



1 15 November 2018
2 EMA/CHMP/299976/2018
3 Committee for Medicinal Products for Human Use (CHMP)

4 **Reflection paper on regulatory requirements for the**
5 **development of medicinal products for chronic non-**
6 **infectious liver diseases (PBC, PSC, NASH).**
7 **Draft**

8

Draft agreed by Gastroenterology Drafting Group	October 2018
Adopted by CHMP for release for consultation	15 November 2018
Start of public consultation	16 November 2018
End of consultation (deadline for comments)	31 August 2019

9
10

Comments should be provided using this [template](#). The completed comments form should be sent to GastroenterologyDG@ema.europa.eu

11
12

Keywords	Non-alcoholic steatohepatitis (NASH), Primary Biliary Cholangitis (PBC), Primary Sclerosing Cholangitis (PSC), clinical trial design, endpoints, conditional marketing authorisation, unmet medical need, liver biopsy
----------	--

13



14 Reflection paper on regulatory requirements for the
15 development of medicinal products for chronic non-
16 infectious liver diseases (PBC, PSC, NASH).
17

18 Table of contents

19	1. Introduction	3
20	2. Scope.....	3
21	3. Legal basis and relevant guidelines	3
22	4. Recommendations	3
23	4.1. General considerations	3
24	4.2. Non-alcoholic steatohepatitis	5
25	4.2.1. Short characterisation of the disease.....	5
26	4.2.2. Selection of patient populations.....	5
27	4.2.3. Study design and endpoints	7
28	4.3. Primary biliary cholangitis	10
29	4.3.1. Short characterisation of the disease.....	10
30	4.3.2. Selection of patient populations.....	11
31	4.3.3. Study design and endpoints	12
32	4.4. Primary sclerosing cholangitis	14
33	4.4.1. Short characterisation of the disease.....	14
34	4.4.2. Selection of patient populations.....	15
35	4.4.3. Study design and endpoints	15
36	4.5. Trials for the symptomatic treatment (PBC and PSC):	17
37	4.6. Safety considerations.....	17
38	4.6.1. Safety in PBC and PSC	17
39	4.6.2. Safety in NASH.....	18
40	4.7. Children and adolescents.....	18
41	4.7.1. NASH in children and adolescents	18
42	4.7.2. PBC: Children and adolescents	20
43	4.7.3. PSC in Children and Adolescents.....	20
44	5. References	20
45		

46 **1. Introduction**

47 Chronic, non-infectious liver diseases are a medical field of high unmet medical needs. At the same
48 time, the specifics of the diseases create major challenges for the development of new medicinal
49 products. This reflection paper restricts the current regulatory approach to 3 different disease entities
50 primary biliary cholangitis (PBC), primary sclerosing cholangitis (PSC), and non-alcoholic
51 steatohepatitis (NASH) for which recent efforts are undertaken to bring new medicinal products to the
52 market.

53 It is anticipated that many of the problems raised and potential solutions described in this reflection
54 paper, may be transferrable to other chronic liver diseases.

55 **2. Scope**

56 As a reflection paper, this guidance document provides a high level description of the requirements for
57 drug development in the field. For all three disease entities dealt with in the paper, the regulatory
58 experience with the licensing of new medicinal product is limited. Therefore, this paper aims at a
59 preliminary definition of development strategies only, which, in the case of several successful MAAs
60 occurring in the future, will have to be refined, and may finally be superseded by full guidance
61 documents.

62 **3. Legal basis and relevant guidelines**

63 This document should be read in conjunction with the introduction and general principles and part I
64 and II of the Annex I to Directive 2001/83/EC as amended. Applicants should also refer to other
65 relevant EU and ICH guidelines (in their current version) and regulations, especially the following:

- 66 • Reflection paper on assessment of cardiovascular safety profile of medicinal products
67 (EMA/CHMP/50549/2015)
- 68 • Reflection paper on the use of extrapolation in the development of medicines for paediatrics.
69 (EMA/189724/2018)
- 70 • Guideline on clinical development of fixed combination medicinal products.
71 (EMA/CHMP/158268/2017)
- 72 • ICH E9(R1) Draft Addendum on estimands and Sensitivity Analysis in Clinical Trials
73 (EMA/CHMP/ICH/436221/2017)

74 **4. Recommendations**

75 **4.1. General considerations**

76 Chronic liver disease is a slowly developing process, and many patients do not develop relevant
77 disease sequelae, and/or symptoms over even a considerable time of observation, and the
78 development of end-stage liver disease may be a process of years, if not decades. All three diseases
79 under consideration will be difficult to be studied for long-term outcomes over a reasonable time span
80 (the term “long-term outcome” is used in the following for events such as liver transplantation and
81 death, as well as clinical events of decompensation of liver cirrhosis which are otherwise also termed
82 “hard outcomes”).

83 An acceptable regulatory strategy for companies developing new agents in the disease area, may be to
84 look for intermediate endpoints for which a reasonable assumption for the prediction of long-term
85 outcomes can be made. These reasonable assumptions are usually based on associations with regard
86 to risk factors for the long-term outcomes in observational natural history cohorts and the biological
87 plausibility attributed (the term “intermediate endpoint” will be used throughout in the following for
88 events otherwise also termed “interim” or “surrogate” endpoint).

89 Strictly speaking, however, such endpoints are not validated in the sense that positive changes for the
90 surrogate as well as the long-term outcome have repeatedly and consistently been demonstrated for
91 therapeutics. Due to the largely unmet medical need in the field, a strategy to obtain an early approval
92 of new compounds based on these intermediate endpoints, however, could be considered. This
93 strategy will require the confirmation of efficacy (and safety) of the compound after approval (including
94 availability on the market) documenting the effects on long-term outcomes. Such a strategy could be
95 acceptable as long as an unmet medical need can still be reasonably concluded¹ If such a strategy is
96 intended, however, the evidence at the time of evaluation of the intermediate endpoints has to be such
97 that it allows the conclusion of a positive risk-benefit ratio independent from the presence of an unmet
98 medical need.

99 In the situation of unmet medical need, the use of placebo as comparator would be the only acceptable
100 way to demonstrate efficacy. However, the authorisation of new substances may trigger a revision of
101 the acceptance of placebo as comparator in the future.

102 The acceptance of the mentioned regulatory strategy has to be regarded to be a case by case issue.

103 For the three disease entities, this document will display endpoints which can currently be considered
104 acceptable surrogates for the manifestation of end-stage liver disease (the intermediate endpoints), as
105 well as those deemed suitable for the confirmation of these surrogates (the long-term endpoints). The
106 specifics to confirm (and thus validate) these surrogates will be dealt with in the relevant chapters.

107 These intermediate endpoints (as well as the long-term endpoints) are currently partly or mainly based
108 on the histological evaluation of liver biopsies. Liver biopsy and histology have been widely criticized
109 for sampling error and intra- and inter-observer variability². However, potential non-invasive methods
110 do currently have insufficient, and especially insufficient disease specific, validation data available, and
111 therefore histology is in most cases still regarded to be the state of the art for the diagnosis, and
112 especially for the follow-up of the course of the diseases, in particular for the purpose of clinical trials.

113 Liver biopsy, however, is also unwanted due to its patient burden, invasiveness, and the associated
114 risks of morbidity³ and potentially even mortality.

115 Therefore, this reflection paper also calls for the further development of non-invasive methods to
116 replace liver histology in the future, be it serological markers, or imaging methods. It is therefore
117 recommended that future applicants should use development programmes aimed at producing
118 evidence for the approval of new medicinal products also for the further validation of novel methods,
119 intending to replace histology in the future.

120 Possible targets of estimation that define treatment effects of interest in the three disease entities are
121 also considered.

122 **4.2. Non-alcoholic steatohepatitis**

123 **4.2.1. Short characterisation of the disease**

124 Non-alcoholic steatohepatitis (NASH) is considered the progressive phenotype of non-alcoholic fatty
125 liver disease (NAFLD), which itself is the most prevalent chronic liver disease worldwide with an
126 estimated prevalence in the Western world of around 25%⁴⁵, and it is estimated that about 20% of
127 these suffer from NASH⁶. The progression is related to the development of liver cell stress, subsequent
128 inflammation, and fibrosis with the potential development of cirrhosis, and end-stage liver disease.
129 NASH is also a relevant risk factor for the occurrence of hepatocellular carcinoma. NASH associated
130 cirrhosis/end-stage liver disease is expected to represent the highest share of patients referred for
131 liver transplantation in the future⁷. NAFLD as well as NASH are associated with other comorbidities and
132 risk factors such as obesity, arterial hypertension, diabetes mellitus type 2 (T2DM), atherogenic
133 dyslipidaemia, and others. The disease – although genetic factors have also been identified⁸ – is
134 thought to be the consequence of hyperalimentation, and has been regarded to be the hepatic
135 manifestation of the so-called metabolic syndrome⁹.

136 From a diagnostic point of view, the diagnosis of NASH is one of exclusion (involving the exclusion of
137 relevant alcohol intake, and infectious and non-infectious liver disease) as well as positive confirmation
138 of the features by liver biopsy and histology, the latter relating to the pathognomonic features of
139 steatosis, hepatocellular ballooning, lobular inflammation, and fibrosis¹⁰. Although health-related
140 quality of life may be impaired¹¹, symptoms do not play a relevant role in (non-cirrhotic) NASH. Also
141 the awareness with regard to the disease and of the associated risks is poor¹².

