

ARTICLE



Multi-institutional development and validation of a novel histologic grading system for colonic graft-versus-host disease

Ayesha Farooq^{1,5}, Iván A. González^{2,6}, Kathleen Byrnes², Sarah M. Jenkins³, Christopher P. Hartley⁴ and Catherine E. Hagen⁴✉

© The Author(s), under exclusive licence to United States & Canadian Academy of Pathology 2022

Graft-versus-host disease (GVHD) remains a major complication for patients who have undergone hematopoietic stem cell transplantation. The Lerner system is the most widely used histologic grading score for gastrointestinal GVHD but its clinic utility is debated. The aim of our study was to develop a novel histologic grading system for gastrointestinal GVHD that incorporates independent evaluation of both apoptotic counts and crypt destruction. Colonic biopsies taken to assess for GVHD were retrospectively assessed for: Crypt damage (No crypt dropout or ulceration–0; crypt dropout without ulceration–1; ulceration–2) and crypt apoptotic counts (No apoptosis–0; 1–6 apoptotic bodies per 10 contiguous crypts–1; >6 apoptotic bodies per 10 contiguous crypts–2). The two scores were added together to get an overall grade (0–4). Alternative apoptotic cutoff points were examined. An apoptotic cutoff of >9 apoptotic bodies per 10 contiguous crypts marginally improved the area under the curve (AUC), but the AUCs from the resulting novel grade calculations were not significantly different ($p = 0.10$). Lerner grading was also applied. The study group consisted of an initial analysis cohort ($n = 191$) and a second validation cohort from a separate institution ($n = 97$). In the initial analysis cohort, our histologic grading system provided prognostic stratification for GVHD-related death within 6 months ($p = 0.0004$, AUC = 0.705). The Lerner system performed similarly in terms of providing prognostic stratification for GVHD-related death ($p = 0.0001$, AUC = 0.707). In the external validation cohort, our histologic grading system was not associated with GVHD-related death ($p = 0.14$, AUC = 0.621), but the Lerner system was associated with GVHD-related death ($p = 0.048$, AUC = 0.663). While our grading system may have some advantages compared to the Lerner system, due to lack of reproducibility we do not currently recommend widespread adoption of this system. Nonetheless, we present a standardized tool for assessing both apoptosis and crypt damage. Future studies assessing alternative histologic grading systems with external validation and further examination the lower apoptotic threshold for GVHD diagnosis are warranted.

Modern Pathology (2022) 35:1254–1261; <https://doi.org/10.1038/s41379-022-01065-z>

INTRODUCTION

Graft-versus-host disease (GVHD) remains a major complication for patients who have undergone hematopoietic stem cell transplantation, with lower gastrointestinal tract involvement showing a high association with non-relapse mortality¹. Diagnosis of GVHD requires consideration of clinical symptoms, endoscopic findings, and histologic features with exclusion of alternative diagnoses.

Histologically, gastrointestinal GVHD is characterized by crypt apoptosis, crypt dropout, and ulceration². The Lerner system is the most widely used histologic grading system for gastrointestinal GVHD and is a four-tiered system based on the extent of crypt destruction^{3,4}. The clinical utility of using the Lerner system for histologic grading of gastrointestinal GVHD is debated for several reasons including the criteria for establishing a diagnosis of grade 1 GVHD are not well established and the Lerner grade is inconsistently a predictor of clinical outcome. Furthermore, the Lerner system does not take into account the degree of apoptotic activity within biopsies^{2,5–11}. As such, evaluation of new histologic grading systems for gastrointestinal GVHD is warranted.

The aim of our study was to develop a novel histologic grading system for gastrointestinal GVHD that incorporates independent evaluation of both apoptotic counts and crypt destruction and to validate this system on a cohort of patients at a second institution.

MATERIALS AND METHODS

Initial analysis cohort and development of novel grading system

Approval of this study was obtained by the Institutional Review Board. A set of colon biopsies taken at our institution from 2008–2018 to assess for GVHD was retrospectively reviewed by one or two pathologists, at least one of whom has special expertise in gastrointestinal pathology (CEH, AF), for the maximum number of apoptotic bodies per 10 contiguous crypts, crypt dropout, and ulceration. Apoptotic counts were performed in the “hotspot” area of the slide. Cases that were reviewed by two pathologists were reviewed simultaneously at a multiheaded microscope. A novel histologic grading system was developed and applied to this cohort of patients. The grading system consisted of two scores, crypt damage and apoptotic counts, which were added together to get an overall grade (0–4).

¹Department of Pathology, Medical College of Wisconsin, Milwaukee, WI, USA. ²Department of Pathology and Immunology, Washington University School of Medicine, St. Louis, MO, USA. ³Department of Quantitative Health Sciences, Mayo Clinic, Rochester, MN, USA. ⁴Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA. ⁵Present address: Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA. ⁶Present address: Department of Pathology, Children’s Hospital of Philadelphia, Philadelphia, PA, USA. ✉email: Hagen.catherine@mayo.edu

Received: 2 March 2021 Revised: 25 February 2022 Accepted: 27 February 2022

Published online: 1 April 2022

Table 1. Summary of histologic scoring system.

