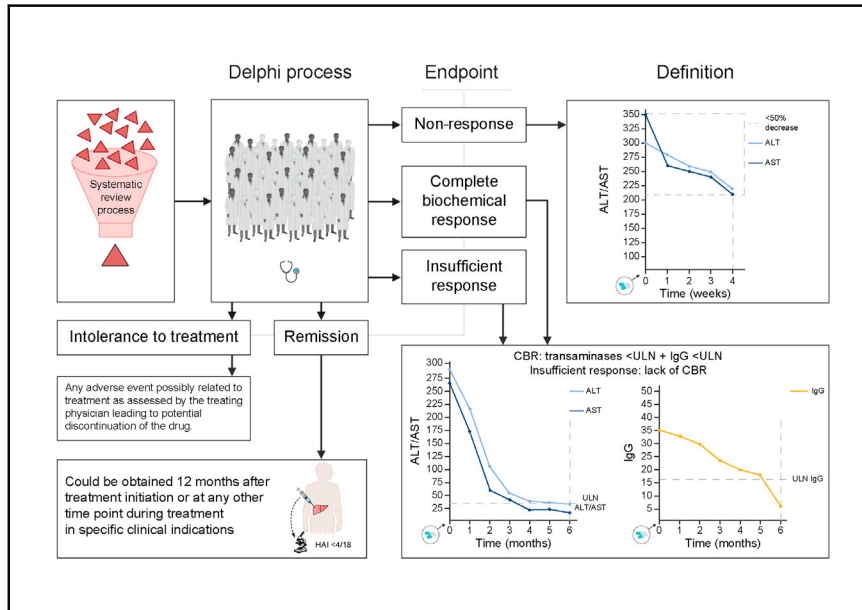


# Systematic review of response criteria and endpoints in autoimmune hepatitis by the International Autoimmune Hepatitis Group

## Graphical abstract



## Authors

Simon Pape, Romée J.A.L.M. Snijders, Tom J.G. Gevers, ..., Christoph Schramm, Joost P.H. Drenth, Michael A. Heneghan

## Correspondence

michael.heneghan@nhs.net (M.A. Heneghan).

## Lay summary

Consensus among international experts on response criteria and endpoints in autoimmune hepatitis is lacking. A consensus on endpoints is urgently required to set a global standard for the reporting of study results and to enable the comparison of results between clinical trials. Therefore, the International Autoimmune Hepatitis Group (IAIHG) herein presents a statement on 5 agreed response criteria and endpoints: complete biochemical response, insufficient response, non-response, remission, and intolerance to treatment, which can be used to guide future reporting.

## Highlights

- Consensus among experts on response criteria in autoimmune hepatitis is lacking.
- A consensus on endpoints is required to set a standard for reporting study results.
- Herein, the IAIHG presents a statement on 5 agreed response criteria and endpoints.



# Systematic review of response criteria and endpoints in autoimmune hepatitis by the International Autoimmune Hepatitis Group

Simon Pape<sup>1,14,†</sup>, Romée J.A.L.M. Snijders<sup>1,14,†</sup>, Tom J.G. Gevers<sup>1,2,14</sup>, Oliver Chazouilleres<sup>3,14</sup>, George N. Dalekos<sup>4</sup>, Gideon M. Hirschfield<sup>5</sup>, Marco Lenzi<sup>6</sup>, Michael Trauner<sup>7,14</sup>, Michael P. Manns<sup>8,14</sup>, John M. Vierling<sup>9</sup>, Aldo J. Montano-Loza<sup>10</sup>, Ansgar W. Lohse<sup>11,14</sup>, Christoph Schramm<sup>11,12,14</sup>, Joost P.H. Drenth<sup>1,14</sup>, Michael A. Heneghan<sup>13,14,\*</sup>, on behalf of the International Autoimmune Hepatitis Group (IAIHG) collaborators<sup>‡</sup>

<sup>1</sup>Department of Gastroenterology and Hepatology, Radboud University Medical Center, Nijmegen, The Netherlands; <sup>2</sup>Division of Gastroenterology and Hepatology, Maastricht University Medical Center, Maastricht 6229HX, The Netherlands; <sup>3</sup>Hepatology Department, Reference Center for Inflammatory Biliary Diseases and Autoimmune Hepatitis, Saint-Antoine Hospital Assistance Publique-Hôpitaux de Paris, Paris, France; <sup>4</sup>Department of Medicine and Research Laboratory of Internal Medicine, University of Thessaly Medical School, Larissa, Greece; <sup>5</sup>Toronto Centre for Liver Disease, University Health Network, University of Toronto, Toronto, Canada; <sup>6</sup>Department of Medical and Surgical Sciences, Sant'Orsola-Malpighi, University of Bologna, Bologna, Italy; <sup>7</sup>Division of Gastroenterology and Hepatology, Department of Internal Medicine III, Medical University of Vienna, Vienna, Austria; <sup>8</sup>Department of Gastroenterology, Hepatology and Endocrinology, Hannover Medical School, Hannover, Germany; <sup>9</sup>Departments of Medicine and Surgery, Baylor College of Medicine, Houston, TX, USA; <sup>10</sup>Division of Gastroenterology and Hepatology, University of Alberta Hospital, Edmonton, Canada; <sup>11</sup>1<sup>st</sup> Department of Internal Medicine, University Medical Center Hamburg-Eppendorf, Hamburg, Germany; <sup>12</sup>Martin Zeitz Centre for Rare Diseases, University Medical Centre Hamburg-Eppendorf, Hamburg, Germany; <sup>13</sup>Institute of Liver Studies, King's College Hospital, London, United Kingdom; <sup>14</sup>European Reference Network on Hepatological Diseases (ERN RARE-LIVER)

**Background & Aims:** Autoimmune hepatitis (AIH) has been well characterised and codified through the development of diagnostic criteria. These criteria have been adapted and simplified and are widely used in clinical practice. However, there is a need to update and precisely define the criteria for both treatment response and treatment.

**Methods:** A systematic review was performed and a modified Delphi consensus process was used to identify and redefine the response criteria in autoimmune hepatitis.

**Results:** The consensus process initiated by the International Autoimmune Hepatitis Group proposes that the term 'complete biochemical response' defined as 'normalization of serum transaminases and IgG below the upper limit of normal' be adopted to include a time point at 6 months after initiation of treatment. An insufficient response by 6 months was a failure to meet the above definition. Non-response was defined as '<50% decrease of serum transaminases within 4 weeks after initiation of treatment'. Remission is defined as liver histology with a Hepatitis Activity Index <4/18. Intolerance to treatment was agreed to stand for 'any adverse event possibly related to treatment leading to potential drug discontinuation'.

**Conclusions:** These definitions provide a simple and reproducible framework to define treatment response and non-response, irrespective of the therapeutic intervention. A consensus on endpoints is urgently required to set a global standard for the reporting of study results and to enable inter-study comparisons. Future prospective database studies are needed to validate these endpoints.

**Lay summary:** Consensus among international experts on response criteria and endpoints in autoimmune hepatitis is lacking. A consensus on endpoints is urgently required to set a global standard for the reporting of study results and to enable the comparison of results between clinical trials. Therefore, the International Autoimmune Hepatitis Group (IAIHG) herein presents a statement on 5 agreed response criteria and endpoints: complete biochemical response, insufficient response, non-response, remission, and intolerance to treatment, which can be used to guide future reporting.

