12 March 2024
EMA/CHMP/111546/2024
Committee for Medicinal Products for Human Use (CHMP)

Overview of comments received on 'Reflection paper on regulatory requirements for the development of medicinal products for chronic non-infectious liver diseases (PBC, PSC, NASH)'(EMA/CHMP/299976/2018)

Interested parties (organisations or individuals) that commented on the draft document as released for consultation.

Stakeholder no.	Name of organisation or individual
1	Allergan
2	AstraZeneca Pharmaceuticals
3	European Association of Hospital Pharmacists (EAHP)
4	EFPIA
5	The Forum for Collaborative Research- PSC Forum
6	GENFIT
7	Gilead Sciences International Ltd
8	Intercept Pharmaceuticals, Inc.
9	Medicines Evaluation Board, the Netherlands
10	Novartis
11	PRA Health Sciences
12	Hepatology Committee, European Society Paediatric Gastroenterology, Hepatology And Nutrition,
13	Global Liver Institute (GLI)
14	The Liver Forum
15	Merck Sharp & Dohme (Europe), Inc.
16	PSC Patients Europe
17	UK-PSC and PSC Support
18	Pfizer

${If.End}${If.App.PowerPoint}
${If.End}

30 Churchill Place ● Canary Wharf ● London E14 5EU ● United Kingdom
Telephone +44 (0)20 3660 6000 **Facsimile** +44 (0)20 3660 5555
Send a question via our website www.ema.europa.eu/contact

1. General comments – overview

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
1	Recommend dividing the reflection paper into 3 sections for clarity. One for each disease or condition (NASH, PBC and PSC).	Agreed. Two reflection papers are now published.
2	Based on the very different prevalence and aetiologies between NASH and PBC and PSC, we suggest that a separate reflection paper/ guideline is developed for NASH. There is a misconception in the paper about what constitutes a surrogate biomarker. Also liver histology is a surrogate biomarker. The need for a non-invasive diagnostic surrogate biomarker for NASH is huge, but it is not possible to develop such a biomarker within the framework of currently ongoing development programmes. For that reason, it is suggested that the agency closely follows the on-going attempts, e.g. IMI2 LITMUS, NIMBLE). For the time being, most important is to find biomarkers that predict improved histology and liver outcome; biomarkers identifying responders to the developed therapy.	Agreed. See above. The comment is noted, and the activities of IMI2, LITMUS and NIMBLE are followed.
3	EAHP would like to complement EMA on the very well written reflection paper.	Thanks for the comment.
4	**Comment:** The EMA and FDA NASH guidance have been released at around the same time (EMA released draft guidance 19 November 2018,	No comment necessary

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	FDA released draft guidance 3 December 2018), which leads to a unique opportunity for closer alignment in this emerging area of research and development. After assessing both draft guidance documents there is concern that the EMA draft guidance may be too conservative, and may create unnecessary hurdles to the research and development of future NASH therapies. Areas of particular concern include: **Proposed NASH Ph3 endpoints (238-248 and 282 - 305):** Requirement to demonstrate both resolution of NASH without worsening of fibrosis and improvement in fibrosis without worsening of NASH as co-primary endpoint sets a high bar that may not be attainable in monotherapy. Treatments may show benefit in only one of these two treatment benefits i.e. NASH resolution or fibrosis improvement. Correlation of histological improvement of NASH with long-term clinical outcomes remains to be established and important treatments options may be missed if the therapeutic threshold is set too high (line 238 – 248). Furthermore, for new substances primarily targeting fibrosis, a two-stage improvement in fibrosis at interim is a very strict requirement and may not be attainable. Accordingly important treatments may be missed or may not be developed if this development hurdle is maintained.	Not agreed with. The explanation for the requirements is given in the reflection paper. Partially agreed. The two-stage fibrosis requirement has been deleted.

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	Therefore, a primary endpoint of either NASH resolution without fibrosis worsening or fibrosis improvement by one stage or more without NASH worsening or both, under consideration of the mechanism of action of the drug, is proposed (see proposal below for line 238 – 248 and line 294 - 300).	Not agreed. Both outcomes are considered relevant. In case only one can be met, the results would need to be convincing in order to overcome the fact that: - In case fibrosis improvement is met only, the ongoing insult (liver cell stress with ballooning and inflammation, as well as steatosis) may outweigh the short-term benefits in the long run - In case the resolution of NASH endpoint is met only, a missing effect on fibrosis may not indicate improvement of prognosis.
	Duration of NASH trials (310-311): Based on existing data, there is no clear rationale why the evaluation of intermediate outcomes should require 2-years of interim evaluation. Alternative language in alignment with FDA draft guidance, allowing for more flexibility i.e. "clinical trials should be of sufficient duration (e.g. one year)" under consideration of study design, is proposed (see proposal below for line 310 – 311).	Partially agreed. The new wording allows flexibility depending on the availability of adequate phase 2 data.
	NASH combination treatment (363-367): At this early stage in the evaluation of new NASH therapeutic options, there is concern about limiting combination options to only 2nd line usage. It should be the clinical utility that should determine the stage of clinical use. It should be included that there is also potential benefit of combinations in first line use in F2/F3 patients if sufficient clinical evidence is available, see proposal below for line 362 - 367.	Partially agreed. Revised wording does not restrict development to 2nd line only, but requests adequate justification.

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	Cardiovascular safety (719-720): It should be clarified that dedicated cardiovascular outcome trials are not expected in NASH drug development. Cardiovascular outcome trials would impose an unnecessary burden, and may hamper drug development in NASH. Please see text proposal for lines 719 – 720 below). Lack of consistency when detailing NAS score. For example, line 171 lists a score of at least 5, and a score of 4; line 205 lists ☐ 5 and ☐ 4. It is recommended that a NAS score of ☐ 4 is used throughout for consistency. See also comment on lines 164-176.	Partially agreed. Since it is referred to the CV risk exclusion reflection paper, there is no need to especially refer to non-necessary outcome trials. Usually, this is also a case-by-case decision which depends on the safety profile. Therefore a general statement about need or non-need is not considered appropriate. The statements are considered consistent. There is general acceptance of NAS minimum of 5, but of 4 only under certain conditions.
5	We note that the International PSC Study Group recommendations seem to carry undue weight in this discussion. These recommendations were based on the discussion amongst 10 hepatologists, in which ALP, histology and elastography were all recommended. The endpoints paper (Reference 54) recommends TE or biopsy along with ALP but this is not adequately described. The key characteristic for any measure to be a treatment response endpoint is that a change in the measure reflects a change in the disease. None of the endpoints listed, especially ALP, biopsy, or cholangiography, has been validated for this purpose. In particular, it is unclear if there is data to support that a change in TE can be measured in a timeframe of two years (duration of most PSC therapeutic trials to achieve interim endpoints). Corpechot et al,	The recommendations of the Reflection paper reflects the regulatory position above all with several references to support that, including some recommendations of the PSC Study Group. It is agreed that none of the endpoints recommended in PSC trials have been validated. There is some support for the surrogacy of ALP and histology. Transient elastography is currently not recommended as a primary endpoint.

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	2014 paper in Gastroenterology showed little change in TE in years 0-2 of follow-up (figure 3) with most of the changes occurring in years 8-14 of follow-up, a timeframe that is unrealistic for clinical development. Thus, TE may not be an appropriate primary endpoint for PSC clinical development. Regulatory approval will depend on the "totality of data" supporting the safety and efficacy of an investigational agent, not just whether a trial meets the pre-defined phase 3 primary endpoint. The endpoint for the pivotal trial may depend in part of the MOA of the investigational agent. While ideally the endpoint should demonstrate improvement in markers of both cholestasis and liver disease, investigational agents with an anti-fibrotic MOA may demonstrate no improvement in markers of cholestasis, yet have meaningful clinical benefit to patients by improving fibrosis Dominant strictures and CCA are less prevalent in children. There is no routine screening for CCA; however, screening for IBD is highly recommended, mostly with stool calprotectin, as IBD symptoms are more quiescent. Small duct PSC is more prevalent and may cross over with other genetic conditions such as PFIC 3 and CF. Overall, UDCA and oral Vancomycin are used more frequently in children than in adults.	It is agreed that for a rare disease as PSC, evidence will depend on the totality of data. It is not agreed that an effect on histology alone, without an effect on cholestatic markers, would be clinically meaningful. Agreed - For paediatric age CCA is very rare condition and screening is not indicated. Screening for IBD using fCP is recommended in clinical practice Partially agreed, but genetic testing of each patient cannot be recommended Can be agreed, but no need for such of statement in GLs
9	Comments made at the stakeholder interaction meeting on the development of medicinal products for chronic non-infectious liver	In general, this statement is agreed with.

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	diseases at EMA on 3 December 2018 should be taken into account with respect to the adjustment of current version of the reflection paper. Many topics with respect to the design, endpoints, and analyses of studies with respect to non-alcoholic steatohepatitis (NASH), primary biliary cholangitis (PBC), and primary sclerosing cholangitis (PSC) are discussed in the reflection paper. It is a challenge to provide concise, comprehensive guidance for each of aforementioned non-infectious liver diseases. With respect to this, several adjustments should be made: - The reflection paper should be split up in two or three different reflection papers with NASH in a separate reflection paper. NASH Is considered a different type of disease, and thus a separate paper is preferred. - All relevant aspects with respect to a particular non-infectious liver disease (NASH, PBC, and PSC) including estimands, paediatric aspects, and safety considerations should be discussed in the chapter/ reflection paper on respective non-infectious liver disease. This approach is preferred instead of mentioning safety and paediatric patients in separate paragraphs. - For each non-infectious liver disease, the study design and recommended endpoints should be discussed in separate sections. In this way, particular recommendations for the study design and endpoints can be found more easily.	Agreed. There will be two reflection papers (one on NASH and another one on PBC and PSC). Partially agreed. There is, e.g. the fact that PBC does not occur in children, while PSC is very much relevant for children. It therefore makes sense that for each of the diseases the specifics are discussed separately. Agreed. The revised structure addresses this comment.

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	- It is advised to formulate requirements for clinical trials on aforementioned conditions more clearly. In addition, shorter sentences will help to provide clearer guidance.	Partially agreed. Concise and clear requests have been drafted.
	It should be discussed in the reflection paper whether extrapolation of study results on PBC is possible to external populations in clinical practice who may be less or more severely affected by PBC. The same should be done with respect to PSC.	Not agreed. The transfer of clinical trial data into "real world" is a general challenge not specific to the diseases dealt with in the reflection paper.
	Some additional information on the following topics should be added to section 4.4.1 on PSC: - Paediatric manifestations of PSC (with a cross-reference to section 4.7.3) - Natural course of PSC (e.g. average time span between diagnosis and detection of cirrhosis, and end-stage liver disease) - The medical need for development of new medicinal products for PSC should be stated more clearly in the reflection paper.	Partially agreed. The Reflection paper has been updated with information regarding paediatric manifestation of PSC and the natural course of PSC.
	More guidance should be provided on endpoints for clinical studies on medicinal products. Specific study endpoints should be recommended taking into account the curative or symptomatic aim of medicinal products. With respect to these endpoints, it should be discussed how these endpoints should be evaluated. Measurement properties, validity, reliability, strengths and limitations of diagnostic instruments (e.g. Fibroscan) and evaluation methods (e.g. liver function tests) should be addressed.	Partially agreed. However, since there is no clear accepted full spectrum of methods, only general recommendations are given. It should also be noted that a reflection paper is a preliminary regulatory position statement only in a situation when regulatory experience with licensing in the intended clinical indication is limited (or non-existing).

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	Fatigue and pruritus are important symptoms of PBC and PSC (Dyson et al. 2018, Lleo et al. 2017). Patients consider pruritus one of the most distressing symptoms of their cholestatic disease and report a significant decrease in quality of life as a result of pruritus (Weisshaar & Dalgard 2009, Kremer et al. 2008). Refractory pruritus, experienced by 5-10% of patients with cholestatic disease, can lead to sleep deprivation, depression, and suicidal ideation (Kremer et al. 2008, Mells et al. 2013). Fatigue, pruritus and maybe other symptoms should therefore be evaluated in clinical studies on PBC and PSC. This was also remarked at the stakeholder interaction meeting organized by EMA in December 2018. The importance of this symptom evaluation depends on the main treatment goal i.e. disease modification, or symptomatic treatment. If the main treatment goal is disease modification, symptomatic responses may be evaluated as key secondary endpoints. However, if medicinal treatment is primarily aimed at achieving symptomatic remission, (co-)primary endpoints may be defined in terms of symptomatic remission or response.	It is agreed that fatigue and pruritus are important symptoms. The Reflection paper states that symptom evaluation should be part of any trial in PBC and PSC. Furthermore, the paper has now added a separate section on drug development for the indication of cholestatic pruritus specifically as separate from disease modification.
	In addition, recommendations with respect to the evaluation of these symptoms should be provided (e.g. in sections 4.3.3, 4.4.3, 4.5, and in the paediatric sections (4.7.2 and 4.7.3). With respect to this, the validity, reliability, and other psychometric properties of available diagnostic tools should be taken into account.	Partly agreed. Recommendations are limited due to the fact that currently no fully validated scales exist. Development of patient reported outcome tools is encouraged.
	The relevance of other symptoms and signs (e.g. jaundice, hyperpigmentation, xanthelasmas)(Lleo et al. 2017) with respect to	Partly agreed. Levels of bilirubin for inclusion should be defined. Bilirubin is also recommended to be considered for

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	patient selection for clinical trials should also be discussed. Minimum requirements with respect to relevant symptoms and signs should be specified and substantiated, if possible.	enrichment of patients in trials with the objective of disease modification.
	Responder analyses with respect to endpoints are recommended (e.g. proportions of patients with particular decreases (e.g. 20, 40, 60 U/L) in ALP levels, proportions of patients with mild, moderate, and large decreases in fatigue scores). Responder analyses are important, as they provide insight into the extent of treatment efficacy.	Agreed. A responder analysis is recommended in the Reflection paper.
	The section on NASH should approach the condition primarily from a cardiovascular perspective. It is expected that the vast majority of the events during long-term trials will be cardiovascular in nature. The medical need in NASH should be discussed further. Although NASH may become an important reason for liver transplantation, most patients with T2DM or obesity will not develop NASH.	The comment is noted, and partially agreed. However, since the intended indication is NASH, a focus will remain on the liver part of the development.
	As NASH is an asymptomatic condition with sequelae only after many years, it is highly unlikely that liver biopsies will be taken routinely in clinical practice to identify patients for treatment. The development program should target patients that can be identified non-invasively.	The comment is noted, but not agreed with. In the clinical trial setting, there is a clear need for liver biopsies in cases where interim evaluation and licensing via CMA is intended. All cases which could go without histology are discussed adequately.
	Reversal of fibrosis in patients with F4 may be more realistic (based on the experience with hepatitis C) than currently assumed by many. Subjects with F4 should be investigated in the clinical development programs for NASH.	Agreed. A respective paragraph on development/endpoints in compensated cirrhosis has been included.

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	The guideline should address if and how substances that affect usual concomitant diseases (obesity, T2DM) can be indicated for NASH. Is it a problem if histological improvement of the liver comes together with weight loss or improved glucose regulation?	The comment is noted and partially agreed. The problem of standardisation of concomitant disease and medication is a general one, not specific to NASH. See Section 5.2.1.
	The MELD component of the proposed long-term endpoint may be unsuitable for NASH, as this could be easily be influenced by heart failure or medications therefore.	The comment is noted, and has been addressed by relevant paragraphs in order to require an adequate assignment of etiology.
	Target of estimation (estimand) paragraphs: These paragraphs should be kept consistent. Use consistent wordings in these paragraphs. Cross-references may be added, if needed. It would be helpful for the reader to provide directions and options to choose a strategy. This should help the reader to decide on a strategy.	Since two different papers are now developed, this is less relevant, and the revised wording takes account of the need for consistency.
	When recommending strategies, provide solid justifications why certain strategies were advised. Please mention clearly all attributes of the estimand. The paragraphs on estimands should also include the following items as advise to the reader: - The chosen strategies should be explicit and specific written down.	Partially agreed. The estimand framework, however, is implemented within the reflection papers at high level, complying with their character of being an "initial" guidance. Please refer to Chapter 5.3.5 of the Reflection paper for NASH and Chapter 5.2.1 of the NASH reflection paper tries to address the general considerations on this. See also Chapters 4.2.4 and 5.2.4 of the PBC/PSC reflection paper.

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	- Clear proposal for sensitivity analyses belonging to the same estimand or targeting a different one. - Identify the most clinically relevant intercurrent events (e.g. AE, intake of co-medication due to lack of efficacy, etc). Missing data are not intercurrent events. - Make distinction between concomitant medication, rescue medication, background medication and concomitant interventions (lifestyle interventions and dietary changes). - Collect reasons why people discontinue the study or stop treatment, because there is valuable information in these reasons to have more understanding.	
10	**Comment:** The EMA and FDA NASH guidance have been released at around the same time (EMA released draft guidance 19 November 2018, FDA released draft guidance 3 December 2018), which leads to a unique opportunity for closer alignment in this emerging area of research and development. After assessing both draft guidance documents there is concern that the EMA draft guidance may be too conservative, and may create unnecessary hurdles to the research and development of future NASH therapies. Areas of particular concern include: **Proposed NASH Ph3 endpoints (238-248 and 282 - 305):** Requirement to demonstrate both resolution of NASH without worsening of fibrosis and improvement in fibrosis without worsening	Comment noted. Not agreed. The reasons for introducing rather "strict" requirements are given in detail in the NASH paper.

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	of NASH as co-primary endpoint sets a high bar that may not be attainable in monotherapy. Treatments may show benefit in only one of these two treatment benefits i.e. NASH resolution or fibrosis improvement. Correlation of histological improvement of NASH with long-term clinical outcomes remains to be established and important treatments options may be missed if the therapeutic threshold is set too high (line 238 – 248). Furthermore, for new substances primarily targeting fibrosis, a two-stage improvement in fibrosis at interim is a very strict requirement and may not be attainable. Accordingly important treatments may be missed or may not be developed if this development hurdle is maintained. Therefore, a primary endpoint of either NASH resolution without fibrosis worsening or fibrosis improvement by one stage or more without NASH worsening or both, under consideration of the mechanism of action of the drug, is proposed (see proposal below for line 238 – 248 and line 294 - 300). **Duration of NASH trials (310-311):** Based on existing data, there is no clear rationale why the evaluation of intermediate outcomes should require 2-years of interim evaluation. Alternative language in alignment with FDA draft	Partially agreed. The two-stage criterion has been deleted. However, both outcomes are considered relevant. In case only one can be met, the results would need to be convincing in all other aspects in order to overcome the fact that: - In case fibrosis improvement is met only, the ongoing insult (liver cell stress with ballooning and inflammation, as well as steatosis) may outweigh the short-term benefits in the long run - In case the resolution of NASH endpoint is met only, a missing effect on fibrosis may not indicate improvement of prognosis. Partially agreed. The revised wording opens the requirements for the duration depending on phase 2 results and other factors.

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	guidance, allowing for more flexibility i.e. "clinical trials should be of sufficient duration (e.g. one year)" under consideration of study design, is proposed (see proposal below for line 310 – 311). **NASH combination treatment (363-367):** At this early stage in the evaluation of new NASH therapeutic options, there is concern about limiting combination options to only 2nd line usage. It should be the clinical utility that should determine the stage of clinical use. It should be included that there is also potential benefit of combinations in first line use in F2/F3 patients if sufficient clinical evidence is available, see proposal below for line 362 - 367. **Cardiovascular safety (719-720):** It should be clarified that dedicated cardiovascular outcome trials are not expected in NASH drug development. Cardiovascular outcome trials would impose an unnecessary burden, and may hamper drug development in NASH. Please see text proposal for lines 719 – 720 below). For details, see comments below.	Partially agreed. There is no clear restriction to second line any more. Not agreed. Since the paper refers to the "reflection paper on assessment of cardiovascular safety profile of medicinal products" (EMA/CHMP/505049/2015)" there is no need to fix the details further. The need for CV outcome studies is a case by case decision, and has to be decided upon based on pre-clinical and clinical safety data.
12	***4.7. Children and adolescents*** **4.7.1. NASH in children and adolescents** Similar to other aspects of the obesity/"metabolic syndrome" epidemic, non-alcoholic fatty liver disease (NAFLD), as well as NASH have been identified to present an increasingly significant	Agreed and noted. See relevant paragraph in the GL. With increased numbers of metabolic/obese patients in

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	health burden in children and adolescents. The prevalence of NAFLD in children is estimated to be around 10-14% depending on age. Whereas 2-4 year old children are expected to suffer from NAFLD at only very low rates, the prevalence in adolescents almost reaches adult levels. Assuming a similar rate of developing NASH from the presence of NAFLD in children/adolescents as in adults, it is clear that NASH is a relevant health problem also in the young age group, perhaps even more so than in adults because of the expectancy of more life years. Obviously, the development of late-stage disease may take years in children/adolescents similar to that in adults and thus might be expected to manifest primarily after reaching adulthood. However, rapid progression to advanced liver disease in childhood has been described. Therefore, there is a relevant medical need to develop treatments for NASH also in children. The diagnosis of NASH is currently considered to require the conduct of liver biopsy with histological evaluation, and the conduct of clinical trials should be mainly based on repeated biopsy results. Yet, it should be realized that liver biopsy remains the "imperfect reference standard" (V. Nibili). Also in childhood/adolescence, the diagnosis of NASH is based on histology. However, the conduct of repeated biopsies in clinical trials requires increased awareness of the potentially associated ethical and procedural problems when children are concerned, and the need for non-invasive outcomes in this population is therefore considered to be of even higher priority. The histological evaluations available have shown distinct features of paediatric NASH as compared to adults, with the presence of a relevant proportion of patients developing a unique histology pattern with presence of portal-based chronic inflammation (and fibrosis) (as opposed to the lobular inflammation found in adults) and less ballooning. The clinical meaning of this distinct type of histology in children is currently unknown, and consequently, a different histological scoring system may be needed for the paediatric population.	paediatric population represents NAFLD/NASH serious problem. Progression to advanced liver disease appears already in younger population, therefore development of effective treatments represents an medical need Agreed, partially reflected in GL. Diagnosis is based on histology and efficacy of treatment is evaluated by repeated biopsies. Since we don't have validated non-invasive test/markers, we need to rely on histology. Histological evaluation differs from adults and paediatric scoring system need to be validated

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	The development of new medicinal products for the treatment of NASH in children therefore requires first of all the collection of new data and the evaluation of existing data with regard to the natural history of the disease. Drug development in children will also require determining the adequate age range to be studied. Young children (e.g. below the age of 6-10 years), might still be early in the disease process, and therefore be appropriate candidates for non-pharmacological interventions, such as life-style and dietary changes, of which success rates (with regard to weight loss) are usually higher than in adults. Consequently, the potential for regression of inflammatory changes is similarly considered to be higher. The availability of more data on natural history, as well as on data of new compounds in adults might enable to more precisely determine the level of extrapolation that can be applied (see draft: Reflection paper on the use of extrapolation in the development of medicines for paediatrics. EMA/199678/2016). Taking into account all these considerations, the conduct of therapeutic trials in children is considered to be relevant, keeping in mind the potential for enhanced regression of NASH. Besides the necessary investigation of the appropriate dose (under full consideration of the potential differences in pharmacokinetics in obese and NASH adolescents compared to adults), and development of age-appropriate formulations, the conduct of placebo-controlled trials, including endpoints based on histology, and thus, repeated liver biopsies may still be required in order to fully account for the differences between childhood/adolescent and adult NASH. Even if from adult studies, an intermediate endpoint method such as an early histology evaluation endpoint, imaging methods, or	Agreed, collection of data about the natural history of the disease is necessary which is included in the reflection paper. There is no clear cut-off age for interventional studies. Nonpharmacological interventions are preferential approach in childhood population. No update needed Agreed, noted. Validation of non-invasive methods in children in addition to biopsy/histology, development of age appropriate formulations and dose finding are priorities

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	biomarkers, have partly been validated, it can be anticipated that these would have to undergo further validation in children. The conduct of studies with histology endpoints should take full account of the potential for the ethical problems associated with any more than minimally invasive procedures, and may need a careful approach with regard to the patient selection (e.g. older age groups, more advanced disease, etc.). However, the large unmet medical need together with the specific arguments detailed above, necessitate that children must be included into the perspectives of developing and testing relevant novel treatments, **4.7.2. PBC: Children and adolescents** The youngest reported age of a confirmed disease onset has been in a 15-year old post-menarche adolescent, and it is thought that PBC is not a truely paediatric disease. Potential applicants developing new substances in the treatment of PBC would therefore be expected to apply for a waiver for a paediatric programme for this specific disease. **4.7.3. PSC in Children and Adolescents** Paediatric PSC is a rare disease, even compared to adult PSC, which itself is classified as orphan. However, it is estimated that the risk in patients with IBD to develop PSC within 20 years after diagnosis is at least as high, if not higher, in the paediatric population as compared to adults. It should be realized that 20 years after diagnosis, these patients are only in their third or fourth decade of life. The development of complications of PSC, including the need for liver transplantation, does not show a plateau over 20 years after initial diagnosis. Therefore, the lifetime risk of complications in patients with a paediatric onset of disease is expected to be very high, and, because of more life-years ahead, likely higher than in patients with adult onset of disease. The incidence of IBD increases in children, as it does in adults, supporting the expectation that patients would be available for the controlled assessment of novel	Agreed and noted Agreed and reflected in GL waiver is accepted Agreed, comment is noted and reflected in GL Peter Agreed and noted

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	therapies. Distinct from adult PSC, there is a higher overlap of PSC with other syndromes in children, especially AIH (PSC-AIH-overlap syndrome or Autoimmune Sclerosing Cholangitis - ASC). The investigation of new compounds, also for children is therefore needed. Since a relevant amount of data on natural history has already been collected ,clinical trials in paediatric PSC can reasonably be undertaken , also with patients suffering from overlap conditions (especially AIH-PSC). The inclusion of patients should be based on the identified risk factors, which are distinct from adult PSC, such as elevated gamma-glutamyl-transferase (GGT) and aspartate aminotransferase-to-platelet ratio index (at diagnosis). Subgroups of patients can be differentiated with different risks on complications (for example, based on GGT levels at 1 yr after diagnosis), and this needs to be taken into account upon designing trials. Besides the need to fully explore the PK profile in the respective population, the recommendations to be given with regard to the design of trials, and endpoints to be used are not expected to be (very) different from those in adult patients. Obviously, consultation with the agency early in the drug development (scientific advice and submission of PIP) is therefore advisable.	Agreed and reflected in relevant paragraph No update needed

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
13	The Global Liver Institute (GLI) appreciates the opportunity to comment on the European Medicines Agency (EMA or Agency) on the Draft reflection paper on regulatory requirements for the development of medicinal products for chronic non-infectious liver diseases (PBC, PSC, NASH). As a nonprofit patient advocacy organization committed to improving the lives of all impacted by liver disease like nonalcoholic steatohepatitis (NASH), we applaud the agency's recognition of the need to open approval pathways and increase treatment options for this life threatening disease. Nonalcoholic steatohepatitis, or NASH, has been called an epidemic, a ticking time bomb, and a silent tsunami. It is the progressive form of nonalcoholic fatty liver disease (NAFLD), and affects more than 115 million people worldwide. By 2030 it is estimated that more than 128 million people will be affected by NAFLD/NASH.	The comments are noted and welcomed.

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	Also NAFLD and NASH are major risk factors for other health conditions: more than 70% of patients are obese, up to 75% have type 2 diabetes, and anywhere from 20-80% have hyperlipidemia. Unchecked, NASH can lead to severe health complications associated with the liver including end-stage liver disease, and hepatocellular carcinoma (HCC) and death. Presently, the "gold standard" option to diagnose NASH is to conduct a liver biopsy. Because it is a painful, invasive procedure, liver biopsy should be a diagnostic of last resort. It also plays a role in the high costs associated with the care for NAFLD and NASH independent of its metabolic comorbidities. The largest increases in health care utilization that may account for the increased costs in NAFLD and NASH are represented by liver biopsies, and hospitalizations. Paired with NASH being asymptomatic, both NAFLD and NASH are underdiagnosed and underreported. There is also no consensus around a single non-invasive test (NIT) to diagnose NASH and replace liver biopsy. However, there are active attempts to develop and validate NITs to reduce the risks and costs associated with biopsy. Patients need alternatives to increase the number of appropriate diagnoses allowing more patients with NASH to be identified and treated. Finally, as you are well aware, there are currently no approved treatments available for NASH. Liver transplantation is the only recourse for people with end-stage liver disease and/or NASH-related HCC. NASH was the fastest growing reason for liver transplantation between 2002 and 2011. We commend the agency's acknowledgement of the lack of treatment options, the connection between NASH and other diseases, and the need for non-invasive diagnostic options. We also appreciate the agency's understanding that it is crucial to proceed with a patient focus while acting quickly to meet the unmet medical need in NASH.	Comment noted. The final aim of developing reliable, fully validated biomarkers potentially replacing histology is also expressed in the NASH reflection paper. Comment noted Currently, there is no finally validated, regulatorily approved biomarker available.

