GASTROENTEROLOGY 2011;140:1472–1480

CLINICAL ADVANCES IN LIVER, PANCREAS, AND BILIARY TRACT

Comparability of Probable and Definite Autoimmune Hepatitis by International Diagnostic Scoring Criteria

ALBERT J. CZAJA

Division of Gastroenterology and Hepatology, Mayo Clinic College of Medicine, Rochester, Minnesota

BACKGROUND & AIMS: The diagnostic scoring systems for autoimmune hepatitis categorize some patients as having probable disease; this designation can affect treatment strategies and recruitment to clinical studies. A retrospective study was performed to determine the bases for the classification of probable autoimmune hepatitis and its clinical importance. **METHODS:** The study included 185 adult patients who had been assessed at presentation for findings common to both international diagnostic scoring systems. **RESULTS:** Seventeen patients (9%) were graded as probable autoimmune hepatitis by the revised original scoring system, and 28 patients (15%) were similarly designated by the simplified scoring system. These patients were distinguished from those designated as definite autoimmune hepatitis by male sex, concurrent immune diseases, lower serum γ-globulin and immunoglobulin G levels, and lower titers of autoantibody. Patients with definite or probable designations by either scoring system responded similarly to conventional corticosteroid regimens during comparable intervals of treatment. Full, partial, or nonresponses and treatment dependence were evident in all diagnostic categories with similar frequencies. Twenty-seven patients designated as probable autoimmune hepatitis by one system were designated as definite autoimmune hepatitis by the other system. **CONCLUSIONS: The designation of probable autoimmune hepatitis by the international scoring systems is based on differences in clinical manifestations and does not reflect differences in the validity of the diagnosis or its treatment response. Large multicenter prospective studies are necessary to establish these observations.**

Keywords: Clinical Phenotypes; Treatment Responses; Diagnostic Classifications; Scoring Systems.

The diagnosis of autoimmune hepatitis has been codified by an international panel,[1] and diagnostic scoring systems are available to quantify the strength of the diagnosis before[1,2] and after corticosteroid therapy.[1] The revised original scoring system is a comprehensive template that grades multiple clinical, laboratory, and histological features,[1] and the simplified scoring system assesses 4 features deemed important by multivariate analyses.[2] These systems have not been validated by prospective studies, and the simplified scoring system does not assess treatment response.[3] Nevertheless, they have each been incorporated into diagnostic algorithms.[4,5]

Both the revised original and simplified diagnostic scoring systems render diagnoses of either definite or probable autoimmune hepatitis.[1,2] The nature and outcomes of patients with a probable diagnosis by one or both scoring systems are unknown, and it is unclear if they can be included in clinical studies containing patients with scores indicative of definite disease. Furthermore, it is uncertain that patients with a probable diagnosis by one system are similar to those classified as probable by the other system. Patients with a probable diagnosis of autoimmune hepatitis may have nonclassical features that warrant their designation as a separate syndrome.[6,7] Such patients may not have the same outcomes as patients with definite autoimmune hepatitis, and they should be studied and treated separately. Alternatively, patients with probable autoimmune hepatitis may have bona fide autoimmune hepatitis but with less pronounced immune manifestations.[8]

The revised original and simplified diagnostic scoring systems each grade the serum level of IgG and the degree of autoantibody production, and these factors can vary spontaneously during the course of the disease or reflect host-related differences in the intensity of immune expression.[9–11] The difference between definite and probable autoimmune hepatitis might simply reflect these spontaneous variations or host-related differences rather than the nature of the disease. Such patients might well

Abbreviations used in this paper: AMA, antimitochondrial antibodies; ANA, antinuclear antibodies; IIF, indirect immunofluorescence; SMA, smooth muscle antibodies.

0016-5085/$36.00
doi:10.1053/j.gastro.2011.02.010

be included in clinical studies of treatment outcomes and treated with vigor and confidence in clinical practice. The designation of probable autoimmune hepatitis might wrongly impugn the legitimacy of an otherwise appropriate diagnosis. The elimination of an imprecise designation that might adversely impact on recruitment to clinical trials and patient care would improve each scoring system.

The goals of this retrospective study are to define the clinical phenotype of probable autoimmune hepatitis as defined by each scoring system, assess the distinctions between patients with definite and probable diagnoses by the same scoring system, and evaluate the responses of patients with each designation to conventional corticosteroid therapy. In this fashion, the nature and treatment outcomes of probable autoimmune hepatitis will be determined for each scoring system, and the need for revision of the scoring systems assessed.

Materials and Methods

Study Population

One hundred and eighty-five patients satisfied the codified clinical criteria for the diagnosis of autoimmune hepatitis at presentation,[1] and they comprised the study population. Study patients had been selected from 310 patients because they were adults (aged 18 years or older) and each had been assessed at presentation for histological features and other findings common to the revised original and simplified diagnostic scoring systems.[1,2] The accession interval was between 1967 and 2005, and 99 patients (54%) were assessed after 1990.

Of the 125 patients excluded from the analysis, 7 were younger than 18 years; 6 did not undergo histological assessments at presentation; 18 did not have determinations of serum IgG levels at presentation; and 77 underwent serological assessments by enzyme immunoassay rather than indirect immunofluorescence (IIF). Seventeen patients had multiple exclusion factors, including determinations of autoantibodies by enzyme immunoassay and absence of serum IgG assessments (10 patients), age younger than 18 years and serological determinations by enzyme immunoassay (4 patients), and lack of histological examination at presentation and no serological assessments by IIF (3 patients).