© 2022 The Authors. Published by Elsevier B.V. on behalf of European Association for the Study of the Liver. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## Introduction

Autoimmune hepatitis (AIH) is a rare liver disease with multiple forms of presentation.<sup>1</sup> It is characterised by the presence of auto antibodies, hypergammaglobulinaemia and abnormalities in liver histology, measurable by specific criteria.<sup>2,3</sup> There are no pathognomonic features, the pathogenesis is partly unknown, and it remains a diagnosis of exclusion.<sup>4,5</sup> Initial treatment consists of steroid induction therapy, where the choice depends on histological severity and the fibrosis stage, followed by

Keywords: autoimmune hepatitis; endpoints; remission; complete biochemical response; insufficient response; non-response; intolerance.

Received 9 April 2021; received in revised form 18 November 2021; accepted 11 December 2021; available online 20 January 2022

\* Corresponding author. Address: Institute of Liver Studies, King's College Hospital, Denmark Hill, London SE5 9RS, United Kingdom; Tel.: +44-203-2993369, fax: +44-203-2993167.

E-mail address: [michael.heneghan@nhs.net](mailto:michael.heneghan@nhs.net) (M.A. Heneghan).

† Shared co-first authorship.

<https://doi.org/10.1016/j.jhep.2021.12.041>



maintenance therapy with a steroid sparing agent.<sup>6</sup> The ultimate treatment goal is to reduce long-term liver-related morbidity and mortality and to enhance quality of life. For some patients, liver transplantation will still be a necessity in the acute or chronic setting.<sup>1</sup>

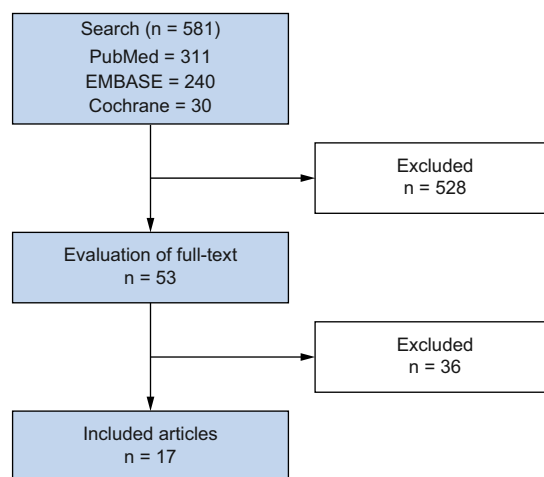
The International Autoimmune Hepatitis Group (IAIHG) has provided guidance and consensus statements in relation to codifying and characterising the disease, as well as providing guidance in case of serological and immunological testing. In its first iteration, in 1993, the first IAIHG provided diagnostic criteria in addition to definitions relating to therapy.<sup>7</sup> This led to the introduction of the terms 'complete response', 'partial response', 'no response', 'treatment failure', and 'relapse'. These definitions of treatment responsiveness have been the basis for defining response in AIH over the past 30 years. However, the definitions left room for multiple interpretations. To add complexity, the American Association for Study of Liver Disease (AASLD) redefined these terms and added 'incomplete response' as a response criteria or endpoint when some or no improvement in the clinical, laboratory and histological features existed despite compliance over a 2 or 3 year period.<sup>8</sup>

Expert opinion has also been a guiding principle in the field, and indeed, terms such as 'good response' and 'insufficient response' have led to more confusion.<sup>9</sup> Therefore, the IAIHG considered the harmonisation of definitions a priority for 3 reasons. Firstly, it would lead to the development of surrogate endpoints in AIH treatment, akin to what is described in other rare diseases, such as primary biliary cholangitis and primary sclerosing cholangitis.<sup>10–12</sup> Secondly, for patients with AIH, it would contribute to a better understanding of outcomes based on patient data, in both randomised studies and large case series of well-defined patients. Thirdly, the need for consensus endpoints that can be reproduced and which are useful for clinical trials based on current evidence, expert guidance and international consensus, was considered critical for the advancement of the field. The identification of these outcome measures will help to better define the inflection points in management algorithms for patients with AIH. For that reason, we have studied the results of a systematic review and the Delphi process, to identify and present what we believe are the most important measures of treatment response in AIH. Finally, we sought external validation of the developed endpoints, non-response and complete biochemical response, in a large multicentre cohort of patients with AIH.

## Materials and methods

### Literature search

Systematic searches of MEDLINE, Embase, Web of Knowledge, and the Cochrane database were performed to identify systematic reviews of trials reporting outcomes of and/or therapeutical interventions in AIH, and published guidelines and consensus statements, or recommendations, regarding the management of AIH. Consequently, this process collated all relevant endpoints that were used in any of the original trials. Only reviews published from January 1, 2010 onwards, were included to reduce the risk of using outdated or obsolete measures. Two authors (S.P. and T.G.) independently identified titles and abstracts of potentially eligible studies. Discrepancies were resolved by consultation between the initial 2 researchers. When the 2 parties failed to agree, a third researcher could be consulted. Adjudication was never needed. Reference lists of relevant



**Fig. 1. Flowchart of included studies after literature search.**

clinical studies, as well as review articles, were explored for additional studies. A detailed description of the MEDLINE search strategy is presented in Table S4. A total of 581 potential studies were retrieved, and 53 articles were evaluated in full-text after the screening of abstracts (Fig. 1). We excluded 36 studies after full-text evaluation, and our final count included 16 articles. We identified 11 (early) clinical trials in AIH.<sup>13–23</sup> Hand searching of reference lists of included studies left us with a further 5 potentially relevant articles: 1 meeting report,<sup>7</sup> and 4 guidelines<sup>8,24–26</sup> that met the inclusion criteria. A variety of endpoints and definitions were used over the years (Table S1). Using definitions published in the literature, various options for each endpoint of interest (non-response, complete biochemical response, insufficient response, remission, and intolerance to treatment) were drafted for entry into the Delphi survey (Table S2). Publications identified in the literature search were used to inform the Delphi process.

### Definition of surrogate endpoints in AIH treatment

The IAIHG, which is an international, non-governmental, non-profit scientific organization based on voluntary participation, providing a platform for carrying out project-based collaborations on AIH and AIH-related diseases, initiated a consensus process. At first, we used a modified Delphi approach. For each endpoint of interest (remission, complete biochemical response, insufficient response, non-response and intolerance to treatment), a number of possible definitions were drafted by a writers' group consisting of 11 members from the IAIHG. Definitions from existing guidelines were used to draft the queries for the survey. All members of the writers' group agreed upon the list of final items. We used Google Forms (Google, Mountain View, CA, United States) to create and execute the survey.

We invited all known members (active and inactive) of the IAIHG to complete the survey. We used the following template for most statements '<remission/complete response/insufficient response/non-response/intolerance to treatment> in AIH is best defined as <item option>' (Table S1). For insufficient response, the distinction was made between insufficient response to first- and second-line AIH therapy, in order to investigate whether physicians would accept different definitions or time points in relation to different therapies. Respondents could rate every

question on a 9-point Likert scale (from 'completely disagree' (n = 1) to 'completely agree' (n = 9)). The median rating of each item was calculated and categorized as inappropriate (median rating 1.0–3.4), uncertain (3.5–6.5) or appropriate (6.5–9.0). A disagreement index (DI) was calculated to establish whether there was consensus on a definition (Table S1). A DI <1 indicated consensus on an item, while a DI >1 indicated non-consensus (Table S3).<sup>27,28</sup> The survey consisted of 2 rounds. Definitions without an 'appropriate' rated item, or definitions without consensus in the first round, were re-rated in the second round.

After the second round, results of the survey were presented and discussed during 2 panel discussions at the IAIHG research workshop on July 1<sup>st</sup> and 2<sup>nd</sup>, 2019 in Vienna. Discussion and voting took place for all endpoints that were still rated 'uncertain' after the second survey round in order to reach a final consensus.

