May 2019
EMA/873574/2019
Human Medicines Development and Evaluation

Report of the stakeholder interaction meeting on the development of medicinal products for chronic non-infectious liver diseases (PBC, PSC, NASH)

European Medicines Agency, London, 03 December 2018

Introduction

On 03 December 2018 The Agency held a one-day workshop on difficulties and opportunities for drug development in the field of chronic liver disease.

The aim of the workshop was to support the public consultation process on the draft Reflection paper on regulatory requirements for the development of medicinal products for chronic non-infectious liver diseases (PBC, PSC, NASH) published 16 November 2018 (EMA/CHMP/299976/2018). Speakers were invited based on their expertise in the field and the workshop was open for the public. Summarized statements are therefore reflecting views within the expert community and should not be attributed to the European Medicines Agency or its scientific committees.

Experts in hepatology (N=13) together with European drug regulators (N=13), 5 patient representatives, representatives from 25 pharmaceutical or consulting companies and non-EU regulators discussed and compiled current scientific evidence on study endpoints, suitable study populations, adequate trial designs and specifics of Paediatric drug development in the field. FDA representatives participated in discussions via virtual connection.

Each session of the workshop was dedicated to one of the three diseases Primary Biliary Cholangitis (PBC), Primary Sclerosing Cholangitis (PSC) or Non-alcoholic steatohepatitis (NASH) starting with state of the art presentations and followed by dedicated discussions. Taking into account the limited regulatory precedence discussions were held in an open setting but chairs were advised to guide their session with the help of a set of predefined questions reflecting current challenges in drug development. The list of questions had been prepared and circulated to all participants in advance of the meeting (see annex of this report). Key messages are summarized alongside these questions within the annex.

Presentations are available on the EMA webpage together with this report and the recorded broadcasting of the event.

Official address Domenico Scarlattilaan 6 ● 1083 HS Amsterdam ● The Netherlands
Address for visits and deliveries Refer to www.ema.europa.eu/how-to-find-us
Send us a question Go to www.ema.europa.eu/contact **Telephone** +31 (0)88 781 6000 An agency of the European Union

Discussions of this workshop and written comments awaited for the public consultation period of the draft reflection paper should lay the ground for drafting the final reflection paper.

Session 1: Primary Biliary Cholangitis (PBC)

The First session, dedicated to Primary Biliary Cholangitis, was chaired by Elmer Schabel from the German drug regulatory agency (BfArM) and Mari Thoern from the Swedish drug regulatory agency (Läkemedelsverket). The first speaker was Robert Mitchell-Thain Head of Education and Development of the PBC Foundation giving a patient perspective on patient priorities on treatment and unmet medical need.

Joachim Musaus from the EMA scientific and regulatory department gave an overview about the marketing authorisation precedence in PBC of Obeticholic acid and the involved regulatory and scientific challenges during the evaluation and the consecutive post authorisation requirements.

Prof Gideon Hirschfield from the Toronto Centre for Liver Disease gave a summary on definitions, natural history and current therapeutic interventions in PBC also touching upon Autoimmune Liver Diseases in general. The following two presentations on "Historical outcome parameters used in PBC and the search for potential alternatives" and "Potential trial designs and suitable study populations" were held by Prof Bettina E Hansen from Erasmus University Medical Center Rotterdam.

Summary of discussion:

It was acknowledged by experts and patient representatives that it would be an important goal to develop first line treatments superior to UDCA. It may also be considered for future trial designs to target individual patient's needs and profiles rather than necessary failure of treatment lines.

Due to UDCAs proven efficacy and safety trials vs. first line treatment may pose recruitment problems and acceptability challenges from patients, doctors, ethic committees and regulators. Therefore Sponsors may quantify for patients the risk not receiving UDCA for a defined time and outline potency of their drug to allow patients an informed decision. As first line registration on promising drugs is an important goal the reflection paper may motivate sponsors to put forward proposals how to perform these trials in a safe way for the patients and without jeopardizing the validity of the trial results.

Experts acknowledged that with more products authorised the employment of a placebo arm will be more and more challenging. Existing data should be reused. Databanks might be used particular on power analysis on new data and selection of patients in high / low risk. Furthermore historical controls may be considered in the future. An active controlled confirmatory trial measuring clinical outcomes would take about 8-15 years follow up and about 500 patients. Matched control arms (prospective registry or historical control) reusing gained knowledge could significantly reduce study-time and approval for patients. Downsides as selection bias, heterogeneity and quality bias could be addressed by stringent inclusion/exclusion criteria, weights to stabilize differences (IPTW) and adequate quality control measures (e.g. site visits / inspection, expert board review).

