Sponsor:	Dr. Falk Pharma GmbH
Study title:	Double-blind, randomised, placebo-controlled, multi-centre phase III clinical study comparing the combination of ursodeoxycholic acid capsules plus budesonide capsules to ursodeoxycholic acid capsules plus placebo in the treatment of primary biliary cirrhosis
Study number:	BUC-56/PBC
[redacted]	[redacted]
SAP version:	Version 3.0, 12-jun-2017

Trial Biostatistician:

_____ _____
Signature Date

Name: [redacted]
Function: Lead Biostatistician
Company: [redacted]
City: [redacted]
Country: [redacted]

_____ _____
Signature Date

Sponsor's representative:

Name: [redacted]
Function: Project Management Clinical Research
Company: Dr. Falk Pharma GmbH
City: Freiburg
Country: Germany

TABLE OF CONTENTS

1 LIST OF ABBREVIATIONS

Abbreviation or special term	Explanation
AE	Adverse Event
ALT	Alanine aminotransferase
AP	Alkaline Phosphatase
AST	Aspartate aminotransferase
BMI	Body Mass Index
CRF	Case Report Form
CSP	Clinical Study Protocol
DXA	Dual-energy x-ray absorptiometry
FAS	Full Analysis Set
GGT	γ-Glutamyltransferase
ICH	International Conference on Harmonisation
IDMC	Independent Data Monitoring Committee
INR	International normalized ratio
FPFV	First Patient first Visit
FPI	First Patient Intake
LOCF	Last Observation Carried Forward
LPLV	Last Patient last Visit
mHAI	modified Hepatic Activity Index
mPP	Modified Per-Protocol
PBC	Primary Biliary Cirrhosis
PP	Per-Protocol
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
TEAE	Treatment Emergent Adverse Event
VAS	Visual Analogue Scale

2 INTRODUCTION

This statistical analysis plan (SAP) was developed according to the "ICH guideline E9 'Statistical Principles for Clinical Trials' (London 1998)".

The final statistical analysis will be performed at [redacted] according to the following documents:

Clinical Study Protocol (CSP) – Version 2.0 / 19-Feb-2008,

Local Study Protocol Amendment Austria No. 1 / 15-Aug-2008,

Local Study Protocol Amendment Denmark No. 1 / 08-Jan-2009

CSP Amendment No. 1 - Final / 14-Aug-2008,

CSP Amendment No. 2 - Final / 22-Jun-2009 and

CSP Amendment No. 3 - Version 3.1 / 23-Sep-2011 and

CSP Amendment No. 4 - Version 1.0 / 19-May-2014 and

CSP Amendment No. 5, Version 1.0 dated 12 Feb-2015.

All specifications described in the CSP, which were not changed in one of the Amendments, will be used for the analysis.

The analysis will be performed after database lock.

3 STUDY OBJECTIVES, ENDPOINTS AND DESIGN

3.1 Objectives

Objectives of this study are:

- To compare the efficacy and tolerability of a combination therapy with ursodeoxycholic acid (12-16 mg/kg body weight/d) plus budesonide (9 mg/d) vs. ursodeoxycholic acid (12-16 mg/kg body weight/d) plus placebo in the treatment of primay biliary cirrhosis (PBC).
- To study safety and tolerability in the form of adverse events and laboratory parameters.
- To assess patients' Quality of Life.

3.2 Endpoints

3.2.1 Primary endpoint

Primary efficacy endpoint is the response rate defined as rate of patients with improvement of liver histology with respect to inflammation (at least 3 points in the modified Hepatic Activity Index (mHAI) sum score or no inflammatory activity according to Ishak et al., 1995[1]) and no progression of fibrosis (staging according to Ludwig[2]) at the individual last patient visit within the study compared to baseline.