One hundred and forty-seven patients in the study population (79%) were women, and the mean age of the study group was 48 ± 1 year (range, 18−82 years; median age, 49 years). Thirty-six patients (19%) had smooth muscle antibodies (SMA) only; 51 patients (28%) had antinuclear antibodies (ANA) only; and 93 patients (50%) had both SMA and ANA at presentation. Five of 168 patients who were tested (3%) had antibodies to liver kidney microsome type 1 (anti-LKM1). Three patients with anti-LKM1 had only this marker; one patient had anti-LKM1 and SMA; and one other patient had anti-LKM1 and ANA. The study had been approved by the Institutional Review Board of the Mayo Clinic.

Clinical, Laboratory, and Scoring Assessments

Clinical examinations had been performed in accordance with a previously published protocol by one physician (AJC).[12] Concurrent extrahepatic disorders of an immune nature had been systematically sought in all patients.[12] Conventional laboratory tests of liver inflammation and function had been performed at each evaluation, and serum IgG concentrations had been assessed by immunonephelometry.[13] Smooth muscle antibodies had been determined by IIF on tissue sections of murine stomach and kidney in all patients; ANA had been assessed by IIF on HEp-2 cells in all patients; and anti-LKM1 had been evaluated by IIF on combined mouse kidney/stomach sections and confirmed by IIF of mouse liver sections in 168 patients (91%).[14] Antimitochondrial antibodies (AMA) had been determined in all patients by IIF of murine kidney and stomach tissue in 183 patients (99%) and by a previously reported enzyme immunoassay in 2 patients.[15] Hepatitis B surface antigen and antibodies to hepatitis C virus had been assessed in all patients by second-generation enzyme immunoassays. Stored frozen (−70°C) serum samples obtained at accession were tested for ANA, SMA, anti-LKM1, and hepatitis C virus in those patients who had accessed before the availability of the current assays. Diagnoses of definite autoimmune hepatitis, probable autoimmune hepatitis, or nondiagnostic chronic hepatitis were rendered pretreatment by applying the revised original diagnostic scoring system (Table 1) and the simplified diagnostic scoring system (Table 2).[1,2]

Histological Assessments

Liver tissue examinations had been performed at accession in all patients, and the liver specimens had been examined by members of the liver pathology working group at the Mayo Clinic. The pathological diagnoses were rendered in accordance with pre-established criteria.[16] Previous validation studies have indicated that the reproducibility of the histological interpretations by this method is 94%.[17] All tissue specimens had been judged to be typical or compatible with the diagnosis of autoimmune hepatitis.

Treatment Regimens

One hundred and fifty-eight patients (85%) had been treated with either prednisone in combination with azathioprine (96 patients) or a higher dose of prednisone alone (62 patients) in accordance with a previously published protocol.[18] Prednisone (30 mg daily) in conjunction with azathioprine (50 mg daily) or prednisone (60 mg daily) constituted the induction phase of treatment. Medication doses were then decreased according to a standardized protocol until maintenance doses of medi-

Table 1. Revised Original Diagnostic Scoring System of the International Autoimmune Hepatitis Group[a]

Variable	Result	Points	Variable	Result	Points
Sex	Female	+2	HLA	DR3 or DR4	+1
AP/AST (or ALT) ratio	>3	−2	Immune disease	Thyroiditis, colitis, others	+2
	<1.5	+2			
γ-Globulin or IgG level above normal	>2.0	+3	Other markers	Anti-SLA, actin, LC1, pANCA	+2
	1.5–2.0	+2			
	1.0–1.5	+1			
	<1.0	0			
ANA, SMA, or anti-LKM1 titers	>1:80	+3	Histological features	Interface hepatitis	+3
	1:80	+2		Plasmacytic	+1
	1:40	+1		Rosettes	+1
	<1:40	0		None of above	−5
				Biliary changes	−3
				Other features	−3
AMA	Positive	−4	Treatment response	Complete relapse	+2
					+3
Viral markers	Positive	−3			
	Negative	+3			
Drugs	Yes	−4	Pretreatment aggregate score:		
	No	+1	Definite diagnosis >15		
			Probable diagnosis 10−15		
Alcohol	<25 g/day	+2			
	>60 g/day	−2			

anti-LC1, antibodies to liver cytosol type 1; anti-LKM1, antibodies to liver kidney microsome type 1; anti-SLA, antibodies to soluble liver antigen; AP/AST (or ALT) ratio, ratio of alkaline phosphatase level to aspartate or alanine aminotransferase level; pANCA, perinuclear antineutrophil cytoplasmic antibodies.

[a]Reprinted from Alvarez F et al. International Autoimmune Hepatitis Group Report: review of criteria for diagnosis of autoimmune hepatitis. (table 5, page 934) J Hepatol 1999;31(5):929-938, with permission from Elsevier BV.

cation were achieved after 6 weeks.[18] The maintenance treatment schedule was prednisone (10 mg daily) in combination with azathioprine (50 mg daily) or prednisone alone (20 mg daily). Each treatment regimen had been shown earlier to be of comparable efficacy.[19] Twenty-seven patients (15%) had received investigational therapies, and these patients were not included in the analysis of treatment responses.