Experts noted the rather strict inclusion criteria of the BEZURSO trail (Corpechot C et al. N Engl J Med 2018) in which the placebo response could be reduced close to zero which may reduce trial size in view of demonstrating statistical significance.

Furthermore ALP and Bilirubin were considered powerful stratifiers of patient's risk. It should however be considered that requiring an ALP above 2 ULN whilst excluding patients without increase in Bilirubin (as recommended in the reflection paper) would mean to exclude about 91% of patients with PBC from the trials rendering the trial population less representative for the patient population in need of early treatment. Therefore it was suggested to focus on the right ALP cut as an inclusion criterion rather than excluding patients without Bilirubin elevations a priori.

However experts also acknowledged that inclusion criteria on Bilirubin elevation or ALP would require stratification for 3 groups with adequate sample size to avoid missing an effect in either of these groups. This may pose problems in reaching the adequate sample sizes in all groups. It was suggested to consider two trials covering different disease stages and duration (early stages on high risk vs. advanced).

Expert noted that the continued improvement in care pathways may lead to fewer patients with insufficient treatment response therefore overly strict inclusion criteria may also render clinical trials unfeasible.

The experts noted that PBC can be diagnosed and managed by clinicians reliably without histology. For entering a trial a combination of cholestasis and anti-mitochondrial / antinuclear antibodies would be sufficient to confirm diagnosis. For other patients liver biopsy would be required to confirm PBC.

Patient's acceptability of histological evaluation would likely be largely positive when used to validate non-invasive endpoints is an additional aim of the trial.

The expert noted that PBC is in itself not a homogenous disease. Biochemistry such as Bilirubin and ALP are changing in the course of the disease with Bilirubin increasing at later stages and ALP potentially falling at very end-stage disease. The biology of an individual drug response should be properly understood when designing the clinical trial and ALP might not always be the ideal outcome measure for any drug mechanism of action and PBC disease phenotype.

Patients with normal ALP but elevated AST or falling platelet counts may be at risk of progression but may not be eligible for trials using standard inclusion criteria; subsequent treatment may also not be reimbursed for these patients when trials are based only on ALP elevation.

For anti-fibrotic treatment designed to treat more the portal hypertensive phenotype an inclusion criterion on ALP may not be the most appropriate and transjugular portal pressure or liver stiffness might be considered as endpoint. As with other liver diseases, equally valuable may be the use of non-invasive serum markers of liver fibrosis.

The draft reflection paper proposes as interim outcome criteria (trial duration 1-2 years) an ALP reduction below 1.5 ULN and 40% reduction compared to baseline and normalized bilirubin. This was discussed in the context of cohort studies evaluating ALP and Bilirubin as surrogates. Experts noted that for ALP the lower the value the lower the risk of liver transplantation/death (Lammers et al. Gastroenterology 2014). Risk continues to be reduced visibly until 1xULN whilst 1.5 is not significantly different to 1.67. For Bilirubin the hazard ratio increases sharply over 1xULN but this increase is already visible at 0.71 which could make a cut of even below 1 reasonable.

The experts considered that ALP is in general a suitable surrogate as it follows the disease process being related to bile acids and because there appear to be little placebo response on ALP in PBC (changes may be around 10% in the course of a trial designed to show efficacy on surrogate endpoints). This may prove relevant when choosing a 15% or 40% cut off for ALP reduction. It needs also to be considered that the disease process in PBC is not linear. Also there are significant number of patients with lower ALP still at risk which challenges requiring a too ambitious cut of as primary efficacy criterion. The patient representative stressed that symptoms response in particular itch should be considered in development programs.

Experts noted that demonstration of efficacy on long term outcomes may need large populations and trial durations of 8-15 years confounding events by competing risks and therapies for other diseases. Having manifestation of cirrhosis as part of the confirmatory endpoints might reduce the necessary follow up time to about 5-8 years but feasibility issues need to be considered. Also the Rotterdam

Severe Disease Stage, could be explored as long term outcome criterion as it may capture events earlier.

Adjusted natural history control populations for long term evaluation might also be considered as means to provide confirmation based on secondary endpoints relying on serological and non-invasive evaluations. It would be optimal, but not necessarily easily deliverable, to have biopsy data outcomes being matched with other outcomes. Data from the POISE trial in phase 4 as soon as available could be used to test the quality of historical control to confirm their proof of concept / quality.

Session 2: Primary Sclerosing Cholangitis (PSC)

The second session, dedicated to Primary Slerosing Cholangitis, was chaired by Prof Michael Trauner from the Vienna University of Medicine and Peter Szitanyi Prague University of Medicine.