Individual Components	Overall sum of individual components
Crypt damage No crypt dropout or ulceration – 0 Crypt dropout without ulceration – 1 Ulceration-2	Indeterminate for GVHD Overall score 1 Low Grade Overall score 2
Crypt apoptotic counts No apoptosis – 0 1–6 apoptotic bodies per 10 contiguous crypts – 1 >6 apoptotic bodies per 10 contiguous crypts – 2	Intermediate Grade Overall score 3 High Grade Overall score 4

The criteria for scoring were as follows: Crypt damage (No crypt dropout or ulceration – 0; crypt dropout without ulceration – 1; ulceration – 2) and crypt apoptotic counts (No apoptosis – 0; 1–6 apoptotic bodies per 10 contiguous crypts – 1; >6 apoptotic bodies per 10 contiguous crypts – 2) (Table 1). The cases were also graded according to the Lerner system^{3,4}. If a patient had biopsies from more than one site within the colon, the highest overall grade was used for analysis. If a patient had more than one set of biopsies performed during the study period, the score from the first biopsy \geq 14 days post-transplant was used for analysis. Biopsies performed < 14 days post-transplant were excluded from this study.

We and others have previously shown that patients with \leq 6 apoptotic bodies per 10 contiguous crypts may or may not represent GVHD and have proposed the category indeterminate for GVHD (iGVHD) for such cases, whereas cases with > 6 apoptotic bodies per 10 contiguous crypts are consistent with definitive GVHD. This cutoff point was therefore selected to align with previous studies^{6–8}. Alternative apoptotic cutoff points were examined. An apoptotic cutoff of > 9 apoptotic bodies per 10 contiguous crypts marginally improved the area under the curve (AUC), but the AUCs from the resulting novel grade calculations were not significantly different ($p = 0.10$). Likewise, a more detailed categorization of crypt damage was also assessed to include: no crypt dropout, focal crypt dropout (involving one biopsy fragment), multifocal crypt dropout (involving more than one biopsy fragment), diffuse crypt dropout (involving all biopsy fragments), or ulceration. This categorization again did not significantly alter the performance of the grading system based on the AUC value and therefore to simplify the system, the crypt damage analysis system with fewer categories was selected (Table 2).

External validation cohort

To assess reproducibility in a separate patient population, the histologic scoring system was retrospectively applied to a second cohort of patients from a second institution who underwent colonic biopsy to assess for GVHD from 2010–2015. The histologic scoring system was applied by two additional pathologists (IG, KB) who also reviewed the cases simultaneously at a double-headed scope. A Lerner grade was also assigned to these cases.

Clinical information

Corresponding clinical information including patient age, sex, underlying disease, type of transplantation, time from transplantation, endoscopic findings, evidence of extraintestinal GVHD, treatment for GVHD, and survival information was collected from chart review at each corresponding institution.

Statistical analysis

Patient and biopsy characteristics were summarized with frequencies and percentages or medians, ranges, and interquartile ranges (25th and 75th percentiles), as appropriate. The primary outcome of GVHD-related death within 6-months was assessed among patients with at least 6 months of post-biopsy follow-up time, or death within 6-months. GVHD-related death was considered when clinical notes contributed the patient's death in part or fully to complications of GVHD. Within each of the initial and external validation cohorts, distributions of patient and biopsy characteristics were compared between those without versus with a GVHD-related death within 6-months with Fisher's exact tests or Wilcoxon rank-sum tests for categorical or ordinal variables, respectively. The percentage of patients

with GVHD-related death was reported along with 95% confidence intervals (CI) using the score method. Logistic regression models were also used to examine associations with the primary outcome, and the AUCs were reported. AUCs were compared using the DeLong method for the overall test for equality of areas. P -values less than 0.05 were considered statistically significant. Analyses were performed using SAS version 9.4 (SAS Institute Inc., Cary, NC) or R (R Core Team, 2019, Vienna, Austria).

RESULTS

Initial evaluation of histologic scoring system

The initial analysis cohort from our institution consisted of colonic biopsies from 191 patients (median age 58, range 20–75; M:F 1.1:1). The indication for stem cell transplantation was the following: acute myeloid leukemia ($n = 59$, 31.0%), multiple myeloma ($n = 44$, 23.0%), non-Hodgkin's lymphoma ($n = 39$, 20.4%), myelodysplastic syndrome ($n = 17$, 8.9%), acute lymphoblastic leukemia ($n = 15$, 7.9%), myelofibrosis ($n = 8$, 4.2%), chronic myelomonocytic leukemia ($n = 5$, 2.6%), aplastic anemia ($n = 2$, 1.0%), acute promyelocytic leukemia ($n = 1$, 0.5%), and classical Hodgkin's lymphoma ($n = 1$, 0.5%). The median time from transplantation to biopsy was 61 days (range 14–3650) and the median follow up time post-biopsy was 1.4 years (range 0–9.8).

Twenty-one patients within this cohort had biopsies that were called negative for GVHD (i.e., no apoptosis or crypt dropout; histologic score of zero). Of these patients, one patient later developed GVHD and died a GVHD-related death within 6-months of biopsy. The remaining patients with biopsies negative for GVHD were all alive at last follow up or died due to causes other than GVHD. Patients with biopsies negative for GVHD had a GVHD-specific survival at 6 months similar to patients with an overall histologic score of 1 or 2 (Table 2).