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	However, we have serious concerns with the unintended consequences of requiring co-primary intermediate endpoints for stage 2 and 3 fibrosis drug therapy development. As NASH patients who have been neglected for far too long, we can not stress enough how important it is to consider all options, and strategies for approval. We can not risk losing the crucial therapies that are close to approval because they favor one intermediate endpoint over another. This is also why we strongly believe consideration should be given to combination therapies allowing for a timely varied response to NASH. We ask EMA to please consider harmonizing their guidance with the United States' Food and Drug Administration. Differences between FDA and the EMA affect drug development planning and procedures. In this case specifically, the FDA has positively decided to not require co-primary intermediate endpoints, and instead given the opportunity to consider either/or a resolution in NASH, or an improvement in fibrosis. As patients for whom access to treatment to this disease is literally a life-and-death issue, it is both encouraging to know that EMA acknowledges the role they can play in protecting patient lives, but concerning to see some of the barriers this reflection paper as written would put in place. The mission of the EMA to foster scientific excellence in the evaluation and supervision of medicines is one that we appreciate and share. The goal should always be to approve safe, sustainable and innovative treatments for patients. We hope EMA recognizes this point and moves swiftly to address the unmet need of current NASH patients, along with the incoming burden of future patients.	Not agreed. See previous comments. Comment noted. Recommendations for combination treatment have been adapted. Specific reasons have been given for the need to be more "strict" for the interim evaluation. These are valued higher than the aim of harmonisation for the time being. Comments noted.

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	We thank the EMA for the opportunity to comment on this critical reflection paper, and forward to working with the agency to improve this document. Specific, detailed comments on the text are included below. We would be pleased to provide further input or clarification of our comments, as needed.	
14	The Liver Forum is submitting the attached comments for consideration, in response to the November 15, 2018 notice of EMA/CHMP/299976/2018. The Liver Forum is a part of the Forum for Collaborative Research, an initiative of the University of California Berkeley School of Public Health which aims to advance the regulatory sciences for the treatment of NASH and liver fibrosis by providing an independent and neutral venue for ongoing multi-stakeholder dialogue to identify and address barriers to drug development. The Liver Forum is comprised of members from academia, industry, regulatory agencies, patient community, and professional societies. We applaud the EMA's commitment to clarifying the drug development pathway for patients with PBC, PSC, and NASH, and we greatly appreciate the opportunity to submit these comments on the reflection paper. These comments are limited to NASH, as the PSC Forum will be submitting comments separately.	Comments are noted and welcomed. The document was split into two separate documents.

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	As an overall comment, we suggest the clarity of the document might be enhanced if split into three distinct sections, one for each disease.	
17	UK-PSC is a collaborative organisation, encompassing patient groups, doctors and scientists; with shared interest in defining the underlying mechanisms of disease, how phenotypic diversity impacts disease progression, and to ultimately improve quality of life and care delivery across for people with PSC across the UK. Our partnerships with industry are essential to drug development and ongoing clinical trial design. Collectively we seek to better understand unmet need in PSC, and conduct studies of new therapies as carefully constructed interventions that deliver specific, measurable, achievable, relevant and time‐cost limited outputs. A cornerstone of activity within our group is to “de‐risk” drug development pathways where possible, but maximize opportunity to advance therapy for patient benefit in a timely way. PSC Support is an active non-profit patient organisation based in the United Kingdom (www.pscsupport.org.uk), with worldwide reach, and an established PSC research funding programme. PSC Support advocates for the needs of people with primary sclerosing cholangitis (PSC). Our vision is to see a world without PSC. PSC Support has captured and published reports on patient views on unmet needs and attitudes to research and potential treatments, has been hosting patient meetings since 1995 and moderates online forums for thousands of people affected by PSC. We are in a strong position to provide patient input into this consultation.	Comments noted. No comment needed

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	For the purposes of this response, PSC Support will limit comments to primary sclerosing cholangitis only, and not PBC/NASH. UK-PSC and PSC Support thank the European Medicines Agency for drafting this reflection paper, and inviting feedback through public commentary.	
18	There is clinical overlap between the two intermediate composite endpoint definitions with respect to disease etiology. A given therapeutic mechanism of action may be directed predominantly at fibrosis with the goal of reversing fibrotic processes (but not necessarily addressing the presumed underlying metabolic etiology of NASH) or it may address the presumed underlying metabolic etiology of NASH and thereby reduce/reverse downstream fibrotic processes through mitigation of proximate insults (e.g. chronic steatosis/inflammation). In this scenario, reduction of fibrotic process is an objective of both albeit over potentially different time courses. Given this, wouldn't it be more appropriate to consider that evidence for one or the other composite endpoints could be a sufficient intermediate outcome to define a responder provided that the subsequent Outcomes study demonstrates appropriate reduction in longer term outcomes (e.g. the composite of all cause death, decompensation of liver disease (with a complete listing), histological diagnosis of liver cirrhosis and MELD > 14)? Considering the importance of patient experience data in medical product development and regulatory decision making as well as recent attention to patient-focused drug development, Pfizer	The comments are noted but only partially agreed. Both intermediate endpoints have to be regarded to be non-validated surrogates for the time being. As implemented in the final NASH reflection paper, the special requirements for conditional approval will make rather strong efficacy criteria necessary. The missing of one of the intermediate endpoints always still leaves the opportunity to demonstrate efficacy in a long-term study with clinical outcomes. Agreed. A respective paragraph has been implemented.

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	suggests that the agency considers including opinion about the role of patient-reported outcomes in new drug development for NASH. Recent qualitative and quantitative studies conducted with patients (references are provided below) suggest that NASH has a notable impact on how patients feel and function. Although the most commonly reported symptoms, such as abdominal discomfort and fatigue, may not be NASH-specific, they have a significant impact on patients' functioning and overall well-being. In addition, the Green Park Collaborative initiative, aimed to develop multi-stakeholder consensus on a "core set" of outcomes to be used in pivotal and post-approval clinical trials of therapies for NASH, identified patient experience data as an important outcome for several stakeholders. Further, inclusion of patient-reported outcomes in the POC (proof of concept) and pivotal trials can help better characterize the patient population that would most benefit from the new NASH treatments as well as facilitate the interpretation of the primary efficacy endpoints and more accurately characterize the clinical meaningfulness of changes in histological endpoints such as reduction of inflammation from the patient's perspective. References: Doward LC, Balp MM, Stewart KE, Cryer D, Langford A, Twiss J, et al. Exploring the patient perceived impact of non-alcoholic steatohepatitis. Journal of Hepatology. 2017;66:S422-S3.	

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	Twiss J, Balp M, Doward L, Slota C, Cryer D, Langford A, et al. PGI39 - Development of a New Patient-Reported Outcome Measure for Non-Alcoholic Steatohepatitis: Nash-Check. Value in Health. 2017;20(9):A638. Palsgrove A, Halzra S, Ferguston B, Cheng R, Dombroski J, Cole J.C. PRM130-Development of Conceptual Framework for Assessing Disease-Specific Patient-Reported Outcomes in Nonalcoholic Steatohepatitis. Presented at ISPOR 26th Annual International Meeting 2016. David K, Kowdley KV, Unalp A, Kanwal F, Brunt EM, Schwimmer JB, et al. Quality of life in adults with nonalcoholic fatty liver disease: baseline data from the nonalcoholic steatohepatitis clinical research network. Hepatology.49(6):1904-12. Sayiner M, Stepanova M, Pham H, Noor B, Walters M, Younossi ZM. Assessment of health utilities and quality of life in patients with non-alcoholic fatty liver disease. BMJ Open Gastroenterology.3. Dan AA, Kallman JB, Wheeler A, Younoszai Z, Collantes R, Bondini S, et al. Health-related quality of life in patients with non-alcoholic fatty liver disease. Alimentary Pharmacology & Therapeutics.26(6):815-20. Kennedy-Martin T, Bae J, Paczkowski R, Freeman EC. Health-related quality of life burden of nonalcoholic steatohepatitis: a robust pragmatic literature review. Journal of Patient-Reported Outcomes. 2018;2(28).	

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	Younossi ZM, Stepanova M, Henry L, Racila A, Lam B, Pham HT, et al. A disease-specific quality of life instrument for non-alcoholic fatty liver disease and non-alcoholic steatohepatitis: CLDQ-NAFLD. Liver International.37(8):1209-18. Newton JL, Jones DEJ, Henderson E, Kane L, Wilton K, Burt AD, et al. Fatigue in non-alcoholic fatty liver disease (NAFLD) is significant and associates with inactivity and excessive daytime sleepiness but not with liver disease severity or insulin resistance. Gut. 2008;57(6):807-13. Sobhonslidsuk A, Satitpornkul P, Sornmayura P. Excessive daytime sleep disorder and fatigue in non-alcoholic fatty liver disease in comparison to cirrhosis. Journal of Gastroenterology and Hepatology.5:164.	

2. Specific comments on text

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
General			Based on stakeholders' comments it was decided to split the original Reflection paper into a reflection paper on PBC/PSC, and another reflection paper on NASH. The proposed textual adjustments were considered for each of the new, separated reflection papers.
	9	It is suggested to mention the meaning of an abbreviation once in the text. Alternatively, a list of abbreviations may be introduced.	Agreed.
	9	It is recommended to indicate bilirubin levels throughout the reflection paper in terms of conjugated bilirubin levels. <u>Motivation</u>: Outside the liver, bilirubin is available in its unconjugated form. Unconjugated bilirubin is conjugated with glucuronic acid in the liver. Hence, elevated conjugated bilirubin levels indicate cholestatic or hepatocellular diseases (Kwo et al. 2017). Since PBC concerns an hepatic disease, bilirubin biomarker levels should be expressed in terms of conjugated bilirubin levels throughout the reflection paper.	Not agreed. Total bilirubin is most commonly measured as standard in drug development and used for example in defining Hy´s Law and in the MELD score. Different terminology such as direct vs conjugated bilirubin makes unconjugated bilirubin less suitable. Furthermore, in chronic liver disease, in particular advanced disease, there is typically a mixture of elevated conjugated and unconjugated bilirubin.
76-88	9	Comment: Since chronic non-infectious liver diseases develop slowly, appropriate definitions of intermediate and long-term endpoints are essential for clinical	Partially agreed and addressed in the final Reflection papers.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		studies. For this reason, these endpoints should be defined without brackets. Moreover, chosen definitions of the endpoints should be maintained throughout the reflection paper. Proposed change (added text: underlined, removed text: ~~striked through~~): Chronic liver disease is a slowly developing process, and many patients do not develop relevant disease sequelae, and/or symptoms ~~over~~ even over a considerable time of observation, and the development of end-stage liver disease may be a process of years, if not decades. All three diseases under consideration ~~will b~~are difficult to be studied for long-term outcomes over a reasonable time span. ~~(t~~The term "long-term ~~outcome~~endpoint" is used in the following for events such as liver transplantation and death, as well as clinical events of decompensation of liver cirrhosis which are otherwise also termed "hard ~~outcomes~~endpoints"~~)~~. An acceptable regulatory strategy for companies developing new agents in the disease area, may be to look for intermediate endpoints for which a reasonable assumption for the prediction of long-term ~~outcomes~~endpoints can be made. These reasonable assumptions are usually based on associations with	The introductory paragraph has been reworded and reformatted. Pls. See chapter 4 of the NASH reflection paper. No need to insert the proposed detailed corrections.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		regard to risk factors for the long-term ~~outcomes~~endpoints in observational natural history cohorts and the biological plausibility attributed. ~~(t~~The term "intermediate endpoint" will be used throughout in the following for events otherwise also termed "interim" or "surrogate" endpoint~~)~~. ~~Strictly speaking, h~~However, such endpoints are not yet validated in the sense that positive changes for the ~~surrogate~~intermediate ~~—~~as well as the long-term ~~outcome~~endpoint have repeatedly and consistently been demonstrated for therapeutics. (…)	
78-94	8	Comment: Section 4.1 "General considerations" states that "[a]ll three diseases under consideration will be difficult to be studied for long-term outcomes over a reasonable time span," and yet goes on to say that "confirmation of efficacy (and safety)…after approval" based on long-term outcomes will be required. Due to the slow progression of each disease and the length of time needed for such outcome studies, such studies may be challenging to complete and sponsors may determine not to invest in further development in these disease areas. Accordingly, the Agency is encouraged to continue to think creatively about how therapies can be pragmatically developed based on the best scientific data and tools available without placing the bar so high as to discourage development in these areas of clear unmet medical need.	Agreed and addressed in the final version of both reflection papers. Since the reflection papers have been split between the orphan diseases (PBC and PSC) and NASH, the issue is dealt with somewhat differently.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		Proposed change (if any): See recommendations below at Line no. 261 and 465.	
Lines 99-101	3	Comment: Regarding the use of placebo as comparator, it should be added in the guideline that such use should only be accepted on a case-by-case basis until new substances have been authorised. Proposed change (if any):	This paragraph has been deleted as there are now two separate Reflection papers. Chapter 5.3.1 of the NASH reflection paper has a new paragraph on placebo which accounts for the comment. Use of placebo is also addressed in relevant chapters of the Reflection paper on cholestatic liver diseases.
107-114	9	Comment: Some editorial adjustments are proposed. Proposed change (added text: underlined, removed text: ~~striked through~~): These intermediate endpoints (as well as the long-term endpoints) are currently partly or mainly based on the histological evaluation of liver biopsies. Liver biopsy and histology have been widely criticized for sampling error and intra- and inter-observer variability2. Liver biopsy is also unwanted due to its patient burden, invasiveness, and the associated risks of morbidity and potentially even mortality (3). However, potential non-invasive methods ~~do~~ currently have insufficient, and especially insufficient disease specific, validation data available~~,~~. ~~and~~ ~~t~~Therefore, histology is in most cases still regarded to be the state-of-the-art for the diagnosis, and ~~especially for~~	This paragraph has been deleted as there are now two separate Reflection papers. However, the comment has been implemented partially.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		the follow-up of ~~the course of~~ the diseases, in particular for ~~the purpose of~~ clinical trials. ~~Liver biopsy, however, is also unwanted due to its patient burden, invasiveness, and the associated risks of morbidity3 and potentially even mortality.~~	
Line 107-114 **Patient perspectives from PSC Support**	17	Comment: Patients support the need to validate non-invasive intermediate end points with a view to replacing liver biopsy/histology. We welcome the acknowledgement that 'liver biopsy is unwanted due to its patient burden, invasiveness, and the associated risks of morbidity and potentially even mortality.' However, PSC patient surveys have shown that PSC patients are prepared to undergo liver biopsies in clinical trials if there is clear justification for their use, for example as a means to validating a non-invasive evaluation method (Walmsley et al., 2019; www.pscsupport.org.uk, 2016). Proposed change (if any):	This paragraph has been deleted as there are now two separate Reflection papers. The patient perspective is, however, acknowledged.
Section 4 Recommendations 4.1 General Considerations Lines 107-119	13	Comment: We applaud the reflection paper's language acknowledging the clear deficiencies attached to liver biopsy. As mentioned within the paper, and as liver patients, we agree that liver biopsy is unwanted due to its burden, invasiveness and the associated risks of morbidity and potentially even mortality. Most importantly, we appreciate the calls for future drug approval applicants to further development of non-invasive	This paragraph has been deleted as there are now two separate Reflection papers.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		methods to replace liver histology in tandem with developing their new medicinal products. Proposed change (if any):	
Lines 113-114	7	**Comment:** It would be worthwhile highlighting the inaccuracies associated with liver biopsy and the need to ensure access to a pathologist with expertise in reviewing liver biopsies. **Proposed Change (if any):** Liver biopsy, however, is also unwanted due to its patient burden, invasiveness, and the associated risks of morbidity and potentially even mortality, ***and is susceptible to sample and reader variability***.	This paragraph has been deleted as there are now two separate Reflection papers. Inaccuracies of liver biopsies are mentioned in the NASH reflection paper (Chapter 5.1) in the context of the need to develop non-invasive surrogates.
Lines 115-119:	5	Comment: The call for further development of non-invasive measures to replace liver histology in the future is strongly supported by patients. Proposed change (if any):	Comment welcomed and agreed with. This paragraph has been deleted as there are now two separate Reflection papers. NASH reflection paper includes a relevant paragraph in chapter 5.1.
115-119	8	Comment: Given validation of novel methods to replace histology is called for by the Agency and would likely be incorporated in future Phase 3 development programs conducted by sponsors, this reflection paper should be expanded to provide	The need for novel methods is acknowledged and reflected in the current Reflection papers. This paragraph has been deleted as there are now two separate Reflection papers.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		further details regarding the level of evidence required of sponsors to utilize such novel methods.	
Line 115-119 Patient perspectives from PSC Support	17	Comment: The call for the further development of non-invasive measures to replace liver histology in the future is strongly supported by patients. Proposed change (if any):	Comment welcomed and agreed with. This paragraph has been deleted as there are now two separate Reflection papers. A quite similar paragraph is still included in the NASH reflection paper.
117	9	Comment: please remove the word 'should' Proposed change (if any): … applicants ~~should~~ use …	This sentence has been deleted as there are now two separate Reflection papers.
NASH Section 4.2. Lines 131-135	6	Comment: It addition, it is also acknowledged that "*NAFLD as well as NASH are associated with other comorbidities and risk factors such as obesity, arterial hypertension, diabetes mellitus type 2 (T2DM), atherogenic dyslipidaemia, and others*" and that "*the disease ...has been regarded to be the hepatic manifestation of the so-called metabolic syndrome"*. We believe this confirms that treatments should be required to have no negative action and ideally a beneficial one on these parameters. We consider that this cardio-metabolic beneficial effect should be specifically mentioned at least on the balance benefits/risks.	Comment noted and partially agreed. However, treatments for NASH are or should be primarily intended to treat liver disease. Cardiovascular disease is specifically dealt with in the safety chapters.
Section 4.2 Non-	13	Comment: We appreciate the agency's inclusion of the connection between NASH and other	Comment noted.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
alcoholic steatohepatitis 4.2.1 Short characterizations of the disease Line 129 4.2.3 Study design and endpoints Lines 214-219		diseases, along with severe health complications such as HCC and death. Obesity, diabetes, metabolic syndrome, chronic kidney disease and cardiovascular disease (CVD) are all impacted by and interconnected with NAFLD and NASH. CVD for example, is the leading cause of death among people with NASH. Proposed change (if any):	
Lines 139-140	7	**Comment:** Recent research indicates that although "mild" (F1) patients tend to be asymptomatic and less impacted by NASH, moderate to severe (F2-F3/F4) patients are symptomatic and more highly impacted psychologically. Pain and tiredness are the most common symptoms experienced by F2 to F4 NASH patients. **Proposed Change (if any):** ~~Although h~~***H***ealth-related quality of life may be impaired, ***and despite being perceived as asymptomatic, non-specific symptoms such as tiredness and pain are common even in non-***	Comment partially agreed, and a respective paragraph on patient reported outcomes has been inserted. Since a more general approach has been taken, the specific comments were not adopted.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		cirrhotic NASH patients. ~~symptoms do not play a relevant role in (non-cirrhotic) NASH~~	
139 - 141	4	**Comment:** - It is considered inconsistent to state that HRQOL is affected and there are no symptoms. The impact of HRQOL is currently being assessed (e.g. Younossi, development of the NASH-CHECK PRO). **Proposed change (if any): "~~Although~~** Health-related quality of life may be impaired, **and research to understand the role of symptoms ~~do not play a relevant role~~** in (non-cirrhotic) NASH **is ongoing**."	Comment noted. See above.
139 - 141	10	**Comment:** It is considered inconsistent to state that HRQOL is affected and there are no symptoms. The impact of HRQOL is currently being assessed (e.g. Younossi, development of the NASH-CHECK PRO). **Proposed change (if any): "~~Although~~** Health-related quality of life may be impaired, **and research to understand the role of symptoms ~~do not play a relevant role~~** in (non-cirrhotic) NASH **is ongoing**."	Comment noted. See above.
Line 142	11	**Comment:** NASH natural history is not fully elucidates and not all NAFLD subjects develop NASH as well as not all NASH subjects progress to decompensated cirrhosis and it`s complications. In the lack of broad NASH disease progression epidemiological data an early fibrosis development markers (both invasive and non-invasive) are highly needed.	Comment noted and partially agreed. The proposed paragraph has been edited and shortened.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		Proposed change (if any): The natural history of NASH has not been fully elucidated, and further efforts are needed to clarify important aspects, e.g. overlap of progression and regression, and especially potential early markers of fibrosis progression for determination of both NAFLD and NASH subpopulations at higher risk of disease progression.	
Lines 143-144	7	**Comment**: Consider strengthening language around fibrosis as a predictor of clinical outcomes (Angulo P, Kleiner DE, Dam-Larsen S, Adams LA, Bjornsson ES, Charatcharoenwitthaya P, et al. Liver Fibrosis, but No Other Histologic Features, Is Associated With Long-term Outcomes of Patients With Nonalcoholic Fatty Liver Disease. Gastroenterology 2015;149 (2):389-97 e10.; Dulai PS, Singh S, Patel J, Soni M, Prokop LJ, Younossi Z, et al. Increased Risk of Mortality by Fibrosis Stage in Nonalcoholic Fatty Liver Disease: Systematic Review and Meta-Analysis. Hepatology AASLD Abstracts 2017;65 (5).) **Proposed change (if any):** "~~The risk of progression to end-stage liver disease is largely related to the baseline fibrosis grade.~~ In	Partially agreed The last sentence in the introductory chapter 4 takes account of this.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		patients with NASH, fibrosis is the only independent histologic predictor of adverse clinical outcomes, including liver-related death."	
Line 146	7	**Comment:** The rate of progression is discussed but there is no mention that fast progression is seen in some patients. Data from a single failed Phase 2 study that analyzed 477 subjects with advanced fibrosis due to NASH indicated that 21.5% of patients with F3 bridging fibrosis progressed to F4 in 24.9 months of follow-up; 19.0% of those with F4 had a clinical event within 26.7 months of follow-up. (Sanyal A. et al. J Hepatol 2017;66(1):S2–S3) **Proposal:** "The progression of fibrosis is estimated to be slow, and progression of 1 fibrosis stage is estimated to occur at a mean of more than 7 years (7.7 years; 95% CI 5.5-14.8 y) [141516]. ***However, significantly faster progression does occur in a number of patients.***"	Comment noted and partially agreed. Text has, however, been simplified.
147-212	4	**Comment:** - There is a substantial difference in natural history between fibrosis stage 2 and fibrosis stage 3 patients. Fibrosis stages 3 and 4 patients are under higher risk for liver related outcomes. This has been demonstrated in the literature – notably in the paper *'Liver Fibrosis, but No Other Histologic*	Partially agreed. Chapter 5.3.3. takes account of developments in a mixed cirrhotic and non-cirrhotic population. However, the distinctions are still rather based on the principles to be derived from the necessary outcomes to be studied.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		Features, Is Associated With Long-Term Outcomes of Patients With Nonalcoholic Fatty Liver Disease', Angulo P, Kleiner DE, Dam-Larsen S, Adams LA, Bjornsson ES, Charatcharoenwitthaya P, Mills PR, Keach JC, Lafferty HD, Stahler A, Haflidadottir S, Bendtsen F, Gastroenterology. 2015 Aug;149(2):389-97.e10. It would be more appropriate to distinguish between fibrosis stage 2 and fibrosis stage 3 patients in the guideline instead of combining them, to avoid any disconnect between the natural history of the disease and the regulatory patient grouping.	Also, from a medical point of view, preventing the worsening of fibrosis and eventually the manifestation of cirrhosis remains a valid approach to treat patients early.
Lines 151-157	7	**Comment**: NASH is a histopathologic diagnosis, despite hepatic histological examination having inherent limitations such as invasiveness, procedural risk (e.g. bleeding, death), sampling and evaluation error, but fibrosis can be staged noninvasively. **Proposed change (if any):** "Histology is currently considered to be ~~the gold~~ standard for finally securing the diagnosis, as well as determining the severity of disease, and is ~~also~~ recommended as part of clinical practice ***for some patients***. ***However, liver fibrosis burden can be determined accurately using noninvasive tests.*** Increased fibrosis burden, whether determined by biopsy or NITs clinical outcomes (Angulo et al, Boursier et al, Parkes, et al) is associated with	Partially agreed. However, the proposed changes are no longer applicable, since the chapter on patient selection has been relevantly shortened.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		increased likelihood of hepatic decompensation events. A selection of patients….”	
154-157	4	**Comment:** - Please clarify that biopsy is not required for early exploratory studies i.e. those assessing biomarkers. **Proposed change (if any): For early phase 2 trials, ~~a~~** selection of patients on the basis of **either a known histological diagnosis of NASH or a combination of biochemical criteria and/or imaging evidence of steatosis/steatohepatitis/fibrosis in addition to known risk factors for NASH is appropriate. ~~symptoms is usually not possible, and the (long-standing) presence of features of the metabolic syndrome can only be used as a trigger to identify potential study participants.~~**	Recommentation partially agreed. Chapter 5.2.2 includes a wording with conditions in case no histology based evaluation is done in early trials.
154-157	10	**Comment:** Please clarify that biopsy is not required for early exploratory studies i.e. those assessing biomarkers. **Proposed change (if any): For early phase 2 trials, ~~a~~** selection of patients on the basis of **either a known histological diagnosis of NASH or a combination of biochemical criteria and/or imaging evidence of steatosis/steatohepatitis/fibrosis in addition to known risk factors for NASH is appropriate. ~~symptoms is usually not possible, and the (long-standing) presence of features of the metabolic syndrome can only be used as a trigger to identify potential study participants.~~**	See above.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
Line 157	11	**Comment:** NASH is a diagnosis of exclusion of alternative liver disease aetiology in metabolic syndrome state. In such a metabolic syndrome clinical state potential confirmation of steatosis is achievable by imaging methods (US, MRI), liver inflammation can be confirmed by cytolysis (AST/ALT elevation) monitoring and it is the extent of liver fibrosis and fibrosis progression risk that have to be assessed by liver histology in the current lack of validated non-invasive methods of liver fibrosis assessment. **Proposed change (if any):** A selection of patients on the basis of symptoms is usually not possible, and the (long-standing) presence of features of the metabolic syndrome can only be used as a trigger to identify potential study participants. In the case of confirmed metabolic syndrome, liver steatosis and chronic liver inflammation (ie cytolysis) it is the liver histology assessment necessary to determine the extent of liver fibrosis and, potentially, the risk of disease progression.	Partially agreed. However, the new structure of the document does not fit with the proposal
160 - 161	4	**Comment:**	