142 The natural history of NASH has not been fully elucidated, and further efforts are needed to clarify
143 important aspects, e.g. overlap of progression and regression. The risk of progression to end-stage
144 liver disease is largely related to the baseline fibrosis grade¹³. The progression of fibrosis is estimated
145 to be slow, and progression of 1 fibrosis stage is estimated to occur at a mean of more than 7 years
146 (7.7 years; 95% CI 5.5-14.8 y)¹⁴¹⁵¹⁶.

147 **4.2.2. Selection of patient populations**

148 The usual principles of the selection of study population, such as being representative of the target
149 population, and being well balanced with regard to demographic characteristics and co-morbidities are
150 of course applicable to NASH. The more specific requirements are dealt with in the following:

151 As mentioned above, the diagnosis of non-alcoholic steatohepatitis is a diagnosis of mixed exclusion of
152 other relevant diseases, as well as a positive diagnosis which is mainly reliant on liver biopsy with
153 histology. Histology is currently considered to be the gold standard for finally securing the diagnosis,
154 as well as determining the severity of disease, and is also recommended as part of clinical practice. A
155 selection of patients on the basis of symptoms is usually not possible, and the (long-standing)
156 presence of features of the metabolic syndrome can only be used as a trigger to identify potential
157 study participants.

158 The risk of progression to end-stage liver disease, liver transplantation and death has been
159 demonstrated to be independently associated with the stage of liver fibrosis only, with only minimally
160 increased risk for stage 1 patients¹⁷. Fibrosis stage 1 patients are therefore currently only
161 recommended for inclusion in therapeutic trials in NASH for exploratory purposes.

162 Therefore, the “natural” selection of patients with an unmet need for treatment in NASH relates to
163 patients with (fibrosis) stages 2-4 NASH.

164 Inclusion of patients in fibrosis stages 2 and 3 should additionally be based on the disease activity /
165 grading because developments of regression and progression may overlap, and a (albeit univariate
166 only) risk of progression has also been associated with higher degrees of ballooning and inflammation.
167 The patient population should be included based on a valid grading system for NASH with minimal
168 requirements for the presence of cell stress (ballooning), as well as inflammation (lobular
169 inflammation). The NASH-CRN (Non-alcoholic steatohepatitis clinical research network) histological
170 scoring system currently appears to be the best validated and most widely accepted system ¹⁸. A total
171 NAS (NAFLD activity score) of at least 5 appears acceptable but a score of 4 can be accepted as well, if
172 it is not based on a high contribution of the steatosis grade alone and minimal requirements for
173 relevant ballooning and lobular inflammation are fulfilled (scoring of at least 1 in each of these 2
174 components). Although the NASH-CRN grading system is the recommended grading system, patients
175 may also be included based on potentially other grading systems for NASH, provided the validation of
176 respective grading systems is substantiated.

177 In patients with manifest cirrhosis (=fibrosis stage 4), the presence of such a rigorous minimal grade is
178 less critical, because the risk of (clinical) progression is thought to be high based on the presence of
179 cirrhosis alone. Nevertheless, in so-called burnt-out NASH cirrhosis or patients initially diagnosed with
180 cryptogenic cirrhosis¹⁹, if definite NASH is not present, all of the following should be available in order
181 to make the diagnosis NASH sufficiently likely: historical biopsies with presence of unequivocal NASH,
182 a high likelihood of NASH based on non-invasive testing (biomarker and imaging), and presence of
183 associated co-morbidity (e.g. obesity with T2DM).

184 Patients with decompensated cirrhosis represent a particularly vulnerable subset of patients. A relevant
185 amount of mechanistic, as well as clinical efficacy and safety data on an investigational compound may
186 be required before the inclusion of such patients into clinical trials. Due to the fact of increased risks of
187 biopsies in this population, historical biopsies (with presence of cirrhosis) together with symptoms of
188 decompensation may be used as inclusion criteria in this population.

189 The multi-stakeholder composed Liver Forum has recommended that histology should always be
190 available, also in early clinical trials ^{20,21}, and inclusion of patients should always be based on
191 histological evaluation (grading and staging). Deviations for exploratory clinical trials, e.g. using
192 imaging methods, or biomarkers (or a combination of those) only, are possible if based on sound
193 scientific principles, for which the uncertainties can be quantified and later stage trials be planned
194 accordingly. ²²

195 The positive influence of weight reduction on NASH has clearly been demonstrated²³. Therefore, before
196 inclusion of respective patients into clinical trials for NASH, it is recommended that patients should
197 have undertaken at least one unsuccessful attempt with weight-reducing diet. Co-morbidities, such as
198 e.g. diabetes, hypercholesterolaemia, and hypertension should adequately and stably be treated at the
199 time of inclusion.

200 Important factors to be considered in all populations are the presence of co-morbidities (e.g. diabetes),
201 and stratification for these factors could be advisable to allow a balanced evaluation of these
202 covariates.

203 In summary, for the purpose of therapeutic clinical trials, NASH may be considered in three broad
204 categories:

- 205 a. Definite NASH based on histology with demonstration of $NAS \geq 5$ (or $NAS \geq 4$ with all components of
206 at least 1) and fibrosis stage 2-3
- 207 b. Compensated NASH-cirrhosis based on histology with fibrosis stage 4 and NASH diagnosis based
208 on either $NAS > 5$ (or $NAS \geq 4$ with all components of at least 1) or the availability of historical

209 histology proving NASH, non-invasive tests pointing to NASH (serological markers, imaging), and
210 relevant co-morbidity risk-factors (obesity and type 2 diabetes mellitus (DM))

211 c. Decompensated NASH Cirrhosis: Presence of historical biopsy data showing unequivocal NASH as
212 well as cirrhosis; symptoms of decompensation.

213 **4.2.3. Study design and endpoints**

214 The natural history of NASH is assumed to end with the manifestation of cirrhosis in the liver, and the
215 subsequent development of portal hypertension and its sequelae, and decompensation of liver
216 function, which ultimately results in liver associated death, or liver transplantation. Because NASH is
217 also associated with a multitude of risk factors for cardiovascular disease (hypertension, obesity,
218 atherogenic dyslipidaemia, and type 2 diabetes), a relevant proportion of patients will also be prone to
219 causes of death other than liver related ones, mainly cardiovascular.

220 The “natural” long-term endpoint in clinical trials for NASH would therefore be the combination of all-
221 cause mortality, liver transplantation, and the manifestation of decompensation (variceal bleeding,
222 ascites, encephalopathy etc.).

223 **Stage 2 and 3 fibrosis:**

224 The time to manifestation of long-term outcomes is currently largely unknown, and reasonably sized
225 trials in patients with the earlier stages of disease (such as fibrosis stage 2 and 3) with the primary
226 aim to demonstrate an effect on survival free of liver transplant and decompensation events might be
227 unfeasible. Therefore, efficacy endpoints reflecting a substantial increase in the risk of disease
228 progression (to the events described) are needed. The histological diagnosis of cirrhosis has been
229 proposed to represent such an endpoint, and is regarded to be an acceptable surrogate and can
230 therefore be part of the long-term endpoints. Similar arguments have been accepted for a model for
231 end-stage liver disease (MELD) score above the threshold of 14. The long-term outcome for the
232 demonstration of efficacy in NASH is therefore proposed to be a composite endpoint with the
233 components all-cause death, decompensation of liver disease (with a complete listing), as well as
234 (histological) diagnosis of liver cirrhosis and MELD>14

235 However, due to feasibility issues in long term studies and the unmet medical need in NASH, an earlier
236 evaluation of efficacy, with an overall shorter duration of clinical trials is warranted and intermediate
237 endpoints reasonably predicting the long-term outcome have been advocated²⁴.

238 Acceptable intermediate endpoints would consist of two composite endpoints to be evaluated at the
239 individual patient level:

240 1. The resolution of NASH – with the presence of any grade of steatosis, no ballooning, and only
241 minimal (grade 1) lobular inflammation and – at the same time – no worsening of the stage of
242 fibrosis.

243 2. The improvement of fibrosis by at least 1 stage without any worsening of NASH (no worsening of
244 ballooning and lobular inflammation, a 1 grade change in steatosis may be acceptable).

245 Efficacy in these two composites should be demonstrated in co-primary fashion, meaning that both will
246 have to independently demonstrate a statistically significant and clinically relevant difference to
247 placebo. This requirement is thought to take account of the uncertainties associated with a strategy to
248 account for the long-term outcomes later.

249

250 **Stage 4 fibrosis (NASH cirrhosis)**

251 In liver disease where cirrhosis has already been manifested, the use of the above mentioned long-
252 term composite is not possible. An acceptable endpoint for patients with already existing cirrhotic liver
253 disease at inclusion would therefore consist of the composite of all-cause death and liver
254 decompensation events. However, because liver cirrhosis represents a wide spectrum of disease, it is
255 currently unclear whether such an endpoint is feasible. When the intention is to use this long-term
256 endpoint in the cirrhotic population, the study population should be enriched with patients with
257 advanced cirrhosis.