The initial approach was refined during peer review – resulting in a hybrid process between a Delphi consensus and a nominal group process – to reach agreement on points directly raised by the peer reviewers and to sharpen the concept of the definition. The final outcome determined the conclusive definitions for endpoints.

### External validation of surrogate endpoints

We pursued external validation of 2 final surrogate endpoints (non-response and complete biochemical response) using data from a multicentre cohort, consisting of 404 patients with AIH, from 5 European countries. The structure of the cohort, and the inclusion and exclusion criteria have been published elsewhere.<sup>29</sup> In brief, adult patients with AIH and a simplified IAIHG score ≥6 were included in this study cohort, whilst patients with variant syndromes or competing liver diseases were excluded.

We used liver-related mortality or liver transplantation as a composite endpoint, while sex and centre-specific upper limit of normal (ULN) values for alanine aminotransferase (ALT) and aspartate aminotransferase (AST) were used as covariates. Only patients who had available serum IgG and serum transaminases were included in this analysis. In case transaminases were elevated and IgG level was unknown within a 6-month time-frame, we considered this outside complete biochemical response.

### Analysis

We used the definitions resulting from the Delphi consensus, to distinguish between 2 groups within our AIH cohort depending on whether they met with the definitions from the Delphi round.

Two additional descriptive analyses were conducted to distinguish between patients attaining a complete biochemical response within 6 months and those attaining a complete biochemical response within 12 months but not within 6 months, and those attaining normal AST and ALT levels but having elevated IgG. We made univariate comparisons using the chi-square test, Mann-Whitney *U* test, or Student's *t* test as appropriate. We used Kaplan-Meier curves with log-rank testing for the composite endpoint of liver-related mortality or liver transplantation. *p* values <0.05 were considered statistically significant. Statistical analysis was done with SPSS version 25. Ethics approval was waived after review by the local institutional review board.

## Results

### Delphi respondents

An invitation to participate in the survey was sent to 220 email addresses on the IAIHG mailing list. A total of 75 respondents (34%) completed one or both survey rounds. The first round was completed by 58 respondents (26%), and the second round was completed by 50 respondents (23%). Thirty-three respondents (15%) completed both survey rounds. The IAIHG workshops on July 1<sup>st</sup> and 2<sup>nd</sup>, 2019 in Vienna were attended by 50 participants.

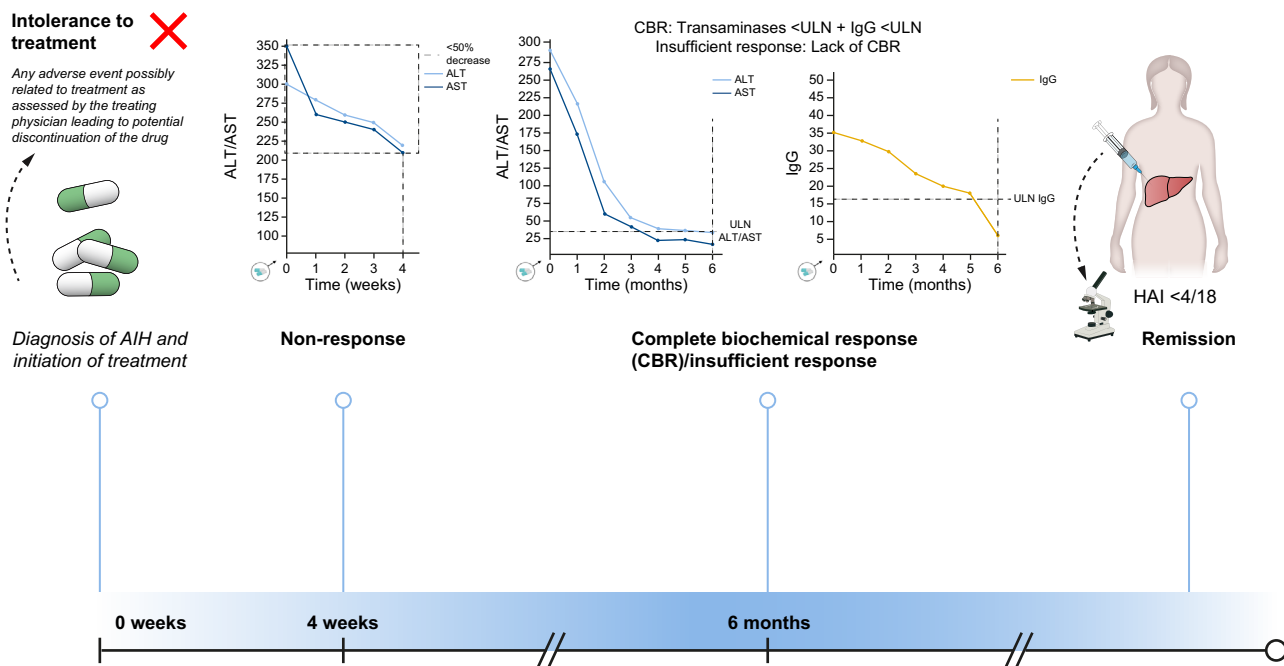
### Complete biochemical response in AIH

In the survey, 6 options for the definition of complete biochemical response were assessed (Table S2). After 2 survey rounds and a discussion, broad consensus was reached (median rating 8, DI 0.16) for the definition 'complete biochemical response' and now defined as 'normalisation of serum transaminases and IgG below the ULN' (Table 1 and Fig. 2). There was a consensus that complete biochemical response should be achieved no later than 6 months after initiation of treatment. The term 'response' was preferred over 'remission', since the term 'remission' reflects a stage in the disease that is less intense, whereas response indicates the transfer to a stage compatible with disease resolution. During the Delphi process we considered normalisation of transaminases below the ULN (without serum IgG) for the definition of complete biochemical response, but this was rated as less suitable and failed to gather consensus (median rating 6, DI 1.29). Normalisation of transaminases and IgG in combination with transient elastography (median rating 7, DI 0.65) was considered promising, based on data where transient elastography enabled the separation of severe from non-severe fibrosis after 6 months of immunosuppressive treatment.<sup>30</sup> Magnetic resonance elastography was not considered, owing to its lack of general availability.

**Table 1. Endpoints for AIH treatment as proposed by the International Autoimmune Hepatitis Working Group after a consensus process.**

Endpoint	Definition
Complete biochemical response	Normalisation of serum transaminases and IgG below the ULN. Should be achieved no later than 6 months after initiation of treatment.
Insufficient response	Lack of complete biochemical response. Should be determined no later than 6 months after initiation of treatment.
Non-response	<50% decrease of serum transaminases within 4 weeks after initiation of treatment.
Remission	Hepatitis activity index <4/18.
Intolerance to treatment	Any adverse event possibly related to treatment as assessed by the treating physician leading to potential discontinuation of the drug.

AIH, autoimmune hepatitis; IgG, immunoglobulin G; ULN, upper limit of normal.



**Fig. 2. Endpoints for AIH treatment as proposed by the International Autoimmune Hepatitis Working Group after a consensus process.** AIH, autoimmune hepatitis; ALT, alanine aminotransferase; AST, aspartate aminotransferase; CBR, complete biochemical response; HAI, hepatitis activity index; ULN, upper limit of normal. (This figure appears in color on the web.)

