Received: 1 November 2021 | Accepted: 30 March 2022

DOI: 10.1002/lt.26482

ORIGINAL ARTICLE

Standardizing the histological assessment of late posttransplantation biopsies from pediatric liver allograft recipients

Stefan G. Hübscher[1,2] | **Sandy Feng**[3] | **Annette S. H. Gouw**[4] | **Hironori Haga**[5] | **Hyo Jeong Kang**[6] | **Deirdre A. Kelly**[7] | **Mina Komuta**[8] | **Andrew Lesniak**[9] | **Benjamin A. Popp**[9] | **Henkjan J. Verkade**[10,11] | **Eunsil Yu**[6] | **Anthony J. Demetris**[9]

[1]Institute of Immunology and Immunotherapy, University of Birmingham, Birmingham, UK

[2]Department of Cellular Pathology, University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK

[3]Division of Transplantation, Department of Surgery, University of California, San Francisco, California, USA

[4]Department of Pathology and Medical Biology, University Medical Center Groningen, Groningen, the Netherlands

[5]Department of Diagnostic Pathology, Graduate School of Medicine, Kyoto University, Kyoto, Japan

[6]Department of Pathology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea

[7]Liver Unit, Birmingham Women's & Children's NHS Trust and University of Birmingham, Birmingham, UK

[8]Department of Pathology, Keio University, Tokyo, Japan

[9]Division of Liver and Transplantation Pathology, Department of Pathology, University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania, USA

[10]Pediatric Gastroenterology/Hepatology, Department of Pediatrics, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands

[11]European Reference Network on Hepatological Diseases (ERN RARE-LIVER), Hamburg, Germany

Correspondence
Stefan G. Hübscher, Department of Cellular Pathology, University Hospitals Birmingham NHS Foundation Trust, Level 1, Queen Elizabeth Hospital, Birmingham, B15 2WB, UK.
Email: s.g.hubscher@bham.ac.uk

Funding information
The Graft Injury Group has received project funding from the European Association for the Study of the Liver (EASL) and from the European Society for Pediatric Gastroenterology, Hepatology and Nutrition (ESPGHAN).

Abstract

Excellent short-term survival after pediatric liver transplantation (LT) has shifted attention toward the optimization of long-term outcomes. Despite considerable progress in imaging and other noninvasive modalities, liver biopsies continue to be required to monitor allograft health and to titrate immunosuppression. However, a standardized approach to the detailed assessment of long-term graft histology is currently lacking. The aim of this study was to formulate a list of histopathological features relevant for the assessment of long-surviving liver allograft health and to develop an approach for assessing the presence and severity of these features in a standardized manner. Whole-slide digital images from 31 biopsies obtained ≥4 years after transplantation to determine eligibility for an immunosuppression withdrawal trial were selected to illustrate a range of typical histopathological findings seen in children with clinically stable grafts, including those associated with alloantibodies. Fifty histological features were independently assessed and, where appropriate, scored

Abbreviations: AI, artificial intelligence; AMR, antibody-mediated rejection; BD, bile duct; CV, central vein; DSA, donor-specific antibody; GIG, Graft Injury Group; HA, hepatic artery; HAI, Hepatitis Activity Index; IPTH, idiopathic posttransplantation hepatitis; iWITH, Immunosuppression Withdrawal for Stable Pediatric Liver Transplant Recipients; LAFSc, Liver Allograft Fibrosis Score; LT, liver transplantation; mHAI, modified Hepatitis Activity Index; N/A, not applicable; NAFLD, nonalcohol-related fatty liver disease; OPV, obliterative portal venopathy; RAI, Rejection Activity Index; UW, unweighted; W, weighted.

semiquantitatively by six pathologists to determine inter- and intraobserver reproducibility of the histopathological features using unweighted and weighted kappa statistics; the latter metric enabled distinction between minor and major disagreements in parameter severity scoring. Weighted interobserver kappa statistics showed a high level of agreement for various parameters of inflammation, interface activity, fibrosis, and microvascular injury. Intraobserver agreement for these features was even more substantial. The results of this study will help to standardize the assessment of biopsies from long-surviving liver allografts, aid the recognition of important histological features, and facilitate international comparisons and clinical trials aiming to improve outcomes for children undergoing LT.

INTRODUCTION

Advances in liver transplantation (LT) have greatly improved the outlook of children with end-stage liver disease. More than 80% of children are now expected to survive over 20 years after LT.[1,2] Optimizing the use of potentially lifelong immunosuppression in these patients remains challenging. A balance needs to be achieved in using adequate immunosuppression to prevent immune-mediated graft injury while simultaneously trying to minimize the harmful adverse effects of immunosuppression.[3] Most late deaths in liver allograft recipients relate to the adverse effects of immunosuppression rather than complications affecting the graft itself.

The role of liver biopsies in monitoring graft injury in long-term survivors following LT is changing. Most centers no longer obtain late protocol biopsies from adult liver allograft recipients. Not only does this relate to the improved outlook of patients with diseases such as hepatitis C, which now rarely recurs following LT, but also to the uncertain clinical significance of histological abnormalities identified in patients with apparent good graft function. The situation is somewhat different in the pediatric population in whom progressive, often subclinical histological, damage is documented more frequently.[2]

Three important histopathological observations have emerged from pediatric centers conducting >12-month post-LT protocol biopsies (summarized in Table 1).[2,4–16] First, histological abnormalities are very common with a prevalence exceeding 90% in some studies. The two most frequently observed abnormalities in >12-month pediatric posttransplantation biopsies are unexplained graft inflammation (idiopathic posttransplantation hepatitis [IPTH]) and graft fibrosis, for which three main patterns are described—periportal, sinusoidal, and centrilobular.[10] While the pathogenesis of late allograft inflammation and fibrosis has yet to be fully understood, there is increasing evidence to suggest that both patterns of graft injury are likely to be at least partly immune mediated. This particularly applies to the pediatric population in whom recurrent disease can largely be excluded as a cause of late graft injury. Current evidence suggests that T-cell–mediated and antibody-mediated mechanisms are both likely to be involved.[2,16–18] Second, studies documenting changes in serial biopsies have shown that the prevalence and severity of abnormal graft histology increase with time after transplantation.[5,7,9,15,19,20] Third, the majority of children studied had normal or near-normal liver biochemistry, which appears to be at variance with the prevalence of abnormal histology. Therefore, it is reasonable to argue that graft injury would not have been detected by routine, noninvasive liver tests alone. Furthermore, the mechanisms underlying common patterns of late graft injury remain incompletely understood and clear guidelines for the management of children with unexplained abnormal graft histology are consequently lacking.

A review of previously published studies indicates that there is a lack of consistency in how histological findings are reported in late posttransplantation biopsies. This applies both to the assessment of graft inflammation, for which several different terms have been used,[21] and to the assessment of graft fibrosis, for which a number of different approaches have been used—these include previously published systems for scoring fibrosis (e.g., those described by METAVIR, Inuyama, and Venturi), as well as other locally devised scoring systems (summarized in Table 1).

The Graft Injury Group (GIG) is an international collaboration involving pediatric hepatologists, surgeons, pathologists, immunologists, and other interested parties working at centers involved in pediatric LT. The group was formed in February 2015 with the view to carrying out detailed observations of long-term outcomes in pediatric liver allograft recipients in the hope that these would provide further insights into areas where uncertainty currently exists.