258 In case the need to use intermediate endpoints in this population is identified, a reasonable endpoint
259 for the general non-decompensated population, could intuitively be the reversal of cirrhosis (e.g.
260 defined as "improvement of liver cirrhosis to non-cirrhotic liver disease (1 or more point improvement
261 in fibrosis stage)"). At this point of time, however, the data available to demonstrate that reversed
262 cirrhosis does indeed also reverse or influence the final prognosis substantially, is considerably less
263 profound than the association shown for progressing disease. Such a trial would therefore need the
264 substantiation of the claim that the prognosis of reversed cirrhosis is similar to the prognosis of
265 (untreated) earlier stages of fibrosis in progressive disease (e.g. from other disease areas such as
266 chronic infectious liver disease; ie Hepatitis C or B). Moreover, this endpoint should be appropriately
267 backed by additional, secondary outcomes, based on non-invasive markers of disease (imaging
268 techniques, determination of liver stiffness, biomarkers) as well as the available (descriptive) data on
269 decompensation events, liver transplantation, and death.

270 In a situation when relevant proportions of patients with advanced cirrhosis are included, an
271 acceptable endpoint would be the occurrence of decompensation events since the prognosis for
272 patients with decompensated cirrhosis is markedly worse than those with compensated cirrhosis²⁵.

273 Other potential endpoints (e.g. lowering of MELD score below a certain threshold, or of the HVPG below
274 10 mm Hg²⁶) are also possible based on specific justifications.

275 The need for addition of post-marketing observations with regard to the manifestation of end-stage
276 liver disease and death (=the long-term outcome observation) will in these cases be assessed based
277 on the overall substantiation of the clinical usefulness of the primary endpoint used and the data on
278 the secondary outcomes.

279 In the special group of decompensated cirrhosis, a therapeutic effect should be demonstrated based on
280 the endpoint all-cause mortality/survival. Liver related death, and liver-related death/ transplantation
281 could be supportive endpoints.

282 **Additional considerations on mode of action**

283 As a simplified pathophysiology of NASH, it has to be assumed that the liver cell toxicity caused by the
284 overload in fat causes inflammation, which itself is the final trigger of fibrosis development. Therefore,
285 it has been assumed that the appropriate target of medicinal products would be mechanisms
286 preventing fat toxicity and/or decreasing inflammatory activity, which would finally lead to beneficial
287 effects in fibrosis. However, new substances primarily targeting the development of fibrosis are
288 currently under development, and it is therefore considered important to reflect whether a decrease in
289 fibrosis stage without any or only minor influence on the fat accumulation in the liver, liver cell stress
290 (ballooning) and inflammation could be appropriate as treatments and benefit patients in the long
291 term. This is considered an uncritical question as long as long-term endpoints are used as objectives in
292 clinical trials. However, in case an intermediate endpoint strategy is followed, the above mentioned two
293 composite endpoints may be impossible to be used due to the fact that a resolution of NASH endpoint
294 is not within reach of such compounds. If an intermediate endpoint strategy is used in such

295 compounds, it is currently recommended to use a stronger endpoint denoted as a composite at the
296 individual patient level such as “fibrosis regression of at least 2 stages without worsening of NASH”, in
297 which stage 2 fibrosis patients would need to achieve complete resolution of fibrosis, and patients with
298 stage 3 would need to regress into a stage associated with only minimally increased risk for disease
299 progression (“no worsening of NASH” could be defined similar to the above).

300 This requirement similarly applies to patients with cirrhosis. Although it might be possible to show that
301 the reversal of cirrhosis benefits patients with other liver diseases (e.g. with data from Hepatitis C or B
302 trials) in the long-term, there remain important questions with regard to the ongoing primary insult
303 (the fat associated necro-inflammation), which cannot be solved with the data from infectious liver
304 disease, because these have – in their vast majority – been derived from patients with sustained viral
305 response, and thus an almost complete suppression of the inflammatory insult.

306 **Duration of trials**

307 The currently published phase 2 data for substances under development have mostly evaluated parts
308 of the above proposed endpoints only. Therefore, uncertainty exists with regard to the duration of
309 trials, both in terms of the time needed for interim evaluation with the intermediate endpoints, as well
310 as for the time needed to show relevant effects on the long-term composite endpoint. As a general
311 rule, a two-year interim evaluation, and a 5-year final evaluation may be considered appropriate.
312 However, this can be modified with factors like size of the trial, activity of the investigational
313 compound, patient characteristics, and the requirements with regard to statistical rigor. The final
314 evaluation would be expected to be usually planned with an event-driven evaluation, and therefore, a
315 fixed duration may not be appropriate to be planned with.

316 **Target of estimation (estimand)**

317 The scientific question(s) of interest, i.e., what the trial seeks to address and ultimately, the target of
318 estimation (estimand) should be specified. The trial planning, design, conduct, analysis and
319 interpretation must be aligned with the estimand. It is referred to ICH E9(R1) Draft Addendum on
320 estimands and Sensitivity Analysis in Clinical Trials (EMA/CHMP/ICH/436221/2017).

321 In order to determine the appropriate strategy for a trial in NASH, a full review of potential intercurrent
322 events is necessary. Relevant intercurrent events expected are those associated with almost all clinical
323 trials, such as treatment discontinuation and use of additional medication. Contrary to other fields of
324 development, the use of rescue medication may – for the time being – not be relevant because no
325 specific treatments are available, but could become more relevant in the future. However, a change in
326 background medication (including excessive life-style changes with weight loss, or uptake of relevant
327 alcohol intake) may relevantly affect the outcome, and may need to be considered.

328 For the intermediate endpoints, the outcome regardless of the occurrence of intercurrent events is of
329 primary interest (i.e. a treatment policy strategy discussed in the addendum). Therefore, data with
330 regard to the outcomes of interest should be collected independently from the occurrence of an
331 intercurrent event. Data that is nevertheless not collected, for example in case the endpoint is based
332 on liver biopsy and the biopsy is missing or not evaluable, results in a missing data problem with
333 regard to subsequent statistical inference.

334 Choices made regarding statistical analysis, including the handling of missing data, must be aligned
335 with the target of estimation. Considering a patient with missing data as a non-responder usually
336 results in a conservative estimate of the treatment effect with regard to the question of primary
337 interest, but alternative handling of missing data may also be acceptable (possibly taking occurrence of
338 intercurrent events and the reason for missing data into account). For example, for patients on
339 treatment who refuse biopsy, replacing missing data using multiple imputation based on response

340 probability of patients still on treatment (possibly taking additional covariates into account) could be
341 considered.

342 The outcome regardless of occurrence of intercurrent events is also of primary interest for the hard
343 endpoints. Aiming at a complete follow-up for the outcome events is of particular importance as
344 patients that are not completely followed are likely to have a different prognosis than patients who
345 complete the study, implying that censoring such patients is probably informative and leads to bias. As
346 a biopsy during the follow-up is only scheduled if there is a high likelihood of a cirrhosis (e.g. based on
347 surveillance with non-invasive methods such as fibroscan), non-performance of a scheduled biopsy
348 should be considered as an event.

349 **Combination treatment**

350 It has been advocated, based on the results of currently available phase 2 trials, that a satisfactory
351 treatment of NASH might only be possible, if new investigational compounds are combined, ideally with
352 a combination of two different principles of action, such as e.g. anti-fibrotic, and anti-inflammatory²⁷.
353 Whereas such a strategy can be followed from a theoretical point of view, potential applicants should
354 move forward carefully with such development programmes in a situation with no established therapies
355 available.

356 The main considerations of the Guideline on clinical development of fixed combination medicinal
357 products (EMA/CHMP/158268/2017) will have to be taken into account when embarking on a co-
358 development, and ultimately on a fixed-dose combination product.

359 The expectations from the regulatory side would be that the combination is based on valid therapeutic
360 principles, but also that for each of the substances involved, the contribution to the therapeutic effect
361 is demonstrated.

362 It will also be expected that the properties of the single substances are fully explored and described
363 either before or during the development of the combination treatment. Also, referring mainly to other
364 disease areas, it will be expected that either a second line treatment is investigated, which has to
365 include the establishment of a definition of an insufficient response to a standard treatment (or at least
366 one of the combination partners), or – in case an initial combination treatment is aimed at – the
367 definition of a patient group with a very high risk of progression.

368 **4.3. Primary biliary cholangitis**

369 **4.3.1. Short characterisation of the disease**

370 Primary biliary cholangitis (PBC), previously known as primary biliary cirrhosis²⁸, is a chronic, slowly-
371 progressive autoimmune cholestatic liver disease²⁹. The disease is mainly diagnosed in female patients
372 with a ratio of about 10:1. PBC is a rare disease, with incidence and prevalence reported at variable
373 rates (0.33 to 5.8 100,000/year for incidence; 1.91 to 40.2 per 100,000 for prevalence). Whereas an
374 increase in the incidence has been reported for the last decades, newer global data also indicate
375 changes in the diagnosis and course of the disease (irrespective of treatment) with older ages at
376 diagnosis, and slower progression over time³⁰. The pathogenesis of the disease is not fully
377 understood, with environmental, infectious, and genetic predispositions, and with an inflammatory
378 process targeting biliary epithelial cells, and resulting changes of bile-acid metabolism, and
379 enterohepatic circulation being involved.