Treatment Responses

Treatment responses were assessed in a uniform fashion at 6-month intervals in all patients according to a previously published protocol.[18] Treatment responses were classified as full response, partial response, nonresponse, and treatment-dependent. Patients who responded fully to treatment had achieved absence of

Table 2. Simplified Diagnostic Scoring System of the International Autoimmune Hepatitis Group[a]

Variable	Result	Points
Autoantibodies		
Antinuclear antibodies or smooth muscle antibodies	≥1:40	+1
	≥1:80	+2
Antibodies to liver kidney microsome type 1	≥1:40	+2
Antibodies to soluble liver antigen	Positive	+2
Immunoglobulin level		
IgG	>Upper limit of normal	+1
	>1.1 times upper limit of normal	+2
Histological findings		
Morphological features	Compatible with autoimmune hepatitis	+1
	Typical of autoimmune hepatitis	+2
Viral disease		
Absence of viral hepatitis	No viral markers	+2
Pretreatment aggregate score		
Definite diagnosis		≥7
Probable diagnosis		6

[a]Reprinted from Hennes EM et al. Simplified criteria for the diagnosis of autoimmune hepatitis (table 2, page 171) Hepatology 2008;48(1): 169-176, with permission from Elsevier.

Table 3. Clinical Features of Definite and Probable Autoimmune Hepatitis by Scoring System

Clinical features at presentation	Revised original scoring system		Simplified scoring system	
	Definite AIH (n = 168)	Probable AIH (n = 17)	Definite AIH (n = 145)	Probable AIH (n = 28)
Age (*y*), mean ± SEM	48 ± 1	48 ± 4	47 ± 1	48 ± 3
Female/male, n	141/27[a]	6/11[a]	116/29	21/7
Concurrent immune features, n (%)	72 (43)[k]	1 (6)[k]	58 (40)	8 (29)
AST (*U/L*) (nl, ≤31 U/L), mean ± SEM	527 ± 33	486 ± 154	527 ± 34	512 ± 126
Bilirubin (*mg/dL*) (nl, ≤1.1 mg/dL), mean ± SEM	3.3 ± 0.3	3.5 ± 1	3.5 ± 0.4	2.4 ± 0.7
Alkaline phosphatase (*U/L*) (nl, 55−142 U/L), mean ± SEM	323 ± 17	429 ± 90	344 ± 21[l]	245 ± 29[l]
γ-globulin (*g/dL*) (nl, 0.7−1.7 g/dL), mean ± SEM	3.03 ± 0.1[b]	1.79 ± 0.2[b]	3.3 ± 0.1[c]	1.7 ± 0.1[c]
IgG (*mg/dL*) (nl, 700−1500 mg/dL), mean ± SEM	2788 ± 99[d]	1457 ± 159[d]	3070 ± 96[e]	1251 ± 43[e]
IgG ≥twice ULN, n (%)	67 (40)[i]	1 (6)[i]	68 (47)[f]	0 (0)[f]
AMA >1:40, n (%)	13 (8)	4 (24)	16 (11)	1 (4)
SMA >1:80, n (%)	84 (50)[h]	1 (6)[h]	80 (55)[g]	4 (14)[g]
ANA >1:80, n (%)	91 (54)	77 (46)	79 (54)	17 (61)
LKM1-positive, n (%)	4/152 (3)	1/16 (6)	5/133 (4)	0/24 (0)
SMA, ANA or LKM1 >1:80, n (%)	132 (79)[j]	8 (47)[j]	118 (81)	20 (71)
DRB1*03 or DRB1*04, n (%)	125/147 (85)	13/17 (76)	109/128 (85)	19/24 (79)

AIH, autoimmune hepatitis; AST, serum aspartate aminotransferase level; LKM1, antibodies to liver kidney microsome type 1; nl, normal range; SEM, standard error of mean; ULN, upper limit of normal range.
Significantly different from each other at level of:
[a,b,c,d,e,f,g]$P < .0001$.
[h]$P = .0005$.
[i,j]$P = .0007$.
[k]$P = .003$.
[l]$P = .04$.

symptoms, normal serum aspartate aminotransferase and γ-globulin levels, and normal histological features.[20] Patients who were partially responsive to treatment had absence of symptoms, laboratory improvement (serum aspartate aminotransferase level abnormal but <2-fold the upper limit of the normal range), and histological features that disclosed mild inflammatory activity (mild portal hepatitis or mildly active cirrhosis).[20] Patients who were nonresponsive to conventional treatment had worsening of clinical, laboratory, and/or histological features despite compliance with therapy, and they required treatment with high-dose prednisone alone (60 mg daily) or a lower dose (30 mg daily) in conjunction with azathioprine (150 mg daily).[21] Patients in whom conventional treatment was ongoing because a predefined response had not been achieved were classified as treatment-dependent.[18]

HLA Determinations

One hundred and sixty-four patients (89%) had been evaluated for the class II (DR locus) HLAs by polymerase chain reaction with sequence-specific oligonucleotide probes (72 patients), restriction fragment length polymorphism (89 patients), or microlymphocytotoxicity (3 patients) in accordance with methods described previously.[22,23] Only the frequencies of HLA DRB1*03 and HLA DRB1*04 were analyzed.[24]

Statistical Analyses

Fisher exact probability test was used to compare categorical variables, and the unpaired *t* test was used to compare differences in the means of continuous variables with a normal distribution. The Mann–Whitney test was used to compare differences in the means of quantitative variables with a skewed distribution. The variables for comparison had been formulated a priori and then assessed systematically, and an unadjusted *P* value <.05 was used to determine statistical significance. The mean ± standard error of the mean for the quantitative variables is presented in the tables and text.