This paper describes a study carried out by the GIG Pathology Working Group, the main aim of which

TABLE 1 Histological findings in liver biopsies obtained >12 months after transplantation from pediatric liver allograft recipients

Center	Number of biopsies	Time post-LT	Abnormal histology	Main histological diagnoses	Pattern(s) of fibrosis assessed	Scoring system used for staging fibrosis
Paris (Fouquet et al. 2005[4])	67	>10 years	73%	Chronic rejection (42%), centrilobular fibrosis (22%), biliary cirrhosis (4%), other (4%)	Centrilobular	None
Birmingham (Evans et al. 2006[5])	113, 135, 164	1, 5, 10 years	32% at 1 year, 55% at 5 years, 69% at 10 years	Chronic hepatitis with or without fibrosis (64%), biliary fibrosis (2%), recurrent primary sclerosing cholangitis (2%), other (2%)—at 10 years	Portal	Other
Chicago (Ekong et al. 2008[6])	63	>3 years	97%	Fibrosis (97%), inflammation (70%)	Portal	METAVIR
Groningen (Scheenstra et al. 2009[7])	77, 64, 66, 55	1, 3, 5, 10 years	34% at 1 year, 48% at 3 years, 65% at 5 years, 69% at 10 years	Fibrosis (69%)—at 10 years	Portal	Other
Osaka (Ueno et al. 2011[8])	24	>1 year	≥71%	Fibrosis (71%), inflammation (58%)	Portal	Inuyama
Kyoto (Miyagawa-Hayashino et al. 2012[9])	67	>5 years	≥84%	Fibrosis (84%), inflammation (58%)	Sinusoidal/ centrilobular	Other
Brussels (Venturi et al. 2012[10])	38	7 years	94%	Fibrosis (94%), inflammation (74%), ductal proliferation (26%), steatosis (26%)	Portal/sinusoidal/ centrilobular	Venturi (LAFSc)
Tokyo (Tomita et al. 2013[11])	59	0.2–15 years (median 6 years)	≥86%	Fibrosis (86%), inflammation (39%), steatosis (10%)	Portal	METAVIR
Helsinki (Kosola et al. 2013[12])	54	>3 years	≥43%	Steatosis (43%), ductular reaction (43%), fibrosis (39%), inflammation (22%)	Portal	METAVIR
Hamburg (Briem-Richter et al. 2013[13])	60	>1 year	40%	Fibrosis (33%), mild acute rejection (20%), steatosis (17%), early chronic rejection (3%)	Not specified	N/A
King's College Hospital, London (Dattani et al. 2014[14])	56	>1 year	84%	Hepatitis (41%), bridging fibrosis/cirrhosis (27%), nodular regenerative hyperplasia (16%), biliary problem (12.5%), rejection (4%), other (11%). Note: No protocol biopsies, but many children had normal liver function test results	Not specified	N/A
Tochigi (Sanada et al. 2014[15])	89, 55	2, 5 years	>42%	Inflammation (42%), fibrosis (34.5%) at 5 years	Portal	METAVIR
Multicenter (Feng et al. 2018[16])	157	>4 years (mean 8.9 years)	>80%	Portal inflammation (64%), lobular inflammation (24%), interface inflammation (22%), perivenular fibrosis (80%), periportal fibrosis (39%)	Portal/sinusoidal/ centrilobular	Venturi (LAFSc) and Ishak

Note: All studies included protocol biopsies with >50% of children having normal or near-normal liver biochemistry. The final two columns summarize the different approaches that were used in these studies to assess the pattern and severity of fibrosis in late biopsies.

Abbreviations: LAFSc, Liver Allograft Fibrosis Score; LT, liver transplantation; N/A, not applicable.

was to design a schema for systematically recording histological findings in late posttransplantation biopsies from pediatric liver allograft recipients. The study was designed to have three main components: first, to agree on which histological features were relevant for the assessment of late biopsies; second, to assess and improve intra- and interobserver reproducibility among participating pathologists; and third, to use the agreed schema to record changes in late biopsies from children enrolled in centers participating in the GIG study in a standardized manner. The results of the first two parts of the study are presented here. It is hoped that our findings will be useful, not only for documenting histological changes seen in protocol biopsies from the GIG study, but also in providing guidelines that may be useful for standardizing the assessment of late posttransplantation biopsies by the broader LT community.

PATIENTS AND METHODS

Study organization and case selection

The GIG Pathology Working Group met on three occasions: The first two meetings, held in Birmingham in February 2015 and Brussels in September 2015, were attended by five pathologists (A.J.D., A.S.H.G., H.H., S.G.H., and M.K.) and by other clinicians and scientists who provided input into the study design. One other pathologist (H.J.K.) also came to the third meeting held in Barcelona in March 2017.

At the first meeting, members of the pathology group agreed to review slides from 10 late posttransplantation biopsies. The biopsies were obtained from patients <18 years old and ≥4 years after primary living or deceased donor LT for nonviral and nonautoimmune liver disease at ≤6 years of age who underwent screening liver biopsy for immunosuppression withdrawal as part of a trial conducted at 12 pediatric LT centers in North America (Immunosuppression Withdrawal for Stable Pediatric Liver Transplant Recipients (iWITH) trial; NCT01638559).[16] Institutional review board approval was obtained at all participating centers as data collection was retrospective; assent and/or informed consent were not required as there were no study interventions. Participants were required to have alanine aminotransferase and gamma-glutamyltransferase levels consistently <50 IU/L based on medical record review by the site principal investigator and to be stably maintained on calcineurin inhibitor monotherapy without rejection during the preceding 2 years. A scoring proforma devised by one of the authors (A.J.D.) for a study of chronic antibody-mediated rejection (AMR) was used for the first round of assessments.[22] Using illustrative examples where appropriate, a consensus was reached regarding the criteria that should be used to designate scores for the individual histological features. Slides were then reviewed as digitized images of formalin-fixed paraffin-embedded sections stained with hematoxylin and eosin and Masson's trichrome (to assess fibrosis) via the University of Pittsburgh Telepathology website.

At the second meeting, scores for the first 10 cases were reviewed and discussed. The histology scoring proforma was revised by removing or collapsing features that had poor interobserver agreement. Scoring guidelines were written for each of the remaining features with the aim to improve observer reproducibility. The histological features that were assessed and scored are summarized in Table S1 and listed in Tables 2 and 3. Participants then assessed the utility of the revised scoring proforma by rescoring the original 10 cases, plus an additional 21 late pediatric liver allograft biopsies, equaling 31 cases in total. Statistical analysis was carried out using "percent agreement" and Shrout and Fleiss intraclass correlation methods.

Findings including statistical analyses were circulated to members of the pathology group and discussed at the third meeting. Further consideration was given to features that produced unsatisfactory observer agreement scores and scoring proforma guideline definitions were revised with an emphasis on improving reliability of the assessments. Details of the criteria used for scoring are provided in Table S2. After a "washout period" of 3 months, slides were reshuffled, recirculated, and rescored using the same scoring proforma. The two rounds of assessments of the full set of 31 slides were carried out by six pathologists (A.J.D., A.S.G.H., H.H., S.G.H., M.K., and H.J.K.).

Statistical analysis

Following discussion with the GIG clinicians, statisticians, and other members of the GIG Steering Committee, it was decided that the final statistical analyses of observer agreement should be carried out using kappa values (unweighted and weighted). Where appropriate, associated significance and 95% confidence intervals were also calculated to compare the unweighted and weighted approaches. Weighted kappa (variations of Cohen kappa), used in many high-profile histopathology reliability studies,[23,24] accounts for the "closeness" of agreement and takes into consideration the variable type of data used—both numerical and categorical. For numerically rated data, such as biopsy length or portal tract number, a linear graded penalty system was employed where the median score was considered to be ground truth; thus, a score within 20% of the median was regarded as equal, 21%–40% as 1 grade of difference, 41%–60% as 2 grades of difference, 61%–80% as 3 grades of difference, and >80% as maximum difference. For categorical data scored on a scale of 0 to 3, 4, or 6, weighting was applied such that a discrepancy of 1 point was regarded as equal, while scores that were discrepant by ≥2 points

TABLE 2 "Heat map" showing interobserver agreement expressed as kappa scores (unweighted and weighted) for the first and second rounds of slide reviews