3.2.2 Secondary endpoints

Secondary efficacy endpoints are:

- Rate of patients presenting with cirrhosis or esophageal varices and/or ascites at the end of treatment or patients registered on the liver transplant waiting list or patients with liver related death during up to 3 years of treatment,
- Rate of patients with improvement of liver histology with respect to stage (according to Ludwig[2]),
- Rate of patients with improvement of liver histology with respect to grade (according to Desmet et al.[3]) and stage (according to Ludwig2),
- Rate of patients with improvement of liver histology with respect to grade (according to Desmet et al.[3])
- Normalisation of serum levels of alkaline phosphatase (AP) or reduction of baseline AP levels by at least 40 %,
- Improvement of serum levels of AP and bilirubin,
- Absolute and relative change of serum levels of AP, bilirubin, AST and GGT
- Course of pruritus (measured by visual analogue scale [VAS]),
- Course of fatigue (measured by PBC-40),
- Course of the Mayo Risk Score adapted,
- Assessment of inflammatory activity[1,3],
- Quality of Life by PBC-40,
- Global assessment of efficacy by patient and investigator,
- Fall in platelet counts < 135.000/mm^3,
- Fall in the prothrombin ratio (INR) of more than 25% and to a value of prothrombin time < 70%,
- Fibrosis marker (hyaluronic acid),

- Doubling in the serum concentration of hyaluronic acid to more than 80 µg/L or increase to a value of more than 100 µg/L,
- Proportion of patients with serum bilirubin levels more than 50 µmol/L (3.0 mg/dL),
- Proportion of patients with fall in serum albumin count of 10% and to a value of less than 34 g/L
- Rate of patients with AP<=1.67 ULN AND >=15% decrease in AP AND total bilirubin within the normal limits at 1year, 2years, 3years (and 1year LOCF, 2 years LOCF, 3 years LOCF)
- Globe score

3.3 Study design

3.3.1 Assessments and study duration

This was a double-blind, randomised, placebo-controlled, multi-centre, comparative, phase III clinical trial. The study was conducted with two arms in the form of a parallel group comparison and serves to compare a combination therapy (ursodeoxycholic acid plus budesonide) with ursodeoxycholic acid therapy only. Patients with PBC were treated. The study was performed according to a 2-stage group-sequential adaptive design with possible sample size adjustment after the planned interim analysis.

The recruitment of patients was started in the first quarter of 2008 (First Patient first visit (FPFV: 29-Jan-2009) und FPI (date of randomization: 18-Feb-2009)). The treatment period for patients who were in this study prior to Protocol Amendment No.5 was 36 months, if not premature termination took place. After the interim analysis the IDMC recommended to treat the ongoing patients for at least 1year and then to perform a termination visit. The schedule of the trial procedures is contained in the following study flow-chart. The clinical trial was finished in the third quarter of 2015 (Last Patient last medication intake: 05-Aug-2015 und Last Patient last visit (LPLV, Biopsy procedure: 19-Oct-2015).

DIAGRAM OF TRIAL PROCEDURES (Flow-Chart)					
		STUDY VISITS			
Procedure	Specification	**Baseline visit** week 0 (day -28 to 0) **V1**	**Interim visits** week 2, week 4 **V2 and V3**	**Interim visits** months 3, 6, 9, 12, 18, 24, 30 **V4 – V10***	**Final or withdrawal visit***
Patient selection		●			
Patient information and informed consent		●			
Demographic data	age, height	●			
Case history		●			
Inclusion/exclusion criteria		●			
Study diagnosis	confirmation	●			
Pregnancy test	Central	●			●
if applicable	local	monthly self-testing by the patient[3]			
Physical examination	documentation / changes	●		●	●
Vital signs		●	●	●	●

Lab assessments	central blood tests (haematology, serum chemistry)	●	●	●	●
	central blood tests (serology)	●			
	local tests (urinalysis, coagulation)	●	●	●	●
Fibrosis marker	hyaluronic acid	●		● (only V7, V9)	●
Bone metabolism, Suppression of adrenal function		●		● (only V7, V9)	●
Bone mineral density (DXA)		●		● (only V7)	●
Liver Biopsy		●[1]			●
FibroScan[2]		●			●
Pruritus		●		●	●
Fatigue (PBC-40)		●		● (only V7, V9)	●
Quality of Life (PBC-40)		●		● (only V7, V9)	●
Mayo Risk Score		●		● (only V7, V9)	●
Adverse events			●	●	●
Concomitant therapy		●	●	●	●
Randomisation		●			
Trial medication	dispensing	●		●	
	return, drug accountability			●	●
Diary	dispensing	●		●	
	return, check			●	●
Compliance			●	●	●
Appointment for the next visit		●	●	●	
Assessment of tolerability	investigator / patient			●	●
Assessment of efficacy	investigator / patient			●	●
Withdrawal	date, reasons, if applicable		●	●	

[1] if last liver biopsy dates back more than 6 months

[2] facultative: if equipment is available at the centre and patient has given consent

[3] in Austria only

*Patients who have completed more than 12 months of the study will terminate the study at their last regular visit (V8, V9, V10, or V11) or by 31 August 2015 at the latest.