Results

Frequency of Probable Autoimmune Hepatitis by Each Scoring System

Seventeen of the 185 patients (9%) who satisfied the international clinical criteria for autoimmune hepatitis at presentation were graded as having probable autoimmune hepatitis by the revised original scoring system (Table 3). The mean pretreatment score for autoimmune hepatitis by this system was 13.8 points (range, 10−15 points; median, 14 points). In contrast, 28 patients (15%) were graded as having probable autoimmune hepatitis by the simplified scoring system, and 12 patients (6%) were graded as nondiagnostic (Table 3). The 28 patients graded as probable autoimmune hepatitis by the simplified system each scored 6 points; whereas the 12 patients who failed to register as probable autoimmune hepatitis by the simplified scoring system each scored 5 points.

Table 4. Treatment Responses of Definite and Probable Autoimmune Hepatitis by Scoring System

Treatment responses	Revised original scoring system (n = 158)		Simplified scoring system (n = 150)	
	Definite AIH (n = 148)	Probable AIH (n = 10)	Definite AIH (n = 129)	Probable AIH (n = 21)
Full response, n (%)	19 (13)	1 (10)	17 (13)	1 (5)
Partial response, n (%)	69 (46)	6 (60)	62 (48)	10 (47)
No response, n (%)	19 (13)	0 (0)	17 (13)	2 (10)
Treatment dependent, n (%)	41 (28)	3 (30)	33 (26)	8 (38)
Duration of treatment until full, partial, or no response (*mo*), mean ± SEM; median (range)	22 ± 2; 15 (1−180)	14 ± 4; 28 (2−29)	22 ± 2; 15 (1−180)	22 ± 6; 18 (2−72)
Duration of treatment in treatment-dependent patients (*mo*), mean ± SEM; median (range)	43 ± 9; 12 (0−202)	4 ± 2; 6 (1−6)	41 ± 10; 12 (0−204)	39 ± 17; 14.5 (1−120)

AIH, autoimmune hepatitis; SEM, standard error of mean.

Distinctive Features Between Patients With Definite and Probable Autoimmune Hepatitis

The 17 patients with probable autoimmune hepatitis by the revised original scoring system were distinguished from the 168 patients designated as having definite autoimmune hepatitis within this same system by male sex, fewer concurrent immune diseases, lower serum γ-globulin and IgG levels, and lower frequency of high serum titers (>1:80) of SMA, ANA, or anti-LKM1 (Table 3). The frequency of cirrhosis at presentation was similar in patients with probable and definite autoimmune hepatitis at presentation (47% vs 26%, $P = .09$), and serum γ-globulin (3 ± 0.2 g/dL vs 2.9 ± 0.1 g/dL; $P = .4$) and IgG levels (2741 ± 180 mg/dL vs 2637 ± 112 mg/dL, $P = .6$) were comparable in patients with and without cirrhosis. Other features that might challenge their designation as autoimmune hepatitis, such as AMA positivity, high serum bilirubin and alkaline phosphatase levels, and low frequency of HLA DRB1*03 or DRB1*04, were similar to those with definite autoimmune hepatitis (Table 3).

The 28 patients with probable autoimmune hepatitis by the simplified scoring system were distinguished from the 145 patients designated as having definite autoimmune hepatitis within this same system by lower serum alkaline phosphatase, γ-globulin, and IgG levels and lower frequency of high titer (>1:80) SMA (Table 3). The frequency of cirrhosis at presentation was similar in the patients with probable and definite autoimmune hepatitis at presentation (14% vs 30%; $P = .1$), and serum γ-globulin (3.2 ± 0.2 g/dL vs 3 ± 0.1 g/dL; $P = .2$) and IgG levels (2908 ± 182 mg/dL vs 2726 ± 113 mg/dL; $P = .4$) were comparable in patients with and without cirrhosis. Other features that might challenge their designation as autoimmune hepatitis, such as AMA positivity, high serum bilirubin level, and HLA phenotype, were similar to those with definite autoimmune hepatitis (Table 3).

Comparison of Treatment Responses Between Definite and Probable Autoimmune Hepatitis

One hundred and fifty-eight patients with definite (148 patients) or probable (10 patients) autoimmune hepatitis by the revised original scoring system underwent conventional corticosteroid therapy, and treatment responses were similar between patients with these designations (Table 4). The frequencies of a full response ($P > .9$), partial response ($P = .5$), and no response ($P = .6$) were comparable between patients with definite and probable designations, and the treatment durations until a full, partial, or no response were also similar between the groups ($P = .4$) (Table 4). The number of patients undergoing active therapy who had not yet achieved a treatment response (treatment-dependent patients) was comparable in each diagnostic category ($P = .6$), and the duration of therapy was similar between treatment-dependent patients with definite or probable diagnoses of autoimmune hepatitis by the revised original scoring system ($P = .4$), albeit numbers in the latter diagnostic category were small (Table 4).