Forty-seven patients had evidence of crypt dropout and 32 had evidence of ulceration. The median apoptotic count per 10 contiguous crypts for all patients was 6 (range 0–76). The crypt damage score ($p = 0.002$, AUC = 0.686) and the apoptotic score ($p = 0.02$, AUC = 0.651) were each significantly associated with GVHD-related death within 6 months. The overall 5-category histologic score ($p = 0.0005$, AUC = 0.705) and the categorized (low, intermediate, high) histologic score ($p = 0.0004$, AUC = 0.705) were also associated with GVHD-related death within 6 months (Table 2).

Increasing histologic grade was significantly associated with an increasing rate of ulcer or exudate present on endoscopy ($p < 0.0001$), evidence of GVHD at extraintestinal sites ($p = 0.03$), treatment for GVHD ($p = 0.0002$), and GVHD-related death within 6 months ($p = 0.0004$) (Table 3).

Validation cohort from second institution

The validation cohort from a second institution consisted of colonic biopsies from 97 patients (median age 50.7 years, range 1.4–72.2; M:F 1.3:1). The indication for stem cell transplantation was the following: acute myeloid leukemia ($n = 48$, 49.5%), non-Hodgkin's lymphoma ($n = 18$, 18.6%), acute lymphoblastic leukemia ($n = 10$, 10.3%), myelodysplastic syndrome ($n = 8$, 8.2%), hemoglobinopathy ($n = 5$, 5.2%), other inherited bone marrow deficiency ($n = 2$, 2.1%), aplastic anemia ($n = 2$, 2.1%), Hodgkin's lymphoma ($n = 1$, 1.0%), acute promyelocytic leukemia ($n = 1$, 1.0%), chronic myelomonocytic leukemia ($n = 1$, 1.0%), and hemophagocytic lymphohistiocytosis ($n = 1$, 1.0%). The median time from transplantation to biopsy was 69 days (range 15–709) and the median follow up time from biopsy was 0.4 years (range 0–8.9).

Twenty-seven patients had evidence of crypt dropout and 28 had evidence of ulceration. The median apoptotic count per 10 contiguous crypts for all patients was 18 (range 1–51). Neither the crypt score ($p = 0.05$, AUC = 0.660) nor the apoptotic score ($p = 1.0$, AUC = 0.505) was associated with GVHD-related death within 6 months in this cohort. The overall 5-tier histologic grade also

Table 2. GVHD-related death within 6 months among all cases based on initial biopsy findings in analysis cohort and external validation cohort.

	Analysis Cohort (N = 191)				External Validation Cohort (N = 97)				
	# Patients	GVHD death (N = 26)	95% CI	P-value ^a	AUC ^b	# Patients	GVHD death (N = 19)	P-value ^a	AUC ^b
Apoptotic score [version 1]^c									
No apoptosis [0]	22	1 (4.5%)	(0.8%, 21.8%)	0.02	0.651	0	0	1.0	0.505
1–6 apoptotic bodies [1]	77	6 (7.8%)	(3.6%, 16.0%)			11	2 (18.2%)		
>6 apoptotic bodies [2]	92	19 (20.7%)	(13.6%, 30.0%)			86	17 (19.8%)		
Apoptotic score [version 2]^d									
No apoptosis [0]	22	1 (4.5%)	(0.8%, 21.8%)	0.002	0.692	0	0	0.76	0.523
1–9 apoptotic bodies [1]	98	7 (7.1%)	(3.5%, 14.0%)			22	5 (22.7%)		
>9 apoptotic bodies [2]	71	18 (25.4%)	(16.7%, 36.6%)			75	14 (18.7%)		
Crypt score									
No crypt dropout [0]	112	8 (7.1%)	(3.7%, 13.5%)	0.002	0.686	42	6 (14.3%)	0.05	0.660
Crypt dropout [1]	47	8 (17.0%)	(8.9%, 30.1%)			27	3 (11.1%)		
Ulceration [2]	32	10 (31.3%)	(18.0%, 48.6%)			28	10 (35.7%)		
Crypt score: modified									
No crypt dropout	112	8 (7.1%)	(3.7%, 13.5%)	0.0005	0.688	42	6 (14.3%)	0.03	0.649
Focal crypt dropout	15	2 (13.3%)	(3.7%, 37.9%)			12	0 (0.0%)		
Multifocal crypt dropout	28	6 (21.4%)	(10.2%, 39.5%)			11	2 (18.2%)		
Diffuse crypt dropout	3	0 (0.0%)	(0%, 56.1%)			4	1 (25.0%)		
Ulceration	32	10 (31.3%)	(18.0%, 48.6%)			28	10 (35.7%)		
Novel grade [version 1]^c									
0	21	1 (4.8%)	(0.8%, 22.7%)	0.0005	0.705	0	0	0.07	0.628
1	65	5 (7.7%)	(3.3%, 16.8%)			10	1 (10.0%)		
2	36	2 (5.6%)	(1.5%, 18.1%)			32	5 (15.6%)		
3	41	9 (22.0%)	(12.0%, 36.7%)			28	4 (14.3%)		
4	28	9 (32.1%)	(17.9%, 50.7%)			27	9 (33.3%)		
Novel grade category [version 1]^c									
Low (score 0–2)	122	8 (6.6%)	(3.4%, 12.4%)	0.0004	0.705	42	6 (14.3%)	0.14	0.621
Intermediate (score 3)	41	9 (22.0%)	(12.0%, 36.7%)			28	4 (14.3%)		
High (score 4)	28	9 (32.1%)	(17.9%, 50.7%)			27	9 (33.3%)		
Novel grade [version 2]^d									
0	21	1 (4.8%)	(0.8%, 22.7%)	0.0002	0.719	0	0	0.15	0.604
1	80	5 (6.3%)	(2.7%, 13.8%)			18	2 (11.1%)		
2	25	2 (8.0%)	(2.2%, 25.0%)			26	5 (19.2%)		
3	39	10 (25.6%)	(14.6%, 41.1%)			27	4 (14.8%)		
4	26	8 (30.8%)	(16.5%, 50.0%)			26	8 (30.8%)		
Novel grade category [version 2]^d									
Low (score 0–2)	126	8 (6.3%)	(3.3%, 12.0%)	0.0002	0.710	44	7 (15.9%)	0.29	0.600
Intermediate (score 3)	39	10 (25.6%)	(14.6%, 41.1%)			27	4 (14.8%)		
High (score 4)	26	8 (30.8%)	(16.5%, 50.0%)			26	8 (30.8%)		
Endoscopy findings (categorized)									
Normal/Subtle	124	12 (9.7%)	(5.6%, 16.2%)	0.002	0.657	87	18 (20.7%)	0.68	0.526
Exudate/Ulcer	42	13 (31.0%)	(19.1%, 46.0%)			9	1 (11.1%)		
Lerner grade (categorized)									
0–2	133	9 (6.8%)	(3.6%, 12.4%)	0.0001	0.707	41	6 (14.6%)	0.048	0.663