The first speaker was Martine Walmsley Chair of the UK Charity, PSC Support giving a patient perspective focussing on patient's symptom burden, life expectancy and potential patient friendly trial designs.

Prof Douglas Thorburn from the Royal Free London NHS Foundation Trust gave an overview about definition, natural history and the lack of approved therapeutic interventions in PSC.

Prof Cyriel Ponsioen from the Amsterdam University Medical Centers presented on the topic "Currently proposed endpoints in PSC: the search for reliable surrogate outcome parameters" touching in particular upon interpretability and variability of alkaline phosphatase and liver biopsy.

The following presentation on "Potential trial designs and suitable study populations" was held by Prof Michael Trauner from the Vienna University of Medicine who gave various examples emphasizing the right selection of study population and endpoints impacting on the suitability of the trial design.

Prof Henkjan Verkade from the University Medical Center Groningen rounded up the presentation part of this session with the topic "The need for paediatric developments in PSC, trial duration and endpoints" giving a comparison on paediatric and adult disease course and characteristics and outlining usefulness and limitation of biochemical response surrogate endpoints.

Summary of discussion:

The experts noted that the selection of patients should allow the occurrence of relevant events in the population included. 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).

Fluctuating biochemistry and cholangitis flares complicate inclusion of patients in clinical trials. Once a flare is subsided wide screening windows and re-screening of patients would be appreciated in principle to have drugs approved for the majority of patients. However, whilst for patients with too early disease ALP may not be the best biomarker too late / severe disease will come with dominant strictures and endoscopic treatments. Whilst patients with advanced fibrosis without bile duct stenosis are hard to find the action of a drug may be best shown in F3 and F4 stages as earlier stages are relatively long stable and might therefor not be ideal for demonstrating an effect on clinical endpoints or fibrosis progression. Populations for evaluation could be enriched by ALP as correlation with disease progression and / or with ELF score/Imaging.

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.

In phase 2 small duct and autoimmune hepatitis may be acceptable but as they are associated with a better outcome, in phase 3 patients with more severe disease should be enrolled.

The experts noted that despite its limited evidence of efficacy UDCA is commonly used to treat PSC patients due to lack of effective treatments. Requiring patients to stop UDCA for clinical trials would present a feasibility issue for clinical trials even though its use as concomitant treatment may slow down the development of effective drugs. Patients' acceptance and feasibility of a trial without concomitant UDCA may however be influenced by promising drug candidates.

The experts noted that current clinical trials could show drug impact on biochemistry surrogates but not on clinical outcomes such as death, liver transplant Cholangiocarcinoma due to challenges to power the studies adequately in this rare disease with a natural annual occurrence of these solid clinical endpoints of only 3-6%. Therefore surrogate endpoints are clearly needed.

ALP was used in all relevant drug trials. Based on unpublished data presented it also appears to have limited variability in the natural course of the disease. The experts considered this cholestasis parameter to be a good predictor and stratification tool/inclusion tool but more data on its sensitivity to meaningful change is needed on this not validated surrogate.

Histology (e.g. based on Nakanuma staging, but not necessarily excluding other staging systems) to assess directly the disease activity was considered having the potential to be a robust surrogate endpoint for clinical trials (level 2 endpoint for progression/prevention of cirrhosis). Inter observer variability was considered good, natural variability rather stable (i.e. either stable or worsening) (Muir 2017, Angulo 1999) and sample variability left vs. right liver lobe (Olsson 1995) but could be accommodated with ultrasound guided biopsy.

Whilst biopsy is not needed for PSC diagnosis, experts agreed that combining both endpoints in the absence of a single convincing surrogate would be the right strategy. It was suggested to analyse both endpoints as composite for phase 2A to find potential divergence but to apply a co-primary setting to phase 2B and phase 3 trials to maximise reassurance of the benefit.

The experts considered the histologic degree of fibrosis in non-cirrhotic patients the best predictor for long term clinical outcomes in paediatric patients even though it may need anaesthesia. It is however not usually used in paediatric patients and may profoundly affect recruitment. Fibroscan and liver stiffness is easier to perform but the obstructive component of the disease presents a significant confounder. MRI techniques cannot be considered sufficiently validated at the moment.

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.

Experts reiterated that even though the prevalence of PSC might be higher than currently assumed due to subclinical disease it is difficult to power studies adequately. Clinical endpoints arise over decades (about 5.1 per 100 patient years for death and OLT (liver transplant) and 1.4 per 100 patient years for HPB cancer).