**Insufficient response in AIH**

Five options for the definition of ‘insufficient response’ in relation to different time points were assessed in the survey (Table S2A). After 2 survey rounds and a panel discussion, there was a broad consensus on the definition of ‘insufficient response’, which was defined as: ‘lack of a complete biochemical response’ for both first- (median rating 8, DI 0.26) and second-line (median rating 8, DI 0.16) AIH therapy. The importance of harmonising the definition of response for both first-line, second-line and third-line therapy was considered. In the second round, a consensus was reached on the assessment of ‘insufficient response’ as follows: the determination should be made no later than 6 months after initiation of treatment (first-line therapy: median rating 7, DI 0.16; second-line therapy: median rating 8, DI 0.16). Making this determination at 12 months after initiation of treatment did not reach a consensus or was rated as less appropriate (first-line therapy: median rating 7, DI 1.09; second-line therapy: median rating 6, DI 0.63). Other options for insufficient response in the survey, such as ‘persistence of elevated serum transaminases >2 times ULN’ or ‘increased liver stiffness of repeated measurements with transient elastography’, were rated as less appropriate (median ratings 6 and 4, DI’s 0.52 and 0.65, respectively).

It was agreed that insufficient response to first-line and second-line therapy may only be diagnosed as such after standard therapy has been applied and adherence has been proven (median rating 8, DI 0.13). Conditions for the standard of care first-line therapy in case of AIH, consist of steroid therapy with predniso(lo)ne at a dosage of at least 0.50 mg/kg/day at initiation and a maximum of 10 mg during maintenance. Moreover, budesonide is considered an equivalent treatment for non-cirrhotic patients. In the case of azathioprine therapy,

therapeutic 6-tioguanine levels were deemed mandatory in order to demonstrate adherence. In the case of second-line therapy, in particular regarding patients who do not tolerate azathioprine, mycophenolate mofetil (MMF) was suggested at a dose of at least 1 g/day, but preferably 2 g/day as an appropriate alternative. There was consensus that demonstration of therapeutic 6-tioguanine levels with 6-mercaptopurine therapy (median rating 8, DI 0.13) establishes adherence.

Consensus regarding the definitions for complete biochemical response and insufficient response indicates that any response other than normalisation of transaminases and IgG should be classified as an insufficient response.

**Non-response in AIH**

Nine options were given for the definition of ‘non-response’. After 2 survey rounds, several answer options were regarded as ‘appropriate’ by consensus, however, not one definition was singled out as superior (Table S2). The options ‘no improvement of serum transaminases and IgG’ (median rating 8, DI 0.16) and ‘no improvement of serum transaminases, serum IgG and persistent histological activity (hepatitis activity index [HAI] ≥4/18 or equivalent)’ (median rating 8, DI 0.29) came forward as possible candidates after the 2 voting rounds. During the panel discussion, the definition for non-response was extensively discussed. Both definitions were considered to be insufficiently accurate for defining non-response. Consequently, voting took place using alternative options. It was finally agreed that non-response in AIH should be defined as ‘<50% reduction of serum transaminases after 4 weeks of treatment’. In the context of this definition, serum transaminases should still be above the ULN to be considered as non-response, since transaminases below the ULN indicate potential complete biochemical response.



### Remission in AIH

Four items were assessed as options for defining remission in AIH. 'Normalisation of serum transaminases and serum IgG below the ULN' was proposed as a possible candidate for the definition of remission (median rating 8, DI 0.02). After 2 survey rounds and a panel discussion, it was agreed that remission in AIH could only truly be diagnosed histologically, and it is defined as 'liver histology with an HAI <4/18 or equivalent' (median rating 8, DI 0.32). This indicates that patients with no or only minimal hepatitis (HAI 0–3) are classified as being in remission.<sup>31,32</sup> Another option for the definition of remission that failed to gather consensus included the normalisation of serum transaminases below the ULN (median rating 6, DI 1.04). After peer review, a discussion took place, and it was finally agreed that remission cannot be achieved or tested within a specific timeframe or at a specific time point. However, a liver biopsy could be performed 12 months after treatment initiation or at any other time point during treatment for specific indications.

### Intolerance to treatment in AIH

Three answer options for the definition of 'intolerance to treatment' were evaluated in the survey. The option 'any adverse event possibly related to treatment as assessed by the treating physician, leading to potential discontinuation of the drug' was judged as the best definition for intolerance to treatment (median rating 8, DI 0.29). The options 'severe corticosteroid related side effects (acne, hypertension, psychosis, diabetes, osteoporotic fractures), side effects of immunosuppression leading to discontinuation (pancreatitis, cytopenia, hepatitis, gastrointestinal symptoms, allergic reactions)' (median rating 8, DI 0.29) and 'inability to reach recommended standard dose of treatment due to adverse events' (median rating 7, DI 0.37) were assessed as less suitable.

### External validation of surrogate endpoints

We were able to include 248 and 293 patients of a total of 404 eligible patients with AIH in our external validation of non-response and complete biochemical response, respectively (Fig. S1 and S2). Due to a lack of follow-up biopsies, it was not possible to validate the surrogate endpoint of remission.

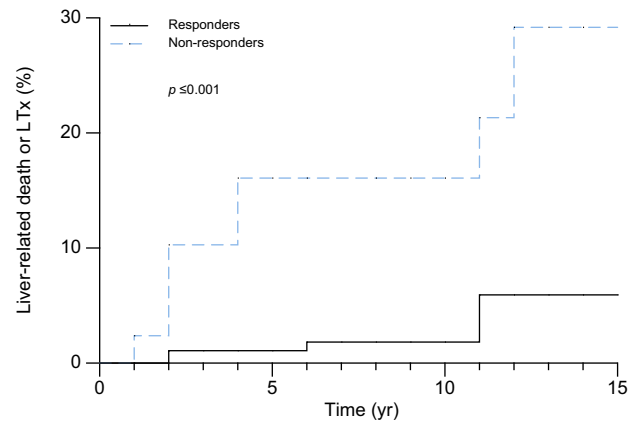
### External validation of surrogate endpoint non-response

Most patients (n = 206; 83.1%) scored as responders within 4 weeks after initiation of treatment (Table S5A). Responders were less likely to have cirrhosis at baseline compared to non-responders (19.4% vs. 46.3%;  $p < 0.001$ ). Liver transplantation or liver-related death occurred less frequently in responders: 2.4% vs. 21.4% (long-rank  $p < 0.001$ ) (Fig. 3 and Table S5A).

### External validation of surrogate endpoint complete biochemical response

One-hundred and thirty-four patients (45.7%) had a complete biochemical response (Table S5B) within 6 months. Responders were less likely to have cirrhosis at baseline compared to non-responders (18.8% vs. 29.5%;  $p = 0.047$ ). Other baseline characteristics were comparable between both groups. Liver transplantation or liver-related death occurred less frequently in responders: 0% vs. 7.5% (long-rank  $p = 0.003$ ) (Fig. 4 and Table S5B).

Additional descriptive analyses showed that liver transplantation or liver-related death occurred less frequently in



**Fig. 3. Kaplan-Meier curve of liver-related death or liver transplantation.** Patients with <50% decrease of serum transaminases within 4 weeks after initiation of treatment (non-responders) are compared with patients with a  $\geq 50\%$  decrease of serum transaminases within 4 weeks (responders) (log-rank  $p < 0.001$ ). LTx, liver transplantation.