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		- Fast progressors as a subpopulation of fibrosis stage 1 patients may be clinically important, and should be mentioned in the guidance. **Proposed change (if any):** Fibrosis stage 1 patients are therefore currently only recommended for inclusion in therapeutic trials in NASH for exploratory purposes. Fibrosis stage 1 patients with **additional risk factors (F1 "fast progressors") may be a clinically important subpopulation for inclusion in clinical trials.**	Not agreed. Identification of "fast progressors" is still regard and "evolving" field on which no clear recommendation seems possible.
160 - 161	10	**Comment:** Fast progressors as a subpopulation of fibrosis stage 1 patients may be clinically important, and should be mentioned in the guidance. **Proposed change (if any):** Fibrosis stage 1 patients are therefore currently only recommended for inclusion in therapeutic trials in NASH for exploratory purposes. Fibrosis stage 1 patients with **additional risk factors (F1 "fast progressors") may be a clinically important subpopulation for inclusion in clinical trials.**	See above.
Line 161	11	**Comment:** In the lack of broad NASH epidemiological data an early fibrosis development markers (including more granular liver histology) is highly needed. **Proposed change (if any):**	Not agreed. The sentence is left unchanged. The purpose of including these patients is left to the discretion of the Sponsors.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		Fibrosis stage 1 patients are therefore currently only recommended for inclusion in therapeutic trials in NASH for exploratory purposes, including for defining more granular histological markers for fibrosis progression risk and for potential fibrosis regression.	
162-174	10	**Comment:** The requirement of a NAS score ≥ 4 as inclusion criterion is of concern. Whereas the diagnosis of NASH relies on presence and pattern of histological abnormalities on liver biopsy, the NAS score was developed and is used as a separate scoring tool to measure changes in NASH during clinical trials. Specific threshold values of the NAS (as used in certain clinical studies) do not correlate well with the histological diagnosis of definitive NASH as noted by Brunt et al (2011). It is therefore proposed to base inclusion of NASH patients on histological confirmation of NASH. **Proposed change (if any):** Therefore, the "natural" selection of patients with an unmet need for treatment ~~in NASH~~ relates to patients with **histologically confirmed NASH with** (fibrosis) stages 2-4 ~~**NASH.**~~ ~~**Inclusion of patients in fibrosis stages 2 and 3 should additionally be based on the disease activity / grading because developments of regression and progression may overlap, and a (albeit univariate only) risk of progression has also been associated with higher degrees of ballooning and inflammation. The patient population should be included based on a valid grading system for NASH with minimal requirements for the presence of cell stress**~~	Not agreed. It is considered that a certain level of disease "activity" (inflammation,ballooning and fat accumulation) should be present in the trial population to allow a more granular evaluation.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		~~(ballooning), as well as inflammation (lobular inflammation). The NASH-CRN (Non-alcoholic steatohepatitis clinical research network) histological scoring system currently appears to be the best validated and most widely accepted system. A total NAS (NAFLD activity score) of at least 5 appears acceptable but a score of 4 can be accepted as well, if it is not based on a high contribution of the steatosis grade alone and minimal requirements for relevant ballooning and lobular inflammation are fulfilled (scoring of at least 1 in each of these 2 components).~~	
164-176	4	**Comment:** - It would be helpful to have a couple of examples of NASH patient inclusion grading systems other than the NASH-CRN grading system that would be acceptable to the Agency. - The language in this section about inclusion of patients in NASH trials could be simplified, and improve understanding. Highlighting the importance of a recent histological diagnosis would add value to the reflection paper. - The requirement of a NAS score ≥ 4 as inclusion criterion is of concern. Whereas the diagnosis of NASH relies on presence and pattern of histological abnormalities on liver biopsy, the NAS score was developed and is used as a separate scoring tool to measure changes in NASH during clinical trials. Specific threshold values of the NAS (as used in	Partially agreed. Other potential grading systems , in particular SAF, are mentioned. Comment noted. The paragraph has been reworded. Not agreed. A certain level of disease activity in all the features assures adequacy of treatment need, and the sensitivity to change.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		certain clinical studies) do not correlate well with the histological diagnosis of definitive NASH as noted by Brunt et al (2011). It is therefore proposed to base inclusion of NASH patients on histological confirmation of NASH. **Proposed change 1 (majority of partners)**: Inclusion of patients **in clinical trials of non-cirrhotic NASH should be based on a histological confirmation of NASH with** ~~in~~ fibrosis stage~~s~~ 2 ~~and~~ **or** 3**, (i.e., a baseline biopsy should be no more than 6-12 months before enrollment).** ~~should a~~ **A**dditionally, **inclusion should** be based on the disease activity/grading because developments of regression and progression may overlap, and a (albeit univariate only) risk of progression has also been associated with higher degrees of ballooning and inflammation. The patient population should be included based on a valid grading system for NASH with minimal requirements for the presence of cell stress (ballooning), as well as inflammation (lobular inflammation). The NASH-CRN (Non-alcoholic steatohepatitis clinical research network) histological scoring system currently appears to be the best validated and most widely accepted system. A total NAS (NAFLD activity score) of **greater than or equal to 4** ~~at least 5~~ appears acceptable ~~but a score of 4 can be accepted as well, if it is not based on a high~~	Agreed. Some simplification inserted. But text changed due to new structure.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		~~contribution of the steatosis grade alone and~~ **provided** minimal requirements for relevant ballooning and lobular inflammation are fulfilled (scoring of at least 1 in each of these 2 components). Although the NASH-CRN grading system is the recommended grading system, patients may also be included based on potentially other grading systems for NASH, provided the validation of respective grading systems is substantiated. **Proposed change 2 (NVS):** Therefore, the "natural" selection of patients with an unmet need for treatment ~~in NASH~~ relates to patients with **histologically confirmed NASH with** (fibrosis) stages 2-4 ~~**NASH.**~~ ~~**Inclusion of patients in fibrosis stages 2 and 3 should additionally be based on the disease activity / grading because developments of regression and progression may overlap, and a (albeit univariate only) risk of progression has also been associated with higher degrees of ballooning and inflammation. The patient population should be included based on a valid grading system for NASH with minimal requirements for the presence of cell stress (ballooning), as well as inflammation (lobular inflammation). The NASH CRN (Non-alcoholic steatohepatitis clinical research network) histological scoring system currently appears to**~~	Partially agreed. Wording changed due to new structure. Requirement for historical biopsy has been set to 6 months.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		~~**be the best validated and most widely accepted system. A total NAS (NAFLD activity score) of at least 5 appears acceptable but a score of 4 can be accepted as well, if it is not based on a high contribution of the steatosis grade alone and minimal requirements for relevant ballooning and lobular inflammation are fulfilled (scoring of at least 1 in each of these 2 components).**~~ **Rationale**: First, it would be helpful to highlight that this paragraph concerns clinical trials in patients with non-cirrhotic NASH. Subsequent paragraphs in the reflection paper discuss trials in patients with cirrhotic NASH. Second, the text should be clarified to emphasize that before assessing disease activity, a histological diagnosis of NASH must be established. As explained by Kleiner and colleagues in their original paper describing the NAS, "It is important to note that the primary purpose of the NAS is to assess overall histological change; it is not intended that numeric values replace the pathologist's diagnostic determination of steatohepatitis (Kleiner et al 2005[i])." Third, highlighting the importance of a recent histological diagnosis would add value to the reflection paper. As noted in the FDA draft guidance, baseline histology is critical for efficacy evaluation; liver biopsies obtained more than 6 months before enrollment may not represent an accurate status of the disease at the	

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		beginning of the trial. Fourth, the language regarding the criteria for disease activity could be simpler. The reflection paper stated that an NAS of 4 is acceptable if minimum requirements for ballooning and lobular inflammation are fulfilled. Thus, there is no reason to recommend an NAS of at least 5. Including such a criterion seems to imply that an NAS of 5 consisting of a steatosis score of 0, a ballooning score of 2 and an inflammation score of 3 is acceptable. This adds unnecessary confusion to the reflection paper because a steatosis score of 0 would only rarely be seen in a patient with histological diagnosis of NASH confirmed by a competent pathologist. The revised text is easier to understand and would be harmonised with the FDA draft guidance.	
166	8	Comment: The presence of steatohepatitis based on pathologist's diagnostic assessment of the overall pattern of injury has been associated with the risk of progression to cirrhosis.	Comment noted. See comments above.
Line 167	11	**Comment:** Liiver histology assessment is a 2-dimensional assessment of both inflammation and fibrosis in NASH which is a result of metabolically driven liver disease. The extent of steatosis and inflammation might be assessed non-invasively but the extent of liver fibrosis needs more granular evaluation criteria to determine	Not agreed. Parameters for insulin resistance, obesity, cholesterol and triglycerides are relevant safety parameters to be evaluated, but may out of the focus of substances not primarily addressing metabolic alterations.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		the risk of progression in addition to the current fibrosis stage. Considering that liver histology criteria for inflammation improvement are secondary to the metabolic state improvement (insulin resistance, hyperlipidaemia, obesity) in NASH the use of already validated clinical criteria for insulin resistance, obesity and serum cholesterol and triglyceride level might be a requisite for addition / alternative of the histological assessment of liver inflammation. **Proposed change (if any):** The patient population should be included based on a valid clinical metabolic syndrome criteria and valid grading system for NASH with minimal requirements for the presence of cell stress (ballooning), as well as inflammation (lobular inflammation).	
170-174	1	Comments: The European Association for the Study of the Liver (EASL) Clinical Practice Guidelines does not include NAS for the diagnosis or definition of NASH3. Additionally, the practice guidance from the American Association for the Study of Liver Diseases (AASLD) defines nonalcoholic fatty liver (NAFL) as the presence of ≥5% hepatic steatosis without evidence of hepatocellular injury in the form of hepatocyte ballooning, and nonalcoholic steatohepatitis (NASH) as the presence of ≥5% hepatic steatosis in addition	Comment noted but only partially agreed. See comments above. The criteria for inclusion based on NAS of 4 or higher have been accepted and implemented in multiple clinical trials.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		to presence of lobular inflammation with hepatocyte injury, exhibited as hepatocellular ballooning1. Importantly, this practice guidance does not reference to NAS in these definitions. The evaluation of NAS was originally designed to assess overall histologic change before and after therapeutic intervention trials and was not intended as numeric values to replace a pathologist's diagnostic determination of steatohepatitis 2. Incorporating a minimum score for NAS for inclusion of study participants may unnecessarily limit the potential study population and may lead to exclusion of patients with histopathologic diagnosis of NASH per a central pathologist's assessment, including those with 5-33% steatosis (grade 1), < 2 foci per 200 x field lobular inflammation (grade 1), with few ballooned hepatocytes (grade 1) and liver fibrosis per NASH CRN, who may benefit from participating in clinical trials. Proposed change (if any):	
Lines 174-179	7	**Comment**: A diagnosis of NASH is made by a pathologist and is not dependent on a certain NAS threshold (Brunt EM, Kleiner DE, Wilson LA, Belt P, Neuschwander-Tetri BA, Network NCR. Nonalcoholic fatty liver disease (NAFLD) activity score and the histopathologic diagnosis in NAFLD: distinct	

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		clinicopathologic meanings. Hepatology 2011;53 (3):810-20.) **Proposed change (if any):** "…..***Patients should have a histologic diagnosis of NASH in the opinion of the pathologist or have a diagnosis of NASH based on medical history, including presence of relevant comorbidities such as diabetes, and exclusion of other causes of liver disease.*** Although the NASH-CRN grading system is the recommended grading system, patients may also be included based on potentially other grading systems for NASH, provided the validation of respective grading systems, ***including quantitative assessments of histological features,*** is substantiated. ***If the anticipated effect of the investigational drug is to reduce fibrosis and fibrosis is assessed using an ordinal staging system, then fibrosis stage should be sufficiently elevated to evaluate a potential response while a certain NAS may not be required.*** In patients with manifest cirrhosis (=fibrosis stage 4), the presence of such a rigorous minimal grade is less critical, because the risk of (clinical) progression is thought to be high based on the presence of cirrhosis alone."	

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
Lines 179-182	7	**Comment**: Specific criteria for T2DM if a historical biopsy is used. **Proposed change (if any):** "….. ~~all~~ ***any one*** of the following should be available in order to make the diagnosis NASH sufficiently likely: historical biopsies with presence of unequivocal NASH, a high likelihood of NASH based on non-invasive testing (biomarker and imaging), ~~and~~ ***or*** presence of associated co-morbidity (e.g. obesity with T2DM).	Not agreed. Diagnostic criteria for T2DM are not needed in a paper on NASH. The requirements implemented for the "burnt-out NASH" population have been amended. Rather strict criteria are used, since it the "diagnostic accuracy" of the Liver Forum classification is unclear (See Chapter 5.24). The primary aim of the proposed criteria was to assure that only "true NASH-cirrhosis" patients are admitted to the trials. Reference is also made to the possibility to make proposals within a Scientific Advice procedure, once accurate data can be presented for less stringent criteria.
179-183	10	**Comment**: Cirrhosis is a clinical diagnosis that occurs usually late and most patients do not have a historical biopsy. To confirm NASH as the etiology, physicians can and do i/ rule out other diseases, as well ii/ perform non-invasive testing in clinical practice, and consider iii/ commonly associated co-morbidities i.e. (e.g. obesity with T2DM). **Proposed change (if any): ~~Nevertheless, in so-called burnt-out NASH cirrhosis or patients initially diagnosed with cryptogenic cirrhosis, if definite NASH is not present, all~~** **In order to diagnose NASH cirrhosis, either** of the following should be available **~~in order to make the diagnosis of NASH sufficiently likely~~**: historical biopsies with presence of unequivocal NASH,	Comment noted. See above.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		or a clinical diagnosis ~~a high likelihood~~ of NASH based on non-invasive testing (biomarker and imaging), and**/or** presence of associated co-morbidity (e.g. obesity with T2DM).	
179-183	4	**Comment**: - Suggest reducing the requirements to 'historical biopsy with presence of unequivocal NASH'; this is considered enough to set the diagnose and stage of decompensated cirrhosis, and it will make recruitment of trial subjects more feasible. - Cirrhosis is a clinical diagnosis that occurs usually late and most patients do not have a historical biopsy. To confirm NASH as the etiology, physicians can and do i/ rule out other diseases, as well ii/ perform non-invasive testing in clinical practice, and consider iii/ commonly associated co-morbidities i.e. (e.g. obesity with T2DM). **Proposed change (if any): ~~Nevertheless, in so-called burnt-out NASH cirrhosis or patients initially diagnosed with cryptogenic cirrhosis, if definite NASH is not present, all …~~ In order to diagnose NASH cirrhosis, either** of the following should be available **~~in order to make the diagnosis of NASH sufficiently likely~~**: historical biopsies with presence of unequivocal NASH, **or a clinical diagnosis ~~a high likelihood~~** of NASH based on non-invasive testing (biomarker and imaging), and**/or**	Comments noted. Similar as above.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		presence of associated co-morbidity (e.g. obesity with T2DM).	
179-188	14	**Comment**: In terms of NASH as an etiology of cirrhosis, we agree that histologic diagnosis of NASH is the reference standard. However, the typical histologic parameters may be absent as NAFLD/NASH progresses to cirrhosis, and many patients with cirrhosis will not have had a historical biopsy. The Liver Forum's NASH Cirrhosis Working Group has engaged a multi-stakeholder group to develop criteria by which to define NASH as the etiology of cirrhosis, and we anticipate that these recommendations will be published in the near term. Lastly, it is important to note that all other etiologies of cirrhosis must be excluded. **Proposed change (if any)**: "Patients should have histological diagnosis of NASH and other causes of liver disease should be ruled out. When histological evidence is absent, evidence of NASH based on non-invasive testing along with the presence of metabolic risk factors, and the ruling out of other causes of liver disease, strengthen the likelihood of NASH as the cause of cirrhosis. These patients may be considered to have a diagnosis of NASH, and specific criteria for inclusion should be discussed with EMA prior to trial."	Similar as above.
180-183	2	**Comment:**	

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		As for non-cirrhotic NASH, exclusion of other causes of cirrhosis is most important to diagnose cirrhosis caused by NASH. It would be problematic if there is a need for an historical biopsy taken at a time-point when the liver was not cirrhotic since that would exclude many patients. Moreover, it is questionable to what extent biomarkers, and especially imaging, could help for diagnosis of cirrhotic NASH, unless the biomarkers are used to exclude other causes of cirrhosis. **Proposed change (if any):** We suggest to keep the text indicating that patients should be included based on a high likelihood of cirrhosis caused by NASH; a diagnosis based on exclusion of other causes of cirrhosis until better biomarkers become available.	The intent of the comment is not fully understood. It is refered to the above comments, and the fact that Chapter 5.2.4 has been reworded.
184-188	14	**Comment**: Similarly to the previous comment, we note that many patients with decompensated cirrhosis will not have had a historical biopsy. **Proposed change (if any)**:	Comment noted. The paragraph on decompensated cirrhosis is kept unspecific only and no clear recommendations given. Experience with development programmes in this population is very limited.
186-188	4	**Comment:** - Decompensated cirrhosis is a clinical diagnosis and most patients do not have a historical biopsy. At the time of decompensation, it can also not be	Comment noted. See above.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		clinically recommended to perform a liver biopsy in all cases. - Due to the risks listed, suggest adding that use of biomarkers may be beneficial. - To confirm NASH as the etiology, physicians can and do i/ rule out other diseases, as well as ii/ perform non-invasive testing in clinical practice, and consider iii/ commonly associated co-morbidities i.e. (e.g. obesity with T2DM). **Proposed change (if any):** Due to the fact of increased risks of biopsies in this population, **a clinical diagnosis of decompensated cirrhosis and one or more additional factors to determine NASH as etiology, i.e. a high likelihood of NASH based on non-invasive testing (biomarker and imaging), or presence of associated co-morbidity (e.g. obesity with T2DM) or** historical biopsies (with presence of cirrhosis) may be used as inclusion criteria in this population.	
186-188	10	**Comment:** Decompensated cirrhosis is a clinical diagnosis and most patients do not have a historical biopsy. At the time of decompensation, it can also not be clinically recommended to perform a liver biopsy in all cases. To confirm NASH as the etiology, physicians can and do i/ rule out other diseases, as well as ii/ perform non-invasive testing in clinical practice, and	Comment noted. See above.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		consider iii/ commonly associated co-morbidities i.e. (e.g. obesity with T2DM). **Proposed change (if any):** Due to the fact of increased risks of biopsies in this population, **a clinical diagnosis of decompensated cirrhosis and one or more additional factors to determine NASH as etiology, i.e. a high likelihood of NASH based on non-invasive testing (biomarker and imaging), or presence of associated co-morbidity (e.g. obesity with T2DM) or** historical biopsies (with presence of cirrhosis) may be used as inclusion criteria in this population.	
Lines 189-193	7	**Comment**: NASH is a histopathologic diagnosis but fibrosis can, and should, be staged noninvasively to mitigate against some risks associated with biopsy. Within earlier Phase trials, patients are increasingly being enrolled into randomised controlled trials that have not been biopsy-diagnosed (thereby reflecting clinical practice and especially relevant when recruiting patients with earlier stages of fibrosis). **Proposed change (if any):** ~~"The multi-stakeholder-composed Liver Forum has recommended that histology should always be available, also in early clinical trials, and inclusion of patients should always be based on histological~~	Comment noted, but not fully agreed. . See above comments on compensated and de-compensated cirrhosis.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		~~evaluation (grading and staging). Deviations for exploratory clinical trials, e.g. using imaging methods, or biomarkers (or a combination of those) only, are possible if based on sound scientific principles, for which the uncertainties can be quantified and later stage trials be planned accordingly.~~ "…..Due to the fact of increased risks of biopsies in this population, historical biopsies (with presence of cirrhosis) together with symptoms of decompensation may be used as inclusion criteria in this population. ***Liver fibrosis can be accurately staged noninvasively and it is encouraged to use noninvasive biomarkers and simple and easy imaging indexes.*** The positive influence of weight reduction…..*"*	
189-194	1	Comments: Early, late and exploratory studies should be defined as there is no clear consensus in clinical trials regarding these definitions, and this would provide clarity in regard to the utilization of histology. Histological changes may not be readily evident in short phase I or exploratory studies. The draft guidance issued by the Food and Drug Administration (FDA) recommend the use of liver biopsy in studies from phase 2b and phase 34.	Comments noted. Currently, the use of non-invasive methods is recommended for enrichment purposes only. Conditions for use of non-invasive methods in early trials are included in Chapter 5.2.2..

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		Proposed change (if any):	
189-194	4	Suggest specifying that endpoints in phase 2 dose finding trials could also be based on biomarkers and or imaging techniques, based on sound scientific principles and evidence, provided endpoints in later stage trials are based on histology	Comment noted. Since no specific recommendations are given for early trials (except for inclusion of patients), no specifics have been included.
189-194	14	**Comment**: We suggest to clarify the sentence "histology should always be available, also in early clinical trials", the Liver Forum paper referred to[1] states that "Liver histology is *ideally* captured for POC trials and is essential for later phase trials". Noninvasive criteria should be considered sufficient for trial entry and endpoint assessment for Phase 1 and 2a trials. We further suggest to clarify what is meant by "early clinical trials", as there is no clear consensus (exploratory vs. Phase 1 vs Phase 2a). We appreciate the flexibility provided, with justification, for noninvasive measures to be utilized in exploratory trials. **Proposed change (if any)**:	Partially agreed. See also above. The text now reads as follows: "Deviations for exploratory clinical trials, e.g. using imaging methods, or biomarkers (or a combination of those) only, are possible if based on sound scientific principles, for which the uncertainties can be quantified and later stage trials be planned accordingly." The term "exploratory trials" makes clear that it is related to both, phase 1 and 2.

[1] https://doi.org/10.1002/hep.29607

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
Line 195	7	**Comment**: Consider 10% as a threshold for significant weight loss. Despite initial weight loss, many NASH patients fail to maintain the loss and therefore NASH progresses. Likewise, there is no evidence of the longer term impact of weight loss on NASH. **Proposed change (if any):** "The positive influence of ***sustained and significant (>10%)*** weight reduction on NASH has ~~clearly~~ been demonstrated."	Comment noted. The requirement for unsuccesful attempts to lose weight before inclusion has been deleted.
195-197	2	**Comment:** It is important to recognize that weight loss is relatively easy to achieve but weight regain is unfortunately almost always seen. This may change with novel pharmacological interventions; however today bariatric surgery is the only treatment with long term weight loss and outcome benefits. In the context of drug development in NASH it is more relevant to include patients that are weight stable and advice a modest lifestyle intervention that has a high probability to be sustained throughout the clinical study. It is unclear what would define an attempt to reduce body weight. Even if a consensus is achieved, it is a	Comment noted and partially agreed. The paragraph with regard to attempts to lose weight has been deleted. Chapter 5.2.1 only includes the requirement for stable weight before inclusion.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		highly subjective from a patient and physicians' perspective what would constitute an attempt to reduce body weight. Also, this rule will result in the consequence that patients with stable and normal weight (BMI ≤25) but with NASH, e.g. for the reason of having familiar NASH, will not be included. **Proposed change (if any):** We suggest that these criteria should be removed	
195-197	4	**Comment:** - It is possible that no standardized programs will be available and therefore should be clarified if local standards can trump any other criteria used - And control for potential confounding factors is always difficult - It would be useful to receive some clarity around the time period and approach that would be acceptable to the Agency to assess unsuccessful weight reduction attempt. We would like to encourage the Agency to recommend how to measure 'one unsuccessful attempt'. Examples of how this is to be included would be helpful. - A healthy diet and exercise plan should be an ongoing feature in clinical trials.	See above.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		- Suggest specifying 'unsuccessful attempt with weight reducing diet'. Moreover, suggest that this requirement could be stage dependent. **Proposed change**: The positive influence of weight reduction on NASH has clearly been demonstrated. ~~Therefore, before inclusion of respective patients into clinical trials for NASH, it is recommended that patients should have undertaken at least one unsuccessful attempt with weight-reducing diet.~~ ***Therefore, Sponsors should consider including diet and exercise counselling for all study participants during the clinical trial, especially for placebo-controlled trials.*** **Rationale:** We agree that lifestyle intervention trials have demonstrated beneficial effects on liver histology (Vilar-Gomez et al 2015[ii]). Diet and lifestyle changes are recommended for all patients with NAFLD in the EASL–EASD–EASO Clinical Practice Guidelines for the management of NAFLD[i]. However, failure to achieve weight loss through dietary adjustments is a subjective measure and is not suitable for an inclusion criterion in a clinical trial. The importance of ongoing attention to a healthy diet and exercise plan should be the focus. For placebo-controlled trials, offering diet and physical activity counselling as part of the trial design is an important ethical consideration. In clinical practice, any future therapies for NASH should be prescribed as an adjunct to diet and exercise. There is a parallel with	

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		the Guideline on clinical investigation of medicinal products in the treatment or prevention of diabetes mellitus (CPMP/EWP/1080/00 Rev. 2), which notes that appropriate life style interventions (i.e. diet and exercise) should be reinforced in all subjects throughout the study. Approved indications for type 2 diabetes medicinal products describe that they should be used as an adjunct to diet and exercise.	
Lines 195-197	7	**Comment**: Regarding weight loss and undertaking of at least one unsuccessful attempt with weight-reducing diet, within standard NASH clinical trials, lifestyle advice is given, however patients will not have had consistently documented efforts before entering the studies. In practice, most patients with metabolic syndrome/ diabetes they will have had extensive advice many times previously from their clinicians / dieticians. If there is a risk of requiring further weight loss attempts before initiating treatment perhaps, especially in advanced disease, the immediate treatment rationale is enhanced if studies demonstrate additional treatment benefits. Stratifying for weight loss could help conclude that improvements are not simply down to weight loss (which in a RCT is likely to be balanced across groups) and therefore the benefit is enhanced and expedited if NASH treatments	See above. In addition weight loss during the trial is dealt with in the chapter on estimands.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		are started simultaneously with lifestyle advice and not deferred pending failure to lose weight. **Proposed change (if any):** N/A	
195-199	14	**Comment**: We believe there is some misinterpretation of this statement, including interpretation that patients must participate in a separate weight loss trial or have an official dietary/ lifestyle intervention before they can be enrolled in a clinical trial for NASH. We suggest that the wording be modified to more clearly define "undertaken at least one unsuccessful attempt". Our understanding of this section is that patients should have been advised that lifestyle changes could positively impact the disease, have made an attempt to lose weight at some point, and despite this attempt, still have NASH. **Proposed change (if any)**:	Comment noted. See above.
197-199	4	**Comment**: - Suggest aligning the specification of adequate and stable treatment of diabetes with the FDA draft NASH guidance, which states that	Comment noted. No specific criteria for (diagnosis and treatment of) comorbidities are given. The final wording refers to "adequate and stable treatment".