Furthermore PSC is a heterogeneous disease and sample size calculation is challenging because of the unknown sensitivity of current surrogates. A cut off change needed to show a difference in clinical effect remains undefined and it remains unknown if studies can be adequately powered.

Therefore a "totality of the data review" might be the last option if the primary endpoint is missed. However, to avoid inflating the type 1 error by cherry picking symptomatic outcomes would clearly need to correlate with the other outcomes. Reversal of cirrhosis, decompensation, bile duct specific endpoints and cholangitis episodes as well as fibrosis and ELF scores could all feed into a totality of the data review as well as PROs.

As IBD activity may drive the biomarkers it could be made part of the trial design and part of a holistic outcome review.

Effective treatment is currently only surgical (liver transplantation, resection, cholecystectomy, ERCP etc.). Whilst clinical trials have so far shown improvements in biochemistry without relevant impact on clinical outcomes liver transplantation is increasing in frequency due to changing approaches towards indication and registration and is the only treatment changing the natural history of the disease significantly.

In the course of a long term clinical trial many confounding therapies need to be taken into consideration. This includes in particular IBD treatment, UDCA (not effective but impacts on ALP and study recruitment), Antibiotics (impacting Cholangitis)

The experts noted that paediatric PSC is considered a seronegative condition based on histological diagnosis. Accordingly outcomes are different than in adults. Whilst the majority of cases are diagnosed in the second decade the disease also exists in younger children (4-8 years). Higher overlap with AIH (PSC-AIH) exists in the paediatric population as well as a higher prevalence of small duct disease. With its overlaps on AIH the disease may have a slightly milder course with longer periods of liver transplant-free survival (about 50% within 20 year (at the age of about 30) (Weismueller 2017) but not all reports confirm a more benign prognosis (Denau et al 2017, 2018).

The experts noted that for patients at this age IBD manifestations and autoimmune disease can be of higher relevance whilst, as disease progresses, cirrhosis complications may become more prominent. However, cirrhosis at the first presentation of PSC disease occurs in a considerable fraction of paediatric patients.

As for candidates of surrogacy in children a GGT below 50 U/L and /or GGT reduction of over 75% after 1 year has been shown to predict a favourable 5 year outcome (event free survival) (Deneau2018). However UDCA lowers GGT without being associated with better outcome on event free survival (Deneau 2018) and consortia delineating the natural history bridging paediatric and adult medicine are needed.

Experts felt that taking into account the mechanism of action of a drug both reversal and prevention of fibrosis / cirrhosis could be considered as endpoints in in trial designs as well as portal hypertension and biliary complications as they are associated with over 50% liver transplantation within 5 years (Deneau 2017). It was further argued that PROs/OoL evaluation should be given consideration in the reflection paper with regards to both paediatric and adult PSC and their relevance even if a drug is not changing the course of the disease significantly.

Session 3: Non-alcoholic steatohepatitis (NASH)

The third session, dedicated to Non-alcoholic Steatohepatitis, was chaired by Prof Deirdre Kelly from the Birmingham Children's Hospital and Peter Mol from the Dutch drug regulatory agency (Medicines Evaluation Board).

The first speaker was Yvonne Gray (ERN-RARE-LIVER) giving a patient perspective on her disease course and interaction with the healthcare system.

Chrissi Pallidis from the EMA Paediatric Medicines Office gave an overview of current PDCO approaches on paediatric investigation plans in particular on trial design and stating age for enrolment in clinical trials outlining also histological differences in paediatric and adult disease.

Prof Frank Tacke from the University Hospital Aachen gave a summary on definitions, natural history and current therapeutic interventions in NASH also touching upon multidisciplinary treatment approaches and the role of liver biopsy for staging and prognosis.

In the following presentation with the title "Outcomes in NASH trials: From histology combined with "hard outcomes" to less invasive reliable surrogates (including biomarkers)?" held by Prof Laurent Castera from the Hôpital Beaujon / Paris Diderot University state of the art knowledge on non-invasive tests and their usefulness for enriching study populations and as surrogate endpoints for treatment response was given before Prof Quentin Anstee from Newcastle University and the Freeman Hospital, Newcastle-upon-Tyne, UK addressed the topic of "Potential trial designs and suitable study populations"

This block of presentations was rounded up by Prof Piotr Socha from the Children's Memorial Health Institute Warsaw with the topic "Paediatrics: Population in need, clinical trial duration and endpoints"

Summary of discussion:

Experts noted that about 20% of patients show an improvement on fibrosis by 1 stage after about 1 year and similar number shows worsening. Genetic factors, age but also alcohol and lifestyle are driving progression of the disease as regression is driven by lifestyle, diet and use of coffee. To control better for such confounders in clinical trials failure of dietary treatments prior to inclusion could be documented by history taking and standardized nutrition counselling and fitness trackers were suggested. The expert further argued that patients may get the diet advice they would normally get to have trial results representative for a real world setting. Currently there is little evidence on efficacy of drugs on NASH but documentation of stable treatment on relevant co-medication should be documented before inclusion.
The expert agreed on the need to include patients with advanced liver fibrosis in clinical trials and noted that the main causes of death in earlier stages of liver diseases (i.e. F2 and F3 groups) were CV events and non-hepatic cancer whereas the main causes of death (or indication for liver translation) once disease had progressed to F4 are hepatic decompensation and/or HCC.
The experts noted that fibrosis stage F3 encompasses a very broad group in terms of histological severity and degree of hepatic fibrosis. Due to the high variability in spontaneous progression and regression a requirement to show an improvement of 2 stages in fibrosis significantly would reduce the placebo response but would also be a very exacting therapeutic effect to demonstrate within the limited duration of a 12/18-month clinical trial. Considering the variability particularly for stage 2 patients the relevance of a 1 point improvement may need to be defined. However, experts were concerned about setting the bar too high for anti-fibrotic treatments as this may lead to false negative trials and suggested to include a 2 stage improvement rather as secondary endpoint.

Whilst regression of cirrhosis (i.e. a 1 stage decrease in fibrosis) was welcomed by the experts as an interim endpoint 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. In that sense stopping the trigger of the disease (NASH) was considered a valid primary outcome even if a simultaneous effect on fibrosis regression could not be demonstrated within the limited 12/18-month timeframe of a clinical trial.

Treating of the underlying disease finally stabilizes the fibrosis but it can currently not be assumed that treating fibrosis only may or may not undermine steatohepatitis. This may require combination therapy

Treating NASH on holding fibrosis is reasonable but pure antifibrotics may pose concerns with regards to Hepatocellular carcinoma.

The experts noted that the best indicator of long term adverse outcomes might be portal Hypertension but there are technical challenges to measure this in bigger trials. For early phase 2 studies improvement of portal hypertension (HVPG <10mmHG) and / or MELD was suggested as study endpoint and for late phase late phase 2 and 3 change in fibrosis score together with HVPG and MELD. Biopsy could be performed as early as at W12 but the timing would need to depend on the MoA of the drug and the anticipated speed of response.

Interim analysis with non-invasive test like MRI-PDFF or MRE elastography could be done in early phase 2 studies (some preliminary data suggests that a 30% relative reduction in MRI-PDFF was associated with a 2 point NAS improvement) but underlying data are still too limited to make these tests sufficiently reliable for pivotal trials.

Expert noted that over 80% of paediatric patients will stay with isolated fatty liver (Harrison S, Clin Gastro Hep 2015) and complications will mainly become apparent in adult age. Selected groups of patients mostly seen by specialised centres may present with fibrosis and progress to cirrhosis I childhood. Genetic factors are connected to the development of NAFLD (even though association with severity is not clear) and should also be considered in paediatric trials (Wood KL, BMJ Open Gastroenterol. 2015). A 8 years cut off was considered reasonable for inclusion of children in clinical trials. Patients could be stratified according to puberty and gender.

There are less clinical indications for biopsy in children which make it more challenging to capture histology in clinical trials and to use it as a primary endpoint. Assessments on advancing liver disease would therefore be ideally done based on good studied surrogates who are unfortunately not yet available. Outcome parameters currently used as clinical trial endpoints are steatosis measured by ultrasound, decrease of ALT, and histology (as too limited biopsy data available) as secondary endpoint. Steatosis measured MRI-PDFF might be the most promising surrogate endpoint for treatment response assessment in liver disease but would need further validation. It was however noted that for MRI elastography sedation is needed in children less than 6 years of age.

Whilst duration of clinical trials in paediatric NASH varies between 4 months to 96 weeks an it was considered to be reasonable to measure effect on steatosis after 4-6 month whilst for trials to show impact on fibrosis a minimum of 1 year would be needed. It was noted that the currently most applied trial design is 12 months double-blind placebo controlled with PK evaluation and primary endpoint on histology.

Experts argued that children usually present with lesser degrees of fibrosis and inflammation and furthermore histopathological differences in paediatric NASH make extrapolation from adults difficult.

A range of biomarkers should be captured alongside biopsy to support validation and to mitigate sampling error. Experts suggested for Phase 3 studies interim analysis at 12 -18 months for conditional MA with a primary outcome based on histology. Long term evaluation to validate histology

against clinical outcomes (event driven) should follow. Studies should always include a variety of biomarkers for validation against histology.