380 The disease is characterised with cholestasis, the presence of specific antibodies (AMA and ANA), and
381 histologic evidence of chronic granulomatous, lymphocytic small bile duct cholangitis. The disease
382 course is progressive ultimately leading to the presence of cirrhosis, and end-stage liver disease. In

383 most patients the progression is slow. In an important subgroup of typically younger patients
384 progression can be more rapid. Usually, PBC is diagnosed on the basis of incident, routine liver
385 transaminase testing at an early stage of disease, without relevant symptoms being present. In
386 addition to the symptoms associated with end-stage liver disease (where present) patients can
387 experience significant systemic symptoms throughout the disease course. Fatigue and pruritus are the
388 most prominent of these symptoms and can be debilitating.

389 PBC is diagnosed in clinical practice based on the findings of careful history taking, exclusion of other
390 immune-mediated diseases, and the presence of specific findings in imaging, and finally, serological
391 tests, including ANA and AMA. Liver biopsy with histology – according to the current European Practice
392 Guideline – is only recommended in cases with ongoing unexplained cholestasis³¹.

393 PBC is the only disease in this reflection paper, for which a standard therapy is available.
394 Ursodeoxycholic acid (UDCA) is currently the established standard of treatment and has been
395 introduced in the 1990s. More recently, obeticholic acid has been licensed in 2016 for the “treatment of
396 primary biliary cholangitis (previously also known as primary biliary cirrhosis) in combination with
397 ursodeoxycholic acid (UDCA) in adults with an inadequate response to UDCA or as monotherapy in
398 adults unable to tolerate UDCA”.

399 **4.3.2. Selection of patient populations**

400 The similar general principles of the selection of study population are also applicable to PBC (see
401 4.2.2). The specific requirements concern the following:

402 Because a standard first-line therapy option, plus a second-line add-on therapy option are available,
403 the inclusion of an adequate patient population depends on the intended place in therapy of the
404 investigational agent.

405 Trials trying to establish a new first-line compound in the disease can include, both, newly diagnosed
406 and/or untreated patients, as well as patients already treated with UDCA. In the latter case, these
407 patients could either have normal liver chemistry (=full responders) (including ALP and bilirubin), or,
408 more adequately, have documented partial response to the agent, which is, however, below the
409 threshold for “unsatisfactory response in one of the established criteria for UDCA response in the
410 literature. The choice of the response criteria will have to be justified. Treatment naive patients should
411 be included based on an unequivocal diagnosis of PBC associated with an (at least) minimal increase in
412 the serological markers of the disease, especially ALP with or without (conjugated) bilirubin elevation,
413 allowing for relevant improvements.

414 Whereas for early-stage trials, the omission of a histological evaluation, including screening as well as
415 endpoint evaluation is considered acceptable, the availability of a baseline histology evaluation (as well
416 as follow-up evaluation, see below), is highly recommended.

417 Trials aiming to justify the add-on treatment on top of UDCA will need to include patients based on an
418 insufficient response to UDCA (for the present time, wherein UDCA is the prevailing therapy). A variety
419 of options has been proposed to define such a population, including the so-called Barcelona, Paris-I, -
420 Toronto, Rotterdam, as well as Paris-II criteria³²³³³⁴³⁵. However, all these criteria were set-up in order
421 to define a population having the best prognosis at long-term follow-up, and not in order to determine,
422 which of these might delineate a population at the highest risk of progression, and thus be most
423 suitable for additional therapy. An analysis of these different proposals, however, has shown that the
424 likelihood to develop endpoints (such as cirrhosis, decompensation events, and liver transplantation
425 and death, see below) during the course of a trial largely depends on the strictness of these inclusion
426 criteria³⁶. It is therefore recommended that the more strict criteria are chosen, allowing only those

427 patients into the trial which have still a relevant alteration of the serological markers of PBC. Currently,
428 best appears to be the combined use of the $ALP \geq 2 \times ULN$, and bilirubin $> 1 \times ULN$ despite an at least 1
429 year therapy with UDCA at the standard recommended dose (10-15 mg/kg b.w./day). Additional
430 criteria with regard to transaminases, albumin, GGT, or Mayo risk score may be applied, if adequately
431 justified.

432 Trials in the add-on-setting may also include patients not tolerating the standard treatment with UDCA.
433 However, it is expected that these form a minority of patients only in these trials. Nevertheless,
434 consistency of the results needs to be demonstrated.

435 **4.3.3. Study design and endpoints**

436 **First-line therapy as alternative to UDCA**

437 Ursodeoxycholic acid (UDCA) has an established efficacy and an acceptable safety record in the disease
438 to be treated. Therefore, the development of alternatives to the first-line therapy have to take into
439 account the level of efficacy, as well as of safety of the standard therapy.

440 The conduct of such trials may include in addition to a direct comparison to UDCA, also a (potentially
441 small; e.g. based on unequal randomisation) placebo arm for assay sensitivity purposes in case non-
442 inferiority will be the aim of such trials. While it is acceptable to demonstrate non-inferiority to the
443 established treatment, as well as an acceptable safety profile for licensing, considering the properties
444 of the current standard of care with moderate efficacy and relatively good and established safety
445 profiles, it might be necessary to aim at superiority in such trials in order to allow a more clear positive
446 conclusion on risk-benefit, especially in case the safety profile does not allow a conclusion on a similar
447 level of acceptability as for UDCA.

448 For an intermediate endpoint evaluation strategy in the first line treatment of PBC, as an alternative to
449 UDCA, one of the previously mentioned response criteria (Barcelona, Paris, Rotterdam etc. criteria; see
450 Chapter 4.3.2) can be used, depending on the included population.

451 In case the untreated population is mainly or solely used, the most obvious endpoint would be the
452 composite of the normalisation of the relevant serological markers, mainly ALP and (total) bilirubin (as
453 composite at the individual level). Any deviation from this stringent definition should be justified.

454 The trial duration would need to be at least 1 year, with extended (controlled) follow-up (see below) to
455 be planned.

456 An evaluation of all potential long-term outcomes is considered to be hardly possible in this population,
457 which would be expected to have a high rate of normalisation of the serological markers at the end of
458 the (primary) observation period, and thus have an even delayed further development of disease
459 deterioration. The necessary follow-up treatment documentation would therefore need to demonstrate
460 a prolonged superiority (or at least non-inferiority) for at least 2 years (potentially to be submitted
461 post-licensing) in the serological (Interim) endpoints, supported by an adequate battery of secondary
462 evaluations, based on non-invasive imaging, additional biomarkers, as well as histology. The clinical
463 relevance of these endpoints should be substantiated.

464 **Add-on therapy to UDCA**

465 The reduction of total bilirubin, as well as for ALP (including % reductions and reductions under certain
466 thresholds) have previously been used and accepted as primary endpoints in trials in the add-on
467 setting. These endpoints have to be regarded to be acceptable intermediate outcomes of efficacy in
468 PBC, because currently, it has only been demonstrated for the natural history as well as for UDCA, that
469 the reduction of these two serological markers leads to an overall improved outcome with regard to the

470 development of end-stage liver disease, decompensation, liver transplantation and death ³⁷. Hence, an
471 endpoint based on these serological markers is considered an adequate intermediate strategy, which,
472 however, for new compounds, would need to be supported by additional long-term outcomes. The
473 choice of adequate thresholds for the definition of response would need to be adapted to the chosen
474 inclusion criteria, but usually, the most clear-cut thresholds close to normalisation would be expected
475 to be evaluated. Previously, the criteria of ALP<1.67xULN, ALP decrease of at least 15%, as well as
476 (total) bilirubin ≤ULN have been thought to be acceptable. However, more stringent definitions of
477 response are advocated here, with the ALP criterion being at least ALP<1.5xULN with an at least 40%
478 decrease, and total bilirubin ≤ULN. Additional criteria with regard to transaminases, GGT, and/or Mayo
479 score may be added, depending on the respective inclusion criteria ³⁸.

480 Because the validity of these intermediate endpoints is not fully established, it would usually be
481 expected that long-term outcome data with respect to the histological manifestation of cirrhosis, the
482 decompensation of cirrhosis, MELD score above a threshold defining a high risk of liver related death
483 (e.g. above 14), as well as liver transplantation and death should form the basis for a long-term
484 follow-up evaluation of efficacy. However, the availability of obeticholic acid as first add-on therapy in
485 PBC on the market makes the conduct of placebo-controlled trials with these long-term outcomes more
486 complicated, and adequate escape procedures may be necessary to be implemented into, in order to
487 allow the ethically acceptable conduct of, such trials in the future.

488 Due to the fact that it is currently not known whether such trials on long-term endpoints finally turn
489 out to be feasible in the disease, the fact that the disease is rare, and the development of later stage
490 disease is slow, the applicants will also have to take care that the best possible evidence with regard to
491 secondary evaluations is also available at the point of interim data evaluation (for the serological
492 endpoint). This should include, but is not restricted to, non-invasive measurements of liver
493 fibrosis/stiffness, serological markers of inflammation and liver damage, as well as histology, including
494 the staging and grading of the disease. The latter item would need to be handled with caution due to
495 the fact that a fully validated histological scoring system for the disease is not available. Historically,
496 Ludwig and Scheuer's classifications, as well as METAVIR have been used in this context, and specific
497 scoring systems seem to be under development³⁹. An early consultation within a Scientific Advice
498 procedure is therefore recommended.

499 Trial durations from 1-2 years have previously been proposed in order to show efficacy on the interim
500 endpoint proposed. From an overall efficacy and safety point of view, but depending on the magnitude
501 of effect to be expected, a study duration of at least 2 years seems to be desirable. A trial extension
502 for the longest possible extend should be aimed at. If indeed studies using long-term outcomes (liver
503 transplantation and death, decompensation events) are intended, these are usually event driven, and a
504 priori determination of the trial duration will not be possible.