One hundred and fifty patients with definite (129 patients) or probable (21 patients) autoimmune hepatitis by the simplified scoring system underwent conventional corticosteroid therapy and treatment responses were similar between patients with these designations (Table 4). The frequencies of a full response ($P = .5$), partial response ($P > .9$), and no response ($P > .9$) were comparable between patients with definite and probable designations, and the treatment durations until a full, partial, or no response were similar ($P = .9$) (Table 4). The number of patients undergoing active therapy who had not yet achieved a treatment response (treatment-dependent patients) was comparable in each diagnostic category ($P = .3$), and the duration of therapy was similar between treatment-dependent patients with definite or probable diagnoses of autoimmune hepatitis by the simplified scoring system ($P = .9$) (Table 4).

Table 5. Clinical Features of Probable Autoimmune Hepatitis Distinguished by Only One Scoring System

Clinical features at presentation	Probable AIH only by revised original scoring system (n = 11)	Probable AIH only by simplified scoring system (n = 22)
Age (*y*), mean ± SEM	46 ± 5	48 ± 3
Female/male, n	5/6[b]	20/2[b]
Concurrent immune features, n (%)	1 (9)	8 (36)
AST (*U/L*) (nl, ≤31 U/L), mean ± SEM	455 ± 175	310 ± 95
Bilirubin (*mg/dL*) (nl, ≤1.1 mg/dL), mean ± SEM	3.9 ± 1	2.2 ± 0.7
Alkaline phosphatase (*U/L*) (nl, 55−142 U/L) , mean ± SEM	527 ± 124[a]	244 ± 29[a]
γ-globulin (*g/dL*) (nl, 0.7−1.7 g/dL), mean ± SEM	1.9 ± 0.3	1.7 ± 0.1
IgG (*mg/dL*) (nl, 700−1500 mg/dL), mean ± SEM	1596 ± 233	1265 ± 49
IgG ≥twice ULN, n (%)	1 (9)	0 (0)
AMA >1:40, n (%)	3 (27)[c]	0 (0)[c]
SMA, ANA, or LKM1 >1:80, n (%)	4 (36)	16 (73)
DRB1*03 or DRB1*04, n (%)	9/11 (82)	15/18 (83)

AIH, autoimmune hepatitis; AST, serum aspartate aminotransferase level; LKM1, antibodies to liver kidney microsome type 1; nl, normal range; SEM, standard error of mean; ULN, upper limit of normal range.
Significantly different from each other at level of:
[a]$P = .006$.
[b]$P = .008$.
[c]$P = .03$.

Comparison of Probable Autoimmune Hepatitis by Each Scoring System

Eleven patients had probable autoimmune hepatitis only by the revised original scoring system; 22 patients had probable autoimmune hepatitis only by the simplified scoring system; and 6 patients had probable autoimmune hepatitis by both scoring systems. The 11 patients with probable autoimmune hepatitis only by the revised original scoring system differed from the 22 patients with probable autoimmune hepatitis only by the simplified scoring system by male sex, higher serum alkaline phosphatase levels, and greater occurrence of AMA (Table 5).

Twenty-eight patients with probable autoimmune hepatitis by either scoring system underwent conventional corticosteroid therapy, including 7 with probable autoimmune hepatitis only by the revised original scoring system, 18 with probable autoimmune hepatitis only by the simplified scoring system, and 3 patients with probable autoimmune hepatitis by both scoring systems. The 7 treated patients with probable autoimmune hepatitis by only the revised original scoring system had similar frequencies of full response ($P = .5$), partial response ($P = .7$), no response ($P > .9$) and treatment-dependence ($P > .9$) as the 18 treated patients with probable autoimmune hepatitis by only the simplified scoring system (Table 6). Treatment durations for patients who achieved a response ($P = .4$) and those still undergoing active therapy ($P = .3$) were also similar between these patients (Table 6).

Discrepant Diagnoses Between Scoring Systems

Diagnoses were concordant between the scoring systems in 146 patients (79%) and discordant in 39 patients (21%). Of the 39 patients with discrepant diagnoses, 5 patients who were graded as probable autoimmune hepatitis by the simplified scoring system were graded as definite autoimmune hepatitis by the revised original scoring system. Six patients with a probable diagnosis of autoimmune hepatitis by the revised original scoring system were graded as nondiagnostic by the simplified scoring system; 22 patients graded as probable autoimmune hepatitis by the simplified scoring system were

Table 6. Treatment Responses of Probable Autoimmune Hepatitis Distinguished by Only One Scoring System

Treatment responses after conventional treatment	Probable AIH only by revised original scoring system (n = 7)	Probable AIH only by simplified scoring system (n = 18)
Full response, n (%)	1 (14)	1 (6)
Partial response, n (%)	4 (57)	8 (44)
No response, n (%)	0 (0)	2 (11)
Treatment-dependent, n (%)	2 (29)	7 (39)
Duration of treatment until full, partial, or no response (*mo*), mean ± SEM; median (range)	15 ± 4; 18 (5−29)	25 ± 7; 18 (3−72)
Duration of treatment in treatment-dependent patients (*mo*), mean ± SEM; median (range)	6; 6	45 ± 19; 20 (1−120)

AIH, autoimmune hepatitis; SEM, standard error of mean.

graded as having definite autoimmune hepatitis by the revised original scoring system; and 6 patients with definite autoimmune hepatitis by the revised original scoring system were categorized as nondiagnostic by the simplified scoring system. Twenty-seven patients designated as probable autoimmune hepatitis by one scoring system were designated as definite autoimmune hepatitis by the other scoring system.