The expert noted the long asymptomatic course of the disease and the lack of validated PROs in NASH

The experts acknowledged that reliably capture the drugs safety profile is challenging in NASH due to the high comorbidity and symptom load in these patients not specific to the liver. Apart from a general safety evaluation including cardiovascular safety a focus should be made according to the mechanism of action of the drug. In phases 3 and 4 event free survival needs to be captured as well as development of cardiovascular disease for safety reasons (such as lipid profiles) as well as DILI (drug induced liver injury).

Annex I to the stakeholder interaction meeting on the development of medicinal products for chronic non-infectious liver diseases (PBC, PSC, NASH)

Session 1: Primary Biliary Cholangitis (PBC)

Questions:

1. Is it relevant to consider the development of alternatives to first-line UDCA treatment? Is it a relevant aim to demonstrate superiority to UDCA for such compounds?

- New treatments would ideally be developed to target individual patient's needs rather than following treatment lines.
- Sponsors may quantify patients risk of not receiving UDCA for a defined time and outline potency of their drug in order to foster acceptance from patients doctors and regulators in first line trials.
- The reflection paper may motivate sponsors to develop trial designs which allow for valid results without jeopardizing patients safety.

2. Would you support the conclusion that long-term efficacy outcome evaluation in PBC (for the first-line therapy) is not possible? If it is: How can it be done, and which endpoints should be used?

- Demonstration of long term efficacy may need trial durations of 8-15 years. Apart from feasibility confounding events and diseases would need to be considered.
- Cirrhosis as long term event might reduce the time to about 5-8 years.
- Adjusted natural history controls could be considered for long-term evaluation.

3. Strict inclusion criteria are proposed for clinical studies in PBC (in the setting of add-on treatment to UDCA), with the intention to enrich the studied population with a "high-risk" population in order to be able to demonstrate effects on hard endpoints. Whereas this increases the "assay sensitivity" of these trials, at the same time, the external validity of the trial may be reduced. Is there a clear

preference in the trade-off between sensitivity and external validity? Are there also potential problems with feasibility (recruitment of patients with this "late-stage" disease to be expected?

- Strict inclusion criteria may reduce placebo response and trial size
- ALP and Bilirubin are considered powerful risk stratifiers
- Requiring ALP above 2 and increased Bilirubin excludes 91% of PBC patients
- Inclusion may focus on optimal ALP cut of rather than on abnormal Bilirubin
- Two trial designs (early stage on high risk vs. late disease) may be considered

4. The reflection paper notes that histology is not part of routine testing in and for diagnosis of PBC. Do you agree that histology should not be compulsory in clinical trials in PBC? Do you also agree to the proposal of the reflection paper that histology (both before and on treatment) should be obtained as frequently as possible.

- PBC can be diagnosed and managed by clinicians reliably without histology
- A combination of cholestasis and AMA / ANA antibodies would be sufficient for diagnosis in clinical trials
- Patients acceptability for liver biopsy in a trial setting would be positively influenced if trials are used to validate non-invasive endpoints

5. Are there other potentially relevant inclusion criteria to be considered for the PBC population (such as non-invasive biomarkers (e.g. ANAs, ELF-score, AST/ALT ratio, elastography, "prediction models" (e.g. GLOBE score or UK-PBC score)) on which risk assessment and inclusion of patients (and stratification) can be based?

- PBC is an inhomogeneous disease with Bilirubin rather increasing and ALP decreasing at later stages.
- Depending on the MoA of the drug ALP might not always be the ideal outcome measure.

6. Are the proposed interim outcome criteria in the "add-on setting" (derived from the benzafibrate trial) realistic, or are they too "strict"?

- ALP is considered a suitable surrogate as it follows reasonably the disease process.
- Placebo response around 10% during the course of the trials should be considered for the efficacy threshold.
- The lower ALP the lower is the risk for hazard. The risk reduces visibly until 1 ULN. Whilst for Bilirubin increase for 0.71 ULN is visible above 1 ULN the risk increases sharply.

7. Is the demonstration of efficacy with long-term outcome data relevant in the add-on setting, and if yes, how should this be accomplished (in the situation when placebo treatment might hardly be possible for prolonged periods)

- Demonstration of long term efficacy may need trial durations of 8-15 years. Apart from feasibility confounding events and diseases would need to be considered.
- Cirrhosis as long term event might reduce the time to about 5-8 years.

- Adjusted natural history control could be considered for long-term evaluation.