3.3.2 Planned sample size

The study was planned with a response rate of the standard treatment arm of approximately 15 % in the patient population selected for the study. It was assumed that in the combination therapy the percentage of patients with improvement in liver histology with respect to inflammation and no progression of fibrosis will be 40 %. At a significance level of $\alpha = 0.025$ (one-sided), a randomisation ratio for budesonide vs. placebo of 2:1 and a statistical power of 80%, a patient number of 76 patients in the budesonide group and 38 patients in the placebo group was estimated (114 patients in both groups, total number of patients: 144 in the full analysis set, accounted for 20 % patients not providing an assessment of the primary efficacy variable, e.g. who do not consent to perform a liver biopsy at their individual last visit within the study).

In Amendment No. 4 - Version 1.0 / 19-May-2014 one interim analysis was introduced due to an unexpectedly low recruitment. The sample size calculation remains valid for a trial with one interim analysis at the information rate of 0.29 and O'Brien and Fleming shaped boundaries as the inflation factor is 1.000 for the specifications used in this trial. The interim analysis will therefore be conducted after the first 42 patients have completed the study. As per the assumptions above, it is expected that 33 of these patients will contribute their primary endpoint to the analysis.

After the interim analysis, the Dr. Falk Pharma GmbH in accordance with the Independent Data Monitoring Committee decided to terminate the recruitment of the study. This is mentioned in the Amendment No. 5, Version 1.0 dated 12 Feb-2015.

3.4 Interim analyses

After the Amendment No. 4 - Version 1.0 / 19-May-2014 the trial was to be conducted according to a 2-stage group sequential test design with possible sample size adaptation after the interim analysis. The interim analysis aimed to verify the assumptions of the sample size calculation or to recalculate the sample size based on the effect size estimations of the interim analysis.
The interim analysis was planned after observation of 42 patients who were evaluable in the FAS (approximately 28 patients in the budesonide group and 14 patients in the placebo group), corresponding to an information rate of 0.29 given that the total planned sample size was 144 patients. The critical values and significance levels to be applied at the 2 stages using the inverse normal method were:

Stage	Critical value	Sign. level (one-sided)
1	3.641	0.0001
2 (final)	1.961	0.0249

If the inverse normal test statistic would have had exceeded 3.641 at the interim analysis, it was possible to stop the study for efficacy. Otherwise, the study could be continued as planned, or with a recalculated sample size. There was no binding stopping for futility boundary implemented. However the study could have been stopped for futility on the recommendation of the Independent Data Monitoring Committee (IDMC), if warranted by the unblinded data.

43 patients were being included into the first interim analysis. At this stage, only a minor trend for a treatment effect was observed and the study was stopped due to the increase in patient numbers needed to confirm the treatment effect and to insufficient patient recruitment seen in this study.

3.5 Final analysis

For the final analysis all randomised patients will be included into stage I (all patients of the first interim analysis + overrun).
The null hypothesis will be tested and the null hypothesis will be rejected, if the inverse normal test statistic exceeds the critical value 3.641. In this case, superiority of budesonide compared to placebo will have been demonstrated.

Additionally a sensitivity analysis will be performed, where all randomised patients of the first interim analysis will be included as stage I and the patients of the overrun as stage II. The null hypothesis will be tested and the null hypothesis will be rejected, if the inverse normal test statistic exceeds the critical value 1.961. This comparison will be exploratory only.

4 ANALYSIS SETS

4.1 Safety Analysis Set

The safety set will include all patients (as treated) who received at least one dose of study medication. Administration of study medication will be checked on CRF-pages 68/69 “Administration of study medication” (and other corresponding pages) and on CRF-page 20 “Return and Dispense of study medication” at Visit 4. If the administration of any study medication is not certain, the patient will be included in the safety set. The analysis of safety will be based on the safety set.