Of the 12 patients graded as nondiagnostic for autoimmune hepatitis by the simplified scoring system, 6 had been graded as definite autoimmune hepatitis and 6 had been graded as probable autoimmune hepatitis by the revised original scoring system. These 12 patients were distinguished from the 173 other patients by lower serum levels of γ-globulin (1.3 ± 0.1 g/dL vs 3 ± 0.1 g/dL; $P < .0001$) and IgG (1084 ± 103 mg/dL vs 2776 ± 96 mg/dL; $P < .0001$) and lower frequency of high serum titers (>1:80) of autoantibodies (17% vs 80%; $P < .0001$). Eight of these 12 patients received conventional corticosteroid treatment. Five of the 8 treated patients responded to therapy (62%), including 2 patients with a full response and 3 patients with a partial response, after 9 ± 3 months of treatment (range, 3−19 months; median, 6 months). Three of the 8 treated patients (38%) continued to undergo therapy (treatment-dependent patients) after 27 ± 21 months (range, 6−68 months; median, 6 months).

Discussion

This study demonstrates that the pretreatment classification of autoimmune hepatitis as definite or probable by either the revised original or simplified diagnostic scoring systems defines patient subsets that differ in sex, frequency of concurrent immune diseases, degree of serum γ-globulin and IgG concentrations, and autoantibody titers but not by treatment response (Tables 3 and 4). Categorizations as definite or probable autoimmune hepatitis based on these differences do not characterize subgroups with disease of different severity or prognosis, and they may be prejudicial if they affect treatment decisions or study design by instilling diagnostic uncertainty. The diagnosis of autoimmune hepatitis does not require a certain number of phenotypic features or particular degree of laboratory abnormality,[4] and low level or absent immune manifestations do not preclude existence of the disease or need for treatment.[25–27] The clinical features at presentation of patients designated as probable autoimmune hepatitis by either scoring system (Table 3) indicated a disease severity that has been associated with a poor prognosis if left untreated.[4] Accordingly, corticosteroid therapy should not be withheld or patients excluded from clinical trials because of their arbitrary designation as having probable autoimmune hepatitis by a diagnostic scoring system.

Autoimmune hepatitis lacks a diagnostic laboratory feature or serological finding, and the diagnosis requires exclusion of other diseases that may resemble it.[4] The lack of a diagnostic phenotype might contribute to the inclusion of patients within the diagnosis who have atypical or variant features, and overlap syndromes of autoimmune hepatitis and primary sclerosing cholangitis or primary biliary cirrhosis have been described.[6,7] Such variants might be designated as probable autoimmune hepatitis by the current diagnostic scoring systems.[8] All patients in this study had histological examinations that supported the diagnosis of autoimmune hepatitis; they lacked prominent clinical or histological features of cholestasis; and the responses to corticosteroid therapy were similar regardless of the scoring system grades.

Variant syndromes are clinical diagnoses often requiring additional diagnostic studies, such as endoscopic retrograde cholangiography[28] or magnetic resonance cholangiography,[29] and their pursuit is typically driven by clinical suspicion. The diagnostic scoring systems have not been developed as discriminative indices to identify these variants; they do not espouse the use of routine cholangiography; they exclude patients with incompatible liver tissue examinations; and they have not performed well when applied to detect the overlap syndromes of autoimmune hepatitis and primary biliary cirrhosis[30] or primary sclerosing cholangitis.[31] This study indicates that patients with the probable designation of autoimmune hepatitis have clinical phenotypes and behavior that are similar to those of definite autoimmune hepatitis by whichever scoring system is applied and that the probable category is unlikely to harbor previously undetected variant syndromes.

Scores based on arbitrary serum levels of laboratory or serological findings assume that certain levels make the diagnosis of autoimmune hepatitis irrefutable or unsustainable. These levels have not been established and they are probably nonexistent.[10] None of the features graded in the scoring systems has disease specificity, prognostic importance, or durability. Even AMA can be accommodated within the diagnosis of autoimmune hepatitis, as evident in this study. AMAs in the absence of cholestatic clinical and histological changes have not altered the diagnosis of autoimmune hepatitis or affected the response to corticosteroid therapy,[15,32–34] and their presence in this study did not distinguish patients with definite or probable autoimmune hepatitis by either scoring system (Table 3). Furthermore, AMA can appear and disappear in patients with autoimmune hepatitis in a fashion similar to SMA and ANA.[10,15]

The revised original and simplified diagnostic scoring systems are intended to support and not prejudice clinical judgment. The shortcomings of the systems are evident in their lack of concordance in 39 patients (21%), including 12 patients (6%) who were scored as definite or probable autoimmune hepatitis by the revised original scoring system and as nondiagnostic by the simplified scoring system. Twenty-seven other patients with discrep-

ant classifications were variously categorized as having definite autoimmune hepatitis by one system or probable autoimmune hepatitis by the other system. None of these discrepant designations impacted on treatment outcomes (Table 6), but they each underscored the arbitrary and imprecise nature of the designations. The lack of concordance between the scoring systems in designating definite and probable autoimmune hepatitis is another reason to consider elimination of this distinction.