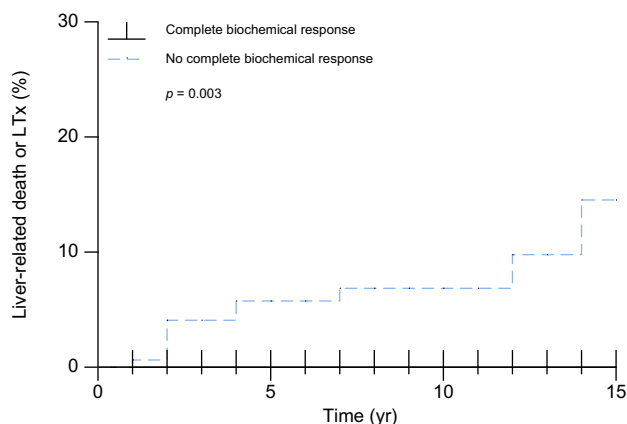
patients with a complete biochemical response within 6 months compared to those attaining a complete biochemical response within 12 months but not within 6 months (0.0% of patients vs. 4.4% of patients) and those attaining normal AST and ALT levels but having elevated IgG (0.0% of patients vs. 6.3% patients).

### Discussion

The original response criteria, published by the IAIHG in 1993, have provided a basis for the management of AIH.<sup>7</sup> Despite the durability of the criteria, it is apparent that the complexity of these original definitions (which allowed 2 options for each individual response criterion) is too complex to apply in everyday clinical practice. Moreover, varied definitions of both response and remission have crept into national and society guidelines.<sup>1,8,33,34</sup> This signals a need for harmonisation and clarification.

We performed a systematic review that informed a Delphi Method Process. Members of the review panel, consisting of both the steering committee and individuals who are general members of IAIHG, met face-to-face on 2 occasions to develop and refine the definitions proposed over an 18-month period in 2018 and 2019. Finally, consensus statements were agreed upon, following a review of the results and discussions in a workshop, hosted in Vienna in July 2019. We propose 5 standardised endpoints in the management of AIH.

These include a definition of a complete biochemical response. A complete biochemical response is defined as 'normalisation of serum transaminase activity and IgG level below the ULN, which should be achieved no later than 6 months after initiation of treatment'. During the assessment of the response criteria, it was found that guidelines derived from older studies<sup>15</sup> suggested that a threshold of transaminase activity of less than twice the ULN could be considered an adequate response.<sup>8</sup> The most recent randomised, double-blind study in the field, applied the primary endpoint of complete biochemical remission without steroid-specific side effects.<sup>20</sup> In this study, a complete biochemical remission was defined as serum AST and ALT within the normal range. As a consequence, and despite the lack of data from randomised clinical trials, the optimal response definition has evolved over time in clinical practice. In real-life practice, an



**Fig. 4. Kaplan-Meier curve of liver-related death or liver transplantation.** Patients with complete biochemical response are compared with patients without complete biochemical response (log-rank  $p = 0.003$ ). LTx, liver transplantation.

isolated elevation of either AST or ALT may be attributed to different aetiologies (for example, alcohol-related liver disease or non-alcoholic fatty liver disease). These factors need to be thoroughly assessed and may require an additional diagnostic trajectory. Thus, although ALT or AST can be elevated due to an obvious alternative aetiology, this will preclude the achievement of a complete biochemical response and will be marked as an insufficient response.

The definition of 'insufficient response' is a failure to achieve a complete biochemical response, using a combination of transaminase activity and IgG levels.<sup>20</sup> There is a need for harmonisation in view of the discordance in the literature about the timing of the first assessment after the start of therapy.<sup>29,35–37</sup> Zachou *et al.* assessed complete biochemical responses at 2 and 3 months,<sup>35,38</sup> whereas Pape *et al.* have recently proposed the concept of a decline in AST level by >80% at 8 weeks of treatment as a useful tool in predicting normalisation of transaminase activity at both 26 and 52 weeks.<sup>29</sup> The response to treatment to a large extent depends on patient-related factors and on the severity of AIH at presentation. An insufficient response does not necessarily imply that the treatment regimen should be altered, but it should alert the clinician since it is likely to have some prognostic value. The label of insufficient response has potential clinical consequences, including patient anxiety, and necessitates a search for possible underlying causes, such as non-adherence to treatment or an alternative explanation of the elevated transaminases.

We simplified the concept of non-responsiveness to entail a decrease of 50%, or less, of serum transaminases within 4 weeks of treatment initiation. The panel considered various definitions for a non-response based on previous guidelines and publications. The current definition is in accordance with the original IAIHG report,<sup>7</sup> but it excludes symptomatic improvement, and no change in inflammatory activity on liver biopsy after 6 months of treatment is required. Additionally, steroid dosage or increasing the dosage during steroid treatment is not specified. Indeed, AASLD guidelines from 2010 had suggested that an incomplete response represented 'some or no response in clinical, laboratory and histological features despite compliance with therapy after 2 to 3 years'.<sup>8</sup> A true non-response after initiation of standard treatment in AIH is a rare event since a favourable response to

steroids is a key characteristic of AIH. In patients with a non-response, adherence should be explored, the AIH diagnosis should be challenged, and the histology should be re-evaluated by an expert liver pathologist. Treatment failure, although helpful as a concept, remains difficult to define precisely, hence the need for accurate prognostic models to stratify patients at an early time point in their treatment course or presentation.<sup>39</sup>

The concept of remission should be simple and reproducible. Previous definitions of a complete response included improvement of symptoms, as well as the return of liver function tests to normal within 1 year of treatment initiation, after which maintenance of normalised liver biochemistry for a further 6 months was mandated.<sup>7</sup> Taking a liver biopsy at some point during this 18-month timeframe was necessary, and in keeping with the definition, minimal disease activity was required on biopsy. We considered the inclusion of biochemical parameters in addition to histological biomarkers to define 'remission'.<sup>40</sup> The panel agreed on the statement that remission in AIH can only truly be diagnosed histologically (median rating 8, DI 0.32), which requires a liver biopsy procedure. However, this is at odds with current clinical practice, where a liver biopsy procedure is only performed at diagnosis, and follow-up biopsies are rarely performed, and if so, only in cases of specific indications. There was consensus that remission could be obtained 12 months after treatment initiation or at any other time point during treatment in specific clinical indications. A follow-up biopsy is recommended in patients with a suboptimal response to treatment, in patients with discrepancies between the transaminase response and IgG response, and ideally before complete cessation of immunosuppression to confirm the complete histological resolution of the disease (HAI <4/18).<sup>34,41</sup> A liver biopsy may add value in detecting concurrent liver diseases, such as steroid-induced steatohepatitis, and offer pivotal information for clinical decision-making that requires a high degree of certainty, such as stopping immunosuppressive therapy.<sup>8,34</sup> The risk that comes with a liver biopsy procedure is a reason to consider non-invasive disease monitoring tools, such as transient elastography. Studies and guidelines have confirmed that there is no available evidence to validate the complete resolution of histologic inflammation.<sup>8,34,42</sup> While transient elastography has gained traction in the community,<sup>43–45</sup> there is currently little evidence to advocate its use in AIH currently other than as a monitoring tool. We interrogated the Delphi panel about the role of transient elastography in AIH based on data that show that transient elastography separates severe from non-severe fibrosis after 6 months or longer of immunosuppressive treatment,<sup>30</sup> but it was rated as less appropriate to identify a remission, and failed to gain consensus.

Three endpoints (complete biochemical response/insufficient response, and non-response) were externally validated in an AIH cohort. Applying the criteria for the 3 endpoints in this cohort resulted in adequate differentiation with respect to liver-related death or liver transplantation. Although we were able to demonstrate a significant difference in liver-specific mortality or liver transplant-free survival, the data should be interpreted with caution because of the relatively small number of events. In addition, as the further analysis comparing patients achieving a complete biochemical response within 6 months with those achieving a complete biochemical response after 6–12 months was based on a relatively insufficient sample size, the assessment of 'insufficient response' after no more than 12 months could be

an alternative option. We were unable to validate the endpoint 'remission' in our AIH cohort due to the lack of follow-up liver biopsies.<sup>34</sup> The endpoints 'complete biochemical response' and 'insufficient response' may serve as the most usable and relevant endpoints in the outpatient clinic.