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		moderately well-controlled diabetes includes an HbA1c level below 9.5%.	
Line 198:	11	**Comment:** If an IP has at the begin of the study for example cholesterol increasing activity an adequate treatment needs to be started if the subjects are randomized in the study. If the patients with hyperlipidemia are pretreated, the treatment effect of lipid lowering medication can be less effective! **Proposed change (if any):** Due to this discussed above, subjects are not eligible in other NASH studies if LDL ≥190 mg/dL and already on a stable dose of statin and/or proprotein convertase subtilisin/kexin type 9 (PCSK9) inhibitor for ≥30 days at Screening.	Comment noted but not agreed. The general recommendations given are independent from the mechanism of action of certain compounds.
Line 200	11	**Comment:** As NASH is a liver sequelae of the metabolic syndrome and the most frequent mortality reason in NASH population is cardiovascular a strong focus on the metabolic disease leading to NASH is proposed to be recommended. Such a focus would vary from research therapies that does not lead to metabolic syndrome worsening (neutral) to research therapies aiming metabolic syndrome improvement as that is the main factor for NASH development and progression.	Comment noted and partially agreed. However, the focuse of treatment of NASH will remain on the liver. Cardiovascular disease will be included in the composite final endpoint evaluation, and has to be dealt with as a main safety aspect (see chapter 5.3.7.2)

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		Proposed change (if any): Important factors to be considered in all populations are the underlying metabolic syndrome and the presence of co-morbidities (e.g. diabetes), and the investigational treatment potential impact on the metabolic syndrome, has to be considered, as well as stratification for the co-morbid factors could be advisable to allow a balanced evaluation of these covariates.	
Lines 203 - 204	7	**Comment:** NASH is not the prognostic factor, however, advanced fibrosis (AF) is. The scientific and medical communities are increasingly moving towards a diagnosis (without mandatory biopsy) of NAFLD with fibrosis/advanced fibrosis as opposed to NASH with advanced fibrosis. **Proposed change (if any):** N/A	Comment noted and partially agreed. Nevertheless, at this point of time not yet accepted.
Lines 203-212	7	**Comment**: To aid global study conduct we would encourage consistent classification of fibrosis stages. The classification of fibrosis stages (early NASH = F0-1, fibrotic NASH = F2-3, NASH-related liver fibrosis = F4) is currently inconsistent with AASLD (American Association for the Study of Liver Diseases), i.e. "advanced" fibrosis referrs specifically to stages 3 or 4, that is, bridging fibrosis or cirrhosis.	Comment noted and partially agreed. However, the categorisation is no longer part of the disease characterisation (which has been shortened relevantly). Categories expressed via the different numbering paragraphs are given as "non-cirrhotic NASH (F2-3)", and compensated and decompensated cirrhosis. Chapter 5.3.3. takes also account of developments in "advanced fibrosis" populations.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		Proposed change (if any): ***In summary, for the purpose of therapeutic clinical trials, NASH may be considered in three broad categories:*** ~~*a. Definite NASH based on histology with demonstration of NAS≥5 (or NAS ≥4 with all components of at least 1) and fibrosis stage 2-3*~~ ~~*b. Compensated NASH-cirrhosis based on histology with fibrosis stage 4 and NASH diagnosis based on either NAS>5 (or NAS ≥4 with all components of at least 1) or the availability of historical histology proving NASH, non-invasive tests pointing to NASH (serological markers, imaging), and 209 relevant co-morbidity risk-factors (obesity and type 2 diabetes mellitus (DM))*~~ ~~*c. Decompensated NASH Cirrhosis: Presence of historical biopsy data showing unequivocal NASH as well as cirrhosis; symptoms of decompensation.*~~ ***a. NASH with no or mild fibrosis (F0-F2);*** ***b. NASH with advanced fibrosis (F3-F4); and*** ***c. NASH with decompensated cirrhosis.***	
205-206	4	**Comment:** - The summary for Section 4.2.2 (Selection of patient populations) should be revised to be	See above.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		made consistent with the proposed changes for lines 164-176. - Criteria needs to be less strict and contemplate a "high likelihood for NASH", with definitive features for worse metabolic profile or metabolic syndrome. **Proposed change** (if any): a. Definite ***non-cirrhotic*** NASH based on histology with demonstration of NAS≥~~5~~***4*** (~~or NAS≥4~~ with at least 1 point each in inflammation and ballooning ~~all components of at least 1~~) and fibrosis stage 2-3 **Rationale**: The summary of Section 4.2.2. provides a helpful list of 3 broad categories for therapeutic clinical trials. By clarifying that the first category refers to non-cirrhotic NASH, the list becomes even more helpful. See previous rationale for justification for simplifying the language regarding inclusion criteria for disease activity.	
205-206	14	**Comment**: We note that patients with 'borderline NASH' (NAS 3 or 4) might be considered for inclusion in clinical trials as well. These patients also have steatosis, ballooning and inflammation, and develop fibrosis at intermediate rates between definite NASH and NAFL. **Proposed change (if any)**:	Comment noted. See comments above on requirements for cirrhotic NASH

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
207-210	4	**Comment**: - For compensated NASH cirrhosis suggest either removing the requirement for NAS score ⩾4 or changing to NAS⩾3.	
207-210	14	**Comment**: We note that the NASH CRN uses the definition of a NAFLD Activity Score (NAS) of 4 points or higher to identify patients with more significant disease activity. However, some of the histological features indicating disease activity may not be readily apparent in NASH patients with cirrhosis, and the NAS can be lower than 4. We again note that when current/historical histological evidence is absent, evidence of NASH based on non-invasive testing along with the presence of metabolic risk factors, and the ruling out of other causes of liver disease, strengthen the likelihood of NASH as the cause of cirrhosis. **Proposed change (if any)**:	Comment noted and partially agreed. Such criteria have been implemented (see above).
Line 211:		**Comment:** TZDs/glitazones, vitamin E and other drugs with NASH-modifying properties can influence the liver histology **Proposed change (if any):** Due to this discussed above for subjects with a historical biopsy, is recommended either not taking	This comment is not identified to relate to line 211

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		TZDs/glitazones or vitamin E, or be on stable doses of TZDs/glitazones or vitamin E for 6 months before Day 1. Furthermore Changes to these drugs with potential NASH-modifying properties are not permitted for the first 18 months of NASH studies.	
NASH Section 4.2. Lines 217-219	6	Comments: While it is acknowledged both that "*NASH is also associated with a multitude of risk factors for cardiovascular disease (hypertension, obesity, atherogenic dyslipidaemia, and type 2 diabetes*)" and that a "*relevant proportion of patients will also be prone to causes of death other than liver related ones, mainly cardiovascular*", we believe that it should be mentioned in the guidance that it is highly recommended for NASH treatments to have a positive action on metabolic/cardiovascular parameters or at least to be neutral on these elements.	Comment noted and in principle agreed with. However, it is mainly referred to the safety chapter on CV safety and the need to refer to the "Reflection paper on assessment of cardiovascular safety profile of medicinal products" (EMA/CHMP/505049/2015),
Line 220	11	**Comment:** NASH is a liver sequelae of the metabolic syndrome and the most frequent mortality reason in NASH population is cardiovascular. **Proposed change (if any):** The "natural" long-term endpoint in clinical trials for NASH would therefore be the combination of all-cause and CV mortality, CV events, liver transplantation, and the manifestation of	Partially agreed only. Demonstrating the benefit of NASH treatments on CV event occurrence was not deemed adequate, since this would refer to a cardiovascular indication. However, including all-cause mortality into the final (composite)endpoint evaluation is considered to sufficiently take account of the cardiovascular risks being present in the patient population. It is also referrred to the chapter on cardiovascular safety.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		decompensation (variceal bleeding, ascites, encephalopathy etc.).	
220-222 and 231-234	1	Comments: To further clarify/highlight this long-term composite endpoint, it is recommended to describe it as a bullet-list and include examples of decompensated liver events. For consistency, consider alignment with FDA guidance for MELD score in the long-term composite endpoint for efficacy4. Long-term composite endpoint for demonstration of efficacy: • All-cause death • Clinical progression to cirrhosis • Histological diagnosis of liver cirrhosis defined as Stage 4 (F4) by the NASH CRN classification • Liver transplantation • Change in MELD score from ≤ 12 to > 15 • Decompensation of liver disease, e.g. o Hepatic encephalopathy o Variceal bleeding o Ascites Proposed change (if any):	Comment and proposal noted.Bullet list not implemented. MELD score >15 not agreed. the current similar or higher than 15 MELD criterion is based on the EASL as well ass AASLD guidelines on liver transplantation.
Lines 220-222	7	**Comment**: Consider including HCC as a "natural" long-term endpoint in clinical trials. **Proposed change (if any):**	Comment noted. HCC has been dealt with in a separate paragraph.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		The "natural" long-term endpoint in clinical trials for NASH would therefore be the combination of all-cause mortality, ***Hepatocellular Carcinoma (HCC),*** liver transplantation, ~~and~~ the manifestation of decompensation (variceal bleeding, ascites, encephalopathy etc.) ***and progression to cirrhosis for those with <F4 at baseline.***	
220-222	8	Comment: The reflection paper should provide more detail on an appropriate long-term endpoint. More specific details on "manifestation of decompensation" is warranted. The reflection paper should also comment specifically on the use of HCC as part of the long-term endpoint.	Comment noted. The three "acceptable " elements of decompensation are mentioned, and e.g. HCC (and other potential events) are dealt with in Chapter 5.3.4
223-248	14	**Comment**: We appreciate the difficulty of determining appropriate endpoints given the uncertainty associated with the surrogate; however, we are concerned that the requirement to demonstrate efficacy in two composite endpoints in co-primary fashion may limit therapeutic development. Requiring both the resolution of NASH with no worsening of fibrosis AND at least 1-stage improvement of fibrosis without worsening of NASH may be too strict a criteria, may not be achievable with monotherapy, and may have the unintended	Comment noted but not agreed. Reasons are given. See above.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		consequence of impeding development of potentially efficacious drugs. We respectfully suggest incorporating additional flexibility regarding the endpoints, and considering either endpoint individually as an acceptable intermediate endpoint. **Proposed change (if any)**:	
228-234	4	**Comment:** - Add liver transplant due to NASH or its complications to the composite endpoint for long-term outcome studies. - 'progression to cirrhosis' is sufficiently a hard outcome (as recognised in line 214) as to be acceptable as a long-term endpoint rather than being classed as a surrogate endpoint. Other text in this paragraph (line 234) supports this - the inconsistency is to say 'acceptable surrogate' here. Cirrhosis is a clinical diagnosis that does not necessitate liver biopsy. - The same listing related to decompensation of liver disease as the one used in the FDA draft guidance *'Noncirrhotic Nonalcoholic Steatohepatitis With Liver Fibrosis: Developing Drugs for Treatment'* of December	Agreed. Liver transplantation is mentioned. Partially agreed. However, cirrhosis as such is not a "hard" endpoint in the strict sense, and there is wide variability in severity of liver disease in different cirrhosis patients. Nevertheless, cirrhosis is accepted as part of the final commposite in non-cirrrhotic NASH. With regard to MELD: See above.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		2018 should be used for consistency, and specified here for endpoint clarity. - It is proposed that MELD>15 (instead of MELD>14) could be used as a surrogate marker of disease progression in NASH, and can be part of the long term endpoints. This proposed cut-off is relevant to define survival free from transplantation and is in line with published literature (e.g. Lau et al. 2013, Roth et al 2017, Merion et al 2005), and the FDA NASH draft guidance. **Proposed change**: "The ~~**histological**~~ diagnosis of cirrhosis has been proposed to represent such an endpoint, and ~~**is regarded to be an acceptable surrogate and**~~ can ~~**therefore**~~ be part of the long-term endpoints. […] The long-term outcome for the demonstration of efficacy in NASH is therefore proposed to be a composite endpoint with the components all-cause death, decompensation of liver disease (with a complete listing, ***e.g., hepatic encephalopathy, variceal bleeding, ascites), liver transplant (due to NASH or its complications)***, as well as (histological) diagnosis of liver cirrhosis and MELD>~~**14**~~**15**.." **Rationale:** Aligning the EMA reflection paper with the December 2018 FDA draft guidance for industry on	

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		Noncirrhotic Nonalcoholic Steatohepatitis With Liver Fibrosis would greatly facilitate global drug development. The FDA draft guidance includes liver transplant as one of the components of the composite endpoint for the demonstration of efficacy in NASH. Although the other components proposed in the EMA reflection paper would usually precede a patient receiving a liver transplant, clinical situations could arise where these other endpoints are not documented prior to liver transplant. For example, patients with NASH are at higher risk of hepatocellular carcinoma (EASD-EASL-EASO guidelines[iii]) and such patients may be candidates for liver transplant if the tumor is small and the risk of recurrence is evaluated to be low (Viveiros et al 2017[iv]). Therefore, we propose that liver transplant due to NASH or its complications should be included in the composite endpoint for long-term outcome studies.	
228 - 234	10	**Comment:** We propose that 'progression to cirrhosis' is sufficiently a hard outcome (as recognised in line 214) as to be acceptable as a long-term endpoint rather than being classed as a surrogate endpoint. Other text in this paragraph (line 234) supports this - the inconsistency is to say 'acceptable surrogate' here. Cirrhosis is a clinical diagnosis that does not necessitate liver biopsy.	Comment noted. See above.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		Proposed change (if any): The ~~**histological**~~ diagnosis of cirrhosis has been proposed to represent such an endpoint, and ~~**is regarded to be an acceptable surrogate and**~~ can ~~**therefore**~~ be part of the long-term endpoints. […] The long-term outcome for the demonstration of efficacy in NASH is therefore proposed to be a composite endpoint with the components […] as well as ~~**(histological)**~~ diagnosis of liver cirrhosis and MELD>14.	
228-234	14	**Comment**: We suggest that 'progression to cirrhosis' is sufficiently a hard outcome as to be acceptable as a long-term/clinical endpoint rather than being considered as a surrogate endpoint. We further suggest that clinical diagnoses of cirrhosis could be used, as this is routinely performed in clinical practice. Diagnosis of cirrhosis can be based on clinical criteria such as patient history, physical examination, laboratory data, liver imaging, and/or endoscopy. Therefore, we suggest non-histologic criteria for the diagnosis of cirrhosis can be proposed by the sponsor and could be acceptable if scientifically supported. **Proposed change (if any)**:	Not agreed. See above. Not agreed. Can be accepted in day-to-day routine care, but not within the clinical trial setting, unless convincing data can be provided on accuracy of this "clinical diagnosis".

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
234	10	**Comment:** It is proposed that MELD>15 (instead of MELD>14) could be used as a surrogate marker of disease progression in NASH, and can be part of the long term endpoints. This proposed cut-off is relevant to define survival free from transplantation and is in line with published literature (e.g. Lau et al. 2013, Roth et al 2017, Merion et al 2005), and the FDA NASH draft guidance. **Proposed change (if any):** The long-term outcome for the demonstration of efficacy in NASH is therefore proposed to be a composite endpoint with […] and MELD>~~14~~**15**.	Not agreed. As mentioned, the chosen threshold is based on the EASL and AASLD guidelines on liver transplantation.
Section 4.2.3 Study design and endpoints Stage 2 and 3 fibrosis Lines 235-248	13	Comment: We have serious concerns with the unintended consequences of requiring co-primary intermediate endpoints for stage 2 and 3 fibrosis drug therapy development. Requiring a resolution in NASH, and an improvement in fibrosis would be devastating to drug development. It greatly reduces chance of approval, and lengthens the period of time before patients receive the treatments they need. It is also important to consider that the mechanism of action for a new therapy may be based on either/or intermediate endpoint, and not designed to consider both in a co-primary fashion.	Comment noted. See above responses.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		Proposed change (if any): We ask EMA to please consider harmonizing their guidance with the United States' Food and Drug Administration. Differences between FDA and the EMA affect drug development planning and procedures. In this case specifically, the FDA has positively decided to not require co-primary intermediate endpoints. Their guidance for industry instead gives drug approval applicants the opportunity to consider either/or a resolution in NASH, or an improvement in fibrosis.	
Line 238	11	**Comment:** As NASH is a liver sequelae of the metabolic syndrome and the most frequent mortality reason in NASH population is cardiovascular research therapies effect on the background disorder is clinically relevant to be assessed in addition to the liver-related clinical outcomes. **Proposed change (if any):** In addition to at least no worsening of metabolic syndrome (serum lipids, hypertension, obesity and T2DM) acceptable intermediate endpoints would consist of two composite endpoints to be evaluated at the individual patient level:	Partially agreed. However, cardiovascular biomarkers and outcomes are regarded primarily as safety parameters. Focus should remain on the liver (with the exception of all-cuase mortaility being part of the composite final efficacy endpoint).
NASH Section 4.2.	6	**Comment:**	Not agreed. The reasons for requesting the co-primary endpoints are mentioned. See above.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
Lines 238-247		The guideline comments that" *Acceptable intermediate endpoints would consist of two composite endpoints to be evaluated at the individual patient level:* *1. The resolution of NASH – ... – and, at the same time, no worsening of the stage of fibrosis.* *2. The improvement of fibrosis by at least 1 stage without any worsening of NASH (...)*. *Efficacy in these two composites should be demonstrated in co-primary fashion, meaning that both will have to independently demonstrate a statistically significant and clinically relevant difference to placebo*." However, the EMA Report of the stakeholder interaction meeting on the development of medicinal products for chronic non-infectious liver diseases, held on 3rd December 2018 (EMA/873574/2019) notes that "...*the evaluation of endpoints on fibrosis regression with resolution of NASH in a co-primary setting was considered too strict. It was argued that showing an effective treatment on steatohepatitis would show regression / non-progression on fibrosis as a natural consequence over time."* The EMA stakeholder interaction report further states that "*...stopping the trigger of the disease (NASH) was considered a valid primary outcome even if a*	

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		simultaneous effect on fibrosis regression could not be demonstrated within the 12/18 months timeframe of a clinical trial". We agree with the above conclusion and respectfully consider that the proposed co-primary requirement based on two composite endpoints is not applicable for treatment products acting on the resolution of NASH.	
238-248	4	**Comment:** - Until more information is understood about new medicines and mechanism of action, it is considered too early to require both components of the proposed composite intermediate endpoint to be achieved in order to support an early approval. - It is unclear why two composite endpoints would be necessary as intermediate endpoints, placing the efficacy bar higher for an anti-fibrotic targeting mode of action. This is inconsistent with the FDA draft guidance *'Noncirrhotic Nonalcoholic Steatohepatitis With Liver Fibrosis: Developing Drugs for Treatment'*. Since fibrosis stage 2 and fibrosis stage 3 patients have a differentiated natural history, as supported by the literature – *'Liver Fibrosis, but No Other Histologic Features, Is*	Comment noted. The reflection paper refers in such cases to obtain Scientific Advice. Comment noted. See above.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		Associated With Long-Term Outcomes of Patients With Nonalcoholic Fatty Liver Disease', Angulo P, Kleiner DE, Dam-Larsen S, Adams LA, Bjornsson ES, Charatcharoenwitthaya P, Mills PR, Keach JC, Lafferty HD, Stahler A, Haflidadottir S, Bendtsen F, Gastroenterology. 2015 Aug;149(2):389-97.e10 – reducing from stage 3 to stage 2 should not be a clinically-relevant endpoints. We would favour alignment with the FDA guidance.	
		- Data from other chronic liver diseases (hepB and C) demonstrate that reversal of cirrhosis is a valid surrogate and can be transferrable to NASH, but will ultimately be influenced by inclusion criteria.	Agreed Reversal of cirrrhosis is included as potentail surrogate for trials in cirrhotic patients.
		- Requirement to demonstrate both resolution of NASH without worsening of fibrosis and improvement in fibrosis without worsening of NASH sets a high bar that may not be attainable in monotherapy. Treatments may show benefit in only one of these two treatment benefits i.e. NASH resolution or fibrosis improvement. Correlation of histological improvement of NASH with long-term clinical outcomes remains to be established and important treatments options may be missed if the therapeutic threshold is set too high.	Not agreed. See above. The possibility that mechanism of action and these requirements do not match are dealt with in the reflection paper and applicants are referred to Scientific Advice.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		- Suggest removing the requirement for composite endpoints requiring both resolution of NASH and improvement of fibrosis. As stated in the section "Additional considerations on mode of action" not all drugs will have a MOA that targets both resolution of NASH and improvement of fibrosis. Likewise, there might also be differences in timing, where improvement of fibrosis does not reverse as quickly as resolution of NASH. - Treatments that can provide clinically relevant improvement of either NASH or fibrosis might benefit a patient segment with an unmet medical need. - De-coupling of co-primary endpoints described on line 240-244 should be harmonized with FDA recommendations - This will facilitate development of different MoAs, contemplating the dynamic pattern of metabolic remodelling vs fibrogenesis individually. - Intermediate endpoints, especially in those with less advanced liver dysfunction could be used to realistically tease out "improvement" or "dynamic changes" of fibrosis mechanisms or separate components beyond all cause mortality.	Not agreed. See above. The previous proposals with regard to mode of action have been deleted. Comment noted and partially agreed. See above. See above Not agreed. No "improvement" category as yet identified which has been proven to be associated with sufficient clinical advantages.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		- On line 246 suggest removing the requirement for composite endpoints requiring both resolution of NASH and improvement of fibrosis. As stated in the section "Additional considerations on mode of action" not all drugs will have a MOA that targets both resolution of NASH and improvement of fibrosis. Likewise, there might also be differences in timing, where improvement of fibrosis does not reverse as quickly as resolution of NASH. - Treatments that can provide clinically relevant improvement of either NASH or fibrosis might benefit a patient segment with an unmet medical need. **Proposed change**: Acceptable intermediate endpoints would consist of ~~two~~ ***one of the following*** composite endpoints to be evaluated at the individual patient level: - The resolution of NASH – with the presence of any grade of steatosis, no ballooning, and only minimal (grade 1) lobular inflammation and – at the same time – no worsening of the stage of fibrosis. ***AND/OR*** - The improvement of fibrosis by at least 1 stage without any worsening of NASH (no worsening of ballooning and lobular	See above. See above Not agreed. See above.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		inflammation, a 1 grade change in steatosis may be acceptable). ~~Efficacy in these two composites should be demonstrated in co-primary fashion, meaning that both will have to independently demonstrate a statistically significant and clinically relevant difference to placebo. This requirement is thought to take account of the uncertainties associated with a strategy to account for the long-term outcomes later.~~ **Rationale:** The two intermediate endpoints in the draft reflection paper have been proposed as surrogates that are reasonably likely to predict clinical benefit. However, at this time, it is unknown which of these intermediate endpoints will predict clinical outcomes. A few long-term longitudinal studies have demonstrated that fibrosis stage, but not other histological features of NASH, independently predicts the progression to clinical outcomes (Angulo et al 2015[v]; Hagstrom et al 2017[vi]). In addition, the presence of NASH is directly associated with an increased risk for fibrosis. A retrospective study in patients diagnosed with NAFLD on liver histology showed that 85% of patients diagnosed with NASH had fibrosis. In contrast, fibrosis was present in only 17% of patients that were identified with non-NASH NAFLD (Angulo, 2015). Thus, these data clearly demonstrate that the presence of NASH significantly increases the risk of fibrosis. As	

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		such, resolution of NASH may decrease the fibrogenic drive, which has been linked to inflammation and ballooning (Schuppan et al. 2018[vii]). Thus, both of the proposed intermediate endpoints are scientifically plausible surrogate endpoints to be validated in clinical outcomes studies. Compounds with a variety of mechanism of actions (e.g. metabolic via weight loss, metabolic via improved insulin sensitivity, anti-inflammatory and anti-fibrotic) are in development for NASH. Some of these compounds may work primarily to resolve NASH, with a hypothetical subsequent effect to induce stabilisation or regression of fibrosis. Anti-fibrotic compounds may not induce resolution of NASH but this may not matter if progression to cirrhosis is arrested. NASH experts have stated that it is "too early to prioritize those drugs or mechanisms that are most promising, as clinical trials thus far have been relatively short (3–18 months)" (Friedman et al 2018 [viii]). Therefore, until the link between histology and clinical outcomes is established, it is very important to allow either of these endpoints to be used for early approval with the requirement that clinical outcomes be assessed for continued marketing. Finally, aligning the EMA reflection paper with the December 2018 FDA draft guidance for industry on Noncirrhotic Nonalcoholic Steatohepatitis With Liver Fibrosis would greatly facilitate global drug development. A requirement in the EU that both intermediate endpoints	

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		are met for early approval could delay the availability of effective therapies for patients with NASH in the EU.	
Lines 238-248	7	**Comment:** To aid global study conduct and regulatory assessments, we would encourage the use of primary endpoints consistent with FDA draft guidelines. Meeting either an endpoint of NASH resolution or of improvement of fibrosis by at least 1 stage without worsening of should be sufficient. Proposed change (if any): Acceptable intermediate endpoints would consist of ~~two composite~~ ***either one of two*** endpoints to be evaluated at the individual patient level: 1. The resolution of NASH – with the presence of any grade of steatosis, no ballooning, and only minimal (grade 1) lobular inflammation and – at the same time – no worsening of the stage of fibrosis. 2. The improvement of fibrosis by at least 1 stage without any worsening of NASH (no worsening of ballooning and lobular inflammation, a 1 grade change in steatosis may be acceptable). Efficacy in ***either of*** these two composites should be demonstrated ~~in co-primary fashion, meaning that both will have to independently demonstrate a statistically significant and clinically relevant difference to placebo. This requirement is thought to take account of the uncertainties associated with a~~	Not agreed. See above. Reasons for non-alignment have been given.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		~~strategy to account for~~ the long-term outcomes ***will be assessed*** later ***for full approval***.	
238-248	8	Comment: Regarding the statement, "Acceptable intermediate endpoints would consist of two composite endpoints to be evaluated at the individual patient level", we would respectfully submit that the evaluation should occur at the "population level." In clinical trials, the intermediate primary endpoints are generally assessed in the study population separately. The results are analysed to assess the percentage of subjects who have improvement of fibrosis of at least 1 stage with no worsening of NASH and the percentage of subjects with resolution of NASH with no worsening of fibrosis (Sanyal 2015, Ratziu 2017). A strong body of literature continues to support the ability of fibrosis to predict all-cause and liver-related mortality (Younossi 2011, Angulo 2015, Ekstedt 2015). Although fibrosis is the strongest predictor of mortality in NASH, natural history studies have also established a relationship between the presence of definite steatohepatitis and increased mortality. The presence of definite NASH, based on the pattern of histopathologic lesions, has been associated with an increased risk of liver transplantation or death, and compared to the absence of definite NASH, histologic presence of definite NASH was associated with	Comment not fully understood. The evaluation is in population level, nevertheless, the evaluation of NASH resolution (and worsening) and development of fibrosis is in the patient level. This was inserted in order to avoid mixing up the intention of co-primary endpoints with an evaluation of the composite of the two components. Not agreed. See above.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		increased rate of liver transplant or death (Angulo 2015, Rafiq 2009). NASH activity has also recently been shown to positively correlate with fibrosis progression (Ratizu 2016). Therefore, fibrosis improvement without worsening of NASH or NASH resolution without worsening of fibrosis can each individually reasonably predict clinical outcomes. Furthermore, based on the time course of the disease it is highly unlikely that both NASH and fibrosis would resolve/improve at the same time. Data actually shows that this is achieved by very few patients and this is based on the biology, not the drug. It is unlikely that steatohepatitis as assessed by the presence of NASH, as well as consequences of steatohepatitis as assessed by fibrosis stage, can be affected by treatment in the same timeframe. References: • Sanyal AJ, Friedman SL, McCullough AJ, Dimick-Santos L; American Association for the Study of Liver Diseases; United States Food and Drug Administration. Challenges and opportunities in drug and biomarker development for nonalcoholic steatohepatitis: findings and recommendations from an American Association for the Study of Liver Diseases-U.S. Food and Drug Administration	Not agreed. Multiple studies have meanwhile shown that this is possible with several different mechanisms of action.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		Joint Workshop. Hepatology. 2015;61(4):1392–1405. doi:10.1002/hep.27678. • Ratziu V. A critical review of endpoints for non-cirrhotic NASH therapeutic trials. J Hepatol. 2018 Feb;68(2):353-361. • Younossi ZM, Stepanova M, Rafiq N, et al. Pathologic criteria for nonalcoholic steatohepatitis: interprotocol agreement and ability to predict liver-related mortality. Hepatology. 2011 Jun;53(6):1874-82. doi: 10.1002/hep.24268. Epub 2011 May 14. • Angulo P, Kleiner DE, Dam-Larsen S, et al. Liver Fibrosis, but No Other Histologic Features, Is Associated With Long-term Outcomes of Patients With Nonalcoholic Fatty Liver Disease. Gastroenterology. 2015;149(2):389–97.e10. doi:10.1053/j.gastro.2015.04.043. • Ekstedt M, Hagstrom H, Nasr P, et al. Fibrosis stage is the strongest predictor for disease specific mortality in NAFLD after up to 33 years of follow-up. Hepatology. 2015 May;61(5):1547-54. doi: 10.1002/hep.27368. Epub 2015 Mar 23. • Rafiq N, Bai C, Fang Y, et al. Long-term follow-up of patients with nonalcoholic fatty liver. Clinical gastroenterology and	

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		hepatology : the official clinical practice journal of the American Gastroenterological Association. 2009 Feb;7(2):234-8. doi: 10.1016/j.cgh.2008.11.005. Epub 2008 Nov 7. • Ratziu V, Francque SM, Harrison S, et al. Improvement in NASH histological activity highly correlates with fibrosis regression. Hepatology. 2016b December;64(6):1140A. Proposed change (if any): Suggest "fibrosis improvement without worsening of NASH" or "NASH resolution without worsening of fibrosis" should each independently be acceptable as the intermediate endpoints of efficacy.	
238-248	10	**Comment:** Requirement to demonstrate both resolution of NASH without worsening of fibrosis and improvement in fibrosis without worsening of NASH sets a high bar that may not be attainable in monotherapy. Treatments may show benefit in only one of these two treatment benefits i.e. NASH resolution or fibrosis improvement. Correlation of histological improvement of NASH with long-term clinical outcomes remains to be established and important treatments options may be missed if the therapeutic threshold is set too high.	Not agreed. See above.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		Proposed change (if any): Acceptable intermediate endpoints would consist of two composite endpoints to be evaluated at the individual patient level: 1. The resolution of NASH – with the presence of any grade of steatosis, no ballooning, and only minimal (grade 1) lobular inflammation and – at the same time – no worsening of the stage of fibrosis. **AND/OR** 2. The improvement of fibrosis by at least 1 stage without any worsening of NASH (no worsening of ballooning and lobular inflammation, a 1 grade change in steatosis may be acceptable). ~~**Efficacy in these two composites should be demonstrated in co-primary fashion, meaning that both will have to independently demonstrate a statistically significant and clinically relevant difference to placebo. This requirement is thought to take account of the uncertainties associated with a strategy to account for the long-term outcomes later.**~~ **The relationship between liver histological improvement and clinical outcomes has not**	

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		been characterized. Sponsors can propose and justify specifically either one of these anticipated to be beneficial based on the mechanism of action of the specific drug under development (i.e., drugs that predominantly address the inflammatory process, treat fibrosis, or both).	
238-248	15	Comment: The Reflection Paper indicates that early approval based on intermediate endpoints will require demonstration of statistically significant and clinically meaningful differences vs. placebo for the two designated histologic endpoints assessing NASH resolution (no worsening of fibrosis stage) and fibrosis regression (no worsening of ballooning and lobular inflammation). As each of these endpoints seems to independently meet the principle of supporting a reasonable assumption of long-term benefit (lines 83-98), we respectfully request that the agency consider that demonstration of benefit for either (not both) of these endpoints be considered as supportive of early approval. For pre-cirrhotic NASH, the fundamental goal of treatment is the prevention of progressive fibrosis, thereby preventing cirrhosis and its associated morbidity and mortality. Therefore, improvement in 1-stage fibrosis reduction without worsening NASH	Not agreed. See above.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		appears highly likely to predict an outcomes benefit. While we acknowledge that persistent underlying steatohepatitis may carry risk independent of fibrosis, we do not think that this possibility undermines the expectation that a demonstrated beneficial effect on fibrosis will reduce clinical events directly caused by progressive fibrosis. For this reason, the 2-stage fibrosis reduction endpoint proposed for consideration in the Reflection Paper (lines 294-299) seems to be an unnecessarily high bar which could hinder innovation and timely approval of promising medications. Steatohepatitis is understood to be the driver of fibrosis in NASH and recently reported analyses from two Ph2 studies conducted by the NASH-CRN have shown a strong association between NASH resolution and fibrosis improvement for two distinct therapeutic mechanisms (PPARγ agonist and FXR agonist) [*Brunt EM, Kleiner DE, Wilson LA, Sanyal AJ, Neuschwander-Tetris BA. Improvements in histologic features and diagnosis associated with improvement in fibrosis in nonalcoholic steatohepatitis. Hepatology, 2019*]. Therefore, both biologic plausibility and available data support the expectation that an improvement in the NASH resolution without worsening fibrosis endpoint should be predictive of outcomes benefit. As agents that primary act on upstream pathogenic drivers (e.g. steatosis, inflammation) are expected to improve NASH	

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		resolution prior to fibrosis reduction, requiring demonstration of fibrosis reduction at the time of an initial MAA may present an unnecessary obstacle to their timely approval. Proposed change (if any): It is proposed to formulate the requirement for a co-primary endpoint as an optionality to demonstrate improvement in either NASH resolution without worsening of fibrosis or ≥1 point reduction in fibrosis stage without worsening NASH.	
	15	Comment: The Reflection Paper indicates uncertainty whether histologically reversed cirrhosis may be reasonably predictive of an outcomes benefit in the compensated NASH cirrhosis population (lines 261-263). Much of the morbidity and mortality resulting from NASH is a direct consequence of progressive fibrosis leading to portal hypertension and deterioration of liver function. Therefore, our perspective is that it seems at least reasonably likely that reversal of cirrhosis will decrease the risk of hepatic decompensation for a patient with compensated cirrhosis. The reflection paper also notes that a trial in the compensated NASH cirrhosis population using reversal of cirrhosis as an endpoint would need to substantiate the expectation that the prognosis of people with	Partially Agreed. Reversal of cirrhosis is implemented as intermediate endpoint in the cirrhotic population. The requirement for demonstration of the relevance of cirrhosis regression is kept although the database has improved (see references in the Reflection Paper).

