Accuracy of the Simplified Criteria for Autoimmune Hepatitis in Children: Systematic Review and Decision Analysis

José V. Arcos-Machancoses, Cristina Molera Busoms, Ecaterina Julio Tatis, María V. Bovo, Javier Martín de Carpi

Sant Joan de Déu Hospital, Barcelona. Department of Pediatric Gastroenterology, Hepatology and Nutrition, Spain

Background/objectives: Several studies have been conducted on the accuracy of simplified criteria for autoimmune hepatitis that were presented in 2008 as an alternative to original criteria. Our purpose is to summarize the evidence available regarding their accuracy in children and to carry out a basic clinical decision analysis based on it. **Methods:** Electronic and manual searches were performed with keywords related to diagnostic validity terms. Data from included studies were extracted, and summary estimates of accuracy measures were calculated. An effect model was chosen depending on heterogeneity, and the presence of publication bias was also studied. Therapeutic threshold was calculated based on the already published data. Through a Bayesian approach, simplified criteria's clinical utility was simulated, taking into account the meta-analyzed indicators and several assumptions on the prevalence of autoimmune hepatitis. **Results:** The search yielded 166 studies, four of which were finally included, providing a total population of 437 patients. Pooled sensitivity and specificity of the simplified criteria for the diagnosis of autoimmune hepatitis in children was 77% and 95%, respectively, with a diagnostic odds ratio of 67. No evidence of publication bias was found. For prevalences ranging from 8.5 to 85.7, the predictive value of either a positive or a negative result moved beyond the therapeutic threshold (estimated at 56%). **Conclusions:** The simplified criteria show high specificity and moderate sensitivity for the diagnosis of autoimmune hepatitis in children. A positive result can justify starting a therapeutic assay, but a negative result does not seem sufficient to rule out this condition. (J CLIN EXP HEPATOL 2019;9:147–155)

Autoimmune hepatitis (AIH) is an inflammatory liver disorder with a wide clinical spectrum that ranges from isolated acute or chronic hypertransaminasemia to acute liver failure (ALF).^{1,2} It is characterized by several analytical features, especially the presence of autoantibodies, high levels of immunoglobulin G (IgG) and, histologically, by liver lymphocytic or

lymphoplasmacytic chronic infiltration that is typically displayed as interface hepatitis.³ AIH tends to naturally evolve toward liver cirrhosis, but when treated, it responds properly in most patients.⁴

With reference to diagnosis, the International Autoimmune Hepatitis Group (IAIHG) proposed, in 1993, some classification criteria to facilitate comparison of diagnostic definitions in clinical studies.⁵ The criteria were revised in 1999 and, as a result, their specificity was improved up to 90%, showing also a very good performance in patients with few or atypical features of AIH.^{6,7} The IAIHG-revised diagnostic criteria have been proven suitable even in an exclusively pediatric population and are considered to be the best diagnostic reference, despite not being a truly gold standard.^{8,9} However, their practical application remains challenging for clinical use because of their complexity, including 13 categories, some of them impractical in children. To overcome these difficulties, the IAIHG proposed a simplified scoring system in 2008 that takes into account only the presence of autoantibodies, IgG levels, histopathology, and absence of viral markers.^{10,11} It ranges from 0 to 8 points, with an optimal cutoff of 6 for probable AIH and 7 for definite AIH. Some validation studies have been

Keywords: autoimmune hepatitis, diagnosis, sensitivity and specificity, meta-analysis

Received: 13.8.2018; **Accepted:** 30.10.2018; **Available online** 8 November 2018

Address for correspondence: J. V. Arcos-Machancoses, Sant Joan de Déu Hospital, Barcelona. Department of Pediatric Gastroenterology, Hepatology and Nutrition. Passeig de Sant Joan de Déu, 2, 08950 Esplugues de Llobregat, Barcelona (Catalonia), Spain.