Table 2. continued

	Analysis Cohort (N = 191)					External Validation Cohort (N = 97)			
	# Patients	GVHD death (N = 26)	95% CI	P-value ^a	AUC ^b	# Patients	GVHD death (N = 19)	P-value ^a	AUC ^b
3	26	7 (26.9%)	(13.7%, 46.1%)			28	3 (10.7%)		
4	32	10 (31.3%)	(18.0%, 48.6%)			28	10 (35.7%)		
Other organ GVHD									
No	98	8 (8.2%)	(4.2%, 15.3%)	0.03	0.619	60	13 (21.7%)	0.60	0.541
Yes	93	18 (19.4%)	(12.6%, 28.5%)			37	6 (16.2%)		

^aP-values comparing death between groups using Fisher's exact test for apoptotic score, crypt score, novel grade category, endoscopy, Lerner grade category, and other organ GVHD; Wilcoxon rank-sum for 5-level novel grade and 5-level modified crypt score.

^bArea under the curve from logistic regression models estimating the probability of GVHD-related death within 6-months.

^cVersion 1 of the apoptotic score; 0 apoptosis = 0, 1–6 apoptosis = 1, >6 apoptosis = 2; this version was utilized in the final version of the scoring system.

^dVersion 2 of the apoptotic score; 0 apoptosis = 0, 1–9 apoptosis = 1, >9apoptosis = 2.

N (row %) shown unless otherwise specified.

Table 3. Comparison of Categorized Novel Grading System with Clinical Variables in Analysis Cohort and Validation Cohort.

Analysis cohort Validation cohort	Low Grade (Grade 0–2) (N = 122) (N = 42)	Intermediate Grade (Grade 3) (N = 41) (N = 28)	High Grade (Grade 4) (N = 28) (N = 27)	P-value
Median age (range)	58.0 (20.0–75.0)	58.0 (22.4–72.0)	60.8 (26.0–71.0)	0.50
	49.7 (1.4–69.1)	53.6 (4.3–70.5)	46.9 (3.1–72.2)	0.62
M:F	1:1.03	1.05:1	2.5:1	0.10
	2:1	1.33:1	1:1.45	0.11
Median days post-transplant (range)	75 (14–3650)	51 (14–363)	62 (20–1218)	0.41
	76 (16–709)	78 (15–358)	52 (16–184)	0.50
Median years follow-up (range)	1.8 (0–9.8)	1.1 (0–6.0)	0.8 (0–8.7)	0.002
	0.4 (0–7.7)	0.2 (0–8.9)	0.6 (0–7.6)	0.57
Ulcer or exudate on endoscopy	8/100 (8.0%)	13/39 (33.3%)	21/27 (77.8%)	<0.0001
	1/42 (2.4%)	2/28 (7.1%)	6/26 (23.1%)	0.02
Treated for GVHD	84 (68.9%)	37 (90.2%)	28 (100.0%)	0.0002
	42 (100.0%)	28 (100.0%)	26 (96.3%)	0.27
Evidence of extraintestinal GVHD	53 (43.4%)	20 (48.8%)	20 (71.4%)	0.03
	16 (38.1%)	12 (42.9%)	9 (33.3%)	0.77
GVHD-related death w/in 6 mos.	8 (6.6%)	9 (22.0%)	9 (32.1%)	0.0004
	6 (14.3%)	4 (14.3%)	9 (33.3%)	0.11

was not significantly associated with GVHD-related death within 6 months ($p = 0.07$, AUC = 0.628) (Table 2).