Feature assessed	Kappa scores (first round of reviews)		Kappa scores (second round of reviews)	
	Unweighted	Weighted	Unweighted	Weighted
Biopsy characteristics				
Length, mm	N/A	0.72	N/A	0.90
Number of portal tracts	N/A	0.48	N/A	0.63
Portal evaluation				
Portal vein branch count	N/A	0.40	N/A	0.49
Number of portal tracts without a portal vein	N/A	0.75	N/A	0.72
Portal tract inflammation overall severity	0.38	0.76	0.32	0.72
Portal tract inflammation distribution	0.35	0.93	0.30	0.90
Primary portal inflammatory cell type	0.74	N/A	0.69	N/A
Secondary portal inflammatory cell type	0.41	N/A	0.50	N/A
Tertiary portal inflammatory cell type[a,b]	0.85	N/A	0.94	N/A
Portal microvasculitis severity	0.61	0.93	0.56	0.89
Biliary senescence[a,b]	0.75	N/A	0.82	N/A
Number of portal tracts without a BD[a,b]	N/A	0.82	N/A	0.77
Interface evaluation				
Interface hepatitis severity	0.66	0.98	0.65	0.95
Ductular reaction	0.21	0.68	0.26	0.65
Vascular evaluation				
Artery pathology[a,b]	0.87	N/A	0.93	N/A
Periportal shunt vessels	0.27	0.70	0.23	0.62
Sinusoidal dilatation	0.26	0.72	0.29	0.64
Hepatic veno-occlusive lesions[b]	0.82	0.98	0.85	1.00
Hepatic vein endothelial inflammation[a,b]	0.85	1.00	0.90	1.00
Lobule evaluation				
Lobular disarray/ballooning[a,b]	0.81	0.96	0.85	0.97
Lobular inflammation severity	0.46	0.84	0.46	0.77
Lobular inflammation distribution	0.46	N/A	0.45	N/A
Primary lobular inflammatory cell type	0.60	N/A	0.64	N/A
Secondary lobular inflammatory cell type[a]	0.81	N/A	0.77	N/A
Sinusoidal leukocytosis	0.77	0.93	0.73	0.92
Central perivenulitis severity	0.68	0.89	0.59	0.79
Fibrosis/architectural disturbance				
Portal tract collagenization	0.16	0.52	0.22	0.51
Periportal fibrosis (Venturi)	0.08	0.75	0.12	0.72
Sinusoidal fibrosis (Venturi)	0.18	0.85	0.19	0.89
Perivenular fibrosis (Venturi)	0.23	0.76	0.13	0.77
Total LAFSc (Venturi)	0.05	0.59	0.05	0.56
Fibrosis stage (Ishak)	0.18	0.70	0.21	0.75
Fibrosis stage (METAVIR)	0.23	0.87	0.28	0.87
Nodular regenerative hyperplasia	0.40	0.58	0.37	0.66
RAI				
Portal tract inflammation (Banff RAI)	0.39	0.99	0.36	0.97

(Continues)

TABLE 2 (Continued)

Feature assessed	Kappa scores (first round of reviews)		Kappa scores (second round of reviews)	
	Unweighted	Weighted	Unweighted	Weighted
BD inflammation (Banff RAI)	0.70	0.96	0.76	0.95
Venous endothelial inflammation (Banff RAI)	0.67	0.97	0.56	0.92
Total RAI score (Banff RAI)	0.29	0.93	0.25	0.91
HAI				
Interface hepatitis (Ishak mHAI)	0.64	0.94	0.64	0.94
Confluent necrosis (Ishak mHAI)[a]	0.89	0.96	0.79	0.88
Lobular inflammation (Ishak mHAI)	0.54	0.99	0.54	0.92
Portal tract inflammation (Ishak mHAI)	0.42	0.97	0.44	0.97
Total HAI score (Ishak mHAI)	0.26	0.88	0.25	0.85
NAFLD scores				
NAFLD steatosis score[a,b]	0.90	0.99	0.89	1.00
NAFLD hepatocyte ballooning score[a,b]	0.98	1.00	0.96	1.00
NAFLD lobular inflammation score	0.56	1.00	0.61	0.91
Total NAFLD score	0.59	1.00	0.62	0.98
NAFLD fibrosis stage	0.17	0.38	0.21	0.42

Agreement

0.81–1.00 (almost perfect)	0.61–0.80 (substantial)	0.41–0.60 (moderate)	0.21–0.40 (fair)	0–0.20 (slight)	N/A (not applicable)

Abbreviations: BD, bile duct; LAFSc, Liver Allograft Fibrosis Score; mHAI, modified Hepatitis Activity Index; N/A, not applicable; NAFLD, nonalcoholic fatty liver disease; RAI, Rejection Activity Index.

[a]Indicates a *first round* feature with infrequent occurrence (<10%).

[b]Indicates a *second round* feature with infrequent occurrence (<10%).

were regarded as being discordant. For any composite score, that is, Ishak Hepatitis Activity Index (HAI), Venturi Liver Allograft Fibrosis Score (LAFSc), Banff Rejection Activity Index, Kleiner nonalcohol-related fatty liver disease (NAFLD) activity score, the weight matrix considered scores discrepant by ≤2 points to be equal. For categorical data, ordered levels were "none," "minimal," "mild," "moderate," and "severe". To further stratify subtle patterns of injury, as well as account for some level of disagreement between any two scores, the associated weighted matrix (1) regarded the comparison of "none" and "minimal" as no difference (i.e., identical, both categories indicating an expression below baseline); (2) established a "lenient" penalty of 0.5 for the difference between "minimal" and "mild"; (3) assigned full penalty weights of 1.0 between adjacent pairs of scores for "mild," "moderate," and "severe" (i.e., the weighted penalty between "mild" and "moderate" was 1.0; the weighted penalty between "moderate" and "severe" was also 1.0); and (4) set penalty weights for nonadjacent pairs equal to the absolute difference between said paired values (e.g., the weighted penalty between "none" and "severe" was 4; "minimal" and "severe" was 3). The level of agreement, according to calculated kappa scores, was then classified as follows: 0–0.20 = slight, 0.21–0.40 = fair, 0.41–0.60 = moderate, 0.61–0.80 = substantial, and 0.81–1.00 = almost perfect. The number of biopsies required to determine observer reproducibility reliably was based on a study by Walter et al, which provides a set of formulae for assessing sample size for kappa significance based on the number of "readers" and assumptions about each "reader," such as expected levels of agreement.[25] Given that the "readers" involved in this study are a population of expert LT pathologists, one can reasonably assume that fair to moderate agreement would be reasonable/expected. Given these assumptions, the study size *N* calculates out to between 29 and 35 for six observers. Therefore *N* = 31 included in this study is appropriate to obtain reliable kappa statistics.

RESULTS

Interobserver agreement

Overall, there were no significant differences in the degree of interobserver agreement (Table 2) between the first and second round of reviews.

TABLE 3 "Heat map" showing intraobserver agreement between the first and second round of slide reviews expressed as kappa scores (unweighted and weighted)

	Kappa scores											
	Pathologist 1		Pathologist 2		Pathologist 3		Pathologist 4		Pathologist 5		Pathologist 6	
Feature assessed	UW	W	UW	W	UW	W	UW	W	UW	W	UW	W
Biopsy characteristics												
Length, mm	N/A	0.96	N/A	0.92	N/A	0.96	N/A	0.60	N/A	0.88	N/A	0.88
Number of portal tracts	N/A	0.60	N/A	0.84	N/A	0.15	N/A	0.52	N/A	0.48	N/A	0.64
Portal evaluation												
Portal vein branch count	N/A	0.40	N/A	0.64	N/A	0.15	N/A	0.56	N/A	0.40	N/A	0.52
Number of portal tracts without a portal vein	N/A	0.76	N/A	0.48	N/A	0.88	N/A	0.72	N/A	0.64	N/A	1.00
Portal tract inflammation overall severity	0.48	0.78	0.60	0.82	0.56	0.72	0.40	0.82	0.35	0.68	0.56	0.72
Portal tract inflammation distribution	0.44	1.00	0.53	1.00	0.70	1.00	0.31	0.91	0.27	0.74	0.61	0.96
Primary portal inflammatory cell type	0.82	N/A	0.89	N/A	0.93	N/A	0.63	N/A	0.59	N/A	1.00	N/A
Secondary portal inflammatory cell type	0.82	N/A	1.00	N/A	0.41	N/A	0.85	N/A	0.63	N/A	0.96	N/A
Tertiary portal inflammatory cell type[a,b]	1.00	N/A	1.00	N/A	0.85	N/A	0.96	N/A	0.96	N/A	0.71	N/A
Portal microvasculitis severity	0.56	0.88	0.84	1.00	0.27	0.90	0.80	0.98	0.80	0.88	0.72	0.88
Biliary senescence[a,b]	0.87	N/A	0.81	N/A	0.94	N/A	0.94	N/A	0.29	N/A	1.00	N/A
Number of portal tracts without a BD[a,b]	N/A	0.61	N/A	1.00	N/A	0.94	N/A	0.68	N/A	0.68	N/A	1.00
Interface evaluation												
Interface hepatitis severity	1.00	1.00	0.83	1.00	0.70	0.96	0.87	1.00	0.87	0.91	0.78	1.00
Ductular reaction	0.31	0.64	0.64	0.92	0.11	0.60	0.19	0.58	0.64	0.82	0.23	0.88
Vascular evaluation												
Artery pathology[a,b]	0.91	N/A	0.87	N/A	0.83	N/A	0.96	N/A	0.96	N/A	0.87	N/A
Periportal shunt vessels	0.23	0.84	0.68	0.90	0.15	0.62	0.84	0.96	1.00	1.00	0.40	0.66
Sinusoidal dilatation	0.84	0.92	0.27	0.74	0.15	0.58	0.64	0.88	0.40	0.76	0.35	0.76
Hepatic veno-occlusive lesions[b]	0.56	0.95	0.95	1.00	0.90	1.00	0.90	1.00	0.95	1.00	0.85	1.00
Hepatic vein endothelial inflammation[a,b]	0.71	1.00	0.90	1.00	0.90	1.00	0.95	1.00	0.90	1.00	0.95	1.00