E-mails: jvicentarcos@gmail.com or cristinamolera@gmail.com or ecaterina02@hotmail.com or mavibovo@gmail.com or javiermartin@sjdhospitalbarcelona.org

Abbreviations: AIH: autoimmune hepatitis; ALF: acute liver failure; CI: confidence interval; ESPGHAN: European Society for Paediatric Gastroenterology Hepatology and Nutrition; IAIHG: International Autoimmune Hepatitis Group; IgG: immunoglobulin G; NASPGHAN: North American Society for Pediatric Gastroenterology Hepatology and Nutrition; QUA-DAS: Quality Assessment of Diagnostic Accuracy Studies; ROC: receiver operating characteristic

<https://doi.org/10.1016/j.jceh.2018.10.006>

carried out in relation to the accuracy of the simplified criteria.^{12–20}

The aim of this study was to conduct a systematic review of studies reporting the main validity indexes for 2008 simplified criteria in children with AIH, diagnosed by expert clinical judgment based on 1999 classical criteria and also, if possible, summarize them through meta-analysis. In the second step, we intended to assess their clinical utility accordingly to the obtained accuracy measures and the prevalence of AIH in a real clinical scenario.

METHODS

The protocol for this systematic review and meta-analysis has been registered in the PROSPERO international prospective register of systematic reviews (www.crd.york.ac.uk/PROSPERO, registry number CRD42017081947). We followed the Preferred Reporting Items for Systematic Reviews and Meta-Analysis statement.²¹

Search Strategy and Study Selection

We carried out a systematic search for studies reporting the diagnostic accuracy of simplified criteria for AIH, since its publication in 2008. An initial search for articles with a similar objective was conducted using MEDLINE and the Cochrane Library. No results were obtained either in children or in adults.

First, the search strategy consisted in looking for articles in MEDLINE, Embase, Scopus, Trip database, Web of Science, and Biblioteca Virtual en Salud. A scheme of the search instructions, keywords, limits, and set operators used in each electronic database are provided in [Supplementary Information File 1](#). Two authors independently reviewed the title and abstract returned by the systematic search and excluded articles by duplicity, by original language other than English, Italian, French, German, Portuguese, or Spanish, or according to their objective, population, relevance, or design.

The selected articles' full text was read to confirm eligibility that was considered if they met these selection criteria: (1) AIH diagnostic reference explicitly based on the IAIHG 1999 reviewed criteria, (2) cross-sectional or case-control design, (3) provision of enough information to assess the study quality and to build the 2 × 2 contingency table to calculate the exact sensitivity and specificity (if needed, the corresponding or the first author was contacted to retrieve additional data), and (4) exclusively pediatric study population (younger than 18 years). Some other criteria were considered to exclude articles: (1) sample size below 20 patients and (2) final diagnosis in non-AIH group insufficiently detailed. In the second step, the citations provided by the remaining studies were also revised and included in the analysis if fulfilled the aforementioned

criteria. Additional relevant reports were identified by handsearching of the *Journal of Pediatric Gastroenterology and Nutrition* (looking for abstracts from European and North American specialty meetings) and the book of abstracts of the *Latin American Society of Pediatric Gastroenterology, Hepatology and Nutrition*. It was undertaken in March and April 2018. To avoid bias in the information abstraction process, primary data collection was also performed independently and double-checked by the same two searchers. Disagreements in each stage were resolved by consensus after discussion or after consultation with a third investigator.

Quality Assessment

The Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool was used to evaluate the methodological quality of the included articles.²² The two reviewers also performed it independently.

Statistical Analysis

It has been shown that labeling between probable and definite AIH according to the simplified criteria scoring (6 or 7 points, respectively) is based on variations in clinical manifestations and does not reflect differences in the reliability of the diagnosis.²³ Nevertheless, we decided to extract the true positive and true negative ratio for the cutoff of 6 points and of 7 points, separately, in each primary study.¹⁰ For this reason, we did not anticipate the need for conducting a threshold effect analysis. As simplified criteria only admit eight possible values, depicting that a receiver operating characteristic (ROC) curve would not draw a true curvilinear trace in the ROC space. In consequence, the meaning of a summary ROC curve, and its underlying area, would be difficult to interpret, so we decided not to calculate them.²⁴