4.2 Full Analysis Set

The Full Analysis Set (FAS) defined according to the intention-to-treat principle will include all randomised patients (as randomised) who received at least one dose of study medication. The intention-to-treat principle will be preserved despite the exclusion of patients who took no study medication, as the decision of whether or not to begin treatment could not be influenced by knowledge of the assigned treatment.

4.3 Per-Protocol (PP) Analysis Set

The per-protocol set will include all patients of the full analysis set, if the following criteria are additionally met:

- all of the major inclusion criteria, none of the major exclusion criteria fulfilled,
- absence of major protocol deviations,
- correct allocation to treatment group,
- efficacy data after baseline under study medication (Baseline and Follow-up liver-histology available),
- duration of treatment at least 12 months,
- premature discontinuations will be included only if the reason for discontinuation was lack of efficacy of the study medication or an adverse event with certain or probable/likely or possible causal relationship with the study medication, or intolerable adverse event that is a result from the deterioration of the study disease,
- sufficient compliance concerning administration of study medication (at least 80 % and maximum 120% of the study medication in the individual interval of treatment),
- documentation of final visit / withdrawal visit within a reasonable time frame after the last study drug administration (75 days for the FU biopsy and/or for the last available visit).

The classification of major and minor protocol deviations and the resulting analysis sets will be described in the evaluability plan. It will be performed prior to unblinding and will be approved by the sponsor.

The evaluation of the primary efficacy variable will be performed for the full analysis set (intention-to-treat analysis) and for the per-protocol set (per-protocol analysis).

The primary analysis will be the intention-to-treat analysis.

The evaluation of the secondary efficacy variable will be performed for the full analysis set (intention-to-treat analysis) and for the per-protocol set (per-protocol analysis).

4.4 Modified Per-Protocol (mPP) Analysis Set

The mPP set will include:

- Patients, who are included into the Per-protocol set;
- Patients, who are excluded from the Per-protocol set because of premature discontinuation due to a CSP Amendment 5 /Study Stop (will be detected by the documented termination cause 'early termination due to amendment 5'), if no other reason (except reason "Efficacy data after baseline not available") leads to an exclusion of the PP analysis set;
- Patients, who are excluded from the Per-protocol set because of missing efficacy data after baseline (no liver biopsy available) and the patient has performed the termination visit and laboratory data are available, if no other reason (except reason "Premature discontinuation according to amendment 5") leads to an exclusion of the PP analysis set

The evaluation of the secondary efficacy variable "Rate of patients with AP<=1.67 ULN AND >=15% decrease in AP AND total bilirubin within the normal limits at 1year, 2 years, 3 years" will be performed for the full analysis set (intention-to-treat analysis), for the per-protocol set (per-protocol analysis) and for the modified per-protocol set (modified per-protocol analysis).

The classification of the relevant protocol deviations will be performed prior to unblinding and will be approved by the sponsor.

A summary of the number of patients per analysis set will be given and reasons for exclusion of a patient from an analysis set will be listed.

5 STATISTICAL CONSIDERATIONS AND ANALYSIS

5.1 Derived variables

Age [years]	Date of written informed consent – date of birth (1st day of month)
Body Mass Index (BMI)	Weight [kg]/(Height [m])^2
Duration of disease [years]	= (Date of baseline visit - Date of first symptoms of PBC + 1)/365
Study duration [months]	= (Date of Final Visit /date of withdrawal – Date of baseline visit + 1)/30 (deviation to the calculation from the interim analysis)
Treatment duration [months]	= (Date of last intake of study medication – Date of first intake of study medication+1)/30
mHAI-score, Desmet score	The recalculated mHAI score will be taken into account. The Desmet score will be derived from mHAI-scores.
Improvement of liver histology with respect to stage (according to Ludwig[2])	A strict negative (<0) change in the Fibrosis subscore (Ludwig score) compared to baseline
Improvement of liver histology with respect to grade (according to Desmet[3])	A strict negative change (<0) in the Inflammatory activity (Desmet score) compared to baseline