Increasing grade was associated with an increasing rate of ulcer or exudate present on endoscopy ($p = 0.02$). However, the overall grade was not significantly associated with evidence of extraintestinal GVHD ($p = 0.77$) or GVHD-related death within 6 months ($p = 0.11$) in this cohort (Table 3).

Comparison of clinicopathologic variables between institutions

The initial analysis cohort from our institution was compared to the cohort from the second institution to determine any differences between the two institutions. The patient population at the second institution included pediatric patients, and the overall age in this cohort was significantly younger compared to our institution (median 50.7 vs. 58.0, $p < 0.0001$). Patients from the second institution were less likely to have ulceration or exudate present on endoscopy (9.4% vs. 25.3%, $p = 0.002$), and more likely to be treated for GVHD (99.0% vs. 78.0%, $p < 0.0001$). According to our new grading system, the cases from our institution tended to have lower overall histologic grade ($p < 0.0001$); and according to

the Lerner grading system, our institution also had lower overall grades compared to the second institution ($p < 0.0001$). The median follow up from biopsy was also longer for cases from the initial analysis cohort (1.4 vs. 0.4 years, $p = 0.001$).

There was no significant difference in patient sex ($p = 0.54$), evidence of extraintestinal GVHD ($p = 0.09$), time from transplantation to biopsy ($p = 0.46$), or rate of GVHD-related death at 6 months ($p = 0.19$) (Table 4).

Comparison to Lerner system

According to the Lerner system with all patients included ($n = 288$), 21 patients (7.3%) were graded as negative for GVHD, 118 patients (41.0%) were graded as grade 1, 35 (12.2%) as grade 2, 54 (18.8%) as grade 3, and 60 (20.8%) as grade 4. According to our grading system, 21 patients (7.3%) were graded as negative for GVHD, 75 (26.0%) cases were graded as grade 1, 68 (23.6%) cases as grade 2, 69 (24.0%) cases as grade 3, and 55 (19.1%) cases as grade 4. Utilizing our novel grading system, 12 (4.2%) cases were downgraded, and 61 (21.2%) cases were upgraded compared to the Lerner system. Forty-three cases that would have been classified as Lerner grade 1, were upgraded to grade 2 utilizing our

Table 4. Comparison of Clinicopathologic Variables between the Analysis Cohort and the External Validation Cohort.

	Analysis Cohort (N = 191)	External Validation Cohort (N = 97)	P-value^a	Total (N = 288)
Male sex	101 (52.9%)	55 (56.7%)	0.54	156 (54.2%)
Age at biopsy, years				
Median (Range)	58.0 (20.0–75.0)	50.7 (1.4, 72.2)	<0.0001	57.0 (1.4, 75.0)
Apoptotic count per 10 contiguous crypts				
Median (Range)	6 (0–76)	18 (1–51)	<0.0001	10 (0–76)
Q1, Q3	2, 17	10, 26		3, 22
Apoptotic score				
No apoptosis [0]	22 (11.5%)	0 (0.0%)	<0.0001	22 (7.6%)
1–6 apoptotic bodies [1]	77 (40.3%)	11 (11.3%)		88 (30.6%)
>6 apoptotic bodies [2]	92 (48.2%)	86 (88.7%)		178 (61.8%)
Crypt score				
No crypt dropout [0]	112 (58.6%)	42 (43.3%)	0.02	154 (53.5%)
Crypt dropout [1]	47 (24.6%)	27 (27.8%)		74 (25.7%)
Ulceration [2]	32 (16.8%)	28 (28.9%)		60 (20.8%)
Novel grade (apoptotic + crypt score) – original apoptotic scoring				
0	21 (11.0%)	0 (0.0%)	<0.0001	21 (7.3%)
1	65 (34.0%)	10 (10.3%)		75 (26.0%)
2	36 (18.8%)	32 (33.0%)		68 (23.6%)
3	41 (21.5%)	28 (28.9%)		69 (24.0%)
4	28 (14.7%)	27 (27.8%)		55 (19.1%)
Endoscopy findings				
Normal	77 (46.4%)	64 (66.7%)	0.0003	141 (53.8%)
Subtle	47 (28.3%)	23 (24.0%)		70 (26.7%)
Exudate	8 (4.8%)	5 (5.2%)		13 (5.0%)
Ulcer	34 (20.5%)	4 (4.2%)		38 (14.5%)
Lerner grade				
0	21 (11.0%)	0 (0.0%)	<0.0001	21 (7.3%)
1	91 (47.6%)	27 (27.8%)		118 (41.0%)
2	21 (11.0%)	14 (14.4%)		35 (12.2%)
3	26 (13.6%)	28 (28.9%)		54 (18.8%)
4	32 (16.8%)	28 (28.9%)		60 (20.8%)
Other organ GVHD				
No	98 (51.3%)	60 (61.9%)	0.09	158 (54.9%)
Yes	93 (48.7%)	37 (38.1%)		130 (45.1%)
Treatment for GVHD				
No	42 (22.0%)	1 (1.0%)	<0.0001	43 (14.9%)
Yes	149 (78.0%)	96 (99.0%)		245 (85.1%)
Days from transplant to biopsy				
Median (Range)	61 (14–3650)	69 (15–709)	0.46	65 (13–3650)
Q1, Q3	28, 182	32, 122		30, 168
Total follow-up years since biopsy^b				
Median (Range)	1.4 (0–9.8)	0.4 (0–8.9)	0.001	1.1 (0–9.8)
Q1, Q3	0.4, 3.2	0.1, 3.0		0.2, 3.2
Status at last follow up, N				
Alive	95	30		125
Deceased	96	67		163
GVHD-related death within 6 months	26 (13.6%)	19 (19.6%)	0.19	45 (15.6%)