As recommended by the Cochrane Diagnostic Test Accuracy Working Group, pooled sensitivity and specificity estimates were calculated by the bivariate random-effect regression model.²⁵ The inconsistency measure I^2 based on the Cochrane's Q statistic, was used to quantify the heterogeneity of the positive and negative likelihood ratios and the diagnostic odds ratio. A p-value less than 0.05 for the Q statistic and a $I^2 \geq 50\%$ was read as indicative of significant heterogeneity. In that case, the DerSimonian-Laird random-effects method was used to calculate these two last validity indicators. Otherwise, fixed-effects method was chosen.²⁶ All accuracy measures were represented with forest plots, and their 95% confidence interval (CI) was provided. The CIs of sensitivity and specificity were calculated using the Snedecor's F distribution method to compute the exact confidence limits for the binomial proportion.²⁷ Pooled sensitivity and specificity were also shown as a summary point in the ROC space with its 95% confidence region. Moreover, we used the

Deeks funnel plot asymmetry test to evaluate potential publication bias, considering as significant a p-value <0.1.²⁸

All data were analyzed with Meta-DiSc® (beta version 1.1.1, Clinical Biostatistics Unit, Ramón y Cajal Hospital, Madrid) and with Stata® (version 14.0, Stata Corporation, Texas) using the Metandi, Midas and Fagan commands.²⁹

Clinical Utility Analysis

Finally, we estimated the post-test probability of AIH based on a Bayesian approach, making use of the summary likelihood ratios obtained in the meta-analysis and the prevalence of AIH already reported by our group, in the setting of a population of children in clinical or analytical situation potentially attributable to AIH.³⁰ Fagan nomogram was used, and a probability-modifying graph was also plotted.³¹ Subsequently, to assess the simplified criteria clinical utility, the therapeutic threshold was calculated according to the model proposed by Pauker and Kassirer.³² Clinical utility was understood as the usefulness of the scoring system to decide whether to treat. Treatment benefits were obtained in terms of the complementary of proportion of patients with clinical, laboratory, and histological worsening, despite the use of immunosuppressant drugs. Costs were considered as the proportion of standard treatment and main adverse reactions. Simplified criteria were considered clinically useful if the therapeutic threshold was a value in between the prevalence of AIH and the post-test probability of AIH after a positive scoring.

RESULTS

Systematic Review of the Literature

The literature search yielded a total of 166 articles, seven of which were abstracts presented in scientific meetings, identified by the title. We studied the full text of 36 studies and contacted the author of one oral communication to gather more information on results and methods. A total of seven articles answering the question about the validity of simplified criteria for AIH diagnosis were excluded because the study was conducted only in patients older than 18 years, four of which were discarded after examining the abstract. Only two studies were dropped owing to language reasons, one in Chinese and another in Japanese. Both of them seemed also inappropriate for the aim of this systematic review. Four studies were finally selected, and no additional articles were added after analysis of references in the retained works (Table 1).^{19,20,30,33} In total, 437 patients were included (203 with AIH and 234 with alternative diagnoses). A flowchart depicting the selection process is included as Figure 1.

Quality Assessment of Primary Studies

The risk of bias and applicability concerns of primary studies are summarized in Figure 2. Patient selection domain is unclear regarding risk of bias because three of the four works included were designed as retrospective case-control studies. Thus, patients in both groups are at risk of being highly selected and hence, not fully representative of the whole clinical spectrum. In any study, the

Table 1 Characteristics of Included Studies.

Study →	Hiejima ¹⁹	Mileti ¹⁸	Gonçalves ³²	Arcos ²⁹
Year of publication	2011	2012	2017	2018
True positive rate (cutoff of 6 points)	11/20	34/37	40/46	72/100
True negative rate (cutoff of 6 points)	31/36	38/40	45/46	108/112
True positive rate (cutoff of 7 points)	8/20	25/31	34/46	45/100
True negative rate (cutoff of 7 points)	33/36	25/25	46/46	112/112
Diagnostic reference	IAIHG 1999 reviewed criteria and Mieli-Vergani's pediatric proposal (2009)	IAIHG 1999 reviewed criteria	IAIHG 1999 reviewed criteria and Mieli-Vergani's pediatric proposal (2009)	IAIHG 1999 reviewed criteria, clinical follow-up and medical reports review
Study design	Phase III: case-control	Phase III: case-control	Phase III: case-control	Phase III: cross-sectional
Information gathering	Retrospective	Retrospective	Retrospective	Ambispective
Patients age range	1–15 years	1 month to 19 years	1–16 years	16 months to 16 years
Non-AIH cases	Mainly chronic hepatitis C. PSC included	Mainly metabolic disorders. PSC included	Mainly chronic hepatitis B. No PSC included	Mainly cryptogenic chronic hepatitis, and Wilson's disease. PSC included
AIH cases	ALF included	ALF included	No ALF included	ALF included

IAIHG, International Autoimmune Hepatitis Study Group; PSC, primary sclerosing cholangitis; ALF, acute liver failure.