Improvement of liver histology with respect to grade (according to Desmet et al.[3]) and stage (according to Ludwig[2])	Improvement of liver histology with respect to grade (according to Desmet et al.[3]) AND improvement of liver histology with respect stage (according to Ludwig[2])
Components of mHAI-Score (baseline)	Will be derived by using the components of the pathologist with the lowest mHAI-scores at baseline. In case of mHAI scores equally assessed by the two pathologists at baseline, the values of pathologist Wendum will be used for the single items of mHAI.
Mayo Risk Score adapted	The original Mayo Risk Score requests information about variceal bleeding. This is not available at the final visit in this study. Therefore the presence of varices at the final visit will be used (instead of variceal bleeding) and the Mayo Risk Score will be adapted with this modification. The Mayo Risk Score adapted (MRS adapt) will be recalculated as: MRS adapt = 0.0295 × (age in years) + 0.5373 × $\log_e$ (total bilirubin in mg/dL) - 0.8389 × (serum albumin in g/dL) + 0.5380 × $\log_e$ (AST in IU/L) + 1.2426 × (point for presence of varices) where - Age is calculated at each visit, where MRS adapted is calculated (age=date of visit – date of birth(1st day of month)), - AST = serum aspartate aminotransferase level, - Point for presence of varices is 0 if none and 1 if present. Point for presence of varices will be derived from information contained in CRF page 9 (Exclusion criterion no. 12 “Sonographic or endoscopic signs of portal hypertension”) and in CRF page 53 (Physical examination at final visit “Presence of esophageal varices?”). At begin of the study, patients should not have cirrhosis because this is an exclusion criteria. If the exclusion criterion no. 12 is not fulfilled, it will be assumed that the patient has no varices (point for presence of varices = 0). If at the final visit, esophageal varices are present, then the patient shows varices at final visit but for these patients the point for presence of varices will be assumed as missing at V7 (visit at 12 months) and V9 (visit at 24 months). If esophageal varices are not present at final visit, it will be assumed that the patient had no varices during the complete study. If esophageal varices are missing or not done at final visit, the corresponding value of MRS adapted will be put as “missing” and the Mayo Risk Score adapted will be missing.

Rate of patients presenting with cirrhosis or esophageal varices and/or ascites at the end of treatment or patients registered on the liver transplant waiting list or patients with liver related death up to 3 years of treatment.	Ludwig-score = Stage 4 or Presence of esophageal varices? = 'Yes' or Presence of ascites? = 'Yes' or SAE which includes the registration on liver transplant waiting list or liver related death up to 3 years of treatment.
First/Last intake of study medication	The date of first intake of study medication (irrespective of the study medication) will be derived by the first date documented on the CRF page 'Study drug administration' in case of 'Total daily dose' > 0. The date of last intake of study medication (irrespective of the study medication) will be derived by the last date documented on the CRF page 'Study drug administration' in case of 'Total daily dose' > 0.
Compliance (based on returned study medication)	Derived by using 'Administration page (68)' and 'Return of study medication (59)': Budesonid/Placebo: $100 * \frac{(No\ of\ dispensed\ capsules - no\ of\ returned\ capsules)x3}{\left((\mathrm{dlast}_j - \mathrm{dfirst}_1 + 1) * (targetdose = 9)\right)}$ Where dfirst_1 is the date of the therapy start for the first line and dlast_j is the date of therapy stop for the last line j of study medication, Given that it is sorted in ascending from (first date in first line (line 1) and last date in last line (line j)) UDCA: $100 * \frac{(No\ of\ dispensed\ capsules - \ no\ of\ returned\ capsules)x250}{\left((\mathrm{dlast}_j - \mathrm{dfirst}_1 + 1) * [\text{individual daily dose }] * 250\right)}$ Where dfirst_1 is the date of the therapy start for the first line and dlast_j is the date of therapy stop for the last line j of study medication, Given that it is sorted in ascending from (first date in first line (line 1) and last date in last line (line j)) The individual daily dose may vary from 2 to 7 capsules (see definition below).
Compliance (based on administrated medication)	Derived by using 'Administration page (68)': Budesonid/Placebo: If the AST (ALT before protocol amendment 2) values are not normalized: $100 * \frac{\sum_{x=1}^{j}((\mathrm{dlast}_x - \mathrm{dfirst}_x + 1) * givendose_x\ )}{\left((\mathrm{dlast}_j - \mathrm{dfirst}_1 + 1) * (targetdose = 9)\right)}$ Where $\mathrm{givendose}_x$ is the total daily dose of the line x, dfirst_x is the date of therapy start of the line x, dlast_x is the date of therapy stop of the line x, dfirst_1 is the date of the therapy start for the first line and dlast_j is the date of therapy stop for the last line j of study medication, Given that it is sorted in ascending from (first date in first line (line 1)