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		reversed cirrhosis is similar to that of untreated people with earlier-stage disease (lines 263-265). We note that a beneficial agent needs to meaningfully improve prognosis, but not necessarily to the extent that it is 'similar' to that of the much lower risk population that has never been cirrhotic. Proposed change (if any): We propose: 1. that histologic reversal of cirrhosis without worsening of NASH be indicated as an acceptable intermediate endpoint for trials of compensated NASH cirrhosis. 2. removal of the indication that agents being developed for compensated NASH cirrhosis must substantiate the expectation that prognosis with treatment will be similar to that of patients who have never been cirrhotic.	
240-248	2	**Comment:** We understand the agency´s concern about approving a drug for non-cirrhotic NASH with an effect on fibrosis and no effect on disease activity as reflected by NASH resolution. This co-primary end-point precludes "pure" antifibrotic therapies for non-cirrhotic NASH, which is probably a good idea in some instances. It should be noted that an effect on	Comment noted, but conclusion not agreed. See above.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		ballooning must also be seen to conclude NASH resolution. Thus, a reasonable conclusion would be that non-cirrhotic NASH drugs should show NASH resolution within 1 year but not necessarily reduction in one stage of fibrosis. However, also a drug with a NASH resolution signal (but not significant) and significant effect on fibrosis should be considered for approval for treatment of non-cirrhotic NASH. Note that the FDA guidance use NASH resolution or reduction in fibrosis. **Proposed change (if any):** We believe it is important that there is a consistency between agencies to facilitate global development programmes in any area of unmet medical need. We therefore think it is reasonable to propose resolution of NASH with no worsening of fibrosis or improvement of fibrosis by one stage with no worsening of NASH.	
243-244	1	Comments: For the Agency's consideration, we recommend that both the resolution and the worsening of steatohepatitis be defined based on histologic features that differentiate steatohepatitis from simple steatosis, i.e. based on lobular inflammation and hepatocellular ballooning. The endpoint for	Comment noted but not fully understood. However, the current wording takes account of the concerns. The current wording of "improvement of fibrosis by at least 1 stage without any worsening of NASH (no worsening of ballooning and lobular inflammation, a 1 grade change in

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		consideration is, therefore, proposed as improvement in liver fibrosis greater than or equal to one stage (NASH CRN fibrosis score) and no worsening of steatohepatitis (defined as no increase in hepatocellular ballooning or lobular inflammation).	steatosis may be acceptable)" makes sufficiently clear that an increase in ballooning and lobular inflammation needs to be excluded.
243-244	4	**Comment:** - The reliance on scoring for ballooning as part of the 'essential' definition of changes in NASH is of concern for two reasons: - Firstly, although closely linked with the definition of NASH, ballooning and scoring of ballooning appears to be the noisiest element of the histologic evaluation of NASH. Ballooning cells are uncommon and therefore the ballooning score can be highly variable based on sampling, with poor agreement on scoring even among experienced hepatopathologists. - Additionally there are already considerations of expanding the ballooning score for NAS, which will introduce an additional complication to reliance on this particular element of the NAS score. Since the NAS score was not actually developed to define NASH, but only to score changes during clinical trials (Brunt et al 2011), reliance on this one element of the NAS score may be too rigid a requirement. For example, the disappearance	Partially agreed. The difficulties with scoring of ballooning are well documented. However, it is not agreed to weaken the criteria for resolution, despite the uncertainties with scoring of ballooning. In the context of a placebo-controlled, randomised trial, this will also potentially increase placebo response rates, which appears undesirable.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		of the typical chicken wire fibrosis from a post-treatment liver sample may be sufficient for an experienced hepatopathologist to declare resolution of NASH. - Therefore, we recommend adding a degree of flexibility in the definition of NASH resolution to include a NAS score of 0-1 for ballooning. *Brunt et al. (2011) The NAS and The Histopathologic Diagnosis in NAFLD: Distinct Clinicopathologic Meanings; Hepatology. 2011 Mar; 53(3): 810–820.* **Proposed change (if any):** The improvement of fibrosis by at least 1 stage without any worsening of NASH **(whereby worsening of NASH is defined as increase by more than > 2 points in NAS due to non-steatotic components).**	
243-244	10	**Comment:** The reliance on scoring for ballooning as part of the 'essential' definition of changes in NASH is of concern for two reasons: Firstly, although closely linked with the definition of NASH, ballooning and scoring of ballooning appears to be the noisiest element of the histologic evaluation of NASH. Ballooning cells are uncommon and therefore the ballooning score can be highly variable based on	Partially agreed. See above.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		sampling, with poor agreement on scoring even among experienced hepatopathologists. Additionally there are already considerations of expanding the ballooning score for NAS, which will introduce an additional complication to reliance on this particular element of the NAS score. Since the NAS score was not actually developed to define NASH, but only to score changes during clinical trials (Brunt et al 2011), reliance on this one element of the NAS score may be too rigid a requirement. For example, the disappearance of the typical chicken wire fibrosis from a post-treatment liver sample may be sufficient for an experienced hepatopathologist to declare resolution of NASH. Therefore, we recommend adding a degree of flexibility in the definition of NASH resolution to include a NAS score of 0-1 for ballooning. *Brunt et al. (2011) The NAS and The Histopathologic Diagnosis in NAFLD: Distinct Clinicopathologic Meanings; Hepatology. 2011 Mar; 53(3): 810–820.* **Proposed change (if any):** The improvement of fibrosis by at least 1 stage without any worsening of NASH **(whereby worsening of NASH is defined as increase by**	

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		more than > 2 points in NAS due to non-steatotic components).	
Lines 243-244	18	Comment: As written, it's unclear as to what a "1 grade change in steatosis may be acceptable" actually is intended to mean. Is a one grade increase really acceptable provided there is at least a 1 grade reduction in fibrosis? Proposed change (if any):	Agreed. Wording amended
245-247	1	Comment: The two composite endpoints described are supported with the current understanding of the disease as well as lack of validated non-invasive markers for efficacy predicting long-term outcomes. The requirement to include them as coprimary endpoints is considered a very strict criterion[5] and not in alignment with other authorities' recommendations [4]. We recommend aligning with those recommendations. This will result in having ANY of the composite endpoints as the primary endpoint - depending on MoA - and include the other endpoint as the secondary endpoint.	Not agreed. See above.
249-299	1	Comment: The risk of progression to end-stage liver disease is largely related to the baseline fibrosis stage and	Partially agreed. The requirement of the 2-stage fibrosis improvement has been deleted. For substances not suitable to fulfill both endpoints (e.g. antfibrotic effects only), no

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		improvement of fibrosis would therefore be relevant as an interim endpoint[6]. The recommendation of an endpoint of at least 2 stage fibrosis improvement without worsening of NASH for new substances that primarily target fibrosis is considered setting the bar too high for anti-fibrotic treatments. This can lead to false negative trials. This endpoint could be recommended as a secondary endpoint in line with the outcome of the 'EMA stakeholder meeting 3 Dec' where this requirement was debated[5] .	clear guidance can be given at this point of time, and Applicants are referred to Scientific Advice.
250-281	2	**Comment:** We agree with the agency that data are available that indicate a relationship between fibrosis and outcome in NASH, but we lack data that reduction in fibrosis improves outcome. However, cirrhosis is a serious condition independent of aetiology. A treatment that reduces fibrosis would reduce portal pressure and improving liver function in cirrhosis would be beneficial. Histology as well as MELD score could be used to describe improvement. **Proposed change (if any):** We think that treatment of cirrhotic NASH should be regarded as treatment in two steps; first rescue the patient from decompensation by reducing fibrosis to non-cirrhotic NASH (an effect associated with reduced portal pressure and improved MELD score), then or concomitantly initiate NASH resolution therapy.	Partially agreed. For cirrhotic NASH, the possibility to use reduction of fibrosis by one grade as intermediate endpoint is given

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
251-255	4	**Comment:** - Distinction needs to be made between compensated and decompensated cirrhosis. In compensated cirrhosis, 1-point improvement in fibrosis without worsening of NASH should still be a feasible intermediate endpoint. The use of HVPG should also be considered as a potential endpoint. **Proposed change (if any):** In liver disease where cirrhosis has already been manifested, **it is important to identify within this population the two subgroups of differential risk and mortality i.e. the compensated and decompensated populations.** ~~**the use of the above mentioned long term composite is not possible**~~. **For compensated cirrhosis the use of 1-point improvement in fibrosis without worsening of NASH is still feasible. Change of the HVPG may also be considered.** An acceptable endpoint for patients with already existing cirrhotic liver disease at inclusion **could also include** ~~**would therefore consist of**~~ the composite of all-cause death and liver decompensation events. However, because liver cirrhosis represents a wide spectrum of disease, it is currently unclear whether such an endpoint is feasible.	Partially agreed, use of 1-point improvement of fibrosis in compensated cirrhosis is includeed. For HVPG, see Chapter 5.3.4. For decompensated cirrhosis, no clear recommendations can currently be given; see Chapter 5.3.2.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
251-255	10	**Comment:** Distinction needs to be made between compensated and decompensated cirrhosis. In compensated cirrhosis, 1-point improvement in fibrosis without worsening of NASH should still be a feasible intermediate endpoint. The use of HVPG should also be considered as a potential endpoint. **Proposed change (if any):** In liver disease where cirrhosis has already been manifested, **it is important to identify within this population the two subgroups of differential risk and mortality i.e. the compensated and decompensated populations.** ~~**the use of the above mentioned long term composite is not possible**~~. **For compensated cirrhosis the use of 1-point improvement in fibrosis without worsening of NASH is still feasible. Change of the HVPG may also be considered.** An acceptable endpoint for patients with already existing cirrhotic liver disease at inclusion **could also include** ~~**would therefore consist of**~~ the composite of all-cause death and liver decompensation events. However, because liver cirrhosis represents a wide spectrum of disease, it is currently unclear whether such an endpoint is feasible.	Partially agreed. See above.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
251-257	8	Comment: We disagree that long-term NASH cirrhosis studies should include advanced cirrhotic patients. Also, this is contrary to FDA guidance which recommends that compensated and decompensated patients be studied separately. Sponsors should have the flexibility to include such patients, but it should not be mandated.	Agreed. Only general remarks as made on decompensated cirrhosis at this point of time.
Line 253	11	**Comment:** NASH is a liver sequelae of the metabolic syndrome and the most frequent mortality reason in NASH population is cardiovascular. **Proposed change (if any):** An acceptable endpoint for patients with already existing cirrhotic liver disease at inclusion would therefore consist of the composite of all-cause and CV death, and liver decompensation and CV events.	Not agreed. Inclusion of CV death into the final composite takes sufficiently account of the fact that indeed mortality in NASH is mainly due to cardiovascular reasons. Making CV events part of the composite would not measure the treatment of liver disease, but of CV risk factors and point to a different indication. The aim of NASH treatment should remain to prevent/treat end-stage liver disease. CV events need to be dealt with as a safety issue, in order to exclude an increased risk (see 5.3.7.2).
255-257	4	**Comment:** - Advanced cirrhosis is not defined. Reference should be made to established systems such as CPT, MELD etc. **Proposed change (if any):** When the intention is to use this long-term endpoint in the cirrhotic population, the study population should be enriched with patients with advanced cirrhosis **as defined by a generally accepted**	Partially agreed. However, new structur and wording do not make the proposed changes necessary. The term advanced cirrhosis has been eliminated.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		qualitative (e.g. Child-Turcotte-Pugh: CTP) or quantitative (e.g. Model for End Stage Liver Disease: MELD) scoring system that assesses the severity of chronic liver disease.	
255-257	10	**Comment:** Advanced cirrhosis is not defined. Reference should be made to established systems such as CPT, MELD etc. **Proposed change (if any):** When the intention is to use this long-term endpoint in the cirrhotic population, the study population should be enriched with patients with advanced cirrhosis **as defined by a generally accepted qualitative (e.g. Child-Turcotte-Pugh: CTP) or quantitative (e.g. Model for End Stage Liver Disease: MELD) scoring system that assesses the severity of chronic liver disease.**	See above.
257, 270-271	2	**Comment:** We think the agency should avoid the terms advanced and less advanced cirrhosis. **Proposed change (if any):** It would be better to use compensated or non-compensated as well as prognostic measures such as HVPG and/or MELD score. We also believe that it is most important to show a fast onset reduction in	Agreed. See above.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		fibrosis e.g. within 2 years going from cirrhosis to non-cirrhosis accompanied by prognostic biomarkers and/or MELD score.	
258-261	4	**Comment:** - Reversal of cirrhosis is an established endpoint and recognised within medical and scientific community. **Proposed change (if any):** In case the need to use intermediate endpoints in this population is identified, a reasonable endpoint for the **~~general non-de~~**compensated population, could **~~intuitively~~** be the reversal of cirrhosis (e.g. defined as "improvement of liver cirrhosis to non-cirrhotic liver disease (1 or more point improvement in fibrosis stage"). **~~At this point of time, however, the data available to demonstrate that reversed cirrhosis does indeed also reverse or influence the final prognosis substantially, is considerably less profound than the association shown for progressing disease.~~**	Partially agreed. Reversal of cirrhosis is given as potential intermediate endpoints in the compensated cirrhosis population. The evidence for the clinical value of reversal of NASH cirrhosis, is, however, still limited.
258-261	10	**Comment:** Reversal of cirrhosis is an established endpoint and recognised within medical and scientific community. **Proposed change (if any):**	See above

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		In case the need to use intermediate endpoints in this population is identified, a reasonable endpoint for the ~~**general non-de**~~compensated population, could ~~**intuitively**~~ be the reversal of cirrhosis (e.g. defined as "improvement of liver cirrhosis to non-cirrhotic liver disease (1 or more point improvement in fibrosis stage"). ~~**At this point of time, however, the data available to demonstrate that reversed cirrhosis does indeed also reverse or influence the final prognosis substantially, is considerably less profound than the association shown for progressing disease.**~~	
Lines 259-261	7	**Comment:** A relevant outcome in cases with Compensated Cirrhosis is also non-progression over for example 12 or 18 months (i.e. stable disease) to Decompensated Cirrhosis. **Proposed change (if any):** In case the need to use intermediate endpoints in this population is identified, a reasonable endpoint for the general non-decompensated population, could intuitively be the reversal of cirrhosis (e.g. defined as "improvement of liver cirrhosis to non-cirrhotic liver disease (1 or more point improvement in fibrosis stage")***, or indeed non-progression to decompensated cirrhosis as indicator of stable disease.***	Comment noted but not agreed to be given as (part of the) primary endpoint. Since progression to decompensation is part of the primary evaluation, non-progression/stable disease can be evaluated as secondary outcome.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
261-263	8	Comment: We disagree that reversal of liver cirrhosis is not an adequate endpoint to support approval. Liver fibrosis is associated with increased adverse clinical outcomes across all stages of fibrosis in the NASH population, and in particular in patients with stage 4 (F4) cirrhosis. In fact, patients with cirrhosis have the highest risk for adverse clinical outcomes and therefore, in non-cirrhotic patients with fibrosis, progression to cirrhosis is considered a relevant clinical outcome. This is consistent with the FDA's draft guidance for NASH fibrosis outcomes studies, in which progression to cirrhosis is considered a clinical endpoint. Therefore, the reversal of cirrhosis to an earlier fibrosis stage in NASH patients with compensated cirrhosis should be considered a relevant clinical endpoint, and studies showing reversal in cirrhosis should be supportive of submissions leading to full approval. This is particularly relevant in this early compensated cirrhotic population which has no evidence of portal hypertension or other physiological changes associated with more advanced stages of cirrhosis. Also, requiring the demonstration of outcomes benefit for approval would take several years and make clinical development unfeasible while preventing	Partially agreed. See above. It is not fully understood why the Draft reflection paper has been understood by so many stakeholders as excluding using the reversal of cirrhosis as primary endpoint. This was clearly included as a potential opportunity which would need, however, adequate justification and back-up.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		access to therapies in the highest unmet need in the NASH population.	
267-268	4	**Comment:** - Severity of portal hypertension can be used to clinically separate less advance from more advanced cirrhosis, but can also be derived from MR, or US based measurements (e.g. liver+speen stiffness) – ultimately combined with invasive assessment of HVPG	Partially agreed. See above .
270-274	14	**Comment**: We suggest that measures of quality of life (QOL) and activities of daily living (ADL) be included to provide a full picture of clinically meaningful outcomes in patients with cirrhosis. We further suggest that 'time to event' be considered. **Proposed change (if any)**:	Agreed. Paragraph on patient reported outcomes has been added.
	14	**Comment**: We appreciate the inclusion of this section and flexibility provided for compounds which target only one part of the composite endpoint (i.e., purely anti-fibrotic compounds). We suggest to clarify that sponsors may propose (with justification) other criteria that might be considered 'a stronger endpoint'. We	Agreed. 2-stage fibrosis endpoint has been deleted. Flexibility is maintained, but no final recommendation given.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		note that a 2-stage improvement in fibrosis may be too high of a bar to be achieved. We further note that the fibrosis benefit index[2] could be useful. **Proposed change (if any)**:	
273-274	4	**Comment**: Use of a certain magnitude of change in MELD as endpoint should be possible alternatively (proposal for flexible language allowing for this below). A HVPG of 12 mmHG instead of 10 mmHg should be considered as below 10 mmHg may be a high hurdle if people are coming with higher values > 16 mmHg. A certain percentage decrease may be more clinically meaningful depending on baseline value. Generally, it appears preferable to keep the requirement flexible to allow for case-by-case assessment**.** **Proposed change (if any):** Other potential endpoints (e.g. lowering of MELD score **or of HVPG** below a certain threshold **or by a certain percentage~~, or of the HVPG below 10 mm Hg~~**) are also **worthy of consideration.**	Not agreed. MELD is given as categorical endpoint with the established threshold of 15. Lowering of MELD is of unclear clinical relevance (unless this is in a population with baseline MELD equal or higher than 15). Lowering of MELD has therefore been deleted. For HVPG: See Chapter 5.3.4 and the recommendation not to use this as an endpoint in confirmatory trials

[2] https://doi.org/10.1016/S2468-1253(17)30005-5

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
273-274	10	**Comment**: Use of a certain magnitude of change in MELD as endpoint should be possible alternatively (proposal for flexible language allowing for this below). A HVPG of 12 mmHG instead of 10 mmHg should be considered as below 10 mmHg may be a high hurdle if people are coming with higher values > 16 mmHg. A certain percentage decrease may be more clinically meaningful depending on baseline value. Generally, it appears preferable to keep the requirement flexible to allow for case-by-case assessment**.** **Proposed change (if any):** Other potential endpoints (e.g. lowering of MELD score **or of HVPG** below a certain threshold **or by a certain percentage~~, or of the HVPG below 10 mm Hg~~**) are also **worthy of consideration.**	See above.
279-281	4	**Comment**: - Potential surrogate endpoints could be considered in decompensated cirrhosis and include event-free survival based on a composite endpoint. **Proposed change (if any):** In the special group of decompensated cirrhosis, a therapeutic effect should be demonstrated based on the endpoint all-cause mortality/survival. Liver related death, and liver-related death/ transplantation could be supportive endpoints. **An acceptable surrogate endpoint in decompensated cirrhosis of event-free survival,**	Comment noted. For the decompensated population, at this point of time, only a general paragraph is included and no dedicated recommendations given due to lack of experience in this field.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		based on a composite clinical endpoint (all-cause mortality, new decompensation events, and MELD score progression are events) could be considered.	
279-281	10	**Comment:** Potential surrogate endpoints could be considered in decompensated cirrhosis and include event-free survival based on a composite endpoint. **Proposed change (if any):** In the special group of decompensated cirrhosis, a therapeutic effect should be demonstrated based on the endpoint all-cause mortality/survival. Liver related death, and liver-related death/ transplantation could be supportive endpoints. **An acceptable surrogate endpoint in decompensated cirrhosis of event-free survival, based on a composite clinical endpoint (all-cause mortality, new decompensation events, and MELD score progression are events) could be considered.**	See above
282-305	4	**Comment:** Two-stage improvement in fibrosis for any compound MoA is a very strict requirement and may not be attainable. Accordingly important treatments may be missed or may not be developed if this development hurdle is maintained. **Proposed change (if any): ~~Additional considerations on mode of action~~ [delete section lines 282 - 305]**	Agreed. See above

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		In addition see proposal above regarding lines 238 – 248.	
NASH Section 4.2. Lines 282-305	6	**Comment:** The guideline acknowledges that" *the liver cell toxicity caused by the overload in fat causes inflammation, which itself is the final trigger of fibrosis development. Therefore, it has been assumed that the appropriate target of medicinal products would be mechanisms preventing fat toxicity and/or decreasing inflammatory activity, which would finally lead to beneficial effects in fibrosis*". The document adds however that "*...new substances primarily targeting the development of fibrosis are currently under development, and it is therefore considered important to reflect whether a decrease in fibrosis stage without any or only minor influence on the fat accumulation in the liver, liver cell stress (ballooning) and inflammation could be appropriate as treatments and benefit patients in the long term. This is considered an uncritical question as long as long-term endpoints are used as objectives in clinical trials."* We consider that this guideline should clearly state, in line with the initial paragraph, that ideally it is best for treatments to act on the cause of the disease (i.e. fat	Not agreed. The possibility to demonstrate benefit for "non-causative pharmacologic action" is kept open. This might especially be relevant for the cirrhotic ppulation.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		accumulation, ballooning and inflammation) rather than on only its consequence (i.e. fibrosis).	
282-305	10	**Comment:** Two-stage improvement in fibrosis for any compound MoA is a very strict requirement and may not be attainable. Accordingly important treatments may be missed or may not be developed if this development hurdle is maintained. **Proposed change (if any): ~~Additional considerations on mode of action~~ [delete section lines 282 - 305]** ***In addition see proposal above regarding lines 238 – 248.***	Agreed. The two-stage fibrosis criterion has been deleted. See also above.
283-287	8	Comment: We disagree with the stated simplified pathophysiology of NASH, which we believe incorrectly assumes that "liver cell toxicity caused by the overload in fat causes inflammation, which itself is the final trigger of fibrosis development." To the contrary, steatosis is not directly correlated with clinical outcomes or survival (Soderberg 2010, Angulo 2015) and may be prone to rapid fluctuations based on lifestyle change. References: • Söderberg C, Stål P, Askling J, Glaumann H, Lindberg G, Marmur J, Hultcrantz R.	Partially agreed. However, pathophysiology and correlation to long-term outcomes might not necessarily be correlated. Anyway, the introductory part with characterisation of the disease has been shortened and simplified further

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		Decreased survival of subjects with elevated liver function tests during a 28-year follow-up. Hepatology. 2010 Feb;51(2):595-602. • Angulo P, Kleiner DE, Dam-Larsen S, et al. Liver Fibrosis, but No Other Histologic Features, Is Associated With Long-term Outcomes of Patients With Nonalcoholic Fatty Liver Disease. Gastroenterology. 2015 Aug;149(2):389-97 e10	
Section 4.2.3 Study design and endpoints Additional considerations of mode of action Lines 287-291 Lines 295-299	13	Comment: We understand the agency acknowledges the use of a “stronger endpoint” for therapies that target one intermediate endpoint over the other. For example, “fibrosis regression of at least 2 stages without worsening of NASH [...] no worsening of NASH.” However, we believe this language leaves too much up for interpretation. By including the initial requirement of co-primary endpoints raises the barrier for treatment approval to unrealistic levels. Proposed change (if any): We can not emphasize enough that additional flexibility is needed regarding endpoints. Each intermediate endpoint must be considered individually, and acceptable.	Agreed. See above.
Section 4.2.3 Study design and endpoints	13	Comment: The reflection paper states that if clinical trials favor long-term endpoints the issue surrounding co-primary intermediate endpoints becomes “uncritical.” The priority given to long-	Partially agreed. The reflection paper is intended to open the regulatory pathway to shorter MA with the possibility to use intermediate endpoints for conditional MA.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
Additional considerations of mode of action Lines 291-292		term endpoints demonstrates a fundamental misunderstanding of the nature of NAFLD and NASH. Consideration must be given to the harmful effects of each stage of the disease, and the burden of having the disease itself. As patients who need some sort of response, the language and barriers placed on all endpoints must be considered critical. Proposed change (if any): We would caution EMA on how the language in this reflection paper can prioritize long-term versus immediate term endpoints, and impact the timely response to treatments for NASH patients.	No preferences are given, but it needs to be considered that conditional MA always includes the ”conditions”.
294-296	4	**Comment:** A regression of 2 stages of fibrosis appears to be a high hurdle, particularly in the absence of approved therapies for treatment of fibrosis. **Proposed change**:-If an intermediate endpoint strategy is used in such compounds, ~~it is currently recommended to use a stronger endpoint denoted as a composite at the individual patient level such as "fibrosis regression of at least 2 stages without worsening of NASH"~~ ***"substantial evidence of fibrosis regression of at least 1 stage without worsening of NASH" should be demonstrated.*** **Rationale:** A regression of 2 stages of fibrosis is a high hurdle as it may be harder to demonstrate fibrosis regression than NASH resolution. Ultimately, the	Agreed. See above.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		clinical outcome is of most importance for patients. Even stabilisation of fibrosis may be sufficient to provide an improvement in patient outcome. In this area of unmet need, a medicinal product showing substantial evidence of fibrosis regression of at least 1 stage should be considered for early approval, to be confirmed with long-term outcome data. Finally, aligning the EMA reflection paper with the December 2018 FDA draft guidance for industry on Noncirrhotic Nonalcoholic Steatohepatitis With Liver Fibrosis would greatly facilitate global drug development.	
306-315	14	**Comment**: We appreciate the acknowledgement of the uncertainty regarding appropriate trial duration, and the flexibility provided by this text. This section seems to be frequently misunderstood and therefore we suggest further clarifying that 2-year interim evaluation duration is not a requirement, and that shorter durations may be proposed by sponsors with justification depending on factors including characteristics of patient population, mechanism of action and characteristics of the investigational compound, and trial size. **Proposed change (if any)**:	Agreed. See Chapter 5.3.4