Figure 1 Flowchart of the literature search, including selection process and reasons for exclusion.

simplified criteria were applied without knowing the final diagnosis of the patients. As mentioned, the IAIHG classical criteria are not a perfect diagnostic tool. Therefore, misapplication and overinterpretation of their results must be avoided.³⁴ Two studies use them as the only standard to confirm AIH, without explicitly mentioning the pediatric-specific criteria proposed by the European Society for Paediatric Gastroenterology, Hepatology and Nutrition (ESPGHAN) and North American Societies for

Pediatric Gastroenterology, Hepatology and Nutrition (NASPGHAN) in 2009.³⁵ Moreover, included studies were published before the ESPGHAN Hepatology Committee recent position statement on pediatric AIH management.³⁶ In consequence, the new diagnostic score for children could have not been applied to enhance AIH diagnosis and autoimmune sclerosing cholangitis exclusion. Each study's QUADAS-2 ratings are detailed and available in [Supplementary Information File 2](#).

Figure 2 Summary of risk of bias and applicability concerns of the included articles following QUADAS-2 tool.

Diagnostic Accuracy of 2008 Simplified Criteria for AIH

Forest plots of all validity indicators for the 2008 simplified criteria are presented as [Supplementary Information Figure 1](#). For the suggested cutoff of 6 points, pooled sensitivity was estimated in 77% (95% CI 71%–83%) with significant heterogeneity and pooled specificity in 95% (95% CI 91%–97%) ([Figure 3](#)). Furthermore, pooled positive and negative likelihood ratios were calculated, respectively, at 13.7 (95% CI 4.5 to 42.2) and 0.2 (95% CI 0.1 to 0.5). Additionally, pooled diagnostic odds ratio was 66.8 (95% CI 13.2 to 339.0), which should be interpreted as the odds of scoring 6 or more points in the simplified criteria, if the patient has AIH, being almost 67 times greater than the odds of the test being positive if the patient does not

have the disease.³⁷ The study providing the sensitivity-specificity pair with the highest weight, and the nearest to the summary point in the ROC space, was conducted by our group ([Figure 4](#)).³⁰

Evaluation of Publication Bias

We found no direct evidence of publication bias as the slope of the regression line in the Deeks asymmetry test was not statistically significant neither for data with cutoff of 6 points nor for that with the cutoff of 7 points (p-value 0.841 and 0.455, respectively). In effect, visual examination of the point cloud in the Deeks funnel plot showed a fairly symmetrical distribution of the observations of the included studies. The entire publication bias analysis is displayed in [Supplementary Information File 3](#).

Figure 3 Forest plots representing the primary studies' sensitivity and specificity and their pooled values with their 95% confidence interval (95% CI).

Clinical Utility Assessment

Data necessary to calculate the therapeutic threshold were obtained from the review by Manns et al.³⁸ about diagnosis and treatment decision in AIH. As 9% of patients are known to develop clinical, laboratory, and histological worsening despite compliance with conventional treatment schedules, treatment utility in AIH was estimated to be 0.91. Spontaneous improvement is possible in 12% of asymptomatic patients with mild disease, so no treatment utility was set in 0.12. On the other hand, immunosuppressive therapy toxicity and deleterious effects of long-term corti-

costeroid are well-known effects of standard treatment in children. Moreover, decision to terminate treatment is considered in a small proportion of pediatric patients, and the subsequent probability of relapse is high. Thus, costs (treatment risks) were estimated to be 1. Both parameters are measured in the same unit (proportion of patients). Therefore, the therapeutic threshold was set at 56%.