Clinical judgment has been the gold standard against which diagnostic scoring systems have been measured for liver diseases that lack a signature abnormality, such as autoimmune hepatitis[1] or drug-induced liver injury.[35] This judgment must continue to supersede the scoring results and remain the principal basis for diagnosis and treatment. The scores should reflect the presence or absence of individual components of the presentation and not grade the degree of particular abnormalities that do not influence treatment outcomes. The current approach implies that the diagnosis is less valid in individuals with appropriate but weakly expressed components of the disease. Future deliberations of the International Autoimmune Hepatitis Group should assess the need for a probable category and consider revision of the scoring systems to better reflect the composite nature of the presentation rather than the strength of its individual components.

Forty-four of the 158 patients who received conventional treatment regimens (28%) did not satisfy criteria for a full, partial, or nonresponse, and they continued to receive their original treatment regimen (Table 4). These patients might ultimately respond in a way that would allow their reclassification to another response category. Importantly, these patients were equally distributed among patients categorized as having definite or probable autoimmune hepatitis by either scoring system and their treatment durations were similar between the diagnostic designations (Table 4). Treatment dependence is a well-recognized outcome in the treatment of autoimmune hepatitis and it warrants distinction from full, probable, and nonresponses because such patients improve with therapy but not enough to consider treatment withdrawal.[4]

Small sample sizes limit confidence in drawing strong conclusions about the lack of difference between the subgroups of definite and probable autoimmune hepatitis. The required sample size to eliminate concern that a true difference between the study populations was overlooked varies according to the comparisons being performed and the degree of difference actually observed.[36] The closer the similarity between the populations, the larger the number of patients needed to exclude an unobserved statistical difference.

The *P* values observed in the laboratory comparisons of Table 3 that were not statistically significant varied from .08 to .9, and the *P* values of the treatment outcomes in Table 4 that were not statistically significant varied from .29 to .99. The number of patients required to fully exclude an unobserved statistical significance in all comparisons exceeds that in this study. Accordingly, the possibility that patients with definite and probable autoimmune hepatitis are actually different cannot be fully excluded. The findings in this study generate a hypothesis that requires a large multicenter prospective investigation to ensure that referral bias and small numbers are not confounding factors. The prospective application of the revised original and simplified scoring criteria to a group of consecutive, previously undiagnosed patients would define the clinical utility of each scoring system and assess the value of the probable and definite categories in a more circumspect fashion. Extensions of such an investigation could evaluate diagnostic differences between medical centers, the importance of liver tissue examinations or other diagnostic features in establishing the diagnosis, and even the need for therapy in patients with probable or definite designations.

In summary, the arbitrary designations of definite and probable autoimmune hepatitis by the revised original and simplified diagnostic scoring systems of the International Autoimmune Hepatitis Group are based mainly on differences in sex and degree of laboratory and serological abnormalities. These features do not impugn the legitimacy of the diagnosis of autoimmune hepatitis, and they do not reflect differences in disease severity or treatment response in this experience from a single institution. The findings suggest that these patients should not be excluded from treatment or clinical studies. Large multicenter prospective studies are necessary to establish these observations.

References

1. Alvarez F, Berg PA, Bianchi FB, et al. International Autoimmune Hepatitis Group Report: review of criteria for diagnosis of autoimmune hepatitis. J Hepatol 1999;31:929–938.
2. Hennes EM, Zeniya M, Czaja AJ, et al. Simplified criteria for the diagnosis of autoimmune hepatitis. Hepatology 2008;48:169–176.
3. Yeoman AD, Westbrook RH, Al-Chalabi T, et al. Diagnostic value and utility of the simplified International Autoimmune Hepatitis Group (IAIHG) criteria in acute and chronic liver disease. Hepatology 2009;50:538–545.
4. Manns MP, Czaja AJ, Gorham JD, et al. Practice Guidelines of the American Association for the Study of Liver Diseases. Diagnosis and management of autoimmune hepatitis. Hepatology 2010;51:2193–2213.
5. Czaja AJ, Manns MP. Advances in the diagnosis, pathogenesis and management of autoimmune hepatitis. Gastroenterology 2010;139:58–72.
6. Czaja AJ. The variant forms of autoimmune hepatitis. Ann Intern Med 1996;125:588–598.
7. Czaja AJ. Frequency and nature of the variant syndromes of autoimmune liver disease. Hepatology 1998;28:360–365.
8. Czaja AJ. Performance parameters of the diagnostic scoring systems for autoimmune hepatitis. Hepatology 2008;48:1540–1548.