505 It has been proposed that – due to potential ethical concerns with regard to prolonged placebo
506 treatment, as well as adherence problems – that an open-label extension should be conducted (e.g.
507 additional 2 years), and the results could be compared to an external natural history cohort derived
508 from the Global PBC Study Group" database"⁴⁰. However, this is currently not recommended as
509 acceptable strategy and must – for the time being – be also considered as supportive endpoint only.

510 Despite the availability of at least one alternative add-on treatment at this point of time, the trials in
511 the add-on setting are recommended to be conducted with placebo-control only. This is related to the
512 safety profile of the potential alternative obeticholic acid, which is potentially leading to relevant un-
513 blinding (high occurrence rate of pruritus), and the currently unconfirmed (in respect to long-term
514 outcomes) efficacy status of the compound.

515

516 **Target of estimation (estimand)**

517 According to ICH E9(R1) Draft Addendum on estimands and Sensitivity Analysis in Clinical Trials
518 (EMA/CHMP/ICH/436221/2017), the scientific question of interest should be specified, and trial
519 features should be aligned.

520 Potential intercurrent events to be taken into account for the outcome in the setting of PBC can be
521 assumed to be lack of adherence to treatment and the intake of rescue medication. The intake of
522 rescue medication should be considered as a treatment failure (expected to occur in first-line settings),
523 i.e. the composite strategy as discussed in the addendum is considered appropriate for this
524 intercurrent event. With regard to other intercurrent events, a treatment policy strategy appears most
525 suitable, i.e. the outcome regardless of the intercurrent event is of primary interest. Therefore, data on
526 outcome should be collected independently from the occurrence of these intercurrent events, which is
527 considered to be feasible especially in this setting because the primary endpoint(s) is/are based on
528 simple blood biomarker evaluations.

529 With regard to the evaluation of long-term endpoints it is referred to the respective paragraph on
530 NASH (see 4.2.3).

531 **4.4. Primary sclerosing cholangitis**

532 **4.4.1. Short characterisation of the disease**

533 Primary sclerosing cholangitis (PSC) is a rare, chronic, heterogeneous, and idiopathic inflammatory
534 disease characterised by intra- and/or extrahepatic stricturing of bile ducts and development of
535 fibrosis. The natural history of PSC includes the development of complications (e.g. bacterial
536 cholangitis), progression of fibrosis to cirrhosis, and ultimately end-stage liver disease with
537 decompensation, liver transplantation, or death. The disease is frequently associated with
538 inflammatory bowel disease (IBD), including both, Crohn's disease (CD), as well as ulcerative colitis
539 (UC). Patients with PSC are at high risk of cholangiocarcinoma and gall bladder cancer, and also have
540 increased risks of colon carcinoma, whereas the presence of an increased risk for hepatocellular
541 carcinoma is controversial ^{41,42,43}.

542 Patients are either diagnosed on the presence of cholestasis when screening at risk patients (e.g. those
543 with IBD), or general health screening. Symptoms usually develop with progression of the disease, and
544 include fatigue, pruritus, and right upper quadrant pain, potentially accompanied by jaundice in later
545 stages. The diagnosis is made based on the serum markers of potential cholestasis and finally on the
546 presence of stricturing cholangiopathy, usually diagnosed with magnetic resonance
547 cholangiopancreatography (MRCP). A final diagnosis also requires the exclusion of relevant secondary
548 cholangitis, particularly IgG4– related disease. Liver biopsy is not regularly performed but regarded to
549 be needed in patients with suspected small duct PSC or in patients with suspected overlap with
550 autoimmune hepatitis. ^{44,45} The age at diagnosis is mostly between 30 and 40 years, but even children
551 can be affected.

552 There are a number of factors relevant for the overall prognosis in patients with PSC: The presence of
553 small duct PSC, and of Crohn's disease are associated with a better outcome, whereas ulcerative
554 colitis, and the occurrence of a so-called dominant stricture are factors associated with negative
555 outcomes (with regard to transplant-free survival) ⁴⁶.

556 The incidence of the disease has been estimated up to 0.4 to 2.0 per 100,000 inhabitants per year with
557 a wide variability, even within Europe. The prevalence has been estimated to be overall less than 50
558 per 100,000 (10 per 100,000 inhabitants) ⁴⁷. The development of the disease is slow, and it has most

559 recently been estimated that the development of end-stage liver disease may regularly take more than
560 20 years⁴⁸.

561 **4.4.2. Selection of patient populations**

562 Similar principles of the selection of study population as above (see 4.2.2 and 4.3.2) are also
563 applicable to PSC. The specific requirements concern the following:

564 As mentioned above, the diagnosis of PSC mainly relies on the profile of elevated ALP and an abnormal
565 cholangiography consistent with PSC, as shown by MRCP, endoscopic retrograde
566 cholangiopancreatography (ERCP), or percutaneous transhepatic cholangiography (PTC). A minimal
567 threshold for ALP and transaminases elevations should be defined. The availability of a liver biopsy
568 consistent with PSC is a compulsory requirement. The presence of overlap (e.g. with AIH) syndromes
569 can be allowed in exploratory clinical trials, but not in confirmatory trials. Other secondary reasons for
570 PSC also need to be excluded.

571 The selection of PSC patients should allow the occurrence of relevant events in the population included.
572 The inclusion of small duct PSC patients would therefore usually require the presence of other,
573 negative risk factors, such as ulcerative colitis etc. If patients have already a dominant stricture at the
574 time of inclusion, patients should not have relevant fluctuations of serum markers historically, and not
575 have relevant cholestasis at inclusion. It may also be sensible to define an upper limit of other markers
576 of liver damage (e.g. for transaminases). The presence of cholangio- and gall-bladder carcinoma, as
577 well as colon carcinoma should be excluded. Usually, patients with active IBD should not be included
578 due to the potential interference with the search for effective medication, and its associated changes.
579 Similarly, concomitant medication for IBD would require to be stable for a relevant time-frame.
580 Occurrence of acute cholangitis should not have occurred for a relevant time-frame, and no concurrent
581 antibiotic therapy should be part of the medication at inclusion.

582 Depending on the aim of the trial (see Chapter 4.4.3. and 4.5), the inclusion of patients having a
583 relevant level of symptoms should be considered. For a disease modifying study, both, symptomatic,
584 as well as asymptomatic patients, can be included. Despite the fact that UDCA is not regarded as
585 recommended medication in PSC, it is in widespread use. Therefore, the inclusion of patients on
586 concomitant UDCA can be allowed, but intake of UDCA should not be altered during the trial.

587 It would usually be expected that the presence of decompensation symptoms should be an exclusion
588 criterion, but cirrhotic patients without signs and/or symptoms of decompensation can be included.

589 **4.4.3. Study design and endpoints**

590 No licensed treatment in PSC is currently available. Therefore, a development strategy aiming at the
591 demonstration of effects at an early time-point using intermediate endpoints, for which the surrogacy
592 has at least been demonstrated by natural history studies, with a later confirmation on long-term
593 endpoints, is regarded to be an acceptable option (See also Chapter 4.1.). Such a strategy is also
594 supported by the fact that previous trials in PSC with UDCA have not demonstrated clear beneficial
595 effects for the long-term endpoints, despite being partly successful with potential surrogates.

596 These trials with UDCA in PSC have been assumed to be largely underpowered, and were – although
597 having demonstrated dose-dependent reductions in ALP –not able to demonstrate relevant effects on
598 the long-term efficacy outcomes such as manifestation of cirrhosis, decompensation clinical events,
599 liver transplantation and death. Whereas, however, the level of ALP – at diagnosis and after follow-up
600 – has repeatedly been demonstrated to be associated with outcomes in PSC⁴⁹⁵⁰, there was an obvious
601 dissociation of ALP and relevant clinical outcomes in the UDCA trials⁵¹⁵². The International PSC Study

602 Group has recently made comprehensive suggestions for the use of intermediate endpoints in PSC⁵³,
603 which has been reflected and partially adopted in the regulatory environment ⁵⁴.

604 Therefore ALP can currently not be accepted as the only intermediate endpoint to be used in this
605 disease. Other endpoints proposed (such as transient elastography and bilirubin) face similar problems
606 as ALP, or have a less robust history of validation. The use of histology in PSC has been discussed
607 controversially (see also 4.1.), however newer research has been shown that – in addition to its
608 obvious face validity – histology can well be used to evaluate the changes.

609 Therefore, a combined use of histology evaluation and ALP changes are regarded to represent an
610 acceptable intermediate endpoint for the disease for the time being.

611 It is again emphasized that intermediate endpoints used for marketing authorisation must be
612 sufficiently reliable to allow the conclusion of a positive benefit risk at time of marketing authorisation.
613 Therefore a co-primary evaluation of these intermediate endpoints should be aimed at. Furthermore it
614 is suggested that a responder-type evaluation based on the criteria of therapeutic response should be
615 the basis, defining serological response as a reduction of ALP to 1.3xULN, or a combination of the
616 reduction to 1.5-1.3xULN with an at least 40% reduction from baseline. For the histological evaluation
617 – best to be based on the newer staging system according to Nakanuma ⁵⁵ – a similar responder-type
618 evaluation is proposed. The response should be defined based on an at least 1 point improvement in
619 the fibrosis stage. Stable disease (no worsening of fibrosis) could be used instead, if adequately
620 justified.