The prevalence of AIH in nontransplanted children who undergo liver biopsy because of signs of acute or chronic hepatocellular damage, regardless of the coexistence of cholestasis and the technique for sample collection (percutaneous, laparoscopic, or transjugular), is 47% in our milieu.³⁰ As can be seen in the Fagan nomogram supplied as [Supplementary Information Figure 2](#), a result more than 6 points in the simplified score moves the prevalence toward the post-test probability of AIH beyond the therapeutic threshold, indicating that starting the treatment in children under suspicion of AIH and positive scoring may be justified under the aforementioned conditions and assumptions. This occurs from prevalences of AIH around 8.5% onwards. In parallel, a negative scoring ceases to deliver a post-test probability of AIH under the therapeutic threshold from prevalences around 85.7% onwards. The probability-modifying plot can be consulted in [Supplementary Information Figure 3](#).

DISCUSSION

To our knowledge, we have conducted the first systematic review with meta-analysis on the topic of the accuracy of simplified criteria for diagnosis of AIH in children. The prevalence of the disease and the validity indicators of any diagnostic system are information needed to perform clinical decision analyses.^{39,40} Accordingly, we aimed to gather the best evidence available to carry out a basic assessment of the utility of the simplified criteria to diagnose and treat AIH in the setting of pediatric specialized care.

Figure 4 Summary point for pooled simplified criteria's sensitivity and specificity, and its 95% confidence region, within the receiver operating characteristic space. Data for cutoff of 6 points. Primary studies, from top to bottom: Mileti 2012, Gonçalves 2017, Arcos 2018 and Hiejima 2011.

A restricted number of primary studies have been recovered through a compilation of the results of an exhaustive search in major databases and in grey literature. The whole process was executed following the guidance of a librarian. Despite having retrieved information from only four works, we have not found signs of publication bias, giving consistency to the pooled estimated indicators.

Primary works were critically appraised according to methodological aspects, and data were extracted independently by different authors. Nonetheless, primary studies' design has given rise to some concerns about the significance of the pooled validity indicators.

First, case-control diagnostic test accuracy studies are particularly prone to selection bias.⁴¹ The inclusion of too clear or typical forms of AIH in the case group may lead to an overestimation of the sensitivity owing to exclusion of doubtful cases. In that respect, the three case-control studies base the confirmation of AIH on classical reviewed criteria, which, despite being the best approach to establish the diagnosis, may cast doubt on the target condition matching the review question.^{9,34} Rightfully, two authors cite and use the pediatric-specific criteria submitted by the ESPGHAN and the NASPGHAN that consider a lower threshold for the autoantibodies titers and suggest performing cholangiographic imaging to rule out overlap syndromes, which are often misdiagnosed for AIH.^{20,33} Considering these pediatric criteria does not completely solve the problem of the absence of a gold standard. However, it can make up this source of bias. The other two primary sources have an unclear answer regarding the reference standard domain within the applicability concerns.

The study providing the highest weight on the final pooled validity estimates includes a retrospective historical wing with a number of cases that did not have cholangiography, which may have an impact on simplified criteria performance.³⁰ Actually, it has been found that 50% of patients with AIH actually have bile duct disease at presentation.⁴² It is possible that clinically relevant autoimmune cholestatic disease is not as high as described in this series, but, nevertheless, the absence of biliary tract imaging makes it impossible to rule out AISC.

One of the aspects considered by the QUADAS-2 regards the blind application of the index test. In our opinion, the IAIHG 2008 simplified criteria do not leave much space to interpretation. Therefore, we have not accounted this item a significant source of bias.

Yet, the pooled sensitivity-specificity pair for the AIH simplified criteria in children has been estimated in 77%/95%. Its clinical translation into the probability of AIH depends, apart from on the positive or negative result, on the prevalence of this condition in the population at risk. In every setting, we need to determine this pretest probability in the overall scenario of children with elevated transaminases without previously known liver disease. Data from

our study group show that AIH can be found in up to 47%.³⁰ However, in other contexts, the prevalence of AIH can be lower, especially in countries with a high incidence of alternative diagnosis such as nonalcoholic steatohepatitis.^{43,44}

In addition, although sensitivity and specificity are test-specific properties, the calculated specificity in the included studies is, in part, depends on the proportion of etiologies in the reference population. For instance, hepatitis B has a clearly defined the underlying cause and can be well established, whereas cryptogenic cirrhosis has an inherently more doubtful definition. In this regard, however, authors of the primary sources have tried to build a representative non-AIH group, and, moreover, the study providing the highest number of patients has been designed ambispectively to outweigh this source of bias.²⁹