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
Lines 307-315	7	**Comment**: Early phase 2 should be of long enough duration to ensure histological effect can be observed if histological endpoint is used. **Proposed change (if any):** The currently published phase 2 data for substances under development have mostly evaluated parts of the above proposed endpoints only. Therefore, uncertainty exists with regard to the duration of trials, both in terms of the time needed for interim evaluation with the intermediate endpoints, as well as for the time needed to show relevant effects on the long-term composite endpoint. As a general rule, ~~a two-year interim evaluation, and a 5-year final evaluation may be considered appropriate~~ ***early Phase 2 trials should be of long enough duration to ensure histological effect can be observed and a relationship to long-term clinical effects can be demonstrated.*** However, this can be modified with factors like size of the trial, activity of the investigational compound, patient characteristics, and the requirements with regard to statistical rigor. The final evaluation would be expected to be usually planned with an event-driven evaluation, and therefore, a fixed duration may not be appropriate to be planned with.	Agreed. As above, see Chapter 5.3.4. However, trial durations for early phase 1 and 2 studies are not given.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
307-315	8	Comment: The proposed durations of a 2-year interim evaluation and a 5-year final evaluation may not be appropriate for new therapies. The appropriate duration of an interim analysis depends upon the surrogate endpoints evaluated, the mechanism of action of the investigational drug and the relative time it takes to reach such surrogate endpoints. Accordingly, the duration required to reverse fibrosis or cirrhosis will be driven by the investigational drug's mechanism of action and may vary from drug to drug. For example, an investigational drug with a dominant antifibrotic effect will more likely meet a fibrotic endpoint, whereas a metabolic modulator or an investigational drug controlling liver cell injury and inflammation will more likely meet a steatohepatitis endpoint. The durations required to reach these endpoints for these different agents may be different. The "final evaluation" of a new investigational drug should be based on clinical outcomes demonstrating the ability of the investigational drug to inhibit the progression to cirrhosis, which results in a reduction in hepatic clinical events (complications of cirrhosis), liver transplantation or death. The measure of clinical outcomes should be driven by the number of clinical outcome events and not the duration of the trial.	Agreed. As above. See Chapter 5.3.4, The 2-year time-fram is still given for reasons of adequate collection of safety data.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
308-311	4 Celgene LLY NN NVS	**Comment:** - Trial duration should be determined by target endpoint, and statistical power to observe intended effect. - Based on existing data, there is no clear rationale why the evaluation of intermediate outcomes should require 2-years of interim evaluation. See proposal to allow for more flexibility below. - Suggest reducing the requirement for interim evaluation after two years; ongoing phase 3 trials in NASH have planned interim analyses after 48-72 weeks. - We would welcome a clarification from the Agency on the evidence supporting a 2-year interim evaluation. Current practice is of one year to 18 months for an interim evaluation. We would suggest aligning with the FDA draft guidance *'Noncirrhotic Nonalcoholic Steatohepatitis With Liver Fibrosis: Developing Drugs for Treatment'* and refer to a trial duration of 12 to 18 months. **Proposed change**: …" As a general rule, **clinical trials should be of sufficient duration (e.g. one year), under consideration of ~~a two-year interim evaluation, and a 5-year final evaluation may be considered appropriate. However, this can be~~**	Agreed. See above.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		~~**modified with factors like**~~ size of the trial, activity of the investigational compound, patient characteristics, and the requirements with regard to statistical rigor." …" Therefore, uncertainty exists with regard to the duration of trials, both in terms of the time needed for interim evaluation with the intermediate endpoints, as well as for the time needed to show relevant effects on the long-term composite endpoint. ~~As a general rule, a two-year interim evaluation, and a 5-year final evaluation may be considered appropriate.~~ **"** **Rationale:** The time required to evaluate histological changes in patients with NASH will depend on the mechanism of action of the compound, the histological endpoint chosen for the trial and the inclusion/exclusion criteria for the trial. In exploratory phase 2 trials histological improvement has been observed in as little as 12 weeks (Harrison et al 2019[ix]). Therefore, for phase 3, an interim analysis even as early as 12 months may be feasible. For clinical outcomes, an event-driven trial design may be the best approach. The time needed to accrue the number of events will depend on the inclusion/exclusion criteria for the trial and the treatment effect size of the compound being studied, among other factors. This may take less than or more than 5 years. Alternative wording has been proposed as there was some concern	

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		that "As a general rule" may be interpreted as something that is more rigid than is scientifically appropriate.	
310-311	10	**Comment:** Based on existing data, there is no clear rationale why the evaluation of intermediate outcomes should require 2-years of interim evaluation. See proposal to allow for more flexibility below. **Proposed change (if any):** As a general rule, **clinical trials should be of sufficient duration (e.g. one year), under consideration of ~~a two-year interim evaluation, and a 5-year final evaluation may be considered appropriate. However, this can be modified with factors like~~** size of the trial, activity of the investigational compound, patient characteristics, and the requirements with regard to statistical rigor.	Agreed. See above.
310-313	1	Comment: From current presented histology defined efficacy in NASH patients it seems possible to show effect in pharmacological studies even after 36 weeks. It is recommended that the general rule for the interim evaluation should be at least 12-18 months to align with both the stakeholder recommendation and other authority guidelines for NASH[4].	Partially agreed. A more flexible approach is given, nevertheless, the 2-years are still given as recommendation.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
334-341	8	Comment: We agree strongly that the handling of missing data and methods of imputation are very important considerations. The binary nature of the endpoint assessments (categorical responder versus non-responder only) leads to dichotomization of response and loss of information with respect to varying degrees of improvement. Due to the invasive nature of the biopsy, recurrent continuous assessment over time is not feasible, limiting the assessment of the histologic benefit to only a single post-baseline measure, while precluding the ability to perform imputations using traditional approaches. This likely results in an underestimation of histologic response in ITT analyses, where missing or inadequate biopsies are often automatically classified as non-responders and active treatment (but not placebo) is penalized. Further, the inherent variability of histologic evaluation makes it an imprecise endpoint that likely dilutes the treatment effect in a clinical trial.	Partially agreed. Estimand strategy description has been revised. Recommendation however, relates to an adequate missing data strategy, and presentation of sensitivity analyses.
345-348	4	**Comment:** - There is currently no published data to support this highly speculative statement outlined within this paragraph. There are many factors that determine the acceptance by a patient or study subject of a second biopsy including relationship with the	Not agreed. In trials with manifestation of cirrhosis being part of the primary endpoint, this can be assumed to be a realistic scenario, which will usually e beased on the evaluation of routine biomarkers, and US-based methods.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		practitioner, experience with the previous biopsy, and symptomatology. **Proposed change (if any):** ~~**As a biopsy during the follow-up is only scheduled if there is a high likelihood of a cirrhosis (e.g. based on surveillance with non-invasive methods such as fibroscan), non-performance of a scheduled biopsy should be considered as an event.**~~	
345-348	10	**Comment:** There is currently no published data to support this highly speculative statement outlined within this paragraph. There are many factors that determine the acceptance by a patient or study subject of a second biopsy including relationship with the practitioner, experience with the previous biopsy, and symptomatology. **Proposed change (if any):** ~~**As a biopsy during the follow-up is only scheduled if there is a high likelihood of a cirrhosis (e.g. based on surveillance with non-invasive methods such as fibroscan), non-performance of a scheduled biopsy should be considered as an event.**~~	Not agreed See above.
Line 349	11	General Comment: More granular and yet concise language on combo therapy in NASH is needed as the currently available	Comment noted and partially agreed. Paragraphs on combination has been revised.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		data are rather vague within the current research trend looking for combination therapies. Adequate defining combo's efficacy and safety is needed. Many drug combinations can have additive or even synergistic effects, it is the degree of synergy that becomes the essential question. Answering it requires benchmarking the two-drug combination against not only the single treatments, but also the best of previously reported drug combinations if any. Testing multiple combinations and analysing the degree of synergy during a combo therapy clinical trial can reveal patterns suggestive of the mechanisms behind drug interactions but this may also require some novel biomarkers to verify. Understanding the mechanism of drug synergy, rather than simply knowing which drugs to combine, enables further optimization of beneficial drug interactions and can offer important insights into the disease understudy in the case of metabolic syndrome and NASH. Proposed change (if any):	
NASH Section 4.2.3 Lines 349-367	6	**Comment:** The guideline mentions: *"...it will be expected that either a second line treatment is investigated, which has to include the establishment of a definition of an insufficient response to a standard treatment (or at*	The section on combination treatment has been revised. The comments are partially agreed, and the previous absolute requirement for second line or high risk patients treatment has been largely abandoned.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		least one of the combination partners), or – in case an initial combination treatment is aimed at – the definition of a patient group with a very high risk of progression". While we understand that the EMA point of view on combination therapy is based …"referring mainly to other disease areas,…" we think that the situation in NASH is somewhat different where at present there is no approved drug to treat this multifactorial disease. We do not agree that first line combination therapy should be limited to patients with a very high risk of progression, considering that drug combinations can potentially employ lower doses to attain benefit(s) with a lessened unwanted side effect profile. As well, drug combinations can provide multiple activities (metabolic, resolution of NASH, and decreased fibrosis) which should be beneficial for NASH patients with different stages of fibrosis.	
Section 4.2.3 Study design and endpoints Combinatio n Treatment	13	Comment: As patients for whom access to treatment to this disease is literally a life-and-death issue we believe that combination therapies must be considered a worthwhile pathway to treatment. While we understand the point of view that combination therapy is generally not recommended as the starting point, with a multi-target disease such as NASH, we disagree that	Partially agreed. Reasons for revision are given in the revised final reflection paper, and has picked up the concerns mentioned.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
Lines 349-367		combination therapy should be limited to 2nd line therapy. Within the reflection paper and highlighted above, we appreciate the acknowledgement of the connection between NASH and other diseases. To answer the unmet need of NASH patients, a robust list of therapies that can meet both endpoints, and may already be approved must be considered. We also have concerns with the recommendation that 1st line combination therapy should be limited to patients with a very high risk of progression. As mentioned earlier, combination therapies can play an important role for all patients at all stages of NASH. Proposed change (if any): We strongly suggest greater consideration be given to combination therapies. We would appreciate additional clarity regarding the level of evidence that would be required for each component of a combination, and if different considerations may be applied if one drug is already available on the market.	
349-367	14	**Comment**: We suggest additional clarity regarding the level of evidence that would be required for each component of a combination, and if different	Partially agreed. Please refer to the revised text in the final reflection paper. See also above.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		considerations may be applied if one component is a drug already available on the market. While we understand the EMA point of view that combination therapy is generally not recommended as the starting point, with a multi-target disease such as NASH, we disagree that combination therapy should be limited to 2nd line therapy. We further disagree that 1st line combination therapy should be limited to patients with a very high risk of progression. Combination treatment can be more effective while also limiting toxicity. **Proposed change (if any)**:	
362-367	1	Comment: Consider adding language regarding the opportunity to accelerate (example adaptive design) phase 2 and phase 3 clinical trials with combination therapies as there is a significant value that combination of drugs could have in treatment of NASH[4,7,8]. Placebo arm data could potentially be extrapolated from corresponding monotherapies development avoiding excessive enrollment with significant ethical, logistic and time execution considerations.	Not agreed. No need to refer to general guidance and best not reer to issues still under discussion in the regulatory field (e.g. see concept paper on platform trials).

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
362-367	4	**Comment**: At this early stage in the evaluation of new NASH therapeutic options, there is concern about limiting combination options to only 2nd line usage. It should be the clinical utility that should determine the stage of clinical use. It should be included that there is also potential benefit of combinations in first line use in F2/F3 patients if sufficient clinical evidence is available. 1. To fully explore the properties of each single substance in a FDC in an indication such as NASH is a tough requirement and may lead to years of delays wrt availability of effective drugs. 2. It is suggested that the demonstration of a contribution in Ph2 is sufficient (ideally based on future biomarkers), and finally the FDC will be required to proof confirmation as a whole (FDC vs Pbo). - Alternately other concepts to consider, e.g.:Mono A vs FDC A&B vs Pbo; Aggressive NASH progressors **Proposed change (if any): Combination therapies could be considered for either first line or second line therapy (**insufficient response to a standard treatment or at least one of the combination partners**) based on their risk/benefit balance.** ~~**Also, referring mainly to other …[…] definition**~~	Partially agreed. Section has been revised. See above.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		~~**of a patient group with a very high risk of progression.**~~	
362-367	8	Comment: We do not agree that combination therapy is only for second line therapy or for patients with a very high risk of progression. A combination of drugs may result in an improved benefit-risk profile by allowing use of lower doses of each component. This, combined with improved efficacy, can lead to better risk-benefit profile than a single agent.	Partially agreed. See above
362-367	10	**Comment:** At this early stage in the evaluation of new NASH therapeutic options, there is concern about limiting combination options to only 2nd line usage. It should be the clinical utility that should determine the stage of clinical use. It should be included that there is also potential benefit of combinations in first line use in F2/F3 patients if sufficient clinical evidence is available. **Proposed change (if any): Combination therapies could be considered for either first line or second line therapy (**insufficient response to a standard treatment or at least one of the combination partners**) based on their risk/benefit balance.** ~~**Also, referring mainly to other …[…] definition of a patient group with a very high risk of progression.**~~	Partially agreed. See above.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
Lines 363-367	7	**Comment**: Recent clinical research of single agents for NASH has suffered a number of Phase 2 and Phase 3 failures. Research efforts are therefore more predominantly focusing on a combination, multi-modal approach to target NASH; and hence a second line treatment approach is unlikely to be feasible. Additionally, despite several "natural history" analyses of prior failed studies, it is not currently possible to define a patient, or patient group, who are likely to be at "very high risk of progression". **Proposed change (if any):** ~~Also, referring mainly to other disease areas, it will be expected that either a second line treatment is investigated, which has to include the establishment of a definition of an insufficient response to a standard treatment (or at least one of the combination partners), or in case an initial combination treatment is aimed at the definition of a patient group with a very high risk of progression.~~	Partially agreed. See above.
371-376	9	Comment: Some editorial adjustments are proposed. With respect to the discussion of the risk factors a reference should be included (e.g. Mantaka et al. 2012).	Partly agreed. References have been added to this section.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		Proposed change (added text: underlined, removed text: ~~striked through~~): Primary biliary cholangitis (PBC), previously known as primary biliary cirrhosis28, is a chronic, slowly-progressive autoimmune cholestatic liver disease29. ~~The disease is mainly diagnosed in female patients with a ratio of about 10:1.~~ PBC is a rare disease, with incidence and prevalence reported at variable rates (0.33 to 5.8 100,000/year for incidence; 1.91 to 40.2 per 100,000 for prevalence). Whereas an increase in the incidence has been reported for the last decades, newer global data also indicate changes in the diagnosis and course of the disease (irrespective of treatment) with older ages at diagnosis, and slower progression over time 30. PBC Is more prevalent in female patients (with a ratio of about 10:1). The pathogenesis of the disease is not fully understood. ~~,~~ ~~with e~~Environmental (e.g. sunlight, toxins), infectious agents (e.g. bacteria, viruses), ~~and~~ genetic predispositions~~, and with~~ may induce an inflammatory process targeting biliary epithelial cells, and resulting in changes of bile-acid metabolism, and enterohepatic circulation ~~being involved~~.	
380-388	9	Comment: Relevant references should be added to proposed text.	References have been added.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		Further, it is advised to include more guidance on the acceptance of itching and pruritus scales, similar this may hold for fatigue. Also include more information on possible surgical interventions and data on the natural course of the disease (e.g. Al-Harthy and Kumagi, 2012; Harms et al. 2019). Proposed change (if any):	Partly agreed. No validated scales for pruritus or fatigue are available. In the meantime, partly validated scales can be included in trials, also for validation purposes.
383	8	Comment: The reflection paper should mention other groups of patients at high risk of rapid PBC progression, e.g., males.	
395	9	Comment: Proposed change (added text: underlined, removed text: ~~striked through~~): …introduced in the 1990s. ~~More recently,~~ In 2016, obeticholic acid has been licensed ~~in 2016~~ for the "treatment of primary biliary cholangitis (previously also known as primary biliary cirrhosis) in combination with….	The text regarding obeticholic acid has been amended.
398	9	comment: It is advised to include some information on the treatment effects of UCDA and obeticholic acid and to explain the need for development of new medicinal products for PBC.	Partly agreed. The medical need for a second-line treatment is stated.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
402-404	9	Comment: Proposed change (added text: underlined, removed text: ~~striked through~~): Because a standard first-line therapy ~~option~~(UDCA), plus a second-line add-on therapy ~~option~~(obeticholic acid) are available, the inclusion of an adequate patient population depends on the intended place in therapy of the investigational agent.	Not agreed. The Reflection paper refers to the standard treatments in detail in other sections.
409	9	Comment: And editorial adjustment is proposed. Recent literature (Harms et al, 2019) showed that patients treated with UDCA, despite an incomplete response, showed a better outcome in transplant free survival, over untreated PBC patients. Proposed change (if any): "unsatisfactory should be changed into "unsatisfactory**"**.	Comment welcomed. Editorial change agreed and the Reflection paper amended accordingly
413	9	Comment: Proposed change (if any): The text in current lines 427-431 on serum biomarkers should be inserted after line 413.	Partly agreed, editorial changes have been made with a different structure of the text.
415-416	4	**Comment**: - To limit the selection of patients to those with biopsy would be critical, as the vast majority of PBC patients do not have baseline biopsy.	Partly agreed. Inclusion of at least a subgroup of patients with baseline histology evaluation is, however, still recommended.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		Proposed change (if any): ~~Whereas~~ **F**or early-stage trials, the omission of a histological evaluation, including screening as well as endpoint evaluation is considered acceptable. ~~**the availability of a baseline histology evaluation (as well as follow-up evaluation, see below), is highly recommended.**~~ **If baseline or follow up histology evaluation is available, it is recommended that these should be collected and recorded.**	
415-416	10	Comment: To limit the selection of patients to those with biopsy would be critical, as the vast majority of PBC patients do not have baseline biopsy. **Proposed change (if any):** ~~Whereas~~ **F**or early-stage trials, the omission of a histological evaluation, including screening as well as endpoint evaluation is considered acceptable. ~~**the availability of a baseline histology evaluation (as well as follow-up evaluation, see below), is highly recommended.**~~ **If baseline or follow up histology evaluation is available, it is recommended that these should be collected and recorded.**	See above.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
417-427	9	Comment: It is proposed to remove this section. The relevance of this section with respect to the selection of study patients is unclear, since in the lines after this section more specific guidance for the selection of study patients is provided. Proposed change (if any): see above.	Not agreed. This information is considered relevant as a background to the recommendations made in the Reflection paper
PBC Section 4.3.2. Lines 423-431	6	**Comment:** The guideline reads "*An analysis of these different proposals, however, has shown that the likelihood to develop endpoints (such as cirrhosis, decompensation events, and liver transplantation and death, see below) during the course of a trial largely depends on the strictness of these inclusion criteria. It is therefore recommended that the more strict criteria are chosen, allowing only those patients into the trial which have still a relevant alteration of the serological markers of PBC. Currently, best appears to be the combined use of the ALP≥2xULN, and bilirubin >1xULN despite an at least 1 year therapy with UDCA at the standard recommended dose (10-15 mg/kg b.w./day).*" On the other hand, the EMA Report of the stakeholder interaction meeting on the development of medicinal products for chronic non-infectious liver diseases, held on 3rd December 2018 (EMA/873574/2019), raised that "*... requiring an ALP above 2 ULN whilst*	Agreed. The requirement for a bilirubin elevation baseline has been removed.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		excluding patients without increase in bilirubin would mean to exclude about 91 % of patients with PBC from trials, rendering the trial population less representative for the patient population in need of early treatment". We consider that these patient criteria are too restrictive given that: • They represent less than 10% of the population of patients with PBC and would not allow capturing the vast majority of patients at need of treatment. There is a need of early treatment. • Recruiting in clinical trials such a small percentage of a patient population suffering from a rare disease appears unrealistic. • It has been shown (Bettina Hansen's presentation at EMA stakeholder interaction meeting on 3 December 2018) that the hazard ratio for transplantation or death is significantly elevated for subjects with ALP value above 1.67xULN as well as for subjects with total bilirubin (TB) values below ULN, as soon as the TB value exceeds 0.6xULN.	
426-431	4	**Comment:** I - Inclusion criteria should be less restrictive in order to enable development of new	Partly agreed. The inclusion criteria have been made more flexible. The role of risk calculators is not established in the clinical trial setting.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		treatments in populations at need. Reference to more recently derived risk calculators such as Globe and PBC-UK should be included. **Proposed change (if any):** It is therefore recommended that the more strict criteria are chosen, allowing only those patients into the trial which have ~~**still a relevant alteration of the**~~ **altered** serological markers of PBC~~**. Currently, best appears to be the combined use**~~ **indicative** of **elevated risk after adequate therapy with UDCA into clinical studies.** ~~**the ALP≥2xULN, and bilirubin >1xULN despite an at least 1 year therapy with UDCA at the standard recommended dose (10-15 mg/kg b.w./day).**~~ Additional criteria with regard to transaminases, albumin, GGT, ~~**or**~~ **Globe or UKPBC** Mayo risk score may be applied, if adequately justified.	
426-431	10	**Comment:** Inclusion criteria should be less restrictive in order to enable development of new treatments in populations at need. Reference to more recently derived risk calculators such as Globe and PBC-UK should be included. **Proposed change (if any):** It is therefore recommended that the more strict criteria are chosen, allowing only those patients into	See above

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		the trial which have **~~still a relevant alteration of the~~ altered** serological markers of PBC**~~. Currently, best appears to be the combined use~~ indicative** of **elevated risk after adequate therapy with UDCA into clinical studies. ~~the ALP≥2xULN, and bilirubin >1xULN despite an at least 1 year therapy with UDCA at the standard recommended dose (10-15 mg/kg b.w./day).~~** Additional criteria with regard to transaminases, albumin, GGT, **~~or~~ Globe or UKPBC~~-~~**Mayo risk score may be applied, if adequately justified.	
427-428	8	Comment: The bar is set too high based on ALP ≥ 2xULN and bilirubin > ULN. Data shows substantial risk of progression are correlated with values lower than this (i.e. anything greater than ULN).	Partly agreed. The requirement for an elevated bilirubin has been removed and inclusion criteria have been made more flexible.
440-442	4	**Comment:** - We suggest to clarify the language to avoid the impression that patients with high risk disease may be randomized to placebo in a 1 year trial, which would not be justifiable. **Proposed change (if any):** **If performed in low risk PBC patients,** the conduct of such trials may include in addition to a direct comparison to UDCA, also a (potentially small; e.g. based on unequal randomisation) placebo arm for assay sensitivity purposes in case non-inferiority will	Partly agreed and the text amended accordingly.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		be the aim of such trials. **In considering the properties of the current standard of care UDCA which has moderate efficacy and a relatively good and well established safety profile, trials of superiority may allow a more clear positive conclusion on risk-benefit of new first-line therapies.**	
440-442	10	**Comment:** We suggest to clarify the language to avoid the impression that patients with high risk disease may be randomized to placebo in a 1 year trial, which would not be justifiable. **Proposed change (if any):** **If performed in low risk PBC patients,** the conduct of such trials may include in addition to a direct comparison to UDCA, also a (potentially small; e.g. based on unequal randomisation) placebo arm for assay sensitivity purposes in case non-inferiority will be the aim of such trials. **In considering the properties of the current standard of care UDCA which has moderate efficacy and a relatively good and well established safety profile, trials of superiority may allow a more clear positive conclusion on risk-benefit of new first-line therapies.**	See above

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
440-447	9	Comment: Proposed change (added text: underlined, removed text: ~~striked through~~): The conduct of such trials, especially in case non-inferiority is aimed for, may also include - in addition to a direct comparison to UDCA – ~~also~~ a (potentially small; e.g. based on unequal randomisation) placebo arm for assay sensitivity purposes. ~~While~~ It~~i~~t is acceptable to demonstrate not-inferiority to the established treatment, as well as an acceptable safety profile for licensing. ~~, c~~Considering the properties of the current standard of care with moderate efficacy and relatively good and established safety profiles, it might be necessary to aim at superiority in such trials in order to allow a more clear positive conclusion on risk-benefit, especially in case the safety profile does not allow a conclusion on a similar level of acceptability as for UDCA.	See above. Text has been amended partly according to the previous comment with suggestion for change.
448-450	9	Comment: It is proposed to remove this section, since recommended criteria have already been specified in lines 427-431. Proposed change (if any):	Agreed. This section has been removed.
456-459	9	Comment: Proposed change (added text: underlined, removed text: ~~striked through~~):	Agreed and the text aligned with the proposal.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		An evaluation of all potential long-term outcomes is considered to be hardly possible in this population, which would be expected to have a high rate of normalisation of the serological markers at the end of the (primary) observation period~~,~~ and thus have ~~an even~~ delayed further ~~development of~~ disease deterioration.	
461	9	Comment: Proposed change (if any): Please change 'interim' into 'intermediate' (see comments above).	Agreed. 'Interim' has been changed into 'intermediate'.
465-467	8	Comment: The use of ALP and bilirubin as endpoints is reasonable overall, but this recommendation needs to take into account the mechanism of the investigational drug. For example, some investigational drugs may lower ALP via transcriptional or translational mechanisms, so that lowering of ALP would be related to on-target effects and not related to amelioration of cholestasis. Also, the reflection paper states that reduction of total bilirubin and ALP has been demonstrated to lead to an overall improved outcome with regard to the development of end-stage liver disease, decompensation, liver transplantation and death. This being the case, these endpoints should no longer be	Not agreed. At present, it has only been demonstrated for the natural history, as well as for UDCA, and not for new investigational agents, that the reduction of ALP leads to an overall improved outcome with regard to the development of end-stage liver disease, decompensation, liver transplantation and death

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		considered intermediate endpoints and should be considered endpoints for full approval. Proposed change (if any): Allow reduction of total bilirubin and ALP to be used as endpoints for full approval for PBC.	
472-479	9	Comment: Required endpoints should be defined more clearly. Please also take into account the comments made at the EMA stakeholders interaction meeting in December 2018. Proposed change (added text: underlined, removed text: ~~striked through~~): The choice of adequate thresholds for the definition of response would need to be adapted to the chosen inclusion criteria~~, but usually,~~. T~~the most clear-cut t~~hresholds close to normalisation would be expected to be evaluated, such as~~. Previously, the criteria of ALP<1.67xULN, ALP decrease of at least 15%, as well as (total) bilirubin ≤ULN have been thought to be acceptable. However, more stringent definitions of response are advocated here, with the ALP criterion being at least~~ ALP<1.5xULN with an at least 40% decrease, and ~~total~~conjugated bilirubin ≤ULN. Additional criteria with regard to transaminases, GGT, and/or Mayo score may be added, depending on the respective inclusion criteria.	Partly agreed. The recommendations for endpoints required have been made clearer in line with the proposal and the comments made at the EMA stakeholder meeting
480-481	9	Comment: Proposed change (added text: underlined, removed text: ~~striked through~~):	Partly agreed. The Reflection paper no longer considers conditional marketing authorisation based on intermediate endpoints to be the most appropriate submission strategy in PBC