^aP-values comparing distributions using chi-square tests for nominal characteristics; Wilcoxon rank-sum for ordinal and continuous characteristics.

^bAll patients who were not deceased within 6 months have at least 6 months of follow-up available. The external validation cohort is more recent (less follow-up time available).

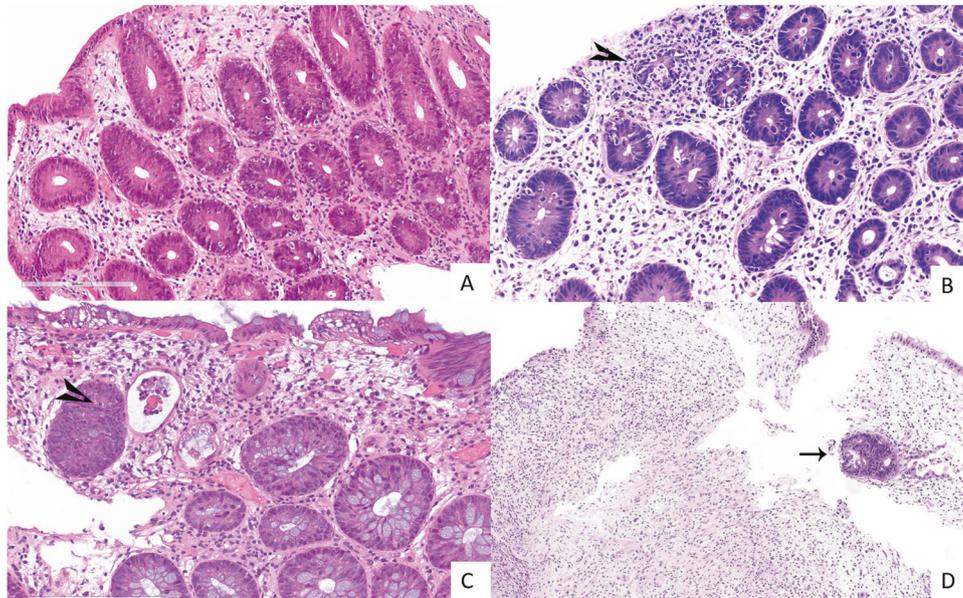


Fig. 1 Histologic examples of cases that changed grade from the Lerner system compared to our system. A case with abundant crypt apoptosis and no crypt dropout. According to the Lerner system this would be called grade 1 but would be classified as grade 2 (still low grade) by our system (A). A case with abundant crypt apoptosis and single crypt dropout (arrowhead). According to the Lerner system this would be grade 2 but would be upgraded to grade 3 by our system (B). A case with contiguous crypt dropout and rare crypt apoptosis (arrowhead). According to the Lerner system this would be grade 3 but would be downgraded to grade 2 by our system (C). A case with extensive ulceration and near-total loss of crypts. Crypt apoptosis is noted in the remaining crypts (arrow) but the overall count is low due to paucity of remaining crypts. According to the Lerner system this would be grade 4 but qualifies as grade 3 by our system (D).

grading system. Of these 43 patients, 5 (11.6%) died a GVHD-related death within 6 months. 18 cases that were classified as Lerner grade 2 were reclassified as grade 3. Of these 18 patients, 1 (5.6%) died a GVHD-related death within 6 months. Seven cases that were graded as Lerner grade 3 were reclassified as grade 2. None of these 7 patients died a GVHD-related death. Four cases that were called Lerner grade 4 were reclassified as grade 3 and one as grade 2 (due to extensive ulceration with no apoptosis). Of these five patients, 2 (40%) patients died a GVHD-related death. (Fig. 1).

The categorized Lerner grade (grades 0–2 combined, grade 3, grade 4) was associated with GVHD-related death within 6 months in both the initial analysis cohort ($p = 0.0001$, AUC = 0.707) and the external validation cohort ($p = 0.048$, AUC = 0.663) (Table 2).

DISCUSSION

GVHD remains a significant source of non-relapse mortality in patients who have undergone hematopoietic stem cell transplantation. Histologically, GVHD is characterized by crypt apoptosis and crypt destruction. The Lerner system has historically been utilized to histologically grade cases of gastrointestinal GVHD and is based on the extent of crypt destruction³. The Lerner system has been criticized for various reasons including poorly defined criteria for grade 1 disease and variable association with clinical outcome. Furthermore, the grading system focuses primarily on crypt destruction without assessment of the degree of apoptotic activity, a key histologic feature of GVHD^{2,5,9–11}.