621 As advocated before, a later evaluation of long-term outcomes is also considered necessary for PSC ,
622 which should be done as a composite endpoint including the manifestation (histological diagnosis) of
623 cirrhosis, a MELD score above 14, decompensation events (such as encephalopathy, variceal bleeding,
624 and ascites), as well as liver transplantation and death.

625 Due to the slow development of fibrotic stages and the low prevalence of the disease, the difficulties
626 for the validation of the proposed intermediate endpoints are acknowledged. Future applicants should
627 therefore also take care that a sufficient amount of supportive evidence for long-term efficacy is
628 available. This should consist of standard evaluations such as imaging modalities, other biomarkers
629 (bilirubin, transaminases, but also e.g. ELF-test and other potential future biomarkers) as well as
630 important clinical events in the course of the disease, such as (number of) bouts of acute cholangitis,
631 occurrence (manifestation) of dominant stenosis, and finally the occurrence of cholangiocarcinoma,
632 and other malignancies. In case the intended long-term outcome endpoints fail to demonstrate a
633 significant difference to placebo, a final conclusion on the benefit-risk ratio would have to be based on
634 the totality of these data.

635 As no effective treatment is currently available, the acceptable comparator is regarded to be placebo.
636 Trial duration is anticipated to be 2 years for the interim endpoints, and should be up to 5 years for the
637 demonstration of the long-term clinical outcomes. This proposed trial duration may need modification
638 based on the mechanism of action, as well as anticipated magnitude of effects of new drug candidates,
639 and the fact that usually, an event driven evaluation will be planned for.

640 **Target of estimation (estimand)**

641 Similar to PBC, with reference to ICH E9(R1) the scientific question of interest should be specified, and
642 trial features should be aligned accordingly.

643 Potential intercurrent events to be taken into account for the outcome in the setting of PSC can be
644 assumed to be lack of adherence to treatment and the occurrence of malignancy. The intake of rescue
645 medication will not play a relevant role for the time being, because no well-established treatments are
646 available.

647 Also similar to PBC, and according to the character of the primary endpoint, treatment policy strategy
648 may thus be most appropriate for the intercurrent events, i.e. the outcome regardless of the
649 intercurrent event is of primary interest.

650 **4.5. Trials for the symptomatic treatment (PBC and PSC):**

651 It has been described that both, PBC, as well as PSC impose a significant and clinically relevant burden
652 of symptoms on patients with the diseases.

653 For these reasons, it is possible to develop new treatments in the two diseases, which address the
654 symptomatic improvement of the patients, without aiming generally at positively influencing the
655 natural disease course (disease modification).

656 Potential drug candidates could involve patients suffering from a variety of symptoms, but at least of
657 the two major features of the disease (fatigue and pruritus). However, if only one symptom of the
658 disease is aimed at, it will usually be expected that effects are also evaluated in other pruritic diseases,
659 in order to claim a general pruritus, or fatigue indication.

660 If the totality of the disease specific symptoms are aimed at with a treatment, it is recommended that
661 disease specific measurements of the symptoms are part of the primary evaluation. The development
662 of such tools (patient-reported outcome tools – PROs) is encouraged. Usually, a claim of efficacy
663 should be based on an instrument measuring the direct symptoms, supported by a more indirect
664 evaluation of the impact of the symptoms, usually to be evaluated with disease-specific Quality of Life
665 scale.

666 Clinical trials with this restricted scope could be planned with a limited duration of (placebo) controlled
667 treatment for 6 months. A sufficient amount of long-term data, in order to demonstrate adequate
668 safety should, however, also be available (reference is made to the ICH E1 guideline).

669 **Target of estimation (estimand)**

670 The evaluation of a symptomatic treatment is expected to be evaluated with a treatment policy
671 evaluation. This is partly due to the different character of the endpoints, but also to the partly different
672 nature of the expected intercurrent events, which at least in the case of pruritus could include a variety
673 of rescue treatments. Also, a complete follow-up of patients, even in the case of study drug
674 discontinuation appears to be possible to a higher extent, also supported by the limited observation
675 period.

676 **4.6. Safety considerations**

677 General safety requirements will apply to trials in chronic liver diseases, similar to other fields of drug
678 development. The general requirements to focus on the known pharmacodynamic effects, including
679 off-target effects known from early development programme will fully apply. The following paragraphs
680 therefore deal with the specifics of safety evaluation with regard to liver in patients with underlying
681 liver disease, and the cardiovascular safety consideration applicable to NASH

682 **4.6.1. Safety in PBC and PSC**

683 The underlying liver disease, as well as fluctuations and flares occurring during the course of clinical
684 trials may hamper the evaluation of hepatic safety due to the overlap in accompanying symptoms, as
685 well as the changes in the routine liver safety biomarkers used, such as transaminases, ALP, and
686 bilirubin. The distinction of fluctuation and flare of the underlying disease, from subclinical liver
687 damage and true drug-induced liver injury (DILI) caused by an investigational agent is therefore the

688 most important feature of the evaluation of liver safety in both disease entities. The distinction of the
689 type of injury pattern, as well as causality assessment (e.g. using the well-established Roussel Uclaf
690 Causality Assessment Method (RUCAM) criteria, as well as expert adjudication), and the search for and
691 potential identification of Hy's law cases are valuable parts of the evaluation of liver safety and
692 potential DILI in clinical trials. In addition, obtaining biopsies whenever possible should be the
693 aim. ⁵⁶⁵⁷.

694 Although a generally increased risk of DILI in patients with underlying liver disease appears to be
695 controversial ⁵⁸ and may depend on the underlying disease⁵⁹, in addition to these general
696 requirements a need exists to define different rules for the safety evaluation during, and after clinical
697 trials with underlying liver diseases. These alternative approaches may include stopping rules, as well
698 as thresholds to define clinically relevant events and the use of novel statistical approaches specifically
699 developed for this purpose⁶⁰. In addition, the inclusion of experimental biomarkers is highly
700 recommended for trials in patients with underlying liver disease ⁶¹, but the influence of the underlying
701 disease on these markers should be known before they are used to help the assessment of safety. It is
702 recommended that all these methods are implemented in addition to the routine liver safety
703 evaluation.

704 **4.6.2. Safety in NASH**

705 Similar to PBC and PSC, the evaluation of liver safety in the field is considered paramount, and at the
706 same time, hampered by the underlying disease process. The principles outlined for PBC and PSC are
707 therefore also applicable in NASH. The higher number of patients that can be expected to be treated
708 might, however, allow more clear conclusions on liver safety.

709 Because NASH is associated with the obesity epidemic, and the liver manifestation of the so-called
710 metabolic syndrome, the patient population included in clinical trials in NASH will be prone to increased
711 risks of adverse events related to concomitant diseases such as arterial hypertension, diabetes
712 mellitus, severe obesity, and hypercholesterolaemia with the associated sequelae cardiovascular
713 events, such as myocardial infarction, stroke, and associated death ⁶²⁶³⁶⁴.

714 Therefore, depending on the mechanism of action, and the pre-clinical data showing potential
715 detrimental effects with regard to cardiovascular safety, the principles of the "reflection paper on
716 assessment of cardiovascular safety profile of medicinal products" (EMA/CHMP/505049/2015), are
717 considered applicable to NASH also, although it is currently not fully clear whether the risk increase for
718 cardiovascular outcomes and the resulting number of events will allow reliable conclusions. Further
719 long-term natural history data, and long-term clinical trials in the field are needed to draw a final
720 conclusion.

721 It is therefore necessary, not only to focus the safety evaluation on the occurrence of the so-called
722 major cardiovascular events (MACE) but also on the off-target effects of the potential investigational
723 products on parameters potentially influencing the overall cardiovascular risk, such as plasma lipids,
724 glucose homeostasis, and (systemic) inflammatory parameters.

725 **4.7. Children and adolescents**

726 **4.7.1. NASH in children and adolescents**

727 Similar to other aspects of the obesity/"metabolic syndrome" epidemic, non-alcoholic fatty liver disease
728 (NAFLD), as well as NASH have been identified to present an increasingly significant health burden in
729 children and adolescents. The prevalence of NAFLD in children is estimated to be around 10-14%

730 depending on age. Whereas 2-4 year old children are expected to suffer from NAFLD at only very low
731 rates, the prevalence in adolescents almost reaches adult levels ⁶⁵.

732 Assuming a similar rate of patients developing NASH from the presence of NAFLD as in adults ⁶⁶, it is
733 clear that NASH is a relevant health problem also in the young age group, although the development of
734 late-stage disease may take years and might be expected to manifest not before reaching adulthood.
735 However, rapid progression to advanced liver disease in childhood has been described⁶⁷.

736 Therefore, there is a relevant medical need to develop treatments for NASH also in children.

737 As outlined above, the diagnosis of NASH is currently considered to require the conduct of liver biopsy
738 with histological evaluation, and the conduct of clinical trials should be mainly based on repeated
739 biopsy results. The diagnosis itself is also based on histology in childhood/adolescence
740 patients^{68,69}. However, the conduct of repeated biopsies in clinical trials requires increased awareness of
741 the potentially associated ethical and procedural problems when children are concerned, and the need
742 for non-invasive outcomes in this population is therefore considered to be of even higher priority.