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		~~Because~~ T~~t~~he validity of these intermediate endpoints is not fully established. Hence, ~~it would usually be expected that~~ long-term outcome data with respect to the histological …	
484-487	9	Comment: The relevance of this section is unclear. Either some more specific guidance should be provided, or this section should be removed. Proposed change (if any):	Agreed. The Reflection paper has been updated in line with the comment.
495	8	Comment: The request for histology evaluation in light of "the fact that a fully validated histological scoring system for the disease is not available" (Section 4.3.3 page 13) seems arbitrary and puts patients at unnecessary risk from biopsy when the utility of such data will be questionable. Liver biopsy has been largely replaced by less invasive measurements such as mathematical models in defining the prognosis of patients with primary biliary cirrhosis. Proposed change (if any): Remove request for histology evaluation.	Partly agreed. There is no longer a request for histology, however, histological evaluations are recommended in at least a sub-group of patients.
PBC Section 4.3.3. Lines 499-502	6 Input from the Working Group is necessary to	**Comment:** The guideline mentions that *"trial durations from 1-2 years have previously been proposed in order to show efficacy on the interim endpoint proposed. From an*	Partly agreed. A submission strategy in PBC aiming at conditional marketing approval based on an intermediate endpoint is no longer considered most appropriate. A trial duration of at least 2 years is therefore desirable

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
	identify stakeholder sending feedback to EMA	*overall efficacy and safety point of view, but depending on the magnitude of effect to be expected, a study duration of at least 2 years seems to be desirable. A trial extension for the longest possible extend should be aimed at.* " Given the low response rate observed in the placebo arm from clinical trials with obeticholic acid (Ocaliva) and elafibranor most notably and compounds currently in development, it does not appear justified - from a ethical perspective - to keep patients in these groups for a duration as long as 2 years. A one-year duration should be sufficient to show efficacy on the interim endpoint and to show safety.	
499-504	9	Comment: Proposed change (added text: underlined, removed text: ~~striked through~~): ~~Trial~~Study durations from 1-2 years have previously been proposed in order to show efficacy on the ~~interim~~intermediate endpoint(s) ~~proposed~~. From an overall efficacy and safety point of view, but depending on the magnitude of effect to be expected, a study duration of at least 2 years seems to be desirable. A ~~trial~~study extension for the longest possible extend should be aimed at. If indeed studies using long-term	See above

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		outcomes (liver transplantation and death, decompensation events) are intended, these are usually event driven, and a priory determination of the ~~trial~~study duration will not be possible.	
500-501	4	**Comment:** - Serologic changes occur relatively rapidly. Studies of shorter duration should be the standard requirement. **Proposed change (if any):** From an overall efficacy and safety point of view, but depending on the magnitude of effect to be expected, a study **of sufficient** duration ~~**of at least 2 years seems to be**~~ **is** desirable.	Not agreed. See above. Serological changes are not yet established as surrogates for long-term clinical outcomes with new investigational agents. In general, for a chronic condition of slow progression where long-term treatment is aimed at, there is need for long-term efficacy and safety data.
500-501	10	**Comment:** Serologic changes occur relatively rapidly. Studies of shorter duration should be the standard requirement. **Proposed change (if any):** From an overall efficacy and safety point of view, but depending on the magnitude of effect to be expected, a study **of sufficient** duration ~~**of at least 2 years seems to be**~~ **is** desirable.	See above
505-509	8	Comment: The Sponsor agrees that the use of natural history controls should be considered. The clinical course of PBC as well as the expanded treatment	Partly agreed. The use of natural history controls/real world evidence data is now addressed in a separate section.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		landscape (i.e., marketed treatments, off-label use of therapies, and proliferation of clinical trials of other investigational agents), makes a long-term placebo-controlled study challenging. Further, the progressive and serious nature of the disease makes the continued participation of patients in a long-term placebo-controlled trial very difficult. Proposed change (if any): Remove the statement, "However, this is currently not recommended as acceptable strategy and must – for the time being – be also considered as supportive endpoint only."	
531	8	Comment: Many of the comments made above for PBC also apply to PSC for which the guidance regarding intermediate endpoints and long-term outcomes studies are the same or very similar.	Comment noted
Line 543	17	Statement: "Symptoms usually develop with progression of the disease" Commentary The available evidence in this domain does not support this statement. Indeed, symptoms relating to fatigue, abdominal pain and pruritus, are multifactorial, may relate to underlying inflammatory bowel disease (IBD) activity, and do not correlate well with disease stage in PSC (Dyson et al., 2015). Moreover, jaundice in PSC	Agreed. The text has been amended in line with the comment.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		may be multifactorial, and relate to acute cholangitis episodes, focal bile duct stenosis, or development of intraductal calculi, none of which relate directly to progression of disease per se. Proposed change: Rephrase to: “Symptoms are broad and heterogeneous, do not correlate with disease stage, and can manifest in patients with even early stage PSC. Acute onset or worsening of existing symptoms may relate to disease progression in some circumstances, but also to underlying IBD activity or episodes of acute cholangitis.”	
556-560	9	Comment: Proposed change (if any): This section should be inserted after the word ‘fibrosis’ in line 535. <u>Motivation</u>: The information in this section concerns general epidemiological information on PSC. This information would rather be stated at the beginning of an introductory paragraph on PSC.	Agreed. A section on epidemiology has been added at the beginning of the chapter on PSC
Line 558	17	Statements: “The development of the disease is slow”	Partly agreed and the text amended in line with the proposal.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		Commentary Published data often indicate 'median' liver transplant free survival times, and approximate 14 years from patients diagnosed in liver transplant centres, to >20 years in non-transplant units. However, rates of disease development and disease progression vary immensely from one individual to the next. Data from the International PSC Study Group, in addition to population-based data from elsewhere, indicate that >20% of patients require liver transplantation or die within less than 5 years from diagnosis (Boonstra et al., 2013; Weismüller* Trivedi* et al., 2017). Proposed change: Rephrase to: Rates of disease progression vary immensely between one individual to the next, and accurately characterising the clinical course which patients experience is challenging.	
Line 564	5	Comment: "Diagnosis…relies on the profile of elevated ALP" –The Reflections Paper relies heavily on ALP for both diagnosis of PSC as well as an endpoint for clinical development; yet, there are significant problems with ALP for both. ALP is normal in a subset of PSC patients without cirrhosis, its utility in advanced PSC patients with cirrhosis is not entirely clear, and its role in patients who develop ALP normalization as cirrhosis progresses are other limitations. At this time, use of ALP as a surrogate	Agreed. The Reflection paper reflects this issue, i.e. diagnosis is not only based on elevated ALP and the ALP as endpoint needs to be supported by other endpoints as well.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		endpoint should be supported by other biomarkers or clinical evidence of benefit.[1] Proposed change (if any):	
Line 564	17	Statement: “Diagnosis relies on the profile of elevated ALP” Commentary: ALP seems to be a heavy focus of the paper, in terms of diagnosis of PSC, in addition to an endpoint for drug development trials. However, ALP is normal in a subset of PSC patients without cirrhosis (particularly young patients, who more often manifest an inflammatory PSC phenotype). Reciprocally, in patients with advanced disease with established cirrhosis, the diagnostic and prognostic utility of ALP is unclear, and overridden by biomarkers of liver fibrosis that are more immediately linked to clinical outcomes. Proposed change: Rephrase to: There is no single diagnostic test for PSC, although serum ALP is elevated in most patients. Importantly however, ALP values exhibit wide inter- and intra-individual variability, may be normal in a subset of (often young) individuals with early yet rapidly progressive disease, in addition to those with established liver cirrhosis.	See above. While it is acknowledged that ALP may be normal in a sub-set of patients, it would not be expected that these patients would be included in clinical trials in PSC. The Reflection paper states that elevation of ALP predominates in most patients.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
Lines 567-568	5	Comment: There are questions about the value of histology for diagnosis. Lines 567-568: "The availability of a liver biopsy consistent with PSC is a compulsory requirement." This is not in accord with the EASL Practice Guidelines[3] of 2009, the AASLD Practice Guidelines of 2010[4], or the American College of Gastroenterology Practice Guidelines of 2015[5]. Proposed change (if any): That statement should be deleted from the paper.	Agreed. The requirement of liver biopsy for diagnosis has been removed.
Lines 567-568	7	**Comment**: In order to enhance trial recruitment in this orphan population, and to prevent unnecessary biopsy, use of a historical biopsy should also be considered (if sample is deemed acceptable for interpretation by a central reader) **Proposed change (if any):** The availability of a liver biopsy ***either at screening, or within 6 months of the screening visit,*** consistent with PSC is a compulsory requirement.	See above. Biopsy is no longer required.
567-568	8	Comment: The statement, "The availability of a liver biopsy consistent with PSC is a compulsory requirement" should be deleted, because it is not	Agreed. A biopsy is no longer needed for diagnosis.

[3] European Association for the Study of the Liver. EASL Clinical Practice Guidelines: Management of cholestatic liver diseases. *Journal of Hepatology*. 2009; 51: 237-267.
[4] Chapman R et al. AASLD Practice Guidelines: Diagnosis and Management of Primary Sclerosing Cholangitis. *Hepatology*. 2010; 51(2): 660-678.
[5] Lindor K et al. ACG Clinical Guidelines: Primary Sclerosing Cholangitis. *American Journal of Gastroenterology*. 2015; 110: 646-659.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		consistent with the AASLD Practice Guidelines, the EASL Practice Guidelines, or the American College of Gastroenterology Practice Guidelines.	
567-568	16	Comment: **Diagnosis of PSC** The availability of a liver biopsy is necessary to exclude AIH, however not a compulsory requirement to diagnose PSC. As stated in the report of the stakeholder meeting: "Diagnosis is clinical (imaging and biomarkers) and presents with considerable variation of biochemistry and symptoms in the course of the disease with poorer long-term survival in patients symptomatic at diagnosis (Broome et al Gut 1996) (median survival from diagnosis until LT or PSC-related death in population based cohorts 21 years and in transplant center based cohorts 13.2 years (Boonstra et al. Hepatology 2013)." See EMA/873574/2019, p4	Agreed and text amended partly in line with the proposal.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		Proposed change (if any): Please change this part in line with the discussion of the stakeholder interaction meeting.	
Lines 567-568	17	Statement: "The availability of a liver biopsy consistent with PSC is a compulsory requirement." Commentary: This comment is confusing, and contradicts clinical practice guidelines from Europe, North America and the United Kingdom (Aabakken et al., 2017; Beuers, 2009; Chapman et al., 2019, Chapman et al. 2010; Lindor et al., 2015). Moreover, liver histological features of PSC may not manifest in younger patients who often present with a more inflammatory PSC phenotype, despite cholangiography being compatible with diagnosis (Gregorio et al., 2001). Proposed change: Liver biopsy is not required to diagnose PSC but when undertaken must have features compatible with a diagnosis of PSC	See above. Availability of liver biopsy is no longer a compulsory requirement.
Line 567-568	17	Comment: 'The availability of a liver biopsy consistent with PSC is a compulsory requirement.' We would suggest that	Agreed. The Reflection paper has been amended in line with the comment. See above.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
Patient perspectives from PSC Support		this is removed as MRCP is the gold standard for diagnosing PSC. Proposed change (if any):	
568	9	Comment: Proposed change (if any): It is assumed that the 'compulsory requirement' in this line pertains to clinical trials. If so, this should be added for clarification.	See above. Liver biopsy no longer a requirement.
Lines 568-569	5	Comment: There is a lack of consensus about the definition of overlap syndrome; further guidance is needed. Proposed change (if any):	Not agreed. Providing a definition of AIH is outside the scope of a regulatory guidance document.
Lines 568-569	17	Statement: "The presence of overlap (e.g. with AIH) syndromes can be allowed in exploratory clinical trials, but not confirmatory trials." Commentary: Overlap syndromes lack codified diagnostic criteria, consensus definition, and it is unclear whether treatment responses in PSC or AIH can be applied herein. Moreover, the term "overlap syndrome," whilst widely used, is not recommended by the international	Not agreed. Patients with AIH should be excluded from confirmatory trials in PSC as a sub-group with different course and treatment AIH/PSC is more frequent in paediatric population, inclusion could be accepted

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		autoimmune hepatitis group, the international PSC study group or UK-PSC. Proposed change: Please rephrase to: There is lack of consensus about the definition of "overlap syndrome" and this terminology should be avoided. The presence of overlapping features (e.g. with AIH) should not preclude entry into exploratory or confirmatory trials, although concomitant usage of immunosuppression and disease phenotype needs to be accounted for.	
Lines 569-570	5	Comment: "Other secondary reasons for PSC…" should be changed to: Proposed change (if any): "Other secondary reasons for sclerosing cholangitis…"	Agreed. Text corrected.
Lines 569-570:	17	Statement: "Other secondary causes for PSC…" Commentary: The simultaneous use of terms "secondary" and "primary (the 'P' in PSC) should be avoided. Proposed change: Please rephrase to: Secondary causes for sclerosing cholangitis need to be excluded before a firm diagnosis of PSC is attributed.	

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
571-588	16	Comment: **Selection on patients** From a patient perspective creative trial designs are encouraged to incorporate co-existing diseases, ages, disease stage and severity, fibrosis, inflammation and colon/bile damage. Although the Reflection Paper represents EMA region, it would be helpful for the global patient community perspective in this rare disease to include consideration of inclusion of patients from around the globe, outside Europe and North America. For example Japan PSC has a relatively large patient community to consider. Proposed change: Please add accordingly	Partly agreed. The Reflection paper does not exclude trials outside EU or the US, however, care has to be taken when including study centres outside these regions considering geographical variations in the disease and clinical practice.
Line 573	17	Statement: “If patients have already a dominant stricture…” Commentary: Consensus definition of dominant strictures is lacking, and strictly this is an ERCP diagnosis. The ERCP	Agreed. The text has been amended in line with the comment.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		definition is not applicable to MRCP, in particular the extrahepatic ducts, given insufficient spatial resolution on MRI and the lack of hydrostatic pressure present during ERCP. Proposed change: Please apply the ESGE and EASL criteria, and clarify the intention of this statement.	
Lines 573 & 575	5	Comment: There is a lack of consensus about the definition of a dominant stricture; some of the definitions are very broad and such that minor, non-clinically relevant, narrowing has been described as a dominant stricture. Other definitions have included narrowing along with clinical symptoms of either cholangitis or substantial increase in bilirubin or serum enzymes. The intentions on lines 573 & 575 as written are unclear and should be clarified. Proposed change (if any):	Agreed. See above.
Line 575	17	Statement: "It may also be sensible to define an upper limit of other markers (e.g. for transaminases)" Commentary: This statement requires clarity, and recognition that younger patients with PSC often manifest serum ALT	Not agreed to change. The reasons for this recommendation relate to liver safety and exclusion of patients with AIH overlap syndromes. In general, clinical trial protocols in various disease areas, including liver disease, define an upper limit for liver enzymes.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		and AST values greater than 200 IU/L (>5 times the upper limit of normal in some laboratories). It is wrong to restrict entry into clinical trials for such patients, who are at risk for disease progression. The Phase 2b study of simtuzumab applied a transaminase cut-off of >10xULN, which seems less conservative, but the precise cut-off is speculative. Proposed change: Please rephrase to: Upper limits of other markers, as per the natural history of the disease, need to be defined, whilst being mindful that younger patients often manifest elevated AST and ALT values.	
577-581	16	Comment: **Concomitant IBD** See EMA/873574/2019, p4 and p5: As stated below the PSC patient community strongly advocates for the inclusion of PSC-IBD in PSC trials. Especially in addition to that the PSC patient population consist for 158pprox.. 70% of patients with concomitant IBD. It would be helpful if future compounds would be indicated for the broad population, including concomitant IBD treatment.	Comment welcomed. Patients with IBD may be included in trials in PSC. The text addressing inclusion of patients with IBD has been expanded with further recommendations.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		"Considering coexisting IBD PSC may also present with a wider group of phenotypes than PBC. Whilst IBD is likely closely related to disease biology patients with active IBD are frequently excluded. Inclusion of IBD patients should therefore likely more depend on the mode of action of the proposed intervention e.g. drugs downregulating inflammation vs. drugs affecting fibrosis. Experts further acknowledged that non investigational co-medication on IBD such as intercurrent biologics may affect the disease course of PSC in a trial. As PSC is a common exclusion criterion in IBD trials little is known about efficacy of authorised IBD medication in PSC activity making currently patients with low IBD activity the preferred choice for clinical trials. To allow studying patients with concomitant IBD more specific discussion is needed how to monitor and treat IBD flares in PSC trials as they present a significant confounder. Given ongoing changes in IBD care, trials need to reflect present-day care e.g. increasing use of biologics. The PSC patient community strongly advocates for the inclusion of PSC-IBD in PSC trials."	

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		Proposed change (if any): Please change this part in line with the discussion of the stakeholder interaction meeting on the development of medicinal products for chronic non-infectious liver diseases (PBC, PSC, NASH), to include considerations on PSC patients with concomitant IBD and concomitant IBD treatment.	
Line 579 Patient perspectives from PSC Support	17	Comment: We strongly support the inclusion of PSC patients with inflammatory bowel disease (IBD) in clinical trials and support the need for stable IBD status and medication. It is important to patients that PSC clinical trial populations reflect the majority, not the minority without compromising clinical trial design. Proposed change (if any):	Comment welcomed. Patients with IBD may be included in trials in PSC. The text addressing inclusion of patients with IBD has been expanded with further recommendations.
Line 580 Patient perspectives from PSC Support	17	Comment: While we understand and acknowledge the need to ensure acute cholangitis should not have occurred for a relevant timeframe, a cholangitis flare is a major episode for PSC patients and as such these individuals are highly motivated to participate in clinical trials. Given that PSC is a rare disease, wider screening windows and opportunities for re-screening for trial participation when it is clear a flare has subsided	Partly agreed. A timeframe of 6 months since last acute cholangitis episode has been defined to make this recommendation clearer.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		should be considered (where appropriate and without compromising trial design). Proposed change (if any):	
Lines 580-581	5	Comment: Patients who have symptomatic strictures should be excluded. This is pointed out in lines 580-581. Proposed change (if any):	Agreed. Patients with relevant strictures needing surgical or endoscopic intervention should be excluded as per the Reflection paper.
Lines 582-585:	5	Comment: We support a pragmatic approach to inclusion in trials whereby the aim of the trial is taken into account in identifying inclusion criteria. Patients support wide inclusion criteria without compromising effective trial design. Proposed change (if any):	Comment noted. The inclusion criteria in PSC trials are not yet established and current recommendations allow some flexibility. In addition it is recommended to seek scientific advice for discussing some aspects of the inclusion criteria.
Line 582-585 Patient perspectives from PSC Support	17	Comment: We support a pragmatic approach to inclusion in trials whereby the aim of the trial is taken into account in identifying inclusion criteria. Patients support wide inclusion criteria without compromising effective trial design. Proposed change (if any):	See above
Lines 584-586:	5	Comment: Patients support the inclusion of individuals on stable dose UDCA. In the absence of	Agreed. Patients on stable dose of UDCA may be included.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		any proven effective treatment of PSC, it is challenging for some patients on long-term UDCA to comprehend stopping it, yet these patients may be otherwise eligible for trials. Therefore, as per medication for IBD, the inclusion of individuals on stable dose of UDCA for a relevant timeframe should be allowed. Proposed change (if any):	
Line 584-586 Patient perspectives from PSC Support	17	Comment: Patients support the inclusion of individuals on stable dose of UDCA. In the absence of any proven effective treatment of PSC, it is challenging for some patients on long-term UDCA to comprehend stopping it, yet these patients may be otherwise eligible for trials. Therefore, as per medication for IBD, the inclusion of individuals on stable dose of UDCA for a relevant timeframe should be allowed. As a patient organisation, we would urge all potential trial participants to consider carefully the benefits and risks of continuing to take UDCA during trials. Proposed change (if any):	See above.
586	9	Comment: Proposed change (if any):	Agreed and text altered in line with the proposal.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		It is advised to replace current phrase 'but intake of UDCA should not be altered during the trial' with 'the dose of UDCA should remain unchanged during the study'.	
Line 588	5	Comment: Agree with the inclusion of cirrhotic patients without signs and/ or symptoms of decompensation. Proposed change (if any):	Comment noted.
Line 590	7	**Comment**: As indicated by COMP, Ursodeoxycholic acid is currently authorised in a number of EU member states (national licences) for the treatment of PSC. **Proposed change (if any):** "~~No licensed treatment in PSC is currently available.~~ ***Ursodeoxycholic acid currently has limited availability in a number of EU member states for the treatment of PSC, and no pan-European licenced treatments are available.*** Therefore, a development strategy….."	Agreed. The text has been amended in line with the proposal.
Lines 590-593:	5	Comment: While there is no licensed treatment for PSC, in order to shorten trial duration (compared to the use of hard endpoints), the use of intermediate (surrogate) endpoints is supported by patients. However, follow-up for as long as possible should be	Not agreed. There is need for long-term data within the frame of a clinical trial. The use of real world evidence is addressed separately within the Reflection paper.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		considered to assure patients of effectiveness and long-term safety. Consideration could be given to simple follow-up strategies such as tracking hospital records (with prior patient consent). Proposed change (if any):	
Line 590-593 Patient perspectives from PSC Support	17	Comment: The use of intermediate (surrogate) endpoints in order to shorten trial duration (compared to the use of hard endpoints) is supported by patients while there is no licensed treatment for PSC, However, follow-up for as long as possible should be considered to assure patients of effectiveness and long-term safety. Consideration could be given to simple follow up strategies such as tracking hospital records (with prior patient consent). Proposed change (if any):	
Line 604	5	Comment: ALP cannot currently be accepted as the only intermediate endpoint to be used in PSC trials. This is supported by PSC Support patient survey results, which found that less than 25% of patients were confident that reduction in ALP correlates with clinical outcome.[6]	Agreed. The Reflection paper reflects this. Histology is proposed as a co-primary endpoint.

[6] Walmsley, M. et al. FRI-062-Identifying research priorities in primary sclerosing cholangitis: Driving clinically meaningful change from the patients' perspective. *Journal of Hepatology*. 2019; 70(1) Supplement: e412-413.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		Proposed change (if any):	
604	9	Comment: Proposed change (added text: underlined, removed text: ~~striked through~~): ~~Therefore~~ ALP Ccan ~~currently~~ not be accepted as the only intermediate endpoint to be used in this disease.	
Line 604 Patient perspectives from PSC Support	17	Comment: Alkaline phosphatase (ALP) cannot currently be accepted as the only intermediate endpoint to be used in PSC trials. This is supported by PSC Support patient survey results, which found that less than 25% of patients were confident that reduction in ALP correlates with clinical outcome (Walmsley et al., 2019; www.pscsupport.org.uk, 2016). Proposed change (if any):	See above
Lines 604-606	5	Comment: The data from studies led by Corpechot et al,[7] [8] [9] regarding the value of TE showing the correlation of clinical response to treatment should be included and critiqued. Proposed change (if any):	Partly agreed. The Reflection paper addressed the use of e.g. transient elastography, however, the use of TE will have to be justified as it is currently not fully validated as an outcome measure. It is acknowledged that this may change in the future.

[7] Corpechot C et al. Baseline values and changes in liver stiffness measured by transient elastography are associated with severity of fibrosis and outcomes of patients with primary sclerosing cholangitis. *Gastroenterology*. 2014; 146(4): 970-979.
[8] Corpechot C. Noninvasive elastography-based assessment of liver fibrosis progression and prognosis in primary biliary cirrhosis. *Hepatology*. 2012; 56(1): 198-208.
[9] Corpechot C et al. Assessment of biliary fibrosis by transient elastography in patients with PBC and PSC. *Hepatology*. 2006; 43(5): 1118-1124.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
604-610	16	Comment: **Biopsies in PSC clinical trials** In addition to the remark that "histology in PSC has been discussed controversially (see also 4.1)", please note the discussion of the stakeholder interaction meeting on the development of medicinal products for chronic non- infectious liver diseases (PBC, PSC, NASH): "It was furthermore emphasized that patient related outcomes should be part of any clinical trial in particular itch and fatigue as well as sufficient exploratory endpoints to allow replacement of biopsy in the future. Data sharing on placebo treated patients and natural history data use was considered particularly important to maximise efficacy of research." Proposed change (if any): - Please emphasize the need for exploratory endpoints to allow replacement of biopsy in	Partly agreed. The Reflection paper recommends PRO as a secondary endpoint in any trial in PSC to support a totality of evidence. The need to replace liver biopsy in the future is acknowledged and validation of non-invasive methods is ongoing. The Reflection paper recommends various non-invasive methods to supplement the primary endpoint. Use of natural history data is addressed in a separate section (recommendation to use as an explorative objective).

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		the future to reduce the burden of biopsies for patients in clinical trials. - also emphasize the need for data sharing on placebo treated patients and natural history data use to maximise efficacy of research	
Lines 606-609	5	Comment: The challenges around histology as an endpoint are further highlighted by the Simtuzumab trial in which ELF and FibroScan appeared to be stable over 96 weeks; yet, there was substantial change in histology (37% worsening, 34% no change, 29% improving >1 stage) suggesting that histology is the outlier. The problems with histology are further underscored in the Simtuzumab trial since the collagen proportional area in histology didn't change over 96 weeks.[10] Proposed change (if any):	Comment welcomed. See above on histology and need to replace in the future.
Line 607	5	Comment: The paper indicates that there is newer research suggesting that histology could be used to evaluate the changes; this statement was not supported and the committee is unaware of references that indicate that histology can be used as a surrogate for improvement in the clinical course of PSC. The Nakanuma system, developed for PBC and	Partly agreed. The Reflection paper refers to the controversy of histology in PSC. Reference has been added (de Vries et al, 2015).