Performance parameters of the simplified score in acute onset AIH are currently under discussion. Both classical and simplified criteria's accuracy in adult and children have proven to be poor in ALF.^{17,20} However, in our cohort, the four cases matching the definition of ALF were correctly classified by the 2008 criteria.³⁰ All of them had mild signs of encephalopathy that recovered after treatment. It is possible that clinical severity of ALF could have an impact on AIH diagnostic criteria reliability, and, as a consequence, previous results could not be strictly inferred to our sample. On this account, caution is advisable when applying the simplified criteria in ALF with more than grade 2 encephalopathy.

Overall, in the real clinical scenario that we defined, a score more than 6 points using the 2008 criteria renders a probability of 92.5% of a true AIH, indicating the good reliability of the simplified scoring system. It is noted that our prevalence is referred to all nontransplanted patients who were biopsied for abnormal liver tests and without any previously known hepatic disorder.²⁹ Once we explored the simplified criteria validity indicators, we conducted a basic clinical decision analyses by obtaining the therapeutic threshold following the model proposed by Pauker and Kassirer.³² We observed that the post-test probability of a positive result of the 2008 simplified criteria moved beyond the estimated threshold from the prior probability, or prevalence. This builds a theoretical basis for justifying a therapeutic trial in children with positive 2008 criteria as its response may add diagnostic information. We are aware that there might be a proportion of patients who show clinical and analytical improvement not related with drug therapy and that there will always be the question of a properly established AIH diagnosis. Nevertheless, we consider this to bear little relevance as improvement after treatment is a typical feature of AIH, and the chronological link between starting medication and liver function recovery is an easy-to-assess causation criterion. Another limitation is the fact that clinical utilities were taken from data reported in general and adult population,

and treatment risk was arbitrary fixed based on the idea of avoiding unnecessary long-term immunosuppressant drugs in children.³⁸ As mentioned, other settings' prevalence, or changes in the treatment benefits and the costs assumed, may have an impact on the clinical decision. But, interestingly, the simplified criteria scores have behaved as a helpful decision tool in a wide range of pretest probabilities (8.5%–85.7%), indicating that may be suitable in most clinical situations.

Additionally, scoring below 6 points gives a probability of 17.1% of a false-negative AIH. In our opinion, this is an unacceptable error rate, especially, considering that AIH is a progressive disease with an excellent clinical, analytical, and histological response to treatment. In this respect, a new scoring diagnostic system has recently been proposed for both pediatric AIH and AISC by the ESPGHAN Hepatology Committee.³⁶ As far as it considers pediatric patient's particularities, it could serve as a more accurate reference for further diagnostic studies. Theoretically, it allows seronegative AIH (the main cause for false negatives) to be diagnosed if the other criteria are met. However, more than half of the children with no autoantibodies do not show high IgG levels. Even when presenting with other conditions suggestive of seronegative AIH, such as aplastic anemia or peripheral thrombocytopenia, these criteria may still have a limited role in this particular phenotype.⁴⁵

As a conclusion, simplified criteria show a moderate sensitivity and a high specificity for the diagnosis of AIH in children. Because there are only four variables to consider in the calculation, it is easier to apply in a real clinical scenario compared with 1999 classical criteria. Beyond that, simplified criteria are a good tool to predict AIH diagnosis in children scoring 6 or more points and to start treatment accordingly, despite a negative result being insufficient to rule out AIH.

AUTHORS' ROLE IN THE SUBMITTED WORK

C.M.B. and J.M.C. proposed the idea. E.J.T. and M.V.B. conducted an initial literature research and contributed to the primary sources' quality assessment. J.V.A.M. and C.M.B. designed the methods, conducted the systematic review, and wrote the article. All the authors participated in the interpretation of the study and in drafting the manuscript. C.M.B. and J.M.C. are the guarantors.

CONFLICTS OF INTEREST

All authors have none to declare.

ACKNOWLEDGMENTS

The authors are grateful to Dr. Vicent Modesto, MD, PhD, for his help in conducting the statistical analysis.