We developed a novel histologic grading system for gastrointestinal GVHD which utilizes assessment of both the degree of apoptosis and extent of crypt destruction. In the initial analysis cohort, our histologic grading system provided prognostic stratification for GVHD-related death within 6 months and was associated with clinical evidence of GVHD including ulcer and/or exudate on endoscopy and evidence of extraintestinal GVHD. The

Lerner system performed similarly in terms of providing prognostic stratification for GVHD-related death within this cohort.

The Lerner system was developed nearly 50 years ago on patients with predominantly severe GVHD³. As such, the grading system focused primarily on the degree of crypt destruction. In recent decades, treatment and prophylactic regimens for GVHD have significantly improved, with patients being less likely to develop severe GVHD. Furthermore those who do develop severe GVHD are more likely to have better outcomes^{12,13}. Not surprisingly, the 2015 NIH consensus statement recommended examining new histologic grading systems based on the degree of apoptotic activity². Myerson et al. recently proposed a five-tiered GVHD grading system based on the average number of apoptotic cells normalized to the number of biopsy fragments. This system predominantly focused on mild cases of GVHD and any cases with crypt destruction were lumped together into activity grade 5¹⁴.

To our knowledge, our grading system is the first to independently assess both apoptotic activity and crypt destruction. As such, we believe our system provides a more global histologic assessment of GVHD and is applicable to GVHD of varying histologic severity. For instance, cases with crypt dropout but little apoptotic activity may represent cases with longstanding injury that is resolving (grade 2) whereas cases with abundant apoptotic activity and crypt dropout (grade 3) likely represent ongoing injury. According to the system by Myerson et al. both cases would be regarded as grade 5¹⁴.

We are not able to make a direct prognostic comparison between our grading system and the one developed by Myerson et al. as our methodologies for counting apoptosis differed¹⁴. However, our system does provide a simplified and less labor-intensive method of assessing apoptotic activity, which certainly is an important consideration for busy, practicing pathologists. Additionally, the main endpoint for evaluation in the study by Myerson et al. was treatment for GVHD. We compare our grading system to several clinical parameters, including

GVHD-specific survival, endoscopic findings, and evidence of extraintestinal GVHD.

Compared to the Lerner system we provide clear histologic criteria for each diagnostic category. Patients with 1–6 apoptotic bodies per 10 contiguous crypts and no crypt damage (overall score 1) would be considered as indeterminate for GVHD, while patients with a score of ≥ 2 would be considered as definitive GVHD. Patients with a score of 1 had a similar survival outcome as patients with a score of 0 or 2. Despite the similar survival of patients with scores 0–2, we believe clearly separating these diagnostic categories is clinically important. We and other authors have previously shown that patients with ≤ 6 apoptotic bodies per 10 contiguous crypts and no crypt damage (score 1) are more likely to be managed conservatively and may have symptom resolution without increased immunosuppression^{6–8}.

Of note, when we were exploring different apoptotic counts for our grading system, a cutoff of >9 apoptotic bodies per 10 contiguous crypts did have a marginally better AUC (0.692 vs. 0.651) for GVHD-related death within 6 months, but the AUC from the resulting novel grade calculation was not significantly different. As such, we chose to keep the apoptotic cutoff of >6 to align with previous studies. Establishing an apoptotic threshold for diagnosis of GVHD requires a balancing act of sensitivity and specificity. In previous studies, the cutoff of >6 was chosen to align with criteria for diagnosis of acute cellular rejection in the small bowel, but at least one follow up study has supported this cutoff point provides the optimal AUC for diagnosis of GVHD (specificity 100%, sensitivity 59.4%) compared to alternative cutoffs. In that study, a cutoff of >9 did provide 100% specificity for the diagnosis of GVHD, but the sensitivity was low (34%)^{6,8}. Choosing a higher apoptotic cutoff could potentially spare patients unnecessary immunosuppression and subsequent opportunistic infection, but with the risk of sacrificing sensitivity. Further studies specifically addressing the lower apoptotic threshold for diagnosis of colonic GVHD are necessary.

One potential limitation of our grading system would be evaluation of biopsy cases with extensive ulceration and near-complete to complete loss of crypts (Fig. 1D). Five cases graded as Lerner grade 4 were downgraded to grade 3 or grade 2 based on our scoring system and of these patients, 2 died a GVHD-related death. It therefore seems possible that apoptotic counts may be underestimated in cases with extensive ulceration. It is hard to draw any major conclusions based on a small number of cases and therefore examination of a larger number of such cases would be necessary to determine how to appropriately classify them.

Unfortunately, our grading system did not provide prognostic stratification in the external validation cohort. Notable differences did exist between the patient populations between the two institutions, including inclusion of pediatric patients at one institution, differing rates of treatment for GVHD, differing follow up lengths, and differing rates of ulcer/exudate on endoscopy. Additionally, the histologic severity of GVHD at our institution tended to be less severe compared to that of the second institution. It is well known that predictive performance of risk models may worsen substantially on external validation. Many authors do not go through the extra step of externally validating their findings by different authors¹⁵. While the lack of reproducibility was a disappointing finding, we feel we have done a thorough job evaluating the utility of our grading system, a measure previous similar studies have not performed.