<table>
<tr><td></td><td>and last date in last line (line j)) in the administration page

If the AST (ALT before protocol amendment 2) values are normalized then the target dose may be adjusted from 9 mg to 6 mg per day and this should be taken into consideration for the calculation of compliance. If the AST/ALT is normal and the given dose is missing, then 6 mg should be considered as target dose.

The differentiation for the date of blood sample between before and after protocol amendment 2 will be based on the date of 22nd June 2009 (if date of blood sample ≥ “22jun2009” then consider AST for compliance calculation, else consider ALT for compliance calculation).

UDCA:
$$100 * \frac{\sum_{x=1}^{j}((\mathrm{dlast}_x - \mathrm{dfirst}_x + 1) * givendose_x\ )}{\left((\mathrm{dlast}_j - \mathrm{dfirst}_1 + 1) * [\mathrm{individual\ daily\ dose}\ ] * 250\right)}$$
Where $\mathrm{givendose}_x$ is the total daily dose of the line x, dfirst_x is the date of therapy start of the line x, dlast_x is the date of therapy stop of the line x, dfirst_1 is the date of the therapy start for the first line and dlast_j is the date of therapy stop for the last line j of study medication Given that it is sorted in ascending from (first date in first line (line 1) and last date in last line (line j))
The individual daily dose may vary from 2 to 7 capsules (see definition below).</td></tr>
<tr><td>Individual daily dose of UDCA</td><td>The definition is given in the Protocol Amendment 3, Version 3.1 / 23 September 2011 as follow:
Based on body weight, the individual daily dose of ursodeoxycholic acid (250 mg capsules) will be:
<table><tr><th>Body weight</th><th>Daily dose</th></tr><tr><td>< 47 kg</td><td>2 capsules</td></tr><tr><td>47 - 62 kg</td><td>3 capsules</td></tr><tr><td>63 - 78 kg</td><td>4 capsules</td></tr><tr><td>79 - 93kg</td><td>5 capsules</td></tr><tr><td>94 - 109 kg</td><td>6 capsules</td></tr><tr><td>≥ 110kg</td><td>7 capsules</td></tr></table>
For this analysis, the mean body weight across all study visits of an individual patient will be used for calculation of the individual daily dose of UDCA.</td></tr>
<tr><td>PBC-40</td><td>Each domain will be calculated as sum of :

Symptoms: Questions: 1-7
Itch: Questions: 8-10
Fatigue: Questions: 11-21
Cognition (Questions: 22-27
Social: Questions: 28-37*
Emotional: Questions: 38-40

*The question 37 will be considered in reverse order (not at all = 5 and very much =0).
If data are missing from a domain (typically missed or duplicated answers) the whole domain should be discarded if <50% of items are</td></tr>
</table>

	completed. If >50% of responses are present then the median value for the completed items in the domain should be ascribed to the missing item.
Modified Globe score	The original globe score as defined by Lammers et al. (4) has to be calculated at one year after start of UDCA therapy. This is not possible for this study, since all patients started UDCA therapy at different months and dosages before enrollment in the study. Therefore a modified globe score at baseline and a modified globe score at the end of the study will be calculated. Modified GLOBE score (baseline/final visit (LOCF)) = 0.044378 * age (baseline/at final visit (LOCF)) [years] + 0.93982 * loge (bilirubin times the upper limit of normal (ULN) (at baseline/at final visit(LOCF)) + (0.335648 * loge (alkaline phosphatase times the ULN (at baseline/at final visit(LOCF)) – 2.266708 * albumin level times the lower limit of normal (LLN) (at baseline/at final visit(LOCF)) – 0.002581 * platelet count per 109/L (at baseline/at final visit(LOCF)) + 1.216865. where - age is calculated at each visit, where modified GLOBE score is calculated (age=date of visit – date of birth(1st day of month)), - “bilirubin times the upper limit of normal (ULN)” is equal to the value of bilirubin divided by the ULN, - “alkaline phosphatase times the ULN” is equal to the value of alkaline phosphatase divided by the ULN - “albumin level times the lower limit of normal (LLN)” is equal to the albumin level divided by the LLN.
Cystatin C values	The parameter 'Cystatin C corrected' will be generated as follows: Cystatin C values measured up to 31 Mar 2013 will be divided by 0.89 as advised by the central laboratory, all other values will be used as reported. Reference ranges as published by Norlund et al. will be applied: Up to 50 years: 0.70 – 1.21 mg/l (or µg/ml) ≥ 51 years: 0.84 – 1.55 mg/l (or µg/ml) In order to apply the same normal range for a patient throughout the study, the age at baseline will be used to decide on the reference range for a patient. Flags for abnormality compared to the reference range will be derived subsequently based on these reference ranges as published in the literature. All summary tables will use 'Cystatin C corrected' and will have the following footnote: 'Cystatin C corrected means that some values were converted to be comparable with other values. Furthermore, generally applicable reference ranges as published in the literature were applied for all patients’.

