, using a 4-grade scale for ordinal data and percentages for numerical values.

Results

Analysis of 42 human livers that were retrieved but not transplanted showed that clinical examination was not reliable for assigning steatosis grades (κ_w , 0.12; 95% CI, -0.06 to 0.30), overestimated steatosis severity, and had an accuracy of 67% for discriminating low- and high-grade steatosis. Digital image analysis had a substantial agreement on absolute fat percentage (intraclass correlation coefficient, 0.76; 95% CI, 0.63–0.84) and steatosis grades (κ_w , 0.70; 95% CI, 0.57–0.82), with 88% accuracy using the revised classification system.

Conclusions

Clinical assessment of steatosis is inaccurate, and relying on this method alone could result in unnecessary discard of livers. Digital image analysis is feasible with higher accuracy and reliability, but further clinical studies are required to evaluate its clinical validity.

Background

Hepatic steatosis is found in up to 30% of deceased donor livers [1-3]. Accurate assessment of steatosis severity is crucial because of the significant association between high-grade steatosis and poor graft outcomes [4,5] and to optimize the utilization of donor grafts. The retrieving surgeon must visually assess steatosis of the donor organ in real time, but this method is highly subjective and inaccurate [6], and studies comparing surgical evaluation with microscopic examination are limited and controversial [7-9].

The current criterion standard for steatosis assessment is based on a semiquantitative visual estimation of hematoxylin and eosin (H&E)-stained parenchymal sections at low-to-medium power magnification by the pathologist [10]. Unfortunately, the reproducibility of this method has also been a matter of debate [10-16]. Alternative microscopic techniques such as digital image analysis (DIA) and stereological point counting are less commonly used despite showing higher reproducibility and accuracy [10,17,18]. One possible reason for the limited utilization of DIA is the measurement discrepancy between this method and visual estimation; fat percentages obtained by standard histology are 1.5- to 4-fold higher than results obtained by DIA [10,11,14,15,17,19,22], which requires standardization of DIA results to match the criterion standard [10,17,19].

This study aimed to compare surgical examination and DIA with the standard H&E assessment by the pathologist for evaluation of steatosis severity in human livers and to evaluate a revised classification system for automated fat measurement.

Materials and Methods

Ethical approval for this study was granted by the North London Regional Ethics Committee 3 (reference no. 10/H0709/70). A total of 42 livers were procured for transplant by different organ retrieval teams across the United Kingdom but found to be unsuitable for transplant were analyzed. Livers were refused because of a combination of donor and graft factors; 60% of the organs originated from donors after circulatory death and were associated with steatosis, fibrosis, poor liver function, and long ischemia times. Appropriate consent was obtained from the next of kin, and the livers were retrieved, flushed, and transported in preservation solution on ice as with standard United Kingdom practice for clinical use [23].

Biopsy Collection and Processing

On arrival at the research laboratory, a wedge biopsy (> 1.5 cm²) was obtained from liver segments 4 and 7 (2 specimens per liver). The number of specimens was based on recommendations that 2 biopsies from different liver segments are required for suspicious liver pathology because of intraliver pathology variation [24]. Biopsy specimens were fixed in formaldehyde and paraffinized. One 4- μ m section was obtained from each paraffinized block and stained with H&E before pathologist and DIA assessment (2 slides per liver).

Standard Histologic Assessment (the Criterion Standard)

Two experienced pathologists (TVL and JW) estimated steatosis severity in whole liver slides using light microscopy (Nikon Eclipse E600 microscope, Nikon UK Ltd, Kingston, United Kingdom) at 10 \times objective magnification.

This assessment yielded the estimated percentage of liver parenchyma occupied by fat, or estimated fat proportionate area (eFPA), which was expressed in percentages on a scale of 0% to 100% and graded semiquantitatively according to the commonly used nonalcoholic fatty liver disease activity score [25] as normal (none) < 5%, mild 5% to 33%, moderate > 33% to 66%, and severe > 66%. Slides from the 2 liver segments were compared for intersite variation within each liver. The 2 pathologists were blinded to each other and the intraoperative grading and DIA results.

Clinical Assessment and Grading System

During retrieval, surgeons routinely graded steatosis as none (0), mild (1), moderate (2), or severe (3) steatosis. Results were collected retrospectively from organ retrieval forms, which accompany all procured organs and record donor information, ischemia times, anatomic characteristics, preservation solution used, and steatosis severity.