In summary, we have developed a novel histologic grading system for gastrointestinal GVHD which separately assesses both apoptotic activity and crypt damage. In the initial analysis cohort, our grading system was associated with clinical evidence of GVHD and provided prognostic stratification. Compared to the Lerner system, our grading system, takes into account degree of

apoptosis and more clearly defines criteria for indeterminate and definitive cases of GVHD. Unfortunately, our grading system did not provide prognostic stratification in the external validation cohort. While our grading system may have some advantages compared to the Lerner system, due to lack of reproducibility we do not currently recommend widespread adoption of this system. Nonetheless we present a standardized tool that could potentially be useful in ongoing research in this challenging area. Future studies assessing alternative histologic grading systems with external validation are warranted.

DATA AVAILABILITY

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

REFERENCES

- Harris, A. C. et al. International, multicenter standardization of acute graft-versus-host disease clinical data collection: a report from the Mount Sinai Acute GVHD International Consortium. *Biol. Blood Marrow Transpl.* **22**, 4–10 (2016).
- Shulman, H. M. et al. NIH Consensus development project on criteria for clinical trials in chronic graft-versus-host disease: II. The 2014 Pathology Working Group Report. *Biol. Blood Marrow Transpl. J. Am. Soc. Blood Marrow Transpl.* **21**, 589–603 (2015).
- Lerner, K. G. et al. Histopathology of graft-vs.-host reaction (GvHR) in human recipients of marrow from HL-A-matched sibling donors. *Transpl. Proc.* **6**, 367–371 (1974).
- Washington, K. & Jagasia, M. Pathology of graft-versus-host disease in the gastrointestinal tract. *Hum. Pathol.* **40**, 909–917 (2009).
- Kreft, A. et al. Consensus diagnostic histopathological criteria for acute gastrointestinal graft versus host disease improve interobserver reproducibility. *Virchows Arch.* **467**, 255–263 (2015).
- Lin, J., Fan, R., Zhao, Z., Cummings, O. W. & Chen, S. Is the presence of 6 or fewer crypt apoptotic bodies sufficient for diagnosis of graft versus host disease? A decade of experience at a single institution. *Am. J. Surg. Pathol.* **37**, 539–547 (2013).
- Rowan, D. J., Hartley, C. P., Carrillo-Polanco, L. F., Oshima, K. & Hagen, C. E. Diagnostic phrasing is independently correlated with the decision to treat for graft-versus-host disease: retrospective review of colon biopsies with rare apoptosis. *Histopathology* **69**, 802–811 (2016).
- Gomez, A. J., Arai, S., Higgins, J. P. & Kambham, N. Clinicopathologic threshold of acute colorectal graft-versus-host disease. *Arch. Pathol. Lab. Med.* **140**, 570–577 (2016).
- Castilla-Llorente, C. et al. Prognostic factors and outcomes of severe gastrointestinal GVHD after allogeneic hematopoietic cell transplantation. *Bone Marrow Transpl.* **49**, 966–971 (2014).
- Ferrara, J. L. M. et al. Regenerating islet-derived 3-alpha is a biomarker of gastrointestinal graft-versus-host disease. *Blood* **118**, 6702–6708 (2011).
- Melson, J., Jakate, S., Fung, H., Arai, S. & Keshavarzian, A. Crypt loss is a marker of clinical severity of acute gastrointestinal graft-versus-host disease. *Am. J. Hematol.* **82**, 881–886 (2007).
- Gooley, T. A. et al. Reduced mortality after allogeneic hematopoietic-cell transplantation. *N. Engl. J. Med.* **363**, 2091–2101 (2010).
- El-Jawahri, A. et al. Improved treatment-related mortality and overall survival of patients with Grade IV Acute GVHD in the modern years. *Biol. Blood Marrow Transpl. J. Am. Soc. Blood Marrow Transpl.* **22**, 910–918 (2016).
- Myerson, D., Steinbach, G., Gooley, T. A. & Shulman, H. M. Graft-versus-host disease of the gut: a histologic activity grading system and Validation. *Biol. Blood Marrow Transpl.* **23**, 1573–1579 (2017).
- Siontis, G. C. M., Tzoulaki, I., Castaldi, P. J. & Ioannidis, J. P. A. External validation of new risk prediction models is infrequent and reveals worse prognostic discrimination. *J. Clin. Epidemiol.* **68**, 25–34 (2015).

AUTHOR CONTRIBUTIONS

C.E.H., A.F., and C.P.H. performed study concept and design; C.E.H., A.F., I.A.G., and K.B. performed data acquisition; C.E.H., S.M.J., and C.P.H. performed data interpretation and analysis; all authors read and provided critical review of the paper.

COMPETING INTERESTS

The authors declare no competing interests.

ETHICS APPROVAL/CONSENT TO PARTICIPATE

This study was approved by the institutional review board (IRB ID PRO00023473) and performed with waiver of informed consent; the study was performed in accordance with Declaration of Helsinki.

ADDITIONAL INFORMATION

Correspondence and requests for materials should be addressed to Catherine E. Hagen.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.