(Continues)

TABLE 3 (Continued)

Feature assessed	Kappa scores											
	Pathologist 1		Pathologist 2		Pathologist 3		Pathologist 4		Pathologist 5		Pathologist 6	
	UW	W	UW	W	UW	W	UW	W	UW	W	UW	W
Lobule evaluation												
Lobular disarray/ballooning[a,b]	0.96	1.00	1.00	1.00	0.52	0.96	0.72	1.00	0.84	0.88	0.96	1.00
Lobular inflammation severity	0.52	0.86	0.68	0.90	0.44	0.74	0.68	0.96	0.48	0.90	0.76	0.96
Lobular inflammation distribution	0.57	N/A	0.61	N/A	0.44	N/A	0.74	N/A	0.61	N/A	0.78	N/A
Primary lobular inflammatory cell type	0.67	N/A	0.82	N/A	0.63	N/A	0.74	N/A	0.67	N/A	0.85	N/A
Secondary lobular inflammatory cell type[a]	0.89	N/A	0.93	N/A	0.82	N/A	1.00	N/A	0.85	N/A	0.85	N/A
Sinusoidal leukocytosis	0.88	1.00	0.96	1.00	0.56	0.82	0.88	1.00	0.84	0.90	0.88	0.98
Central perivenulitis severity	0.52	0.86	0.84	0.98	0.80	0.88	0.84	0.98	0.88	0.92	0.72	0.92
Fibrosis/architectural disturbance												
Portal tract collagenization	0.27	0.62	0.48	0.80	0.31	0.68	0.80	0.94	0.23	0.76	0.48	0.70
Periportal fibrosis (Venturi)	0.18	0.91	0.48	0.96	0.40	0.78	0.35	0.96	0.48	0.91	0.23	0.96
Sinusoidal fibrosis (Venturi)	0.35	0.96	0.74	1.00	0.57	0.96	0.70	1.00	0.61	1.00	0.61	1.00
Perivenular fibrosis (Venturi)	0.31	0.87	0.57	0.96	0.35	0.87	0.57	1.00	0.10	0.87	0.61	1.00
Total LAFSc (Venturi)	0.09	0.78	0.24	0.96	0.24	0.89	0.42	1.00	0.13	0.78	0.13	1.00
Fibrosis stage (Ishak)	0.21	0.81	0.55	1.00	0.47	0.74	0.32	0.89	0.25	0.74	0.28	0.89
Fibrosis stage (METAVIR)	0.27	0.96	0.68	1.00	0.56	1.00	0.48	0.96	0.27	0.80	0.35	0.96
Nodular regenerative hyperplasia	0.56	0.76	0.80	0.88	0.44	0.48	0.64	0.84	0.96	1.00	0.40	0.68
RAI												
Portal tract inflammation (Banff RAI)	0.53	1.00	0.78	1.00	0.78	0.91	0.44	1.00	0.48	0.87	0.53	1.00
BD inflammation (Banff RAI)	0.53	1.00	0.74	1.00	0.74	0.96	0.83	0.96	0.57	0.91	0.87	1.00
Venous endothelial inflammation (Banff RAI)	0.31	1.00	1.00	1.00	0.64	0.82	0.87	1.00	0.78	0.96	0.70	1.00
Total RAI score (Banff RAI)	0.20	0.93	0.60	1.00	0.6	0.93	0.35	1.00	0.46	0.82	0.42	1.00
HAI												
Interface hepatitis (Ishak mHAI)	0.96	1.00	0.84	1.00	0.76	0.92	0.92	1.00	0.76	0.88	0.72	1.00

TABLE 3 (Continued)

Feature assessed	Kappa scores											
	Pathologist 1		Pathologist 2		Pathologist 3		Pathologist 4		Pathologist 5		Pathologist 6	
	UW	W	UW	W	UW	W	UW	W	UW	W	UW	W
Confluent necrosis (Ishak mHAI)[a]	0.96	1.00	0.85	0.96	0.77	0.85	0.89	1.00	0.92	0.92	0.96	1.00
Lobular inflammation (Ishak mHAI)	0.80	0.96	0.80	1.00	0.48	1.00	0.68	1.00	0.52	0.92	0.76	1.00
Portal tract inflammation (Ishak mHAI)	0.35	1.00	0.76	1.00	0.84	1.00	0.44	1.00	0.48	0.84	0.72	0.96
Total HAI score (Ishak mHAI)	0.42	0.97	0.56	1.00	0.39	1.00	0.39	0.97	0.49	0.80	0.42	0.97
NAFLD scores												
NAFLD steatosis score[a,b]	1.00	1.00	0.87	1.00	0.83	1.00	1.00	1.00	1.00	1.00	0.96	1.00
NAFLD hepatocyte ballooning score[a,b]	1.00	1.00	1.00	1.00	0.85	1.00	1.00	1.00	1.00	1.00	0.95	1.00
NAFLD lobular inflammation score	1.00	1.00	0.78	1.00	0.31	0.96	1.00	1.00	1.00	1.00	1.00	1.00
Total NAFLD score	1.00	1.00	0.68	1.00	0.33	1.00	1.00	1.00	1.00	1.00	0.93	1.00
NAFLD fibrosis stage	0.23	0.23	0.72	1.00	0.31	0.84	0.40	1.00	1.00	1.00	0.96	0.96

Agreement

0.81–1.00 (almost perfect)	0.61–0.80 (substantial)	0.41–0.60 (moderate)	0.21–0.40 (fair)	0–0.20 (slight)	N/A (not applicable)

Abbreviations: N/A, not applicable; BD, bile duct; LAFSc, Liver Allograft Fibrosis Score; mHAI, modified Hepatitis Activity Index; NAFLD, nonalcoholic fatty liver disease; RAI, Rejection Activity Index; UW, unweighted, W, weighted.

[a]Indicates a *first round* feature with infrequent occurrence (<10%).

[b]Indicates a *second round* feature with infrequent occurrence (<10%).

For both rounds of reviews, the majority of histological features with unweighted kappa scores indicating substantial or better agreement ($\kappa > 0.60$) were ones rarely seen in late posttransplantation biopsies from apparently healthy recipients, and thus scored accordingly as "zero," "none," "absent," or "N/A" (not applicable). Examples include biliary senescence, artery pathology, bile duct loss, lobular disarray, sinusoidal leukocytosis, hepatic vein endothelial inflammation, hepatic veno-occlusive lesions, bile duct damage, steatosis, hepatocyte ballooning, tertiary portal inflammatory cell type, and secondary lobular inflammatory cell type. Unweighted kappa scores for features related to the severity of late graft inflammation ranged from "fair" to "substantial": portal inflammation (0.38 in the first round, 0.32 in the second round), interface hepatitis (0.66 and 0.65, respectively), and lobular inflammation (0.46 and 0.46, respectively). Unweighted kappa scores for fibrosis were only "slight" or "fair" for all three of the fibrosis staging systems used (METAVIR, Ishak, and Venturi), with scores ranging from 0.05 (total LAFSc, first and second rounds) to 0.28 (METAVIR stage, second round).

For both rounds of reviews, comparison of weighted and unweighted kappa scores yielded statistically significant differences (using a 95% confidence interval) for all evaluated features. This suggests that the levels of "disagreement" suggested by the unweighted kappa scores were relatively minor, representing sub-baseline patterns of injury, and not likely to be clinically significant (see the "Patients and Methods" section). Weighted kappa scores for features related to inflammation and fibrosis were all in the range of "substantial" (0.61–0.80) to "almost perfect" (0.81–1.00). The only features for which weighted kappa scores were less than substantial were portal tract number (first round), portal vein number (both rounds), nodular regenerative hyperplasia (1st round), portal tract collagenization (both rounds), and total liver allograft LAFSc (both rounds).