[10] Muir AJ et al. Simtuzumab for primary sclerosing cholangitis: Phase 2 study results with insights on the natural history of the disease. *Hepatology*. 2019; 69(2): 684-698.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		published in 2010, is of interest. The de Vries paper[11] cited would indicate that the Nakanuma staging system showed the strongest predictive value for prognosis for PSC; however, the focus was solely on prognosis and not on treatment response. We caution over-emphasis on the Nakanuma system based on limited literature that it outperforms other classifications (Line 617) and recommend a more cautious discussion about the different scoring systems for (Nakanuma, Ishak, Ludwig). Additionally, regardless of any actual sampling error, the determinations made as a result of a biopsy are highly dependent upon the location in the liver from which the biopsy was taken. A study of 44 PSC patients by Olsson et al.[12] found significant disagreement in staging, fibrosis assessments and the presence of cirrhosis in biopsies taken on the same day in different zones of the liver. Proposed change (if any):	
Line 607-608	17	Statement: "…however newer research has been shown that – in addition to its obvious face validity – histology can well be used to evaluate the changes."	Not agreed to change the text. See above

[11] de Vries EMG et al. Validation of the prognostic value of histologic scoring systems in primary sclerosing cholangitis: An international cohort study. *Hepatology*. 2017; 65(30): 907-919.
[12] Olsson R et al. Sampling Variability of Percutaneous Liver Biopsy in Primary Sclerosing Cholangitis. *J Clin Pathol*. 1995; 48: 933-935.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		Commentary: This statement is vague and ambiguous, and we are unaware of any data to suggest that any 'improvement' in liver histology associates with an improved clinical course for patients. Whilst fibrosis progression most certainly associates with worse clinical outcomes, the reverse has not been proven in PSC. Secondly, data on paired liver biopsies in patients with PSC indicate that biopsy reports vary depending on the location from which the liver has been sampled, with significant disagreement in fibrosis staging and the presence versus absence of cirrhosis (Olsson et al., 1995). Proposed change: Please rephrase to: …in addition to its obvious face validity, progression of liver fibrosis, histologically, is associated with an increased risk of clinical events. However, there is no evidence currently that improvement in histological indices of disease severity correlate with better clinical course for patients. Moreover, biopsy reports vary depending on the location from which the liver has been sampled, with significant disagreement in fibrosis staging and the presence versus absence of cirrhosis.	
Line 609	11	**Comment:**	

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		Considering the patchy type of distribution of PSC pathological changes in the liver and the liver histology clinical relevance in PSC mainly for diagnosis confirmation adding an imaging method, ie MRCP to the ALP dynamics would allow disease regression/progression changes detailed intermediate monitoring. **Proposed change (if any):** Therefore, a combined use of cholangiopancreatography (MRCP, ERCP, PTC) evaluation, as potential alternative of liver histology evaluation, and ALP changes are regarded to represent an acceptable intermediate endpoint for the disease for the time being.	Not agreed. Imaging methods such as MRCP are recommended as supportive endpoints for demonstration of long-term efficacy
Line 609-610	5	Comment: The guidance document's focus on histology lacks consideration of the level of risk patients are willing to take and discounts the importance of patient experience. The focus on hard clinical endpoints may be detrimental to accepting other surrogates. The Forum strongly supports the use of Patient Reported Outcomes. Proposed change (if any):	Partly agreed. The use of PRO is recommended as secondary endpoint in all clinical trials in PSC to supplement the totality of evidence. The need to develop alternative endpoints based on non-invasive methods for the future is agreed with.
Line 609-610	17	Comment: 'Therefore, a combined use of histology evaluation and ALP changes are regarded to represent an acceptable	

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
Patient perspectives from PSC Support		intermediate endpoint for the disease for the time being.' In line with our previous comments, we would like to see a less rigid statement about PSC endpoints with an emphasis on the need for the development of intermediate endpoints that reflect the mechanism of action of drug candidates, are non-invasive and incorporate Patient Reported Outcomes where appropriate. Proposed change (if any):	
609-613	4	**Comment:** - To allow for more options and promote development in this important orphan indication it is proposed to use intermediate endpoints in a more flexible manner. **Proposed change (if any):** ~~**Therefore, a combined use of histology**~~ **At present, histologic** evaluation and/or ALP changes are regarded to represent ~~**an**~~ acceptable intermediate endpoint**(s)** for the disease ~~**for the time being**~~, **although assessing different pathophysiological processes**. It is again emphasized that intermediate endpoints used for marketing authorisation must be sufficiently reliable to allow the conclusion of a positive benefit risk at time of marketing authorisation. ~~**Therefore a co-primary**~~**The** ~~-~~evaluation of these intermediate	Not agreed. There is currently insufficient evidence to recommend alternative novel and non-validated primary endpoints for trials in PSC.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		endpoints should be ~~aimed at~~ **aligned with the mechanisms employed to achieve clinical benefit for those with PSC.**	
609-613	10	**Comment:** To allow for more options and promote development in this important orphan indication it is proposed to use intermediate endpoints in a more flexible manner. **Proposed change (if any):** ~~**Therefore, a combined use of histology**~~ **At present, histologic** evaluation and/or ALP changes are regarded to represent ~~**an**~~ acceptable intermediate endpoint**(s)** for the disease ~~**for the time being**~~, **although assessing different pathophysiological processes**. It is again emphasized that intermediate endpoints used for marketing authorisation must be sufficiently reliable to allow the conclusion of a positive benefit risk at time of marketing authorisation. ~~**Therefore a co-primary**~~**The** evaluation of these intermediate endpoints should be ~~**aimed at**~~ **aligned with the mechanisms employed to achieve clinical benefit for those with PSC.**	See above
Line 613	7	**Comment**: The association between changes in serum ALP and complications in PSC is controversial, as highlighted	Not agreed. Limitations of ALP are recognised and therefore the need for a co-primary endpoint addressing a different aspect of the disease.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		by the increased risk of complications observed in patients treated with high-dose ursodeoxycholic acid (UDCA), despite improvements in serum ALP, compared to those treated with placebo {Lindor 2009}. The use of a co-primary endpoint is not recognised as the optimal approach within the PSC scientific community and does not reflect a global approach with the U.S FDA favouring a single endpoint. **Proposed change (if any):** N/A	
Lines 613-616	7	**Comment**: ALP normalization may be best for some patients, however, there is a continuum of risk reduction for complications along the spectrum of ALP reduction {De Vries 2016}. **Proposed change (if any):** "Furthermore it is suggested that a responder-type evaluation based on the criteria of therapeutic response should be the basis, defining serological response as a reduction of ALP."	Agreed. A responder analysis is recommended for ALP changes.
Lines 615-616	5	Comment: The definition of a "serological response as a reduction in ALP to 1.3xULN, or a combination of the reduction to 1.5-1.3xULN with at least 40% reduction from baseline" is arbitrary and not supported by existing literature. The Forum is unable	Comment welcomed and partly agreed. Overall, there is limited experience with clinical trials in PSC to form the basis for a clear recommendation. The Reflection paper now suggest a cut-off of ALP<1.5xULN. Alternative suggestions

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		to give a specific reduction in ALP and is concerned about the differences between cirrhotic and non-cirrhotic PSC. The Forum is providing evidence, as an appendix, on the importance of ALP variation and general challenges associated with the use of ALP.[13] Proposed change (if any):	are recommended to be addressed within a scientific advice procedure.
Lines 615-616	17	Statement: "…it is suggested that a responder-type evaluation based on the criteria of therapeutic response should be the basis, defining serological response as a reduction of ALP to 1.3xULN, or a combination of the reduction to 1.5-1.3xULN with an at least 40% reduction from baseline." Commentary: This statement is arbitrary and contentious, lacks an evidence base, and not supported by contemporary developments in the PSC biomarker field. Moreover, there is emerging data from the phase 2b Simtuzumab study in PSC, which shows wide per-patient inter- and intra-individual variability in serum ALP (Trivedi et al., 2019 - full paper available on request). This shows that a significant proportion of patients improve ALP values within these parameters over time, as per the natural	A responder analysis is recommended. See above on the cut-offs for ALP suggested.

[13] Trivedi PJ et al. Utility and Variability of Alkaline Phosphatase and the Enhanced Liver Fibrosis Score as Surrogate Outcome Measures in Primary Sclerosing Cholangitis: Prospective Evaluation of Controlled Clinical Trial Data. No date. *Submitted;* Provided in Confidence.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		history of the disease, without associated improvement in clinical outcomes. Proposed change: Please rephrase to: …it is suggested that a responder-type evaluation based on the criteria of therapeutic response should be the basis, defining serological response as a reduction of ALP. However, as a significant proportion of patients improve ALP values as per the natural history of PSC, specific thresholds cannot be recommended at this stage.	
Lines 616-619	5	Comment: Additional clarity is needed in the definition of histological response. Specifically, if the Nakanuma stage is to be used, there is ambiguity in the statement as to whether this only includes the fibrosis score or all components of the score. The Nakanuma system is based on a total score of 9 (based on features of both cholestasis and fibrosis, used to classify patients as stage 0-4), and theoretically a decrease of one stage by Nakanuma could be achieved by a reduction in markers of cholestasis without a reduction in stage of fibrosis. Yet, it is the stage of fibrosis that is likely most relevant to prognosis in PSC. Thus, staging histology by Ishak or Ludwig would seem a justifiable alternative to Nakanuma. Proposed change (if any):	Agreed. Reference to the Nakanuma staging specifically has been removed from the Reflection paper, one of accepted staging systems are now recommended.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
Lines 617-619	17	Statement: “For the histological evaluation – best to be based on the newer staging system according to Nakanuma – a similar responder-type evaluation is proposed. The response should be defined based on an at least 1 point improvement in the fibrosis stage.” Commentary: The Nakanuma pathology system comprises features of inflammation, cholestasis and fibrosis, and each component carries prognostic weight at different stages of disease. Certainly if advanced fibrosis were to be present, then the other two components may hold less prognostic utility compared to assessment earlier on in the disease course. If the meaning of this statement is to refer to the fibrosis component of Nakanuma only, then staging histology by Ishak (or equivalent) is equally justified (de Vries et al., 2014) Proposed change: Please rephrase to: For histological assessment, a similar responder-type evaluation is proposed. In terms of fibrosis specifically, the response should be defined based on an at least a 1-point improvement in fibrosis stage, by any of the validated scoring systems	Agreed. See above.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		in PSC: e.g. the Nakanuma, Ishak or Ludwig system. When assessing composite measures of disease histologically (including cholestasis or inflammation), evaluation is best made on the newer staging system according to individual components of the Nakanuma system.	
Line 618	11	**Comment:** As bile ducts related (ex. Cholangitis, cholangiocarcinoma) clinical outcomes in PSC are to be considered in addition to the liver cirrhosis decompensation related, potential treatment investigation would monitor both bile ducts and fibrosis progression. Even in the lack of obvious implications of Nakanuma scoring system for treatment effect monitoring the response assessment would combine both Nakanuma bile duct loss and cholangitis activity in addition to the fibrosis stage. **Proposed change (if any):** For the histological evaluation – best to be based on the newer staging system according to Nakanuma [55] – a similar responder-type evaluation is proposed. The response should be defined based on an at least 1 point improvement in the fibrosis stage and no worsening of bile duct loss and cholangitis activity grade. Stable disease (no worsening of fibrosis) could be used instead of 1 point fibrosis improvement, if adequately justified.	Partly agreed. Reference to the Nakamuna staging has been removed and stable disease mentioned as an alternative to improvement, if adequately justified.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
Lines 619-620	5	Comment: Additional clarity is needed on the conditions which would justify using stable disease (no worsening of fibrosis) as the definition of response. The population, drug mechanism, and proposed indication all combine to define the endpoints; thus, drugs targeting fibrosis vs. inflammation vs. bile duct metabolism likely require different endpoints. Proposed change (if any):	Partly agreed. Recommendation for co-primary endpoints supplemented by endpoints relevant to the disease address this issue.
Lines 619 – 620	17	Statement: "Stable disease (no worsening of fibrosis) could be used instead, if adequately justified." Commentary: This statement is ambiguous and lacks clarity. The study cohort, mechanism of action of any drug, and proposed indication for treatment, collectively inform the choice of study endpoint: CTIMPS targeting symptoms, bile acid flow, inflammation and fibrosis all require different endpoints.	Partly agreed. Current evidence does not allow for a more detailed recommendation. This could be addressed by individual sponsor within a scientific advice procedure.
Lines 628-629	5	Comment: "…other biomarkers (bilirubin, transaminases, but also e.g. ELF-test…)". We note the following evidence supports ELF as a surrogate biomarker for PSC clinical development: (i) ELF score is prognostic in patients with PSC[14]	Comment welcomed. Currently ELF is not considered sufficiently validated as a surrogate biomarker of efficacy for clinical trials in PSC

[14] Myers, R "Lessons Learned from the Simtuzumab Primary Sclerosing Cholangitis Program". PSC Forum 1. Washington DC. April 2017.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		(ii) A reduction in ELF >0.19 over 96 weeks appears to be associated with an improved prognosis in PSC.[9] (iii) Change in ELF is correlated with change in stage of fibrosis by both Scheuer and Ishak staging systems.[15] A change in ELF score of 0.52 corresponded to a change in Ishak score of 1 stage. While these data were derived from patients with a variety of liver diseases, these data provide support for a clinically meaningful change in ELF in patients with PSC. (iv) The ELF test has received FDA "breakthrough device designation" and is seeking FDA clearance.[16] Proposed change (if any):	
Line 628 - 629	17	Statement: "Biomarkers (bilirubin, transaminases, but also e.g. ELF-test…)" There is more evidence supporting the utility of ELF as a treatment efficacy endpoint in PSC, being the prognostic value seen in observational studies. Data from the Simtuzumab trial in PSC highlights that reduction ≥0.19 over 2 years associates with improved	Not agreed. Currently ELF cannot be accepted as a surrogate endpoint. As per the Reflection paper, multiple relevant secondary endpoint are recommended to support totality of evidence.

[15] Day, J et al. Derivation and Performance of Standardized Enhanced Liver Fibrosis (ELF) Test Thresholds for the Detection and Prognosis of Liver Fibrosis. JALM. 2019; 3(5): 815-826.
[16] Press Release: FDA Grants Breakthrough Device Designation to Siemens Healthineers Enhanced Liver Fibrosis (ELF™) Test. November 2018. https://www.siemens-healthineers.com/en-us/press-room/press-releases/elftest.html

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		prognosis. Moreover, unpublished data from the same study (Trivedi *et al.* submitted; paper available on request) shows that the intra-individual variability is very low compared to other proposed surrogate markers discussed in the reflection paper – for instance, serum ALP Proposed change: Please include more detailed discussion and references to the ELF score as a surrogate marker for treatment efficacy.	
635-639	9	Comment: Proposed change (added text: underlined, removed text: ~~striked through~~): As no effective treatment is currently available, ~~the acceptable comparator is regarded to be placebo~~placebo treatment as comparative treatment is acceptable. ~~Trial~~study ~~-~~duration is anticipated to be 2 years for the ~~interim~~intermediate endpoints, and should be up to 5 years for the demonstration of the long-term clinical outcomes. This proposed ~~trial~~study duration may need modification based on the mechanism of action, as well as anticipated magnitude of effects of new drug candidates, and the fact that usually, an event driven evaluation will be planned for.	Partly agreed. Trial duration is anticipated to be at least 2 years with longer-term follow-up upto 5 years for clinical endpoints. Based on the mechanism of action, trial duration may need to be modified.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
Lines 636-638	7	**Comment**: The duration of trial is heavily dependent on the natural history of fibrosis progression, coupled with the patient section, i.e. the natural history of fibrosis progression in non-cirrhotic patients with PSC, is generally slow compared to cirrhotic patient progression. **Proposed change (if any):** Trial duration is anticipated to be 2 years for the interim endpoints, and ~~should~~ ***could*** be up to 5 years for the demonstration of the long-term clinical outcomes. This proposed trial duration ~~may need modification based~~ ***is dependent*** on the mechanism of action, ***the selected patient population***, as well as anticipated magnitude of effects of new drug candidates, and the fact that usually, an event driven evaluation will be planned for.	Partly agreed and text has partly been amended in line with the proposal, 5-years duration is anticipated rather than a requirement
Line 660-665	16	Comment: **Patient Reported Outcomes in PSC** From a patient perspective there is a strong need for symptomatic treatment of fatigue, itch and pain and improved QoL	Partly agreed. A recommendation has been added to include PRO evaluating PSC related symptoms in any trial in PSC. IBD related QoL measures are less relevant and are not recommended, but can be included

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		Proposed change: Please add that Patient Related Outcome Measures should be part of any clinical trial in particular itch, fatigue, pain and PSC and IBD related QoL measures	
Lines 661-665	5	Comment: Patients strongly support the development and use of Patient Reported Outcome tools, both symptom-specific (pain, fatigue and pruritus) and PSC/ IBD-specific quality of life measures in clinical trials. Proposed change (if any):	Agreed. Development of PSC specific PROs are encouraged in the Reflection paper. IBD specific PROs are outside the scope of this guidance document.
Line 661-665 Patient perspectives from PSC Support	17	Comment: Patients strongly support the development and use of Patient Reported Outcome tools, both symptom-specific (e.g. pain, fatigue and pruritus) and PSC/IBD-specific quality of life measures, in clinical trials. Proposed change (if any):	See above
683-686	9	Comment: Proposed change (added text: <u>underlined</u>, removed text: ~~striked through~~):	The Reflection paper is now split into two separate documents with different text regarding safety. Comment partly agreed, wording revised. The need for liver biopsy is, however, still indicated as being useful, and to be

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		The underlying liver disease, as well as fluctuations and flares occurring during the course of clinical trials may hamper the evaluation of hepatic safety due to the overlap in accompanying symptoms. ~~, as well as~~ This also applies to the changes in the routine liver safety biomarkers used, such as transaminases, ALP, and bilirubin.	taken "whenever possible" which is considered to be sufficiently flexible.
691-692	4	**Comment**: - Suggest that potential identified Hy's law cases should be followed by adjudication to identify true Hy's law cases. - Suggest downgrading the need for biopsies to diagnose DILI and state that "in rare cases biopsies may be considered".	Agreed. Potential Hy´s Law cases and other serious liver adverse events should use expert adjudication. The need for liver biopsies has been downgraded, however, "may be considered" appears too "weak". Obtaining a liver biopsy "whenever possible" is more accurate.
712	4	**Comment:** - We propose to change 'hypercholesterolaemia' to 'dyslipidaemia'. The typical lipid abnormality in metabolic syndrome and NASH is increased triglycerides and low HDL-C. Cholesterol levels (total cholesterol or LDL-C) may not be increased. **Proposed change (if any):** …such as arterial hypertension, diabetes mellitus, severe obesity, and **dyslipidaemia** ~~**hypercholesterolaemia**~~ with the associated sequelae cardiovascular events…	Agreed. Implemented

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
712	10	**Comment:** We propose to change 'hypercholesterolaemia' to 'dyslipidaemia'. The typical lipid abnormality in metabolic syndrome and NASH is increased triglycerides and low HDL-C. Cholesterol levels (total cholesterol or LDL-C) may not be increased. **Proposed change (if any):** …such as arterial hypertension, diabetes mellitus, severe obesity, and **dyslipidaemia** ~~**hypercholesterolaemia**~~ with the associated sequelae cardiovascular events…	Agreed. See above
714-724	2	**Comment:** The reflection paper indicates that the principles of the "reflection paper on assessment of cardiovascular safety profile of medicinal products" (EMA/CHMP/505049/2015), could be considered applicable to NASH. However, it would be helpful to understand the expected sizing and duration of any cardiovascular outcome study. Also, if available, whether CV safety data generated from other insulin resistant populations (for instance a basket CVOT across NASH, T2DM, and increased IR with previous CVD) could be considered relevant. **Proposed change (if any):**	Partially agreed. However, due to missing relevant experiences, contrary to e.g. T2DM, a more general reference to the CV safety reflection paper only can be given.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
719-720	4	**Comment:** - Cardiovascular outcome trials would impose an unnecessary burden, and may hamper drug development in NASH. **Proposed change (if any):** … Further long-term natural history data, and **monitoring of cardiovascular safety during development of NASH drugs ~~long-term clinical trials in the field~~** are needed to draw a final conclusion. **Dedicated cardiovascular outcome studies are usually not expected** *[Schabel, liver forum 2015]***.**	Partially agreed. A dedicated request for CV outcome studies has not been part of the Draft Reflection paper. However, the principles of the "Reflection paper on assessment of cardiovascular safety profile of medicinal products" (EMA/CHMP/505049/2015) do apply. Paragraph has been reworded. CV safety needs to be addressed in adequate manner, therefore long-term studies will be needed, even without a dedicated need for outcome studies. Important part of life-cycle management, RMP
719-720	10	**Comment:** Cardiovascular outcome trials would impose an unnecessary burden, and may hamper drug development in NASH. **Proposed change (if any):** … Further long-term natural history data, and **monitoring of cardiovascular safety during development of NASH drugs ~~long-term clinical trials in the field~~** are needed to draw a final conclusion. **Dedicated cardiovascular outcome studies are usually not expected** *[Schabel, liver forum 2015]***.**	Partially agreed. See above
727	4	**Comment:** - In Paediatrics, as population will be small, long-term endpoints will be difficult to be measured and programs to be finalized,	Partially agreed, non-invasive markers should be validated, but the population under 10 years of age should be not automatically excluded. Exact cut off for age is unclear and

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		including adequate power and effect size to be measured. - The section should be open to non-invasive biomarker-derived endpoints in this population once a confirmatory proof had been provided in adults. - In addition as patients at ≤10 years or age are expected to have "only" milder forms of NASH pharmacological interventions are proposed to be excluded	therefore a range for lower cut-off is included (6-10 years). No need to change RP
736	2	**Comment:** Early studies limited to adolescents, reasonable to proposal prior to development of validated non-invasive diagnostic tools. **Proposed change (if any):** Consider range up to 24 years based on precedent in T2DM.	Not clear to me range 24 years represents adult population
737 - 742	4	**Comment:** - Biopsy in the vulnerable paediatric population at study entry and for evaluation should not be mandatory. **Proposed change:** As outlined above, the diagnosis of NASH is currently considered to require the conduct of liver biopsy with histological evaluation, and the conduct of clinical trials should be mainly based on	Partially agreed. Following validation, alternative non-invasive methods could be used, but we do not have validated test (eg elastography)

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		repeated biopsy results. The diagnosis itself is also based on histology in childhood/adolescence patients. However, **alternative non-invasive methods may be considered given that** the conduct of repeated biopsies in clinical trials requires increased awareness of potentially associated ethical and procedural problems when children are concerned. ~~**and t**~~**T**he need for **further development of** non-invasive outcomes in this population is therefore considered to be of even higher priority.	
737 - 742	10	**Comment:** Biopsy in the vulnerable paediatric population at study entry and for evaluation should not be mandatory. **Proposed change (if any):** As outlined above, the diagnosis of NASH is currently considered to require the conduct of liver biopsy with histological evaluation, and the conduct of clinical trials should be mainly based on repeated biopsy results. The diagnosis itself is also based on histology in childhood/adolescence patients. However, **alternative non-invasive methods may be considered given that** the conduct of repeated biopsies in clinical trials requires increased awareness of potentially associated ethical and procedural problems when children are concerned. ~~**and t**~~**T**he need for **further development of** non-invasive	See above comment

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		outcomes in this population is therefore considered to be of even higher priority.	
752-755	1	Comment: Given the current knowledge of the NASH disease in children, it would be reasonable to defer paediatric trials until the natural history in this population (< 12 years of age) is better described and relevant pharmacological studies can be designed including valid endpoints - preferably using non-invasive methods -, duration and selection of patients.	reflected in GL, having data on natural history in younger population is crucial, exact cut of age is unclear and validation of non-invasive methods is necessary
758-763	4	**Comment:** Could the Agency clarify the intention behind the timing in the conduct of clinical trials in children, and explain why it would be important to wait for adult data? This may delay development in children substantially – for instance in the case where a conditional approval has been granted based on surrogate endpoints in the adult population, would this mean a Paediatric Investigation Plan needs to be deferred until post-approval in adults?	Adult data are needed to have a sufficient basis to start trials in the more vulnerable population particularly depending on the products specific safety profile. Inclusion of paediatric adolescent patients is recommended to accelerate the process and sufficient number of adolescents could generate the data needed
Lines 791-793	5	Comment: The investigation of new compounds for children is strongly supported by patients, with consideration given to the features of autoimmune hepatitis that present in paediatric PSC. Proposed change (if any):	

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
Line 792-793 Patient perspectives from PSC Support	17	Comment: The investigation of new compounds for children is strongly supported by patients, with consideration given to the features of autoimmune hepatitis that present in paediatric PSC.	Noted and agreed, reflected in relevant paragraph
Lines 794-795	5	Comment: Patients strongly support the need for natural history data to be collected for paediatric PSC Proposed change (if any):	Agreed. The need for collection of natural history data in children is reflected in the Reflection paper.
Line 794-795 Patient perspectives from PSC Support	17	Comment: Patients strongly support the need for natural history data to be collected for paediatric PSC. Proposed change (if any):	agreed The need for collection of natural history data in children is reflected in the Reflection paper
794-796	9	Comment: It is acknowledged that natural history data on PSC in paediatric patients concern useful background information. It is supported that more such data are gathered. It is however not agreed that PSC studies should not be undertaken because of a lack of natural history data. Proposed change (if any): It is proposed to remove the section on natural history data.	Disagreement, natural history data on PSC in children will support performing relevant studies.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
Line 794-800	16	Comment: **Paediatric PSC** Since PSC has "a chronic, progressive and relentless course", children diagnosed with PSC are an important group to consider, since their young lives will be largely impacted by the course of the disease. Therefor the unmet need in this group is high. Proposed change Within this context of paediatric research for PSC, emphasize the need for data sharing on placebo treated patients and natural history data from paediatric patients to maximise the efficacy of research in paediatric patients.	Noted and agreed, reflected in the relevant paragraph
Lines 796 - 797	7	**Comment**: The Reflection Paper recognises the complexity of including adult patients with over-lapping syndromes, including active IBD, due to "potential interference	,No update needed. As the majority of paediatric PSC patients are over-lapping, IBD associated, inclusion them to trials will increase the numbers of available probands, NHD will help to classify different phenotype of paediatric patients

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		with the search for effective medication". However, the paediatric Reflection Paper recommendation, and PDCO preference, is to include patients with over-lapping syndromes in a mixed population. The source of the natural history data is also unclear. Clarity is requested. **Proposed change (if any):** N/A	
Lines 801-802	5	Comment: There is support for use of GGT as a biomarker in pediatric PSC due to elevated and fluctuating ALP levels in pediatric patients related to bone growth. Proposed change (if any):	Agreed, GGT and aspartate aminotransferase/palelet RI are mentioned in pediatric 5.3. paragraph. ALP is not only liver marker, changing with age. No need to reflect it in RP No need to change the text RP
Lines 801-802	5	Comment: All new drugs should take into consideration the unique pharmacokinetics in children. Adolescents are more similar to adults; however, non-compliance is more prevalent. Proposed change (if any):	Noted, well known no need to specifically mention in RP
Section 4.7.3	7	**Comment**: The Applicant would welcome further guidance on the conduct of paediatric PSC trials as PDCO have gained recent experience with various PSC PIP assessments.	PDCO does evaluate study plans PIPs, assessment of specific product data is not in the scope of PDCP. Applicants may ask for scientific advise for more detailed guidance

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
Lines 801-802		**Proposed change (if any):** N/A	
801-804	9	Comment: Proposed change (added text: underlined, removed text: ~~striked through~~): Besides the need to fully explore the PK profile in the respective population, currently no clear recommendations ~~there~~ can ~~currently no clear recommendations~~ be given with regard to the design of trials, and endpoints to be used. Consultation with the agency early in the drug development (scientific advice and submission of PIP) is therefore ~~advisable~~ advised.	Agreed
Overall	16	Comment: In this reflection paper it would be appropriate to state which of the comments are especially important from a Patient Expert Stakeholder Perspective, as is well stated and emphasized in the Report on stakeholder interaction meeting on the development of medicinal products for chronic non- infectious liver diseases (PBC, PSC, NASH)EMA/873574/2019 Proposed changes;	Comment considered. As regulatory guidance primarily addresses sponsors, however, including the perspective of other stakeholders such as patients is not within the primary scope.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		Please add the Patient Stakeholder perspective specifically within all appropriate paragraphs	
General comments	17	The International PSC Study Group recommendations form the basis for many recommendations in this reflection paper. However, this was not representative of the IPSCSG, rather informal discussion of 10 hepatologists – many of whom do not agree with the final statements made. Serum ALP, histology and transient elastography were all recommended in the aforementioned paper. A critical feature for any metric to be classified as a treatment efficacy endpoint is that a value change reflects a change in the disease. None of the endpoints listed, ALP, biopsy, or elastography, has been validated for this purpose. Proposed change: Wording needs to be clarified, to indicate that serum ALP values, histology, cholangiography and transient elastography have demonstrated risk stratification properties in observational cohort studies; but evidence to show that improvement in any of these parameters over time reflects an underlying change in the disease itself is lacking.	Comment welcomed. The recommendations of the Reflection paper now reflect the regulatory position above all with several references to support that, not limited to the recommendations of the PSC Study Group. Limitations of the endpoints are reflected in the reflection paper and thus co-primary endpoints are recommended as well as a number of secondary endpoints to support the primary outcome measures.

[i] Kleiner et al. Design and validation of a histological scoring system for nonalcoholic fatty liver disease. Hepatology 2005;41:1313-1321.

[ii] Vilar-Gomez et al. Weight Loss Through Lifestyle Modification Significantly Reduces Features of Nonalcoholic Steatohepatitis. Gastroenterology 2015: 149:367–378.

[iii] EASL-EASD-EASO Clinical Practice Guidelines for the management of non-alcoholic fatty liver disease. J Hepatol. 2016**:** 64(6):1388-402.

[iv] Viveiros et al. Hepatocellular carcinoma: when is liver transplantation oncologically futile? Transl Gastroenterol Hepatol. 2017; 2: 63.

[v] Angulo et al. Liver Fibrosis, but no Other Histologic Features, Associates with Long-term Outcomes of Patients With Nonalcoholic Fatty Liver Disease. Gastroenterol. 2015: 149:389–397.

[vi] Hagstrom et al. Risk for development of severe liver disease in lean patients with nonalcoholic fatty liver disease: A long-term follow-up study. Hepatology Communications 2017: https://doi.org/10.1002/hep4.1124

[vii] Schuppan et al. Determinants of fibrosis progression and regression in NASH. J. Hepatol. 2018: 68(2): 238-250

[viii] Friedman et al. Mechanisms of NAFLD development and therapeutic strategies. Nat Med. 2018 :24(7):908-922.

[ix] Harrison et al. NGM282 Improves Liver Fibrosis and Histology in 12 Weeks in Patients With Nonalcoholic Steatohepatitis. J. Hepatol. 2019: https://doi.org/10.1002/hep.30590