Digital Image Analysis

The measured fat proportionate area (mFPA) was calculated using DIA in the same slides evaluated by the 2 pathologists. Techniques for image capturing, artifacts exclusion, and binary segmentation were adapted from Hall et al [10,14]. At 10× objective magnification, 3 random and nonoverlapping microscopic fields were selected from each slide and captured using Nikon DS-Fi1 camera (Nikon UK Ltd). Results from the 3 microscopic fields were averaged to obtain whole slide mFPA, which was expressed in percentages on a scale of 0% to 100%. Because of the known discrepancy between DIA-measured and pathologist-estimated fat percentage area, we used a revised mFPA classification system based on findings by Marsman [17] and Hall [14]. We hypothesized an additional cutoff point for normal (nonsteatotic) livers, which was not suggested by the latter researchers to conform to the 4-grade scale used by pathologists. Accordingly, steatosis grades corresponding to mFPA percentages were normal < 2%, mild 2% to 10%, moderate > 10% to 25%, and severe > 25%. Slides from the 2 liver segments were compared for intersite variation within each liver. The operator (HA) was blinded to the results of the other 2 assessment methods.

Data Analysis

Data were analyzed using SPSS Statistics for Macintosh, Version 20.0 (IBM, Armonk, NY, United States). Interrater reliability was assessed using intraclass correlation coefficient (ICC), Cohen's kappa (κ), and weighted kappa (κ_w) for continuous, binomial, and ordinal variables, respectively, and reliability was interpreted according to Landis and Koch [26]. Sensitivity, specificity, and overall accuracy were calculated as measures of diagnostic accuracy after dichotomizing the livers into low- and high-grade steatosis (\leq mild steatosis and \geq moderate steatosis, respectively). Mann-Whitney test was used to analyze intersite variability. The level of significance was set at $P < 0.05$.

Results

Standard H&E Histologic Assessment

In total, 84 slides from 42 livers were analyzed. The agreement between the 2 pathologists was almost perfect on absolute eFPA scores (ICC, 0.90; 95% CI, 0.81–0.94) and substantial on grading steatosis (κ_w , 0.75; 95% CI, 0.60–0.90). The average eFPA was 12% (SD, 16%; range, 0%–58%), and livers were classified according to their steatosis grades as normal ($n = 20$), mild ($n = 17$), moderate ($n = 5$), and severe ($n = 0$) (Fig 1). There was no difference in steatosis severity between liver segments 4 and 7 ($P = 0.6$).

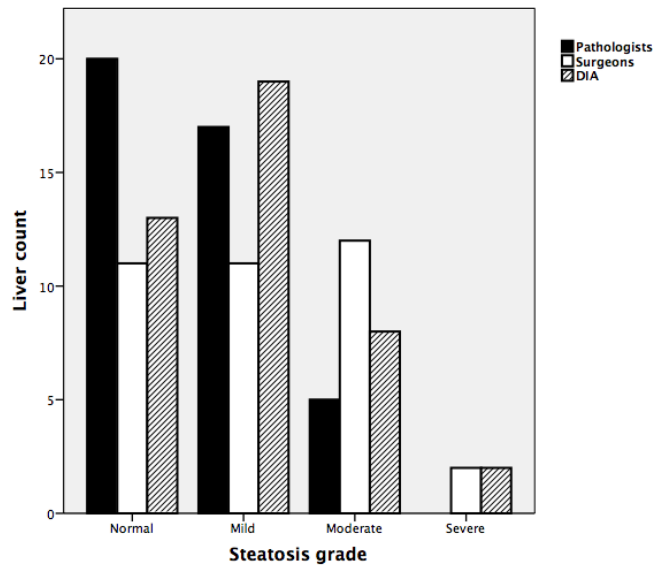


Fig 1 Bar chart showing the number of livers per assessment method and steatosis grade (n = 42). The number of livers graded by surgeons is underrepresented (n = 36), as 6 livers were classified as borderline steatosis (1 mild-moderate and 5 moderate-severe steatosis). DIA, digital image analysis.

Of 42 livers, 10 had macrosteatosis and grade 1 to 2 neutrophil infiltration (perivenular and midzonal, respectively) but no ballooning or perivenular “chicken-wire” fibrosis. Consequently, none of the livers were diagnosed with steatohepatitis [25,27,28]. Neutrophil infiltration was most likely due to ischemia-reperfusion injury or proinflammatory processes in brain-dead donors [29–32]. None of the livers were reported to have microvascular abnormalities.