Intraobserver agreement (Table 3)

Similar to the data for interobserver agreement, unweighted kappa scores were nearly all substantial or better for histological changes that were observed infrequently in the 31 liver biopsies assessed—biliary senescence, artery pathology, bile duct loss, lobular disarray, sinusoidal leukocytosis, hepatic vein endothelial inflammation, hepatic veno-occlusive lesions, bile duct damage, steatosis, hepatocyte ballooning, tertiary portal inflammatory cell type, and secondary lobular inflammatory cell type. Unweighted kappa scores for features related to the severity of late graft inflammation were better than those obtained for interobserver agreement: portal inflammation (mean 0.49, range 0.35–0.60), interface hepatitis (mean 0.84, range 0.78–1.00), and lobular inflammation (mean 0.59, range 0.44–0.76). Unweighted kappa scores for features related to the severity of fibrosis were also better than those obtained for interobserver agreement: METAVIR fibrosis stage (mean 0.44, range 0.27–0.68), Ishak fibrosis stage (mean 0.35, range 0.21–0.47), Venturi periportal fibrosis score (mean 0.35, range 0.18–0.48), subsinusoidal fibrosis score (mean 0.60, range 0.35–0.74), perivenular fibrosis score (mean 0.42, range 0.10–0.61), and total LAFSc (mean 0.21, range 0.13–0.42).

Similar to interobserver agreement, intraobserver agreement greatly improved when weighted kappa scores were applied. Scores for the severity of inflammation (portal, interface, lobular) were "almost perfect" (0.81–1.00) for 13/18 comparisons between the first and second rounds and "substantial" (0.61–0.80) for the other 5. Weighted kappa scores for features related to the severity of fibrosis (METAVIR, Ishak, Venturi [periportal, subsinusoidal, perivenular], and total LAFSc) were "almost perfect" for 25/36 comparisons between the first and second rounds and "substantial" for the remaining 11. The levels of intraobserver agreement using weighted kappa scores were also "substantial" or "almost perfect" for the great majority of the other histological features assessed. Similar to the observations made for interobserver agreement, weighted kappa scores for intraobserver agreement were less than substantial for portal tract number (4/6 observers) and portal vein number (5/6 observers). Conversely, weighted kappa scores for intraobserver agreement were mostly substantial or better for nodular regenerative hyperplasia (5/6 observers) and portal tract collagenization (6/6 observers).

Representative illustrations are provided for important features in assessing long-surviving allografts such as interface activity (Figure 1); portal capillaritis or microvasculitis, obliterative portal venopathy (OPV); and periportal shunt-type vessels (Figure 2); Ishak fibrosis scoring (Figure 3); portal tract collagenization (Figure 4); and perivenular fibrosis (Figure 5).

DISCUSSION

Protocol biopsies continue to play an important role in identifying subclinical graft injury in long-term survivors following pediatric LT and in providing greater insights into mechanisms of indolent, clinically silent, allograft injury. Indeed, several studies have identified a high prevalence of unexplained graft inflammation and progressive fibrosis, which in some cases leads to graft failure.[2,26] Although these studies have provided important insights into the natural history of liver allograft injury, the use of several different methods to assess inflammation and fibrosis makes comparison between previously published studies difficult. Furthermore, histological assessments are semiquantitative in nature and are therefore prone to observer variability. To the best of our knowledge, this is the first study to

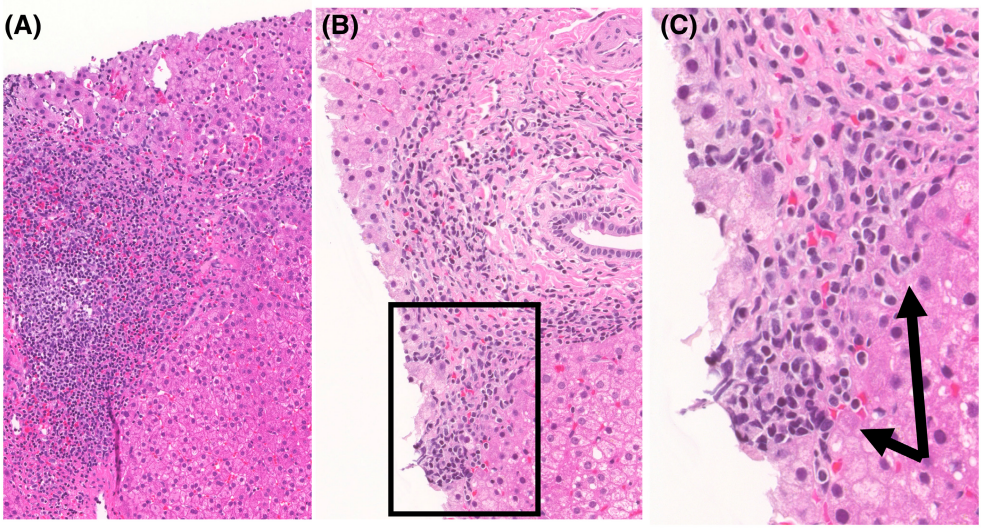

FIGURE 1 Interface necroinflammatory activity. (A) Representative portal tract and interface zone from biopsy in which all participating pathologists agreed that interface activity was present. (B) Representative portal tract and interface zone from a biopsy on which there was disagreement about the presence or absence of interface necroinflammatory activity. The black box highlights the area shown at higher magnification in (C). Higher magnification of one area of the interface zone showing lymphocytes penetrating into the edge of the lobule (arrows). [Colour figure can be viewed at wileyonlinelibrary.com]

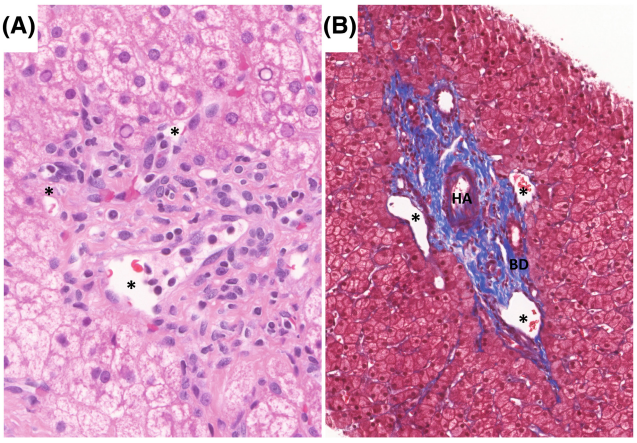

FIGURE 2 Important, but not widely appreciated or evaluated, findings in long-surviving pediatric liver allografts: (A) portal tract from a biopsy for which there was unanimous agreement that portal capillaritis or microvasculitis was present. Note the sludging and margination of mononuclear cells in the lumen of several portal capillaries and/or inlet venules (asterisks), endothelial cell hypertrophy, and opening of periportal shunt-type vessels. Anecdotally, this patient tested positive for multiple DSA with a DSA sum of >27,000. (B) OPV and periportal shunt-type vessels. Scores from this biopsy showed a high agreement on the presence of shunt-type vessels (asterisks). Note also the lack of a portal vein branch (OPV), compensatory enlargement of HA branches compared with the accompanying BD, and opening of periportal shunt-type vessels (asterisks). [Colour figure can be viewed at wileyonlinelibrary.com]

formally assess observer reproducibility across a range of features that are relevant for the evaluation of assessing late posttransplantation pediatric liver allograft biopsies, including features associated with circulating donor-specific antibodies (DSAs) and not commonly assessed in native liver biopsies (e.g., portal microvasculitis and collagenization, and three-compartment fibrosis scoring).

The term "idiopathic posttransplantation hepatitis (IPTH)" was proposed by the Banff Working Group to describe otherwise unexplained inflammatory changes in late posttransplantation biopsies[27] seen in up to 70% of late biopsies from children (Table 1).[28] A number of studies have observed an association between the presence of IPTH and the development of graft fibrosis, the severity of which usually progresses with time.[5,6,12,29,30] Up to 50%–70% of children with graft inflammation progress to bridging fibrosis or cirrhosis by 10 years after LT[5,29] and, in a recent study from London, this is now the commonest indication for late retransplantation (>10 years) in children.[26] However, other studies have failed to show an association between late graft inflammation and fibrosis.[7,8,10,11] It is imperative, therefore, to more precisely identify those patients at risk for fibrosis progression.