743 Furthermore, the histology evaluations available have shown distinct features of paediatric NASH as
744 compared to adults, with the presence of a relevant proportion of patients developing a unique
745 histology with presence of portal-based chronic inflammation (and fibrosis) (as opposed to the lobular
746 inflammation found in adults and less ballooning⁷⁰). The clinical meaningfulness of this distinct type of
747 histology in children is currently unknown, and consequently, a different histological scoring system
748 may be needed for the paediatric population.

749 The development of new medicinal products for the treatment of NASH in children therefore requires
750 first of all the collection of new and evaluation of existing data with regard to the natural history of the
751 disease.

752 Drug development in children will also require the final determination of the adequate age range to be
753 studied. Young children (e.g. below the age of 6-10 years), might still be early in the disease process,
754 and therefore be appropriate candidates for non-pharmacological interventions, such as life-style and
755 dietary changes, of which success rates (with regard to weight loss) are usually higher than in adults.
756 Consequently, the potential for regression of inflammatory changes is similarly considered to be
757 higher⁷¹.

758 The development of new medicinal products for NASH in children would also need a determination of
759 the quantity of data needed to be available for adults, before therapeutic trials are conducted. At this
760 point of time – when there still seems to be a need for more natural history data – it is recommended
761 that relevant clinical trials are deferred until data in adults on long-term endpoints are available (with
762 regard to progression to cirrhosis, liver transplantation and death) at least until the validity of the
763 proposed interim endpoints has been relevantly substantiated.

764 The availability of further data on natural history, as well as on the individual new compound in adults
765 might already enable to more precisely determine the level of extrapolation that can be applied (see
766 draft: Reflection paper on the use of extrapolation in the development of medicines for paediatrics.
767 EMA/199678/2016).

768 Once the above mentioned data are available, and a decision on the possible level of extrapolation can
769 be taken, the conduct of therapeutic trials in children is considered to be relevant, keeping in mind the
770 potential for enhanced regression of NASH. Besides the necessary investigation of the appropriate dose
771 (under full consideration of the potential differences in pharmacokinetics in obese and NASH
772 adolescents compared to adults), and development of age-appropriate formulations, the conduct of
773 placebo-controlled trials, including endpoints based on histology, and thus, repeated liver biopsies may
774 still be required in order to fully account for the differences between childhood/adolescent and adult

775 NASH. Even if from adult studies, an intermediate endpoint method such as an early histology
776 evaluation endpoint, imaging methods, or biomarkers, have partly been validated, it can be anticipated
777 that these would have to undergo further validation in children

778 The conduct of studies with histology endpoints should take full account of the potential for the ethical
779 problems associated with any more than minimally invasive procedures, and may need a careful
780 approach with regard to the patient selection (e.g. older age groups, more advanced disease, etc.).

781 **4.7.2. PBC: Children and adolescents**

782 The youngest reported age of a confirmed disease onset has been in a 15-year old post-menarche
783 adolescent ⁷², and it is thought that a true paediatric disease is not encountered.

784 Potential applicants developing new substances in the treatment of PBC would therefore be expected to
785 apply for a waiver for a paediatric programme in the disease.

786 **4.7.3. PSC in Children and Adolescents**

787 Paediatric PSC is a very rare disease, even compared to adult PSC, which itself is classified as orphan.
788 However, it is estimated that the risk in patients with IBD to develop PSC is doubled in the paediatric
789 population as compared to adults. Therefore, PSC appears to be a major source of morbidity in this
790 population. With the rising incidence of IBD, a clear unmet medical need exists. Also distinct from adult
791 PSC, there is a higher overlap of PSC with other syndromes, especially AIH (PSC-AIH-overlap
792 syndrome or Autoimmune Sclerosing Cholangitis - ASC) ^{73,74}. The investigation of new compounds,
793 also for children is therefore considered to be needed.

794 Although a relevant amount of data has already been collected for paediatric PSC ⁷⁵, there is still a
795 need to collect further natural history data before clinical trials in PSC can reasonably be undertaken.

796 Once these natural history data are available and have been evaluated to a sufficient extend, trials in
797 paediatric PSC may also be conducted with patients suffering from overlap conditions (especially AIH-
798 PSC), if adequate. The inclusion of patients should be based on the identified risk factors, which are
799 distinct from adult PSC, such as elevated gamma-glutamyl-transferase (GGT) and aspartate
800 aminotransferase-to-platelet ratio index (at diagnosis).

801 Besides the need to fully explore the PK profile in the respective population, there can currently no
802 clear recommendations be given with regard to the design of trials, and endpoints to be used.

803 Consultation with the agency early in the drug development (scientific advice and submission of PIP) is
804 therefore advisable.

805 **5. References**

¹ Sanyal AJ et al: Challenges and opportunities in drug and biomarker development for non-alcoholic steatohepatitis: Findings and Recommendations from an American Association for the Study of Liver Disease – US Food and Drug Administration joint workshop. *Hepatology* 2015; 61: 1392-1405.

² Ratzliff V et al: Sampling variability of liver biopsy in nonalcoholic fatty liver disease. *Gastroenterology* 2005; 128: 1898-1906.

³ Seeff LB et al: Complication Rate of Percutaneous Liver Biopsies among Persons with Advanced Chronic Liver Disease in the HALT-C Trial. *Clin Gastroenterol Hepatol* 2010; 8: 877-883.

⁴ Rinella ME et al: Nonalcoholic fatty liver disease. A systematic review. *JAMA* 2015; 313: 2263-2273.

⁵ Younossi ZM et al: Global epidemiology of nonalcoholic fatty liver disease – meta-analytic assessment of prevalence, incidence, and outcomes. *Hepatology* 2016; 64: 73-84.

⁶ Younossi Z et al: Global burden of NAFLD and NASH: trends, predictions, risk factors and prevention. *Nat Rev Gastroenterol Hepatol* 2018; 15: 11-20.

⁷ Cholangkeril G et al: Liver Transplantation for non-alcoholic steatohepatitis in the US: Temporal trends and outcomes. *Dig Dis Sci* 2017; 62: 2915-2922.

⁸ Eslam M et al: Genetics and epigenetics of NAFLD and NASH: Clinical impact. *J Hepatol* 2018; 68: 268-279.