Clinical Assessment

Of 42 livers, 6 were reported to have borderline steatosis according to retrieval surgeons (1 mild-moderate and 5 moderate-severe steatosis). Steatosis was graded in the remaining livers as normal (n = 11), mild (n = 11), moderate (n = 12), and severe (n = 2) (Fig 1).

Agreement between surgeons and pathologists on assigning steatosis grades omitted 6 livers that did not adhere to the 4-grade scale. The agreement for the remaining livers (n = 36) was slight and no better than chance (κ , 0.12; 95% CI, -0.06 to 0.30). The 2 livers reported by surgeons to have severe steatosis were graded as mild and moderate by pathologists.

When livers were dichotomized into low- or high-grade steatosis, surgeons classified 23 livers as low-grade steatosis and 19 livers as high-grade steatosis (n = 42 livers), and there was a fair agreement beyond chance for discriminating low- and high-grade steatosis (κ , 0.28; 95% CI, 0.07–0.49). Sensitivity, specificity, and diagnostic accuracy of clinical assessment were 100%, 62%, and 67%, respectively.

Digital Image Analysis Assessment

A total of 84 slides from 42 livers were analyzed using DIA. The average mFPA was 6% (SD, 7%; range, 0%–29%). Absolute mFPA scores were lower than absolute eFPA (1.5-fold higher for eFPA on average), and this discrepancy became more evident as steatosis severity increased (Fig 2). The agreement between DIA and pathologists on absolute fat percentage was substantial (ICC, 0.76; 95% CI, 0.63–0.84).

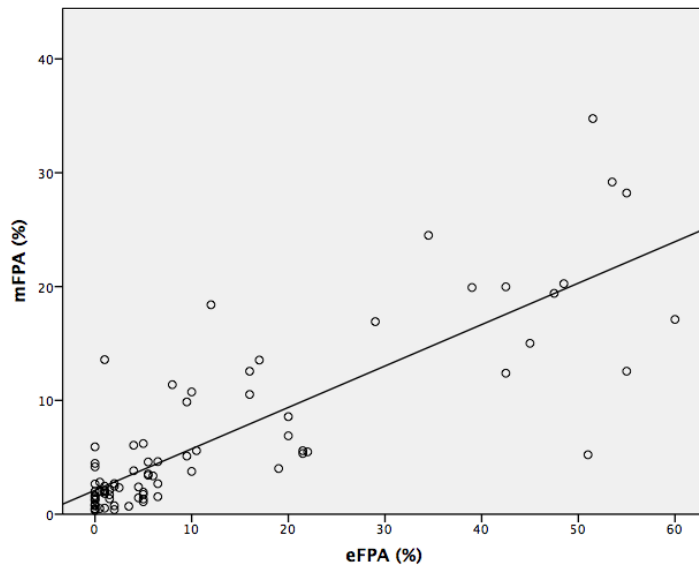


Fig 2 Scatterplot of estimated fat proportionate area (eFPA) and measured fat proportionate area (mFPA). There is a strong overall agreement (ICC, 0.76; 95% CI, 0.63–0.84). The agreement was stronger at lower steatosis levels, and mFPA scores were generally lower than eFPA scores as steatosis severity increased. eFPA, estimated fat proportionate area; ICC, intraclass correlation coefficient; mFPA, measured fat proportionate area.

According to the revised mFPA classification system, livers were graded as normal ($n = 13$), mild ($n = 19$), moderate ($n = 8$), and severe ($n = 2$) (Fig 1). The agreement between DIA and pathologists was substantial on assigning steatosis grades (κ_w , 0.70; 95% CI, 0.57–0.82). The 2 livers classified by DIA as severe steatosis were graded as moderate by pathologists and were different to the 2 severely steatotic livers as per clinical examination.

Sensitivity, specificity, and diagnostic accuracy of DIA were 100%, 87%, and 88%, respectively. Finally, there was no difference in steatosis severity between liver segments 4 and 7 ($P = 0.8$).

Discussion

This is the first study to compare quantitative liver steatosis assessment by standard H&E histology with retrieval surgeons and the newer technique of DIA. Additionally, we evaluated a revised steatosis grading system for automated fat measurement as an objective means of steatosis assessment.