In the present study, a descriptive approach (none, minimal, mild, moderate, and severe) and a semiquantitative scoring system, as described by Ishak et al.,[31] were both used for the assessment of inflammation. For portal inflammation the level of interobserver agreement was higher using the Ishak system than the descriptive approach—kappa scores for the second round of reviews were 0.44 (unweighted) and 0.97 (weighted) for Ishak versus 0.32 and 0.72 for the descriptive approach. The Ishak system likewise achieved higher kappa scores for the assessment of lobular inflammation (0.54 unweighted, 0.92 weighted) compared

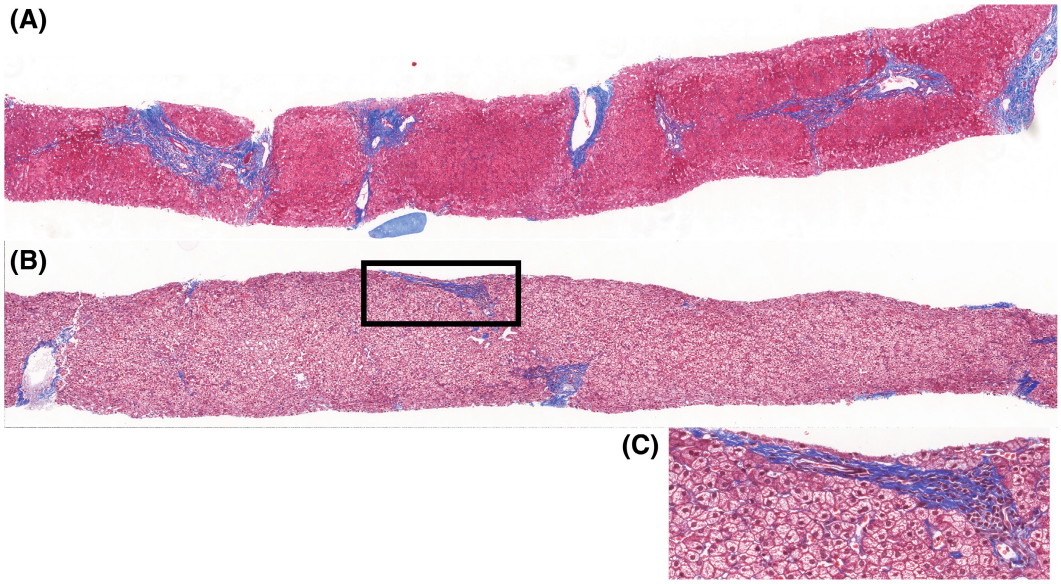

FIGURE 3 Fibrosis scoring. (A) This biopsy was unanimously scored as showing bridging fibrosis (Ishak fibrosis score = 3) by all pathologists. By contrast, the scores for the biopsy shown in (B) showed some disagreement among pathologists about the Ishak fibrosis stage. This presumably occurred because of the fibrous tract highlighted by the black box, which is shown at higher magnification in (C). Several pathologists thought that this represented a longitudinally sampled fibrous tract, as evidenced by the longitudinally sampled HA branch that it encases, whereas others scored it as bridging fibrosis. [Colour figure can be viewed at wileyonlinelibrary.com]

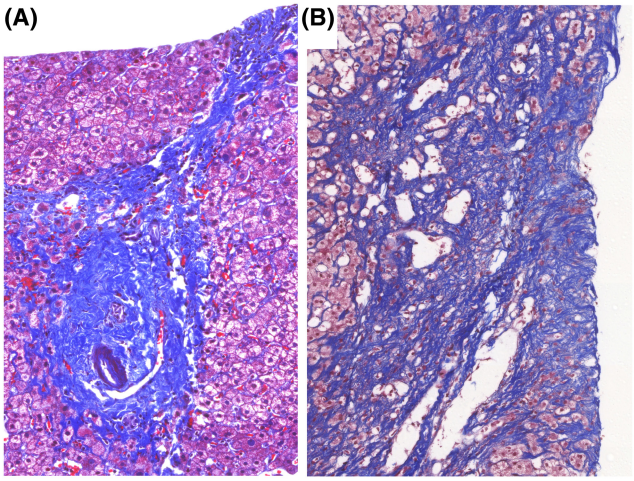

FIGURE 4 Portal tract collagenization: Representative examples of biopsies where (A) all pathologists agreed or (B) where there was some level of disagreement. Note the dense, almost keloid appearance of the portal tract connective tissue in (A) as compared with (B), which shows significant fibrosis but a basket-weave pattern of portal and periportal fibrosis. [Colour figure can be viewed at wileyonlinelibrary.com]

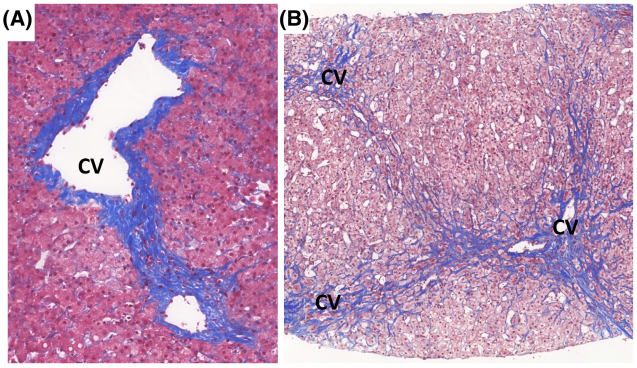

FIGURE 5 Examples of biopsies where there was a high level of agreement for (A) moderate perivenular fibrosis and (B) severe perivenular fibrosis with central-to-central bridging fibrosis. Note that the fibrosis in and around the CVs also extends into the perivenular sinusoids in a star-shaped pattern. [Colour figure can be viewed at wileyonlinelibrary.com]

with the descriptive approach (0.46 unweighted, 0.77 weighted). Both approaches had similar kappa scores for the assessment of interface hepatitis—Ishak kappa scores were 0.64 (unweighted) and 0.94 (weighted) compared with scores of 0.65 and 0.95 using the descriptive approach. The higher reproducibility that was achieved using the Ishak system to assess inflammation may relate to the fact that the participating pathologists were all familiar with this system, which has been widely used to assess the severity of inflammation in chronic inflammatory diseases involving the native liver, particularly hepatitis C. Other advantages of the Ishak system are that (1) it provides a wide range of possible scores, which increases granularity for assessing the severity of graft inflammation and (2) the confluent necrosis component provides a method to score the severity of centrilobular necroinflammatory activity (central perivenulitis), which is likely to be a manifestation of alloimmune injury in the liver allograft.

Three main patterns of fibrosis have been recognized in late posttransplantation biopsies—portal, sinusoidal,

and centrilobular.[10] A semiquantitative staging system proposed by the Brussels group (Venturi et al.) scores each compartment on a scale of 0 (none) to 3 (bridging) to produce a total liver allograft fibrosis (LAF) score of 0–9.[10] This system showed good observer interagreement and correlated better with morphometric fibrosis quantitation in pediatric liver allograft biopsies than METAVIR or Ishak scores. Subsequent studies have shown that the Venturi system is useful for assessing the dynamics of fibrosis progression in serial biopsies and have identified different risk factors associated with the individual patterns of fibrosis.[19,20,32,33]

For the assessment of graft fibrosis, three semiquantitative scoring systems were used. Two systems (METAVIR and Ishak) were originally designed to assess chronic inflammatory diseases associated with periportal fibrosis in the native liver, mainly hepatitis C. The third system (Venturi) allows the assessment of two other patterns of fibrosis that are commonly seen in late pediatric biopsies. Unweighted kappa scores for all three components of the Venturi scoring system were all in the range of "slight" interrater agreement (0.01–0.20) for the first round of slide reviews. This contrasts with an adequate level of interobserver agreement observed in the original study by Venturi et al.[10] Possible explanations for these discrepant findings include the fact that pathologists participating in the present study were unfamiliar with using the Venturi system and the use of different statistical methods for assessing observer agreement. For the second round of reviews, observer agreement for perivenular fibrosis improved to "fair," but remained unchanged for periportal and sinusoidal fibrosis. Interestingly unweighted kappa scores for intraobserver agreement were substantially higher, with the majority of scores falling in the "moderate" category. Weighted kappa statistics for all three Venturi components was substantial or near perfect, which suggests that the differences driving lower unweighted kappa scores were minor. Therefore, this approach may be suitable for large-scale studies documenting the natural history of fibrosis progression in late posttransplantation biopsies. Nevertheless, further attempts should be made to more precisely define the fibrosis stages in a more standardized manner.