REFERENCES

1. Krawitt E-L. Clinical features and management of autoimmune hepatitis. *World J Gastroenterol*. 2008;14(21):3301–3305.
2. Vergani D, Mieli-Vergani G. Cutting edge issues in autoimmune hepatitis. *Clin Rev Allergy Immunol*. 2012;42(3):309–321.
3. Mieli-Vergani G, Vergani D. Autoimmune hepatitis in children: what is different from adult AIH? *Semin Liver Dis*. 2009;29(3):297–306.
4. Della Corte C, Sartorelli MR, Sindoni CD, et al. Autoimmune hepatitis in children: an overview of the disease focusing on current therapies. *Eur J Gastroenterol Hepatol*. 2012;24(7):739–746.
5. Johnson PJ, McFarlane IG. Meeting report: international autoimmune hepatitis group. *Hepatology*. 1993;18(4):998–1005.
6. 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(5):929–938.
7. Czaja AJ. Performance parameters of the diagnostic scoring systems for autoimmune hepatitis. *Hepatology*. 2008;48(5):1540–1548.
8. Ebbesen RL, Schreiber RA. Diagnosing autoimmune hepatitis in children: is the International Autoimmune Hepatitis Group scoring system useful? *Clin Gastroenterol Hepatol*. 2004;2(10):935–940.
9. Kochar R, Fallon M. Diagnostic criteria for autoimmune hepatitis: what is the gold standard? *Hepatology*. 2010;51(1):350–351. author reply 351.
10. Hennes EM, Zeniya M, Czaja AJ, et al. Simplified criteria for the diagnosis of autoimmune hepatitis. *Hepatology*. 2008;48(1):169–176.
11. Choi G, Peters MG. The challenge of diagnosing autoimmune hepatitis: less is more. *Hepatology*. 2008;48(1):10–12.
12. Qiu D, Wang Q, Wang H, et al. Validation of the simplified criteria for diagnosis of autoimmune hepatitis in Chinese patients. *J Hepatol*. 2011;54:340–347.
13. Li Y, Peng M, Gong G. Evaluation of the revised versus the simplified scoring system in patients with autoimmune hepatitis. *Exp Ther Med*. 2014;7(1):131–136.
14. Fujiwara K, Yasui S, Tawada A, Fukuda Y, Nakano M, Yokosuka O. Diagnostic value and utility of the simplified International Autoimmune Hepatitis Group criteria in acute-onset autoimmune hepatitis. *Liver Int*. 2011;31(7):1013–1020.
15. Liu F, Pan ZG, Ye J, et al. Primary biliary cirrhosis-autoimmune hepatitis overlap syndrome: simplified criteria may be effective in the diagnosis in Chinese patients. *J Dig Dis*. 2014;15(12):660–668.
16. Gatselis NK, Zachou K, Papamichalis P, et al. Comparison of simplified score with the revised original score for the diagnosis of autoimmune hepatitis: a new or a complementary diagnostic score? *Dig Liver Dis*. 2010;42(11):807–812.
17. 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(2):538–545.
18. Muratori P, Granito A, Pappas G, Muratori L. Validation of simplified diagnostic criteria for autoimmune hepatitis in Italian patients. *Hepatology*. 2009;49(5):1782–1783. author reply 1783.
19. Mileti E, Rosenthal P, Peters MG. Validation and modification of simplified diagnostic criteria for autoimmune hepatitis in children. *Clin Gastroenterol Hepatol*. 2012;10(4):417–421.e1–2.
20. Hiejima E, Komatsu H, Sogo T, Inui A, Fujisawa T. Utility of simplified criteria for the diagnosis of autoimmune hepatitis in children. *J Pediatr Gastroenterol Nutr*. 2011;52(4):470–473.
21. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med*. 2009;6(7):e1000097.
22. Whiting PF, Rutjes AWS, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155(8):529–536.