Hepatic steatosis is a common finding in deceased donor livers [1–3]. Based on United Kingdom transplant figures between 2006 and 2016, steatosis was the most common reason for nonutilization (39%) in deceased donor livers that were retrieved but not transplanted [33]. Because of the prevalence of steatosis and its association with poor outcome in transplanted livers, accurate assessment is crucial to maximize the utilization of livers and reduce the mortality and morbidity associated with high-grade steatosis.

Subjective assessment of liver specimens by the pathologist is the criterion standard for evaluating liver steatosis. This method has been criticized for poor interpathologist reproducibility [11], but our results show excellent interpathologist agreement in estimating fat percentage area and assigning steatosis grades, similar to other reports [10,14,15,18,25]. It is possible that the poor reliability reported previously is related to different steatosis definitions and cutoff values as well as varying levels of experience.

Macroscopic assessment by the retrieving surgeon is highly subjective and inaccurate [6], and studies have reported both underestimation [34] and overestimation [7,8] of steatosis by surgeons. A survey of surgical practice in the United Kingdom and United States indicated that 38% of UK surgeons and 47% of US surgeons performed biopsies if the liver looked macroscopically steatotic, while 50% of UK surgeons did not incorporate microscopic

assessment in their practice at all [6]. In our research, clinical evaluation was inaccurate and overestimated steatosis severity compared with the criterion standard, which was also reflected by the high percentage of false positives in this cohort (38%). Given the shortage in donor livers, this risks unnecessary discard of potentially transplantable grafts based on clinical assessment alone. Clinical evaluation should be viewed as a screening test and should not be used exclusively to discard grafts. When significant steatosis is suspected clinically, microscopic examination before transplant is recommended, as per UK transplantation guidelines [23]. Surgical assessment remains vital for identifying other pathologies such as poor organ perfusion or mottling, fibrosis, focal lesions, and vascular abnormalities.

Although it is common practice for surgeons to visually classify livers according to a 4-tier grading system, this scale was originally adapted from clinical studies demonstrating a significant association between poor graft outcome and steatosis severity based on microscopic steatosis assessment [4,5]. To demand surgeons adhere to a 4-graded pathologic scale is not supported by high-level evidence and might be counterproductive, as demonstrated here by the poor agreement between surgeons and pathologists on assigning steatosis grades. It is more feasible for surgeons to classify retrieved livers into low-grade and high-grade steatosis, which emphasizes the role of clinical evaluation as a screening test that should ideally be followed by a more objective diagnostic investigation. It might be useful to include real-time images of the organ before and after perfusion to enhance the perception of steatosis severity among clinicians, although standardization of image quality and light source stability is required to ensure reproducibility of the results.

Automated measurement of steatosis using DIA has previously shown higher reproducibility and accuracy compared with standard H&E examination [10,14,17,19]. It is likely that the lack of high-level clinical evidence for automated measurement and confusion originating from the discrepancy between measured and estimated fat percentages have restricted the widespread application of this method.

In this study, absolute eFPA scores were higher than mFPA measurements, similar to what was described previously [10,11,14,15,17,19,22], and there was a trend for this discrepancy to increase with increasing steatosis severity. The revised mFPA classification used herein matches the conventional pathology grading system and adjusts for systematic disparities between measured and estimated fat percentage areas. Agreement with pathologists was high, and the scale was accurate in differentiating low- and high-grade steatosis. Besides, DIA images can be transferred to the recipient hospital for senior surgical review. Digital image analysis is accessible to most pathologists and surgeons; a light microscope mounted with a digital camera is available in most hospitals involved in liver retrieval, dedicated software for image analysis can be installed on a personal laptop, and steatosis scoring takes 10 to 15 minutes to complete after uploading the images. Despite these advantages, it is crucial to note that DIA cannot replace standard H&E assessment in liver transplantation yet. The performance of DIA should be evaluated using frozen tissue biopsies, similar to the current clinical practice. Furthermore, the validity of the revised mFPA scale needs to be analyzed in clinical outcome studies, alongside clinical and pathologic assessment methods. It is noteworthy that DIA is a measurement of area and cannot replace routine histopathology for assessment of other necroinflammatory lesions of steatohepatitis, such as inflammation and hepatocyte ballooning.