- ⁹ Bellentani S: The epidemiology of non-alcoholic fatty liver disease. *Liver International* 2017; 37 (Suppl. 1) 81-84.
- ¹⁰ European Association for the Study of the Liver (EASL), European Association for the Study of Diabetes (EASD), European Association for the Study of Obesity (EASO): EASL-EASD-EASO Clinical practice guidelines for the management of non-alcoholic fatty liver disease. *Diabetologia* (2016) 59: 1121-1140.
- ¹¹ Dan AA et al: Health-related quality of life in patients with non-alcoholic fatty liver disease. *Aliment Pharmacol Ther* 2007; 26: 815-820.
- ¹² Sattar N et al: Non-alcoholic fatty liver disease. *BMJ* 2014; 349: g4596 doi: 10.1136/bmj.g4596
- ¹³ Dulai P et al: Increased risk of mortality by fibrosis stage in non-alcoholic fatty liver disease: Systematic review and meta-analysis. *Hepatology* 2017; 65t: 1557-1565.
- ¹⁴ Marengo A et al: Progression and natural history of non-alcoholic fatty liver disease in adults. *Clin Liver Dis* 2016; 20: 313-324
- ¹⁵ Goh GBB and AJ McCullough: Natural history of non-alcoholic fatty liver disease. *Dig Dis Sci* 2016; 61: 1226-1233.
- ¹⁶ Singh S et al: Fibrosis progression in non-alcoholic fatty liver vs non-alcoholic steatohepatitis: A systematic review and meta-analysis of paired-biopsy studies. *Clinical Gastroenterology and Hepatology* 2015; 13: 643-654.
- ¹⁷ Hagström H et al: Fibrosis stage but not NASH predicts mortality and time to development of severe liver disease in biopsy-proven NAFLD. *J Hepatol* 2017; 67: 1265-1273.
- ¹⁸ Kleiner DE et al.: Design and validation of a histological scoring system for nonalcoholic fatty liver disease. *Hepatology* 2005; 41: 1313–21.
- ¹⁹ Siddiqui MS et al: Case definition for inclusion and analysis of endpoints in clinical trials for non-alcoholic steatohepatitis through the lens of regulatory science. *Hepatology* 2018; 67: 2001-2012.
- ²⁰ Patel Y et al: Baseline Parameters in clinical trials for non-alcoholic steatohepatitis: Recommendations from the Liver Forum. *Gastroenterology* 2017; 153: 621-625.
- ²¹ Boursier J et al: Screening for therapeutic trials and treatment indication in clinical practice: MACK-3, a new blood test for the diagnosis of fibrotic NASH. *Aliment Pharmacol Ther* 2018; 47: 1387-1396.
- ²² Vilar-Gomez E and N Chalasani: Non-invasive assessment of non-alcoholic fatty liver disease: Clinical prediction rules and blood-based biomarkers. *J Hepatol* 2018; 68: 305-315.
- ²³ Vilar-Gomez E et al: Weight Loss through lifestyle modification significantly reduces features of nonalcoholic steatohepatitis. *Gastroenterology* 2015; 149, 367-378.
- ²⁴ Ratzu V: A critical review of endpoints of non-cirrhotic NASH therapeutic trials. *J Hepatol* 2018; 68: 353-361.
- ²⁵ D'Amico G et al: Natural history and prognostic indicators of survival in cirrhosis: A systematic review of 118 studies. *J Hepatology* 2006; 44: 217-231.
- ²⁶ Ripoll C et al: Hepatic venous pressure gradient predicts clinical decompensation in patients with compensated cirrhosis. *Gastroenterology* 2007; 133: 481-488.
- ²⁷ Polyzos SA et al: Nonalcoholic fatty liver disease: Is it time for combination treatment and a diabetes-like approach. *Hepatology* 2018; 68: 389-389
- ²⁸ Beuers U et al: Changing nomenclature for PBC: from "cirrhosis" to "cholangitis". *Gut* 2015; 64: 1671-1672.
- ²⁹ Carey EJ et al: Primary biliary cirrhosis. *The Lancet* 2015; 386; 1565-1575.
- ³⁰ Murillo Perez, CF et al: Milder disease stage in patients with primary biliary cholangitis over a 44-year period: A changing natural history. *Hepatology* 2018; 67: 1920-1929.
- ³¹ European Association for the Study of the Liver: EASL Clinical Practice Guidelines: The diagnosis and management of patients with primary biliary cholangitis. *Hepatology* 2017; 67: 145-172.
- ³² Pares A et al: Excellent long-term survival in patients with primary biliary cirrhosis and biochemical response to ursodeoxycholic acid *Gastroenterol* 2006; 130: 715-720.
- ³³ Kumagi T et al: Baseline ductopenia and treatment response predict long-term histological progression in primary biliary cirrhosis. *Am J Gastroenterol* 2010; 105: 2186-2194.
- ³⁴ Kuiper EMM et al: Improved prognosis of patients with primary biliary cirrhosis that have a biochemical response to ursodeoxycholic acid. *Gastroenterology* 2019; 136: 1281-1287.
- ³⁵ Corpechot C et al: Early primary biliary cirrhosis: biochemical response to treatment and prediction of long-term outcome *J Hepatol* 2011; 55: 1361-1367.
- ³⁶ Momah N et al: Optimizing biochemical markers as endpoints for clinical trials in primary biliary cirrhosis. *Liver Int* 2012; 32: 790-795
- ³⁷ Lammers WJ et al: Levels of alkaline phosphatase and bilirubin are surrogate end points of outcomes of patients with primary biliary cirrhosis: An international follow-up study. *Gastroenterol* 2014; 147: 1338-1349.
- ³⁸ Corpechot C et al: A placebo-controlled trial of bezafibrate in primary biliary cholangitis. *NEJM* 2018; 378: 2171-2181.
- ³⁹ Wendum D et al: Primary biliary cirrhosis: proposal for a new simple histological scoring system. *Liver International* 2015; 35: 652-659.
- ⁴⁰ Lammers WJ et al: Development and validation of a scoring system to predict outcomes of patients with primary biliary cirrhosis receiving ursodeoxycholic acid therapy *Gastroenterology* 2015; 149: 1804-1812.
- ⁴¹ Dyson JK et al: Primary sclerosing cholangitis: *The Lancet* 2018; 391: 2547-2559
- ⁴² Karlsen TH et al: Primary sclerosing cholangitis – a comprehensive review *J Hepatol* 2017; 67: 1298-1323.
- ⁴³ Horley-Silva JL et al: An update on cancer risk and surveillance in primary sclerosing cholangitis. *Liver Int* 2017; 37: 1103-1109.
- ⁴⁴ European Association for the Study of the Liver. EASL clinical practice guidelines: management of cholestatic liver disease. *J Hepatol* 2009; 51: 237-267.
- ⁴⁵ Lindor KD et al: ACG Clinical Guideline: Primary Sclerosing Cholangitis. *Am J Gastroenterol* 2015; 110: 646-659.⁴⁶ Weismüller, TJ et al. Patient age, sex, and inflammatory bowel disease phenotype associate with course of primary sclerosing cholangitis. *Gastroenterology*. 2017; 152: 1975–1984
- ⁴⁷ Molodecky NA et al: Incidence of primary sclerosing cholangitis: a systematic review and meta-analysis. *Hepatology* 2011; 53: 1590-1599.
- ⁴⁸ Boonstra K et al: Population-based epidemiology, malignancy risk, and outcome of primary sclerosing cholangitis; *Hepatology* 2013; 58: 2045-2055.
- ⁴⁹ Rupp, C et al: Reduction in alkaline phosphatase is associated with longer survival in primary sclerosing cholangitis, independent of dominant stenosis. *Aliment Pharmacol Ther* 2014; 40: 1292-1301

-
- ⁵⁰ De Vries EMG et al: Alkaline phosphatase at diagnosis of primary sclerosing cholangitis and 1 year later: evaluation of prognostic value. *Liver International* 2016; 36: 1867-1875.
- ⁵¹ Olsson R et al: High-Dose Ursodeoxycholic Acid in Primary Sclerosing Cholangitis: A 5-Year Multicenter, Randomized, Controlled Study. *Gastroenterology* 2009; 129: 1464-1472.
- ⁵² Lindor KD et al: High-dose ursodeoxycholic acid for the treatment of primary sclerosing cholangitis. *Hepatology* 2009; 50: 808-814.
- ⁵³ Ponsioen CY et al: Surrogate endpoints for clinical trials in primary sclerosing cholangitis: Review and results from an International PSC Study Group consensus process. *Hepatology* 2016; 63: 1357-1367.
- ⁵⁴ Ponsioen CY et al: Design and Endpoints for Clinical Trials in Primary Sclerosing Cholangitis. *Hepatology* 2018. E-pub ahead of print. *Hepatology* 2018; 68: 1174-1188
- ⁵⁵ 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: 907-919
- ⁵⁶ Teschke R and G. Danan: Diagnosis and management of drug-induced liver injury (DILI) in patients with pre-existing liver disease. *Drug Saf* 2016; 39: 729-744.
- ⁵⁷ Kullak-Ublick GA et al: Drug-induced liver injury: recent advances in diagnosis and risk assessment. *Gut* 2017; 66: 1154-1164
- ⁵⁸ Teschke R and G Danan: Drug-induced liver injury: Is chronic liver disease a risk factor and a clinical issue? *Exp Op Drug Metabol and Toxicol* 2017; 13: 425-438.
- ⁵⁹ Massart J et al: Role of nonalcoholic fatty liver disease as risk factor for drug-induced hepatotoxicity. *J Clin Transl Res* 2017; 3: (Suppl 1) 212-232.
- ⁶⁰ Kullak-Ublick GA et al: Liver safety assessment in special populations (Hepatitis B, C, and oncology trials). *Drug Saf* 2014 (Suppl 1) 37: S57-62.
- ⁶¹ Letter of support for drug-induced liver injury (DILI) biomarker. EMA/423870/2016
- ⁶² Adams LA et al: Non-alcoholic fatty liver disease and its relationship with cardiovascular disease and other extrahepatic diseases. *Gut* 2017; 66: 1138-1153.
- ⁶³ Leonardo A et al: Hypertension, diabetes, atherosclerosis and NASH: Cause or consequence? *J Hepatol* 2018; 68: 335-352.
- ⁶⁴ Sookian S and CJ Pirola: Systematic review with meta-analysis: risk factors for non-alcoholic fatty liver disease suggest a shared altered metabolic and cardiovascular profile between lean and obese patients. *Aliment Pharmacol Ther* 2017; 46: 85-95.
- ⁶⁵ Doycheva I et al: Nonalcoholic fatty liver disease in adolescents and young adults: The next frontier in the epidemic. *Hepatology* 2016; 65: 2100-2109.
- ⁶⁶ Schwimmer JB et al: Prevalence of fatty liver in children and adolescents, *Pediatrics* 2006; 37: 1202-1219.
- ⁶⁷ Kohli R et al: Rapid Progression of NASH in childhood. *J Pediatr Gastroenterol Nutr* 2010; 50: 453-456.
- ⁶⁸ Vajro P et al: *JPGN* 2012; 54: 700-713.
- ⁶⁹ Vos M et al: NASPGHAN Clinical Practice Guideline for the diagnosis and treatment of nonalcoholic fatty liver disease in children: Recommendations from the expert committee on NAFLD (ECON) and the North American Society of Pediatric Gastroenterology, Hepatology and Nutrition (NASPGHAN), *JPGN* 2017; 64: 319-334.
- ⁷⁰ Schwimmer JB et al: Histopathology of pediatric nonalcoholic fatty liver disease. *Hepatology* 2005; 42: 641-649.
- ⁷¹ Vos M et al (see 73).
- ⁷² Dahlan Y et al: Pediatric-onset primary biliary cirrhosis. *Gastroenterol* 2003; 125: 1476-1479.
- ⁷³ Cotter JM and CL Mack: Primary Sclerosing Cholangitis: Unique aspects of Disease in children. *Clinical Liver Disease* 2017; 10: 120-123.
- ⁷⁴ Mieli-Vergani, G et al: Diagnosis and Management of Pediatric Autoimmune Liver Disease: ESPGHAN Hepatology Committee Position Statement. *JPGN* 2018; 66: 345-360.
- ⁷⁵ Deneau MR et al: The natural history of primary sclerosing cholangitis in 781 children: A multicentre, international collaboration. *Hepatology* 2017; 66: 518-527.