9. Muratori L, Cataleta M, Muratori P, et al. Liver/kidney microsomal antibody type 1 and liver cytosol antibody type 1 concentrations in type 2 autoimmune hepatitis. Gut 1998;42:721–726.
10. Czaja AJ. Behavior and significance of autoantibodies in type 1 autoimmune hepatitis. J Hepatol 1999;30:394–401.
11. Luth S, Herkel J, Kanzler S, et al. Serologic markers compared with liver biopsy for monitoring disease activity in autoimmune hepatitis. J Clin Gastroenterol 2008;42:926–930.
12. Czaja AJ, Davis GL, Ludwig J, et al. Autoimmune features as determinants of prognosis in steroid-treated chronic active hepatitis of uncertain etiology. Gastroenterology 1983;85:713–717.
13. Czaja AJ, Wolf AM, Baggenstoss AH. Laboratory assessment of severe chronic active liver disease during and after corticosteroid therapy: correlation of serum transaminase and gamma globulin levels with histologic features. Gastroenterology 1981;80:687–692.
14. Czaja AJ, Homburger HA. Autoantibodies in liver disease. Gastroenterology 2001;120:239–249.
15. Montano-Loza AJ, Carpenter HA, Czaja AJ. Frequency, behavior, and prognostic implications of antimitochondrial antibodies in type 1 autoimmune hepatitis. J Clin Gastroenterol 2008;42:1047–1053.
16. Czaja AJ, Carpenter HA. Sensitivity, specificity, and predictability of biopsy interpretations in chronic hepatitis. Gastroenterology 1993;105:1824–1832.
17. Soloway RD, Baggenstoss AH, Schoenfield LJ, et al. Observer error and sampling variability tested in evaluation of hepatitis and cirrhosis by liver biopsy. Am J Dig Dis 1971;16:1082–1086.
18. Czaja AJ. Treatment of autoimmune hepatitis. Semin Liver Dis 2002;22:365–378.
19. Summerskill WH, Korman MG, Ammon HV, et al. Prednisone for chronic active liver disease: dose titration, standard dose, and combination with azathioprine compared. Gut 1975;16:876–883.
20. Montano-Loza AJ, Carpenter HA, Czaja AJ. Improving the end point of corticosteroid therapy in type 1 autoimmune hepatitis to reduce the frequency of relapse. Am J Gastroenterol 2007;102:1005–1012.
21. Montano-Loza AJ, Carpenter HA, Czaja AJ. Features associated with treatment failure in type 1 autoimmune hepatitis and predictive value of the model of end-stage liver disease. Hepatology 2007;46:1138–1145.
22. Czaja AJ, Carpenter HA, Santrach PJ, et al. Significance of HLA DR4 in type 1 autoimmune hepatitis. Gastroenterology 1993;105:1502–1507.
23. Czaja AJ, Strettell MD, Thomson LJ, et al. Associations between alleles of the major histocompatibility complex and type 1 autoimmune hepatitis. Hepatology 1997;25:317–323.
24. Donaldson PT, Doherty DG, Hayllar KM, et al. Susceptibility to autoimmune chronic active hepatitis: human leukocyte antigens DR4 and A1-B8-DR3 are independent risk factors. Hepatology 1991;13:701–706.
25. Czaja AJ, Carpenter HA, Santrach PJ, et al. The nature and prognosis of severe cryptogenic chronic active hepatitis. Gastroenterology 1993;104:1755–1761.
26. Gassert DJ, Garcia H, Tanaka K, et al. Corticosteroid-responsive cryptogenic chronic hepatitis: evidence for seronegative autoimmune hepatitis. Dig Dis Sci 2007;52:2433–2437.
27. Heringlake S, Schutte A, Flemming P, et al. Presumed cryptogenic liver disease in Germany: high prevalence of autoantibody-negative autoimmune hepatitis, low prevalence of NASH, no evidence for occult viral etiology. Z Gastroenterol 2009;47:417–423.
28. Perdigoto R, Carpenter HA, Czaja AJ. Frequency and significance of chronic ulcerative colitis in severe corticosteroid-treated autoimmune hepatitis. J Hepatol 1992;14:325–331.
29. Abdalian R, Dhar P, Jhaveri K, et al. Prevalence of sclerosing cholangitis in adults with autoimmune hepatitis: evaluating the role of routine magnetic resonance imaging. Hepatology 2008;47:949–957.
30. Talwalkar JA, Keach JC, Angulo P, et al. Overlap of autoimmune hepatitis and primary biliary cirrhosis: an evaluation of a modified scoring system. Am J Gastroenterol 2002;97:1191–1197.
31. Boberg KM, Fausa O, Haaland T, et al. Features of autoimmune hepatitis in primary sclerosing cholangitis: an evaluation of 114 primary sclerosing cholangitis patients according to a scoring system for the diagnosis of autoimmune hepatitis. Hepatology 1996;23:1369–1376.
32. Nezu S, Tanaka A, Yasui H, et al. Presence of antimitochondrial autoantibodies in patients with autoimmune hepatitis. J Gastroenterol Hepatol 2006;21:1448–1454.
33. Mishima S, Omagari K, Ohba K, et al. Clinical implications of antimitochondrial antibodies in type 1 autoimmune hepatitis: a longitudinal study. Hepatogastroenterology 2008;55:221–227.
34. O'Brien C, Joshi S, Feld JJ, et al. Long-term follow-up of antimitochondrial antibody-positive autoimmune hepatitis. Hepatology 2008;48:550–556.
35. Rockey DC, Seeff LB, Rochon J, et al. Causality assessment in drug-induced liver injury using a structured expert opinion process: comparison to the Roussel-Uclaf causality assessment method. Hepatology 2010;51:2117–2126.
36. Lenth RV. Statistical power calculations. J Anim Sci 2007;85:E24–E29.

Received November 17, 2010. Accepted February 6, 2011.

Correspondence

Address correspondence to: Albert J. Czaja, MD, Mayo Clinic College of Medicine, 200 First Street SW, Rochester, Minnesota 55905. e-mail: czaja.albert@mayo.edu; fax: (507) 284-0538.

Conflicts of interest

The author discloses no conflicts.