5.2 Handling of missing data and/or invalid data and outliers

The definition of the primary efficacy variable as the rate of patients with improvement of liver histology with respect to inflammation and no progression of fibrosis compared to baseline at the individual last patient visit within the study implies using the last observation carried forward (LOCF) method (including baseline). PBC is a progressive disease that leads to cirrhosis and liver failure, i.e. in the present study the patients' condition is expected to deteriorate over time. In this clinical situation the widely used LOCF method to replace missing data might not give a conservative estimate of the treatment effect. To demonstrate robustness of the analyses based on LOCF the following sensitivity analyses of the primary efficacy variable will be performed:

- including all patients of the full analysis set who completed the 3-year treatment period,
- including all patients of the full analysis set who completed < 12 months of the 3-year treatment period (visit Month 9 + remapped visit Month 12 (Month 36/final visit, for patients who discontinued the study after Month 9)),
- including all patients of the full analysis set who completed >= 12 months, but < 36 months of the 3-year treatment period (>= visit 12 (not remapped visit) and <=visit Month 30, + remapped visit Month 36 (Month 36/final visit, for patients who discontinued the study after Month 30)),
- including all patients of the full analysis set with a baseline and a final biopsy,
- including all patients of the full analysis set with a baseline and a final biopsy, who completed the 3-year treatment period.

A patient who terminates the study prematurely due to lack of efficacy will be defined as a non-responder. A patient who does not meet the criterion of the primary efficacy variable at the final/withdrawal visit will be defined as a non-responder as well. Data of other patients who terminate the study prematurely will not be imputed.

The analysis of the primary efficacy variable is based on histological evaluations of two central pathologists regarding the Ludwig score and the mHAI score. Both pathologists will find a consistent assessment. This consent will be performed without knowledge of their primarily assessments.

The main analysis will be based on the values found by consent of the two pathologists. This consent might not be available for all patients (e.g. if the recalculated mHAI scores of the two central pathologists are identical, then values on the consent of the pathologist are not required). In this case the best values of each mHAI-item of the baseline biopsy and the worst values of each mHAI-item of the final biopsy will be taken for the analysis (worst case approach). The date of liver biopsy on the CRF will be taken into consideration (dates on assessment-sheet from 1st pathologist, on assessment-sheet from 2nd pathologist and consent-sheet will not be considered).

Additionally to this main analysis (based on the consensus) the primary endpoint will be analysed for all patients by using the assessments of each of the two pathologists as separate analyses.

Values of the efficacy and safety laboratory parameters will be displayed in the course of the study. Additionally, a visit containing the last measurement (including baseline) will be displayed by using LOCF (last observation carried forward). Values documented under the final visit/month 36 will be analysed under the next visit after premature termination (for example: Visit 2 was the last documented visit before termination → final visit will be analysed as Visit 3; if Visit 3 is not scheduled for a special examination, this visit will be analysed as LOCF visit only).

For the analyses of laboratory values the last available value per visit will be taken into account. Laboratory data with the assessment 'Technical failure' will not be taken into account.

For the calculation of durations missing values concerning day or month of start dates will be replaced by '1'. Missing values concerning day or month of end dates will be replaced by the last possible day of a month or by the last possible month of a year (=December).