Previous studies have suggested that interobserver reproducibility may be improved by pathologists working together to agree on how to apply these definitions.[31,34] For large-scale studies involving a central review of slides, participating pathologists are therefore routinely involved in collaborative discussions and training sessions similar to those that took place as part of the present study. The lack of improvement in kappa scores between the first and second round of reviews suggests that the participating pathologists had already reached agreement about scoring biopsies as far as was possible during the preceding meetings. Nevertheless, as predicted by our clinical colleagues, the disparity observed between unweighted and weighted kappa values illustrates that the discrepancies in scoring driving lower unweighted kappa values are minor and unlikely to be clinically relevant.

A recent consensus paper by the Banff Working Group has proposed criteria for the diagnosis of probable chronic AMR in the liver allograft.[17] Histological features required to diagnose chronic AMR are (1) otherwise unexplained and at least mild mononuclear portal and/or perivenular inflammation with interface and/or perivenular necroinflammatory activity and (2) at least moderate portal/periportal, sinusoidal, and/or perivenular fibrosis. These features are included among the assessments made in the current study and some are illustrated in Figures 1, 3, and 5. Features other than inflammation and fibrosis that have also been identified as relevant for the diagnosis of chronic AMR are portal collagenization (Figure 4) and OPV (Figure 2B).[22,35] Portal collagenization is characterized by the presence of a dense hyalinized (keloid-like) portal stroma, contrasting with the looser stroma seen in normal portal tracts and may be associated with a reduced number of portal microvessels. The relatively poor level of unweighted interobserver agreement that was seen for the assessment of portal collagenization in this study suggests that this pattern of graft injury may be difficult to identify consistently and reliably. OPV is characterized by loss and/or reduction in the caliber of portal vein branches. In the present study, using "number of portal tracts without a portal vein" as a marker for OPV, weighted kappa scores were "substantial" for interobserver agreement (0.75 and 0.72) and "substantial" or "almost perfect" for intraobserver agreement for five of the six observers, suggesting that this feature can be reliably assessed in late posttransplantation biopsies.

A potential limitation of the present study is the relatively small number of biopsies that were assessed. As discussed in the "Patients and Methods" section, the number of biopsies included in this study is adequate to provide reliable kappa statistics for observer reproducibility.[25] Furthermore, the 31 biopsies assessed in the present study is similar to the number of biopsies included in other published studies that have assessed the reproducibility of various liver biopsy findings—examples include 30 biopsies included in a study of chronic hepatitis C,[36] 32 biopsies reviewed in a study of alcoholic hepatitis,[37] and 31 biopsies assessed in a study of NAFLD.[38] Nevertheless, a larger study with an independent group of observers might be helpful to validate the reliability of the observations made in the present study.

There remain a number of areas where current understanding is incomplete and further insights are required. These include the natural history of late graft injury and the role of immunosuppression in preventing or reversing graft injury. Most studies observing an association between graft inflammation and fibrosis are cross-sectional in nature and there is a lack of

information concerning early features that may predict subsequent progressive graft injury. One study suggested that previous portal inflammation may predict the subsequent development of portal fibrosis,[33] but data concerning early histological predictors of progressive fibrosis are otherwise lacking. While there is some evidence to suggest that maintaining or increasing immunosuppression may help to prevent or reverse graft inflammation, there are conflicting data concerning the extent to which these strategies may be effective in preventing the development of graft fibrosis.[12,15,39–41] Well-designed longitudinal studies are therefore required to determine the natural history of late graft injury in pediatric liver allograft recipients and in particular to identify those children who are at risk of developing progressive graft fibrosis. Important future challenges include (1) collecting enough follow-up histopathological data in an era when protocol biopsies are being discouraged; (2) devising a clinically meaningful grading system with reproducible cut-offs for late-onset T-cell–mediated rejection that has an appearance quite similar to other causes of chronic hepatitis—for example, might the Ishak modified HAI (mHAI) system be more appropriate than the Banff system for assessing late graft inflammation that is likely to be immune mediated?; and (3) achieving a better understanding of the contribution of circulating DSAs to late graft injury, such as the development of portal microvasculitis, portal collagenization, and OPV. We believe that a comprehensive multifaceted assessment of all histological features that may be relevant in understanding the pathogenesis of late graft injury, such as the system described in this study, should help to address issues related to the aforementioned second and third points.

Lastly, emerging artificial intelligence (AI)–aided histopathology evaluation is an area that holds potential for the development of tools that enable pathology decision support. Coupled with robust training data, AI algorithms can create and refine models that identify key features from digitized whole-slide images, and present this information for screening, prognostic, or diagnostic utility. While strong correlations have been shown between conventional histological assessments and deep learning results, often the deterministic predictive imaging features cannot be elucidated, thereby limiting utility for both determining the scale and obtaining biological insights into mechanisms of liver injury.[42] Recent advancements in machine learning (such as recurrent neural networks) have begun to incorporate modifications that preserve relevant imaging features from sparse training sets and improve classification accuracy.

In conclusion, we suggest that the histological features that were evaluated in this study should provide a basis for assessing late posttransplantation biopsies from pediatric liver allograft recipients in a comprehensive and standardized manner. In addition, using validated scoring systems is recommended to assess disease severity. We believe that this approach will be particularly valuable for studies documenting the natural history of late graft injury in serial protocol biopsies and will also facilitate comparison of data between different centers. The high prevalence of abnormal graft histology (particularly fibrosis) that was seen in this study of biopsies obtained from a group of children who were clinically well with normal/near-normal liver biochemistry results underscores the importance of protocol biopsies in identifying subclinical graft injury that may impact on long-term graft function. We also hope that the guidelines provided in Table S2 will be relevant and useful to other pathologists who are involved in reporting late posttransplantation biopsies from children as well as biopsies from adults who have undergone liver transplantation as children.

ACKNOWLEDGMENTS

The authors thank Carla Lloyd, Graft Injury Group Project Manager, for providing assistance with organizing meetings of the group.

CONFLICT OF INTEREST

Nothing to report.

ORCID

Stefan G. Hübscher https://orcid.org/0000-0003-1609-8527
Sandy Feng https://orcid.org/0000-0002-2601-4350
Hyo Jeong Kang https://orcid.org/0000-0002-5285-8282

REFERENCES

1. Duffy JP, Kao K, Ko CY, Farmer DG, McDiarmid SV, Hong JC, et al. Long-term patient outcome and quality of life after liver transplantation: analysis of 20-year survivors. Ann Surg. 2010;252:652–61.
2. Kelly D, Verkade HJ, Rajanayagam J, McKiernan P, Mazariegos G, Hubscher S. Late graft hepatitis and fibrosis in pediatric liver allograft recipients: current concepts and future developments. Liver Transpl. 2016;22:1593–602.
3. Ng VL, Mazariegos GV, Kelly B, Horslen S, McDiarmid SV, Magee JC, et al. Barriers to ideal outcomes after pediatric liver transplantation. Pediatr Transplant. 2019;23:e13537.
4. Fouquet V, Alves A, Branchereau S, Grabar S, Debray D, Jacquemin E, et al. Long-term outcome of pediatric liver transplantation for biliary atresia: a 10-year follow-up in a single center. Liver Transpl. 2005;11:152–60.
5. Evans HM, Kelly DA, McKiernan PJ, Hubscher S. Progressive histological damage in liver allografts following pediatric liver transplantation. Hepatology. 2006;43:1109–17.
6. Ekong UD, Melin-Aldana H, Seshadri R, Lokar J, Harris D, Whitington PF, et al. Graft histology characteristics in long-term survivors of pediatric liver transplantation. Liver Transpl. 2008;14:1582–7.
7. Scheenstra R, Peeters PM, Verkade HJ, Gouw AS. Graft fibrosis after pediatric liver transplantation: ten years of follow-up. Hepatology. 2009;49:880–6.
8. Ueno T, Tanaka N, Ihara Y, Takama Y, Yamada H, Mushiake S, et al. Graft fibrosis in patients with biliary atresia after pediatric living-related liver transplantation. Pediatr Transplant. 2011;15:470–5.