Other difficulties concerning the management of AIH, based on previous guidelines, have become apparent. Notably, the previous practice guidelines of AASLD indicated that a period of 2 to 3 years had to pass before meeting the definition of incomplete response.<sup>8</sup> Indeed, original data from the clinical trials of Soloway and Summerskill have informed this thinking.<sup>15,46</sup> British guidelines suggested that in practice, the duration of therapy should be for at least 2 to 3 years, with normal transaminases for at least 18 months, to increase the likelihood of complete remission.<sup>33</sup> In retrospect, it is easy to criticize these attempts to define response/non-response/incomplete response; however, it must be acknowledged that the field has been crippled by a lack of randomised data, and in addition, most therapeutic interventions have been reported retrospectively, rather than prospectively.

We believe that these more user-friendly and time point-derived definitions facilitate the design and delivery of therapeutic procedures within the field, whether they involve first-line, second-line or salvage therapies or not. By unifying these concepts, any patient who has not achieved a complete biochemical response within 6 months of treatment has shown an insufficient response. While recognizing that treatment approaches vary around the world, the most recent practice guidelines of the AASLD suggest that based on the available data, both MMF and tacrolimus-based therapy are appropriate second-line treatments.<sup>1</sup> In contrast, the European Reference Network for hepatological diseases in conjunction with the European Association for the Study of the Liver guidelines suggested that MMF be utilized where thiopurine therapy (azathioprine or mercaptopurine) is not tolerated, and third-line treatment with tacrolimus must be used in cases of an ongoing insufficient response.<sup>34,41</sup> Despite these nuances in clinical practice and the style of practice, it is apparent that our current definitions of response can be used, irrespective of the drug regimen initiated at any given time point.

An agreed upon set of predefined endpoints comes with a number of key advantages. It increases transparency of the therapeutic trajectory, as it signposts the key junctures of the AIH patient journey. This has clear educational benefits. The use of these endpoints contributes to higher uniformity of treatment plans for patients with AIH and allows for better comparison between patients. This will be a boon for research collaboration since disease endpoints can be compared, allowing benchmarking of patient outcomes in AIH. A key advantage is that the use of these endpoints marks the timepoints along the treatment journey, allowing a timely escalation or de-escalation of therapy. It also allows us to identify hitherto unidentified treatment responses, such as relapses or a partial response, which would benefit from a robust definition.

Our study also comes with limitations. We achieved a response rate of 34% over 2 survey rounds. We drew on the mailing list of the IAIHG, and this list contains active contributors, but also a number of non-practicing clinicians and scientists who have since left the field. The IAIHG is highly diverse, comprised of basic scientists and clinicians, including paediatricians and pathologists, who fall outside the remit of our

survey. The endpoints proposed in this paper have been agreed upon by 50 individuals who participated in the discussion during the IAIHG workshops on 1-2 July 2019 in Vienna, which is an accurate reflection of the AIH expert community. We missed an opportunity to define 'loss of response' in AIH, as it was not included in the formal Delphi process. Using a real-world cohort of patients with AIH, the current retrospective database is insufficiently sized to establish that the proposed endpoints are the best early response markers. In view of the retrospective nature of the analysis, a number of relevant data points, in particular IgG, were missing. In addition, the number of events was limited, curtailing our options for a highly robust multivariable analysis.

To establish a research agenda in the future, one of the pressing needs in the field is the incorporation of patient-reported outcome measures (PROMs) or patient preferences into therapeutic guidance. We recognize that improvements in quality of life must be a treatment goal in AIH,<sup>47</sup> but there are currently no PROMs validated in AIH that provide a detailed assessment of quality of life and symptoms. Second, non-invasive markers, such as FibroScan, magnetic resonance elastography, or multiparametric magnetic resonance imaging, correlating with fibrosis stage and/or hepatic inflammation, need to be validated to assess AIH disease activity, as they have not earned a place among treatment and society guidelines to date.<sup>1</sup> Last, confirmation of the proposed endpoints will be required to investigate their reproducibility and validity in both adult and paediatric cohorts. Large prospective database studies are necessary to address these questions. The European Reference Network on Hepatological Diseases (ERN RARE-LIVER) is collecting prospective data from newly diagnosed patients with AIH (R-liver), and this initiative will contribute to a better understanding of the validity of the proposed surrogate endpoints.

## Conclusions

In conclusion, we established consensus among international experts on the definitions of the surrogate endpoints: complete biochemical response, insufficient response, non-response, remission, and intolerance in AIH. We encourage the AIH community to incorporate these endpoints in future studies to facilitate comparison of outcomes between studies, to properly validate these endpoints, and to share and convey data regarding the effect on long-term endpoints, such as liver transplantation and liver-related death.

## Abbreviations

AASLD, American Association for Study of Liver Disease; AIH, autoimmune hepatitis; ALT, alanine aminotransferase; AST, aspartate aminotransferase; CBR, complete biochemical response; DI, disagreement index; ERN RARE-LIVER, European Reference Network on Hepatological Diseases; HAI, hepatitis activity index; IAIHG, International Autoimmune Hepatitis Group; IgG, immunoglobulin G; MMF, mycophenolate mofetil; PBC, primary biliary cirrhosis; PROMs, patient-report outcome measures; PSC, primary sclerosing cholangitis; ULN, upper limit of normal.

## Financial support

This IAIHG meeting (July 2019) was supported by the YAEL Foundation. This work has been generated within the European Reference Network on Hepatological Diseases (ERN RARE-LIVER).



**Conflict of interest**

All authors report no potential conflicts that are relevant to the manuscript.

Please refer to the accompanying ICMJE disclosure forms for further details.

**Authors' contributions**

MH and JD conceived the study. MH, JD, TG and AL supervised the study. TG, OC, GD, GH, ML, MT, MM, JV, AM-L, CS, AL, JD and MH provided intellectual input for the study. SP and RS analysed the data. SP and RS wrote the manuscript. All authors critically reviewed the manuscript.

**Data availability statement**

The dataset used during this study is available from the corresponding author upon reasonable request.