Finally, there was no difference in steatosis severity between 2 distant liver segments, suggesting that 1 liver biopsy is sufficient for steatosis assessment, which agrees with the practice of the majority of UK and US surgeons [6]. This is noteworthy as biopsied steatotic livers are at a higher risk of hemorrhage and subcapsular hematoma following reperfusion [35].

In the current study, pathologists and DIA evaluated the same slides at similar magnification power, and raters were blinded to results from other methods to reduce bias. We used specimens stained with H&E, in preference to fat-specific stains, because it is commonly used in surgical practice. Standard H&E assessment was performed by 2 experienced pathologists routinely involved in liver assessment, and the DIA method used herein was previously validated and reported in previous studies from our center [10,14]. A possible limitation of the study is that the high level of agreement between the 2 pathologists was influenced by working in the same

institution, despite being blinded to each other. It is unlikely that this caused bias, as high agreement between pathologists from different centers has been shown in previous studies [10,14,25]. Another limitation of the study is the lack of clinical outcomes, which could not be assessed, as the livers were discarded and nontransplantable by default. Part of this research was performed on retrospective data from organ retrieval forms, and 6 livers were not assigned a specific grade indicating a degree of diagnostic uncertainty.

In conclusion, clinical assessment of steatosis is inaccurate and could increase the risk of unnecessary discard of livers. While DIA is feasible and more accurate, further clinical studies are required to evaluate its validity.

Acknowledgments

We want to thank Mr Andrew Hall for his invaluable training and advice on digital image analysis. We would also like to thank Mr Peter DeMuylder from Organ Recovery Systems who facilitated the funds used for histology specimen processing and the liver transplant coordinators for their great assistance in this research.

REFERENCES

- [1] Angele MK, Rentsch M, Hartl WH, et al. Effect of graft steatosis on liver function and organ survival after liver transplantation. *Am J Surg* 2008;195:214e20.
- [2] Marsman WA, Wiesner RH, Rodriguez L, et al. Use of fatty donor liver is associated with diminished early patient and graft survival. *Transplantation* 1996;62:1246e51.
- [3] Verran D, Kusyk T, Painter D, et al. Clinical experience gained from the use of 120 steatotic donor livers for orthotopic liver transplantation. *Liver Transpl* 2003;9:500e5.
- [4] D'Alessandro AM, Kalayoglu M, Sollinger HW, et al. The predictive value of donor liver biopsies for the development of primary nonfunction after orthotopic liver transplantation. *Transplantation* 1991;51:157e63.
- [5] Ploeg RJ, D'Alessandro AM, Knechtle SJ, et al. Risk factors for primary dysfunction after liver transplantation: multivariate analysis. *Transplantation* 1993;55:807e13.
- [6] Imber CJ, St Peter SD, Lopez I, Guiver L, Friend PJ. Current practice regarding the use of fatty livers: a trans-Atlantic survey. *Liver Transpl* 2002;8:545e9.
- [7] Fernández ED, Schmid M, Bittinger F, Mauer D. Organ Procurement Central Region. Intraoperative assessment of liver organ condition by the procurement surgeon. *Transplant Proc* 2007;39:1485e7.
- [8] Rey J, Wirges U, Dienes H, Fries J. Hepatic steatosis in organ donors: disparity between surgery and histology? *Transplant Proc* 2009;41:2557e60.
- [9] Yersiz H, Lee C, Kaldas FM, et al. Assessment of hepatic steatosis by transplant surgeon and expert pathologist: a prospective, double-blind evaluation of 201 donor livers. *Liver Transpl* 2013;19:437e49.
- [10] Hall AR, Green AC, Luong TV, Burroughs AK, Wyatt J, Dhillon AP. The use of guideline images to improve histological estimation of hepatic steatosis. *Liver Int* 2014;34:1414e27.
- [11] El-Badry AM, Breitenstein S, Jochum W, et al. Assessment of hepatic steatosis by expert pathologists: the end of a gold standard. *Ann Surg* 2009;250:691e7.
- [12] Bedossa P, Poynard T, Naveau S, Martin E, Agostini H, Chaput J. Observer variation in assessment of liver biopsies of alcoholic patients. *Alcohol Clin Exp Res* 1988;12:173e8.
- [13] Bedossa P. Intraobserver and interobserver variations in liver biopsy interpretation in patients with chronic hepatitis C. *Hepatology* 1994;20:15e20.
- [14] Hall AR, Dhillon AP, Green AC, et al. Hepatic steatosis estimated microscopically versus digital image analysis. *Liver Int* 2013;33:926e35.
- [15] Levene AP, Kudo H, Armstrong M, et al. Quantifying hepatic steatosis - more than meets the eye. *Histopathology* 2012;60:971e81.
- [16] Biesterfeld S, Knapp J, Bittinger F, Gotte H, Schramm M, Otto G. Frozen section diagnosis in donor liver biopsies: observer variation of semiquantitative and quantitative steatosis assessment. *Virchows Arch* 2012;461:177e83.
- [17] Marsman H, Matsushita T, Dierkhising R, et al. Assessment of donor liver steatosis: pathologist or automated software? *Hum Pathol* 2004;35:430e5.
- [18] Franzen LE, Ekstedt M, Kechagias S, Bodin L. Semiquantitative evaluation overestimates the degree of steatosis in liver biopsies: a comparison to stereological point counting. *Mod Pathol* 2005;18:912e6.
- [19] Li M, Song J, Mirkov S, Xiao S-Y, Hart J, Liu W. Comparing morphometric, biochemical, and visual measurements of macrovesicular steatosis of liver. *Hum Pathol* 2011;42:356e60.
- [20] Fiorini RN, Kirtz J, Periyasamy B, et al. Development of an unbiased method for the estimation of liver