9. Miyagawa-Hayashino A, Yoshizawa A, Uchida Y, Egawa H, Yurugi K, Masuda S, et al. Progressive graft fibrosis and donor-specific human leukocyte antigen antibodies in pediatric late liver allografts. Liver Transpl. 2012;18:1333–42.
10. Venturi C, Sempoux C, Bueno J, Ferreres Pinas JC, Bourdeaux C, Abarca-Quinones J, et al. Novel histologic scoring system for long-term allograft fibrosis after liver transplantation in children. Am J Transplant. 2012;12:2986–96.
11. Tomita H, Hoshino K, Fuchimoto Y, Ebinuma H, Ohkuma K, Tanami Y, et al. Acoustic radiation force impulse imaging for assessing graft fibrosis after pediatric living donor liver transplantation: a pilot study. Liver Transpl. 2013;19:1202–13.
12. Kosola S, Lampela H, Jalanko H, Makisalo H, Lohi J, Arola J, et al. Low-dose steroids associated with milder histological changes after pediatric liver transplantation. Liver Transpl. 2013;19:145–54.
13. Briem-Richter A, Ganschow R, Sornsakrin M, Brinkert F, Schirmer J, Schaefer H, et al. Liver allograft pathology in healthy pediatric liver transplant recipients. Pediatr Transplant. 2013;17:543–9.
14. Dattani N, Baker A, Quaglia A, Melendez HV, Rela M, Heaton N. Clinical and histological outcomes following living-related liver transplantation in children. Clin Res Hepatol Gastroenterol. 2014;38:164–71.
15. Sanada Y, Matsumoto K, Urahashi T, Ihara Y, Wakiya T, Okada N, et al. Protocol liver biopsy is the only examination that can detect mid-term graft fibrosis after pediatric liver transplantation. World J Gastroenterol. 2014;20:6638–50.
16. Feng S, Bucuvalas JC, Demetris AJ, Burrell BE, Spain KM, Kanaparthi S, et al. Evidence of chronic allograft injury in liver biopsies from long-term pediatric recipients of liver transplants. Gastroenterology. 2018;155:1838–51.e1837.
17. Demetris AJ, Bellamy C, Hubscher SG, O'Leary J, Randhawa PS, Feng S, et al. 2016 comprehensive update of the Banff working group on liver allograft pathology: introduction of antibody-mediated rejection. Am J Transplant. 2016;16:2816–35.
18. Chong AS, Rothstein DM, Safa K, Riella LV. Outstanding questions in transplantation: B cells, alloantibodies, and humoral rejection. Am J Transplant. 2019;19:2155–63.
19. Venturi C, Sempoux C, Quinones JA, Bourdeaux C, Hoyos SP, Sokal E, et al. Dynamics of allograft fibrosis in pediatric liver transplantation. Am J Transplant. 2014;14:1648–56.
20. Baas M, Gouw ASH, van den Heuvel MC, Hepkema BG, Peeters P, Verkade H, et al. Unique clinical conditions associated with different acinar regions of fibrosis in long-term surviving pediatric liver grafts. Pediatr Transplant. 2017;21:e12988.
21. Hubscher SG. What is the long-term outcome of the liver allograft? J Hepatol. 2011;55:702–17.
22. O'Leary JG, Cai J, Freeman R, Banuelos N, Hart B, Johnson M, et al. Proposed diagnostic criteria for chronic antibody-mediated rejection in liver allografts. Am J Transplant. 2016;16:603–14.
23. Kleiner DE, Brunt EM, Van Natta M, Behling C, Contos MJ, Cummings OW, et al. Design and validation of a histological scoring system for nonalcoholic fatty liver disease. Hepatology. 2005;41:1313–21.
24. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. Phys Ther. 2005;85:257–68.
25. Walter SD, Eliasziw M, Donner A. Sample size and optimal designs for reliability studies. Stat Med. 1998;17:101–10.
26. Neves Souza L, de Martino RB, Sanchez-Fueyo A, Rela M, Dhawan A, O'Grady J, et al. Histopathology of 460 liver allografts removed at retransplantation: a shift in disease patterns over 27 years. Clin Transplant. 2018;32:e13227.
27. Demetris AJ, Adeyi O, Bellamy CO, Clouston A, Charlotte F, Czaja A, et al. Liver biopsy interpretation for causes of late liver allograft dysfunction. Hepatology. 2006;44:489–501.
28. Shaikh OS, Demetris AJ. Idiopathic posttransplantation hepatitis? Liver Transpl. 2007;13:943–6.
29. Herzog D, Soglio DB, Fournet JC, Martin S, Marleau D, Alvarez F. Interface hepatitis is associated with a high incidence of late graft fibrosis in a group of tightly monitored pediatric orthotopic liver transplantation patients. Liver Transpl. 2008;14:946–55.
30. Feng S, Ekong UD, Lobritto SJ, Demetris AJ, Roberts JP, Rosenthal P, et al. Complete immunosuppression withdrawal and subsequent allograft function among pediatric recipients of parental living donor liver transplants. JAMA. 2012;307:283–93.
31. Ishak K, Baptista A, Bianchi L, Callea F, De Groote J, Gudat F, et al. Histological grading and staging of chronic hepatitis. J Hepatol. 1995;22:696–9.
32. Venturi C, Reding R, Quinones JA, Sokal E, Rahier J, Bueno J, et al. Relevance of activated hepatic stellate cells in predicting the development of pediatric liver allograft fibrosis. Liver Transpl. 2016;22:822–9.
33. Varma S, Ambroise J, Komuta M, Latinne D, Baldin P, Reding R, et al. Progressive fibrosis is driven by genetic predisposition, Allo-immunity, and inflammation in pediatric liver transplant recipients. EBioMedicine. 2016;9:346–55.
34. The METAVIR Cooperative Group. Inter- and intra-observer variation in the assessment of liver biopsy of chronic hepatitis C. Hepatology. 1994;20:15–20.
35. Guerra MR, Naini BV, Scapa JV, Reed EF, Busuttil RW, Cheng EY, et al. Obliterative portal venopathy: a histopathologic finding associated with chronic antibody-mediated rejection in pediatric liver allografts. Pediatr Transplant. 2018;22:e13124.
36. Bedossa P, Bioulacsage P, Callard P, Chevallier M, Degott C, Deugnier Y, et al. Intraobserver and Interobserver variations in liver-biopsy interpretation in patients with chronic hepatitis-C. Hepatology. 1994;20:15–20.
37. Horvath B, Allende D, Xie H, Guirguis J, Jeung J, Lapinski J, et al. Interobserver variability in scoring liver biopsies with a diagnosis of alcoholic hepatitis. Alcohol Clin Exp Res. 2017;41:1568–73.
38. Jung ES, Lee K, Yu E, Kang YK, Cho MY, Kim JM, et al. Interobserver agreement on pathologic features of liver biopsy tissue in patients with nonalcoholic fatty liver disease. J Pathol Transl Med. 2016;50:190–6.
39. Haller W, Lloyd C, Hubscher SG, Brown RM, McKiernan PJ, Kelly DA. Is there a role of corticosteroids in preventing graft hepatitis and fibrosis in liver allografts following Paediatric liver transplantation? Hepatology. 2009;50:754a.
40. Miyagawa-Hayashino A, Haga H, Egawa H, Hayashino Y, Uemoto S, Manabe T. Idiopathic post-transplantation hepatitis following living donor liver transplantation, and significance of autoantibody titre for outcome. Transpl Int. 2009;22:303–12.
41. Egawa H, Miyagawa-Hayashino A, Haga H, Teramukai S, Yoshizawa A, Ogawa K, et al. Non-inflammatory centrilobular sinusoidal fibrosis in pediatric liver transplant recipients under tacrolimus withdrawal. Hepatol Res. 2012;42:895–903.
42. Bosch J, Chung CH, Carrasco-Zevallos OM, Harrison SA, Abdelmalek MF, Shiffman ML, et al. A machine learning approach to liver histological evaluation predicts clinically significant portal hypertension in NASH cirrhosis. Hepatology. 2021;74:3146–60.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Hübscher SG, Feng S, Gouw ASH, Haga H, Kang HJ, Kelly DA, et al. Standardizing the histological assessment of late posttransplantation biopsies from pediatric liver allograft recipients. Liver Transpl. 2022;28:1475–1489. https://doi.org/10.1002/lt.26482