**International Autoimmune Hepatitis Group**

P. Almasio (University of Palermo, Palermo, Italy), F. Alvarez (University of Montreal, Montreal, Canada), R. Andrade (University of Málaga, Málaga, Spain), C. Arikan (Koç University School of Medicine, Istanbul, Turkey), D. Assis (Yale School of Medicine, New Haven, Connecticut, United States), E. Bardou-Jacquet (Rennes University Hospital, Rennes, France), M. Biewenga (Leiden University Medical Center, Leiden, The Netherlands), E. Cancado (University of São Paulo, São Paulo, Brazil), N. Cazzagon (University of Padua, Padua, Italy), O. Chazouillères (Saint Antoine Hospital, Paris, France), G. Colloredo (San Pietro Hospital, Ponte San Pietro, Italy), M. Cuarterolo (Hospital Garrahan, Buenos Aires, Argentina), G. Dalekos (General University Hospital of Larissa, Larissa, Greece), D. Debray (AP-HP, Hôpital Necker Enfants Malades, Paris, France), M. Robles-Díaz (University of Málaga, Málaga, Spain), J. Drenth (Radboud University Medical Center, Nijmegen, The Netherlands), J. Dyson (Newcastle upon Tyne Hospitals, Newcastle, United Kingdom), C. Efe (Harran University Hospital, Urfa, Turkey), B. Engel (Hannover Medical School, Hannover, Germany), S. Ferri (University of Bologna, Bologna, Italy), R. Fontana (University of Michigan, Ann Arbor, United States), N. Gatselis (General University Hospital of Larissa, Larissa, Greece), A. Gerussi (University of Milano-Bicocca, Monza, Italy), E. Halilbasic (Medical University of Vienna, Vienna, Austria), N. Halliday (Royal Free Hospital, London, United Kingdom), M. Heneghan (King's College Hospital, London, United Kingdom), G. Hirschfield (University of Toronto, Toronto, Canada), B. van Hoek (Leiden University Medical Center, Leiden, The Netherlands), M. Hørby Jørgensen (Rigshospitalet, Copenhagen, Denmark), G. Indolfini (University-Hospital of Florence, Italy), R. Iorio (University of Naples, Naples, Italy), S. Jeong (Seoul National University College of Medicine, Seoul, South-Korea), D. Jones (Newcastle upon Tyne Hospitals, Newcastle, United Kingdom), D. Kelly (Birmingham Women's and Children's Hospital, Birmingham, United Kingdom), N. Kerkar (University of Rochester Medical Center, Rochester, United States), F. Lacaillle (AP-HP, Hôpital Necker Enfants Malades, Paris, France), C. Lammert (Indiana University School of Medicine, Indianapolis, United States), B. Leggett (University of Queensland, Herston, Australia), M. Lenzi (University of Bologna, Bologna, Italy), C. Levy (University of Miami, Miami, United States), R. Liberal (King's College Hospital, London, United Kingdom), A. Lleo (Humanitas University, Pieve Emanuele, Italy), A. Lohse (University Medical Center

Hamburg-Eppendorf, Hamburg, Germany), S. Ines Lopez (Hospital Garrahan, Buenos Aires, Argentina), E. de Martin (AP-HP Hôpital Paul-Brousse, Villejuif, France), V. McLin (University Hospital of Genève, Genève, Switzerland), G. Mieli-Vergani (King's College Hospital, London, United Kingdom), P. Milkiewicz (Medical University of Warsaw, Warsaw, Poland), N. Mohan (Medanta Medicity Hospital, Gurgaon, India), L. Muratori (University of Bologna, Bologna, Italy), G. Nebbia (Ospedale Maggiore Policlinico, Milan, Italy), C. van Nieuwkerk (Amsterdam University Medical Center, Amsterdam, The Netherlands), Y. Oo (University of Birmingham, Birmingham, United Kingdom), A. Ortega (Hospital Virgen de la Victoria Málaga, Málaga, Spain), A. Páres (University of Barcelona, Barcelona, Spain), T. Pop (University of Medicine and Pharmacy, Cluj-Napoca, Romania), D. Pratt (Massachusetts General Hospital, Boston, United States), T. Purnak (Hacettepe University, Ankara, Turkey), G. Ranucci (Santobono Pausilipon Children's Hospital, Naples, Italy), S. Rushbrook (Norwick and Norfolk University Hospitals, Norwich, United Kingdom), C. Schramm (University Medical Center Hamburg-Eppendorf, Hamburg, Germany), A. Stättermayer (Medical University of Vienna, Vienna, Austria), M. Swain (University of Calgary, Calgary, Canada), A. Tanaka (Teikyo University School of Medicine, Tokyo, Japan), R. Taubert (Hannover Medical School, Hannover, Germany), D. Terrabuio (University of São Paulo, São Paulo, Brazil), B. Terziroli (Ospedale Regionale di Locarno, Locarno, Switzerland), M. Trauner (Medical University of Vienna, Vienna, Austria), P. Valentino (Yale School of Medicine, New Haven, United States), F. van den Brand (Amsterdam University Medical Center, Amsterdam, The Netherlands), A. Villamil (Hospital Italiano de Buenos Aires, Buenos Aires, Argentina), S. Wahlin (Karolinska University Hospital, Stockholm, Sweden), H. Ytting (Rigshospitalet, Copenhagen, Denmark), K. Zachou (General University Hospital of Larissa, Larissa, Greece), M. Zeniya (Sanno Medical Center, Tokyo, Japan).

**Supplementary data**

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jhep.2021.12.041>.

**References**

*Author names in bold designate shared co-first authorship*

- [1] Mack CL, Adams D, Assis DN, Kerkar N, Manns MP, Mayo MJ, et al. Diagnosis and management of autoimmune hepatitis in adults and children: 2019 practice guidance and guidelines from the American association for the study of liver diseases. *Hepatology* 2019.
- [2] **Galaski J, Weiler-Normann C**, Schakat M, Zachou K, Muratori P, Lampalzer S, et al. Update of the simplified criteria for autoimmune hepatitis: evaluation of the methodology for immunoserological testing. *J Hepatol* 2021;74:312–320.
- [3] Mieli-Vergani G, Vergani D, Czaja AJ, Manns MP, Krawitt EL, Vierling JM, et al. Autoimmune hepatitis. *Nat Rev Dis primers* 2018;4:18017.
- [4] Herkel J, Carambia A, Lohse AW. Autoimmune hepatitis: possible triggers, potential treatments. *J Hepatol* 2020;73:446–448.
- [5] Weiler-Normann C, Lohse AW. Autoimmune hepatitis: from immunopathogenesis to diagnostic and therapeutic innovation. *Curr Opin Gastroenterol* 2021;37:86–90.
- [6] Doycheva I, Watt KD, Gulamhusein AF. Autoimmune hepatitis: current and future therapeutic options. *Liver Int* 2019;39:1002–1013.
- [7] Johnson PJ, McFarlane IG. Meeting report: international autoimmune hepatitis group. *Hepatology* 1993;18:998–1005.
- [8] Manns MP, Czaja AJ, Gorham JD, Krawitt EL, Mieli-Vergani G, Vergani D, et al. Diagnosis and management of autoimmune hepatitis. *Hepatology* 2010;51:2193–2213.