steatosis. *Clin Transplant* 2004;18:700e6.

[21] Bawa S, Vickers J, Barakat O, et al. Does a visual analogue colour chart carried by the retrieval team help in assessment of the fat content of donor livers? *Transpl Int* 2000;13:S140e3.

[22] Lee MJ, Bagci P, Kong J, et al. Liver steatosis assessment: correlations among pathology, radiology, clinical data and automated image analysis software. *Pathol Res Pract* 2013;209:371e9.

[23] British Transplantation Society. Transplantation from deceased donors after circulatory death. Cheshire, UK: British Transplantation Society; July 2013.

[24] Frankel WL, Tranovich JG, Salter L, Bumgardner G, Baker P. The optimal number of donor biopsy sites to evaluate liver histology for transplantation. *Liver Transpl* 2002;8:1044e50.

[25] Kleiner DE, Brunt EM, Van Natta M, et al. Design and validation of a histological scoring system for nonalcoholic fatty liver disease. *Hepatology* 2005;41:1313e21.

[26] Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;159e74.

[27] Harmon RC, Tiniakos DG, Argo CK. Inflammation in nonalcoholic steatohepatitis. *Expert Rev Gastroenterol Hepatol* 2011;5:189e200.

[28] Tsutsumi V, Nakamura T, Ueno T, Torimura T, Aguirre- García J. Structure and ultrastructure of the normal and diseased liver. In: Muriel P, editor. *Liver Pathophysiology*. London, UK: Elsevier ; 2017. p. 23e44.

[29] Watson CJ, Jochmans I. From "gut feeling" to objectivity: machine preservation of the liver as a tool to assess organ viability. *Curr Transplant Rep* 2018;5:72e81.

[30] Olinga P, Van Der Hoeven JA, Merema MT, Freund RL, Ploeg RJ, Groothuis GM. The influence of brain death on liver function. *Liver Int* 2005;25:109e16.

[31] Dziondzio T, Biebl M, Pratschke J. Impact of brain death on ischemia/reperfusion injury in liver transplantation. *Curr Opin Organ Transplant* 2014;19:108e14.

[32] Nebrig M, Neuhaus P, Pascher A. Advances in the management of the explanted donor liver. *Nat Rev Gastroenterol Hepatol* 2014;11:489.

[33] NHS blood and transplant. Annual report on liver transplantation. September 2016. Available from: https://nhsbtdbe.blob.core.windows.net/umbraco-assets-corp/1314/organ_specific_report_liver_2016.pdf. [Accessed 12 January 2018].

[34] Yersiz H, Lee C, Kaldas FM, et al. Assessment of hepatic steatosis by transplant surgeon and expert pathologist: a prospective, double-blind evaluation of 201 donor livers. *Liver Transpl* 2013;19:437e49.

[35] Lo IJ, Lefkowitz JH, Feirt N, et al. Utility of liver allograft biopsy obtained at procurement. *Liver Transpl* 2008;14:639e46.