23. Czaja AJ. Comparability of probable and definite autoimmune hepatitis by international diagnostic scoring criteria. *Gastroenterology*. 2011;140(5):1472–1480.
24. Gatsonis C, Paliwal P. Meta-analysis of diagnostic and screening test accuracy evaluations: methodologic primer. *AJR Am J Roentgenol*. 2006;187(2):271–281.
25. Leeflang MMG, Deeks JJ, Gatsonis C, Bossuyt PMM. Cochrane diagnostic test accuracy working group. Systematic reviews of diagnostic test accuracy. *Ann Intern Med*. 2008;149(12):889–897.
26. Deeks JJ. Systematic reviews of evaluations of diagnostic and screening tests. In: Egger M, Smith GD, Altman DG, eds. *Systematic Reviews in Health Care Meta-analysis in Context*. Londres: BMJ Books; 2001:248–282.
27. Leemis LM, Trivedi KS. A comparison of approximate interval estimators for the Bernoulli parameter. *Am Statistician*. 1996;50:63–68.
28. Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *J Clin Epidemiol*. 2005;58(9):882–893.
29. Zamora J, Abraira V, Muriel A, Khan K, Coomarasamy A. Meta-DiSc: a software for meta-analysis of test accuracy data. *BMC Med Res Methodol*. 2006;6:31.
30. Arcos-Machancoses JV, Molera Busoms C, Julio Tatis E, et al. Accuracy of the 2008 simplified criteria for the diagnosis of autoimmune hepatitis in children. *Pediatr Gastroenterol Hepatol Nutr*. 2018;21(2):118–126.
31. Fagan TJ. Letter: nomogram for bayes theorem. *N Engl J Med*. 1975;293(5):257.
32. Pauker SG, Kassirer JP. The threshold approach to clinical decision making. *N Engl J Med*. 1980;302(20):1109–1117.
33. Gonçalves C, Ferreira M, Ferreira S, Nobre S, Gonçalves I. Use of diagnostic scores in children's autoimmune liver disease - IAIHG scores can rule out the diagnosis in clinical practice. In: *Oral Communication. Oporto: LASPGHAN Annual Meeting*. Oporto; 2017.
34. Czaja AJ. Diagnosis and management of autoimmune hepatitis: current status and future directions. *Gut Liver*. 2016;10(2):177–203.
35. Mieli-Vergani G, Heller S, Jara P, et al. Autoimmune hepatitis. *J Pediatr Gastroenterol Nutr*. 2009;49(2):158–164.
36. Mieli-Vergani G, Vergani D, Baumann U, et al. Diagnosis and management of pediatric autoimmune liver disease: ESPGHAN Hepatology committee position statement. *J Pediatr Gastroenterol Nutr*. 2018;66(2):345–360.
37. Glas AS, Lijmer JG, Prins MH, Bossel GJ, Bossuyt PMM. The diagnostic odds ratio: a single indicator of test performance. *J Clin Epidemiol*. 2003;56(11):1129–1135.
38. Manns MP, Czaja AJ, Gorham JD, et al. Diagnosis and management of autoimmune hepatitis. *Hepatology*. 2010;51(6):2193–2213.
39. Stalpers LJA, Nelemans PJ, Geurts SME, Jansen E, de Boer P, Verbeek ALM. The indication area of a diagnostic test. Part I—discounting gain and loss in diagnostic certainty. *J Clin Epidemiol*. 2015;68(10):1120–1128.
40. Djulbegovic B, Desoky AH. Equation and nomogram for calculation of testing and treatment thresholds. *Med Decis Making*. 1996;16(2):198–199.
41. Kohn MA, Carpenter CR, Newman TB. Understanding the direction of bias in studies of diagnostic test accuracy. *Acad Emerg Med*. 2013;20(11):1194–1206.
42. Gregorio GV, Portmann B, Karani J, et al. Autoimmune hepatitis/sclerosing cholangitis overlap syndrome in childhood: a 16-year prospective study. *Hepatology*. 2001;33(3):544–553.
43. Schwimmer JB, Deutsch R, Kahen T, Lavine JE, Stanley C, Behling C. Prevalence of fatty liver in children and adolescents. *Pediatrics*. 2006;118(4):1388–1393.
44. Das K, Das K, Mukherjee PS, et al. Nonobese population in a developing country has a high prevalence of nonalcoholic fatty liver and significant liver disease. *Hepatology*. 2010;51(5):1593–1602.
45. Maggiore G, Socie G, Sciveres M, et al. Seronegative autoimmune hepatitis in children: spectrum of disorders. *Dig Liver Dis*. 2016;48(7):785–791.

SUPPLEMENTARY DATA

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jceh.2018.10.006>.