Primary Sclerosing Cholangitis (PSC):

1. Have the inclusion criteria been defined sufficiently strong in order to result in a population suitable to observe relevant changes for the intermediate endpoints, as well as number of events for the long-term outcomes (e.g. exclusion of small duct disease; presence of additional risk factors required)?

- Considerable fluctuation in biochemistry and cholangitis flares in PSC may need to be considered when defining screening windows.
- Patients with advanced fibrosis are needed to demonstrate clinical effect but prevalence of bile duct stenosis may result in enrolment issues.
- Small duct diseases and AIH should not be enrolled in confirmatory trials due to their association with a better outcome.
- Enrichment based on ALP, ELF score, Imaging could be considered.
- Drugs affecting inflammation rather than fibrosis may perform better in IBD patients but more discussion is needed how to deal with IBD flares in clinical trials.

2. According to the reflection paper a, for the "intermediate endpoint" a combined use of histology evaluation and ALP changes are regarded to represent an acceptable surrogate. A reduction of ALP to 1.3xULN, or a combination of the reduction to 1.5-1.3xULN with at least 40% reduction from baseline are proposed. Is this sufficiently realistic, or too strict?

- Due to the failure previous trials on clinical endpoints a strong surrogate is needed.
- ALP is a good predictor and stratification tool/inclusion tool; more data on its sensitivity to meaningful change is needed.
- Histology has the potential to be a strong surrogate; sample variability could be accommodated with ultrasound guided biopsy.
- In the absence of a single validated surrogate combining ALP based outcomes with histology is a valid development strategy.
- Assessing these endpoints as co-primary may maximize reassurance of the benefit for confirmatory trials.
- Degree of fibrosis in non-cirrhotic paediatric patients is currently the best predictor for long term outcomes but a requirement on liver biopsy may impact on recruitment. Non-invasive methods are currently not sufficiently validated.
- PROs, in particular fatigue and pruritus should be part of clinical trials.

3. Long-term endpoints: The document opens up to an evaluation based on the "totality of data". Does this allow a sufficiently reliable conclusion on the change of disease course?

- Due to the low annual incidence of relevant clinical events necessary sample sizes appear to be too large to be practically feasible.

- Reversal of cirrhosis, decompensation, bile duct specific endpoints and cholangitis episodes, fibrosis and ELF scores and PROs could feed into a totality of the data review.
- Symptomatic and clinical outcomes would need to clearly correlate.
- Type 1 error needs to be controlled.

4. Children and adolescents:

4.1. Is extrapolation from adults possible?

- PSC in children is considered a seronegative condition based on histological diagnosis; outcome parameters are different.
- With its overlap to AIH the disease may have a slightly milder course with longer times of transplant free survival.
- A current candidate of surrogacy in children is on GGT reduction but there are doubts on its relationship between drug-induced reduction and clinical outcomes. Endpoints on prevention/reversal of cirrhosis and biliary complications and PROs/QoL should be given consideration.
- Consortia delineating the natural history bridging paediatric and adult medicine are needed.

4.2. Do you agree that a "mixed population" with "overlap syndromes" should be included into clinical trials?

- Higher overlap with AIH and small duct disease exists in paediatric PSC with longer transplant free survival.
- Due to this overlap and the slow progression of liver disease the autoimmune components may be more relevant to AIH patients in early stages.

Non-alcoholic steatohepatitis (NASH):

1. Inclusion criteria:

- **Inclusion of patients with documented failure of weight loss/dietary treatment: How strong should the evidence be that patients have undergone unsuccessfully other treatments?**
- **Inclusion criteria for patients with NASH cirrhosis: If a sufficient grade of NASH is not or no longer present, how strict should the inclusion criteria be (currently, it is requested in the reflection paper that all of the following should be: historical biopsies with presence of unequivocal NASH, a high likelihood of NASH based on non-invasive testing (biomarker and imaging), and presence of associated co-morbidity (e.g. obesity with T2DM))?**
- **Should patient reported outcome measurements (particularly on fatigue) play a major role in regulatory assessment in early and / or late stage NASH?**
  - Failure of dietary treatments prior to inclusion could be documented by history taking and standardized nutritioning counselling could be applied as in real world setting

- Patients with advanced fibrosis should be enrolled in clinical trials
- There is currently a lack of validated disease specific PRO developed to current regulatory standards for NASH