- [9] Manns MP, Lohse AW, Vergani D. Autoimmune hepatitis—Update 2015. *J Hepatol* 2015;62:S100–S111.
- [10] van Buuren HR, Lammers WJ, Harms MH, Hansen BE. Surrogate endpoints for optimal therapeutic response to UDCA in primary biliary cholangitis. *Dig Dis* 2015;33(Suppl 2):118–124.
- [11] Ponsioen CY, Lindor KD, Mehta R, Dimick-Santos L. Design and endpoints for clinical trials in primary sclerosing cholangitis. *Hepatology* 2018;68:1174–1188.
- [12] Ponsioen CY, Chapman RW, Chazouilleres O, Hirschfield GM, Karlsen TH, Lohse AW, et al. Surrogate endpoints for clinical trials in primary sclerosing cholangitis: review and results from an International PSC Study Group consensus process. *Hepatology* 2016;63:1357–1367.
- [13] Cook GC, Mulligan R, Sherlock S. Controlled prospective trial of corticosteroid therapy in active chronic hepatitis. *Q J Med* 1971;40:159–185.
- [14] Murray-Lyon IM, Stern RB, Williams R. Controlled trial of prednisone and azathioprine in active chronic hepatitis. *Lancet (London, England)* 1973;1:735–737.
- [15] Summerskill WH, Korman MG, Ammon HV, Baggenstoss AH. Prednisone for chronic active liver disease: dose titration, standard dose, and combination with azathioprine compared. *Gut* 1975;16:876–883.
- [16] Tage-Jensen U, Schlichting P, Aldershvile J, Andersen P, Dietrichson O, Hardt F, et al. Azathioprine versus prednisone in non-alcoholic chronic liver disease (CLD). Relation to a serological classification. *Liver* 1982;2:95–103.
- [17] Czaja AJ, Carpenter HA, Lindor KD. Ursodeoxycholic acid as adjunctive therapy for problematic type 1 autoimmune hepatitis: a randomized placebo-controlled treatment trial. *Hepatology (Baltimore, Md)* 1999;30:1381–1386.
- [18] Stellon AJ, Hegarty JE, Portmann B, Williams R. Randomised controlled trial of azathioprine withdrawal in autoimmune chronic active hepatitis. *Lancet (London, England)* 1985;1:668–670.
- [19] Stellon AJ, Keating JJ, Johnson PJ, McFarlane IG, Williams R. Maintenance of remission in autoimmune chronic active hepatitis with azathioprine after corticosteroid withdrawal. *Hepatology (Baltimore, Md)* 1988;8:781–784.
- [20] Manns MP, Woynarowski M, Kreisel W, Lurie Y, Rust C, Zuckerman E, et al. Budesonide induces remission more effectively than prednisone in a controlled trial of patients with autoimmune hepatitis. *Gastroenterology* 2010;139:1198–1206.
- [21] Czaja AJ, Wang KK, Shiels MT, Katzmann JA. Oral pulse prednisone therapy after relapse of severe autoimmune chronic active hepatitis. A prospective randomized treatment trial evaluating clinical, biochemical, and lymphocyte subset responses. *J Hepatol* 1993;17:180–186.
- [22] Stern RB, Wilkinson SP, Howorth PJ, Williams R. Controlled trial of synthetic D-penicillamine and prednisone in maintenance therapy for active chronic hepatitis. *Gut* 1977;18:19–22.
- [23] Hegarty JE, Nouri Aria KT, Eddleston AL, Williams R. Controlled trial of a thymic hormone extract (Thymostimulin) in 'autoimmune' chronic active hepatitis. *Gut* 1984;25:279–283.
- [24] EASL Clinical Practice Guidelines. Autoimmune hepatitis. *J Hepatol* 2015;63:971–1004.
- [25] Bittencourt PL, Cancado EL, Couto CA, Levy C, Porta G, Silva AE, et al. Brazilian society of hepatology recommendations for the diagnosis and management of autoimmune diseases of the liver. *Arquivos de gastroenterologia* 2015;52(Suppl 1):15–46.
- [26] Chinese consensus on the diagnosis and management of autoimmune hepatitis (2015). *J Dig Dis* 2017;18:247–264.
- [27] Mavarakis E, Patel F, Kronenberg DG, Chung L, Fiorentino D, Allanore Y, et al. International consensus criteria for the diagnosis of Raynaud's phenomenon. *J Autoimmun* 2014;48–49:60–65.
- [28] Lantinga MA, Darding AJ, de Sevaux RG, Alam A, Bleeker-Rovers CP, Bobot M, et al. International multi-specialty Delphi survey: identification of diagnostic criteria for hepatic and renal cyst infection. *Nephron* 2016;134:205–214.
- [29] Pape S, Gevers TJG, Vrolijk JM, van Hoek B, Bouma G, van Nieuwkerk CMJ, et al. Rapid response to treatment of autoimmune hepatitis associated with remission at 6 and 12 months. *Clin Gastroenterol Hepatol* 2019.
- [30] Hartl J, Denzer U, Ehlken H, Zenouzi R, Peiseler M, Sebode M, et al. Transient elastography in autoimmune hepatitis: timing determines the impact of inflammation and fibrosis. *J Hepatol* 2016;65:769–775.
- [31] Desmet VJ, Gerber M, Hoofnagle JH, Manns M, Scheuer PJ. Classification of chronic hepatitis: diagnosis, grading and staging. *Hepatology (Baltimore, Md)* 1994;19:1513–1520.
- [32] Ishak K, Baptista A, Bianchi L, Callea F, De Groote J, Gudat F, et al. Histological grading and staging of chronic hepatitis. *J Hepatol* 1995;22:696–699.
- [33] Gleeson D, Heneghan MA, British Society of G. British Society of Gastroenterology (BSG) guidelines for management of autoimmune hepatitis. *Gut* 2011;60:1611–1629.
- [34] European Association for the Study of the L. EASL clinical practice guidelines: autoimmune hepatitis. *J Hepatol* 2015;63:971–1004.
- [35] Zachou K, Gatselis N, Papadamou G, Rigopoulou EI, Dalekos GN. Mycophenolate for the treatment of autoimmune hepatitis: prospective assessment of its efficacy and safety for induction and maintenance of remission in a large cohort of treatment-naïve patients. *J Hepatol* 2011;55:636–646.
- [36] Choi J, Choi GH, Lee D, Shim JH, Lim YS, Lee HC, et al. Long-term clinical outcomes in patients with autoimmune hepatitis according to treatment response in Asian country. *Liver Int* 2019;39:985–994.
- [37] Biewenga M, Verhelst X, Baven-Pronk M, Putter H, van den Berg A, Colle I, et al. Aminotransferases during treatment in AIH aminotransferases during treatment predict long-term survival in patients with autoimmune hepatitis type 1: a landmark analysis. *Clin Gastroenterol Hepatol* 2021.
- [38] Zachou K, Gatselis NK, Arvaniti P, Gabeta S, Rigopoulou EI, Koukoulis GK, et al. A real-world study focused on the long-term efficacy of mycophenolate mofetil as first-line treatment of autoimmune hepatitis. *Aliment Pharmacol Ther* 2016;43:1035–1047.
- [39] Pape S, Schramm C, Gevers TJ. Clinical management of autoimmune hepatitis. *United Eur Gastroenterol J* 2019;7:1156–1163.
- [40] Al-Chalabi T, Heneghan MA. Remission in autoimmune hepatitis: what is it, and can it ever be achieved? *Am J Gastroenterol* 2007;102:1013–1015.
- [41] Lohse AW, Sebode M, Jorgensen MH, Ytting H, Karlsen TH, Kelly D, et al. Second-line and third-line therapy for autoimmune hepatitis: a position statement from the European reference Network on hepatological diseases and the international autoimmune hepatitis group. *J Hepatol* 2020;73:1496–1506.
- [42] Laschtowitz A, Zachou K, Lygoura V, Pape S, Derben F, Jaeckel E, et al. Histological activity despite normal ALT and IgG serum levels in patients with autoimmune hepatitis and cirrhosis. *JHEP Rep* 2021;3:100321.
- [43] Hartl J, Ehlken H, Sebode M, Peiseler M, Krech T, Zenouzi R, et al. Usefulness of biochemical remission and transient elastography in monitoring disease course in autoimmune hepatitis. *J Hepatol* 2018;68:754–763.
- [44] Xu QY, Sheng L, Bao H, Chen XY, Guo CJ, Li H, et al. Evaluation of transient elastography in assessing liver fibrosis in patients with autoimmune hepatitis. *J Gastroen Hepatol* 2017;32:639–644.
- [45] Zachou K, Lygoura V, Arvaniti P, Giannoulis G, Gatselis NK, Koukoulis GK, et al. FibroMeter scores for the assessment of liver fibrosis in patients with autoimmune liver diseases. *Ann Hepatol* 2021;22:100285.
- [46] Soloway RD, Summerskill WH, Baggenstoss AH, Geall MG, Gnitnick GL, Elveback IR, et al. Clinical, biochemical, and histological remission of severe chronic active liver disease: a controlled study of treatments and early prognosis. *Gastroenterology* 1972;63:820–833.
- [47] Janik MK, Wunsch E, Raszeja-Wyszomirska J, Moskwa M, Kruk B, Krawczyk M, et al. Autoimmune hepatitis exerts a profound, negative effect on health-related quality of life: a prospective, single-centre study. *Liver Int* 2019;39:215–221.