2. Non-cirrhotic NASH:

- **Could a decrease in fibrosis stage without any or only minor influence on the inflammatory activity, be appropriate as measurement for treatment benefit? The current proposal requests a more "stringent" approach (at least 2 stage improvement without worsening of NASH). Is this necessary, and, is it also "realistic"?**
  - Due to spontaneous progression and regression of the disease showing 2 stages of fibrosis improvement would significantly reduce the placebo response
  - The clinical relevance for a 1 stage improvement (particularly for stage 2 patients) would need to be defined
  - A 2 stages improvement could be included as secondary endpoint to avoid false negative trials in anti-fibrotic treatments
  - Treating fibrosis may not undermine steatohepatitis. The risk of hepatocellular carcinoma would need to be considered
  - Due to spontaneous pro- and regression of the disease showing 2 stages of fibrosis improvement would significantly reduce the placebo response
  - The clinical relevance for a 1 stage improvement (particularly for stage 2 patients) would need to be defined in long-term follow up studies
  - A 2 stages improvement could be included as secondary endpoint to avoid false negative trials in anti-fibrotic treatments
  - Treating fibrosis alone with a 'pure' antifibrotic agent may not ameliorate steatohepatitis. In this situation the risk of hepatocellular carcinoma would need to be considered during follow up of patients

3. Cirrhotic NASH:

- **What is the current evidence that reversal of cirrhosis in NASH relevantly influences the final prognosis? Are there data from other chronic liver diseases available to sufficiently support the "surrogacy" of the reversal of fibrosis as valid endpoint? Are these fully transferrable to NASH (e.g. as in the case of a purely anti-fibrotic compound having almost no influence on the inflammation; See lines 303-308 of the reflection paper)?**
- **Would you agree to the "dichotomization" of the cirrhotic population in an "advanced population" for which no "intermediate endpoint" evaluation strategy is necessary (but clinical decompensation events, and liver related (or all-cause) death could be the primary endpoint straightforwardly), and a "less advanced" cirrhosis population for which intermediate endpoints might be needed?**
- **What are potential "intermediate" endpoints both in the non-advanced as well as in the advanced population apart from "reversal of cirrhosis"?**

- Interim analysis with non-invasive test like MRI-PDFF or MRE elastography could be done in early phase 2 studies but underlying data are still to limited to make these tests sufficiently reliable for pivotal trials
- Primary outcomes for late phase 2 and 3 trials could be based on improvement of fibrosis, portal hypertension (HVPG <10mmHG) and MELD score.
- Biomarkers should be captured alongside biopsy to support validation and to mitigate biopsy sampling error, if a consistent direction of change was observed in histology and biomarkers.

4. Children and adolescents:

- **Population to be included in clinical studies: Severity and age groups. From what age do patients with NAFLD/NASH need to be studied?**
- **Endpoints to be evaluated: Histology deemed necessary at present, but liver biopsy can be a risky procedure. Other (non-invasive) methods to evaluate efficacy?**
- **Duration of studies: Many applicants for NASH PIPs propose 1 year. Is this sufficient?**
- **Is extrapolation from adults possible?**
  - A 8 years cut off was considered reasonable for inclusion of children in clinical trials. Patients could be stratified according to puberty.
  - Feasibility of biopsy in children is challenging and non-invasive surrogates are still needed
  - MRI-PDFF might be the most promising surrogate endpoint for treatment response assessment in liver disease with regard to degree of steatosis but would need further validation.
  - A 8 years cut off was considered reasonable for inclusion of children in clinical trials. Patients could be stratified according to puberty.
  - Feasibility of biopsy in children is challenging and non-invasive surrogates are still needed
  - MRI-PDFF might be the most promising surrogate endpoint for treatment response assessment in liver disease but would need further validation.

Additional issues discussed

Symptomatic treatment

- **Are developments aiming at symptomatic treatment only (without disease "modification") a viable and necessary option? How far would the influence on specific (PBC- or PSC- related "outcomes" need to be documented, in order to exclude detrimental effects on the overall disease course?**
- **Are the proposed developments sufficiently described?**
- **If only one symptom is targeted: Is it realistic to develop relevant treatments also for other pruritic disease?**
  - The expert noted the long asymptomatic course of the disease and the lack of validated PROs in NASH.

Safety:

- **Are the specifics of liver safety evaluation in the targeted populations (= patients with underlying liver disease) relevantly and sufficiently described. Are there any suggestions for further relevant criteria to be evaluated?**
- **Is the evaluation of cardiovascular safety (e.g. exclusion of CV risk increase) relevant for all NASH products or should it be dependent on the mechanism of action and/or early (pre-clinical and clinical) trial results?**
  - Reliable capturing of the drugs' safety profile is challenging due to the high co-morbidity.
  - A focus of safety evaluation should be made based on the MoA of the drug.
  - In phase 3 and 4 trials survival, development of CV disease and DILI should be capture for safety.