No other missing values will be imputed.

6 STATISTICAL PLAN AND METHODS

The statistical summary tables will be prepared with the actual SAS® version 9.3 or higher and will be presented in the appendix of the statistical report.

Descriptive statistical methods will be used to analyze all variables specified in this section. Whenever appropriate, standard summary statistics (number of valid cases (n), arithmetic mean (mean), 95%-confidence interval (CI), standard deviation (SD), minimum value (minimum), lower quartile, median, upper quartile, maximum value (maximum)) will be calculated. Categorical values will be presented in frequency tables including the observed frequency (n) and relative frequency (% (observed frequency / number of subjects in the corresponding treatment group x 100)).

6.1 Background characteristics

6.1.1 Subject disposition

The following characteristics will be tabulated by treatment group and overall:

- Allocation of patients to centres and country,
- Study duration [months]/ treatment duration [months] and study termination (including reason for early withdrawal),
- Frequency of screened, randomised patients and allocation to analysis sets, including dates for first patient in (i.e., screened), first patient treated and last patient out (i.e., last patient's last visit),
- Frequency of reasons for exclusion from analysis sets,
- Protocol deviations.

6.1.2 Demographic, baseline characteristics and medical history

The following demographic and baseline conditions will be tabulated by treatment group and overall:

- Demographic data: Sex, age, body height [cm], body weight [kg] and BMI [kg/m²],
- Duration of disease [years],
- Results of the Liver Biopsy at baseline: Stage of fibrosis (according to Ludwig[2]), mHAI-Score (including its components), grade of inflammatory activity (according to Desmet[3]),
- Stratification factor 'original inclusion criterion no. 6' (yes/no) (see definition 6.2.1),
- Pruritus [cm, VAS].

The data of 'Medical History' and concomitant diseases (documented on the baseline visit under Physical examination) will be tabulated by treatment group and overall.

Additionally, results of the liver biopsy at baseline will be shown for randomised patients versus not randomised patients (screened patients who were not randomised).

6.1.3 Prior and concomitant medications

Each medication will be coded to a preferred term and an Anatomic Therapeutic Classification (ATC) code using Drug Dictionary.

The medication documented on the 'Prior and Concomitant medication' page will be analysed as follows:

- Prior medication, if the last intake was prior to first intake of study medication,
- Concomitant medication, if the intake was between first intake of study medication (inclusive) and study termination,
- Follow-up medication, if the first intake was on/after study termination visit.

If the day or month of the start date is unknown, the medication will be analysed using the available information. Is this information not sufficient for the division into one of the above mentioned categories, the worst case will be used (e.g. start date = NK/09/2010, end date = 13/09/2010, date of study termination = 12/09/2010 → concomitant medication).

The number and percentage of subjects taking each concomitant medication will be displayed by ATC-level 3 and preferred name for each treatment group. This display will be created for both the Per-protocol analysis set and Full analysis set. Due to the fact, that the Safety set will be the same as the Full analysis set no tables for the Safety set will be displayed.

6.1.4 Compliance

Patients who received study medication for at least one time will be used for the analysis. A total of 4 different compliance values will be calculated per patient: compliance based on returned study medication for Budesonid/Placebo and for UDCA and compliance based on administered study medication for Budesonid/Placebo and for UDCA.

Patients will be assessed as compliant regarding Budesonid/Placebo if the calculation (based on returned or administered medication Budesonid/Placebo) results in a compliance rate between 80% and 120% (compliance=80% and compliance=120 means patient is compliant). Patients will be assessed compliant regarding UDCA if the calculation (based on returned or administered medication UDCA results in a compliance rate between 80% and 120% (compliance=80% and compliance=120 means patient is compliant).

Summary tables for compliance values based on returned study medication, compliance values based on documentation of administered medication and compliance categories (<80; 80-120; >120%) for both calculations (based on returned medication and based on documentation of administered medication) will be provided both for Budesonid/Placebo and UDCA groups.

6.2 Efficacy analysis

6.2.1 Primary efficacy analysis

The primary efficacy variable is the rate of patients with response defined as improvement in liver histology with respect to inflammation (at least 3 points in the mHAI sum score or no inflammatory activity) and no progression of fibrosis compared to baseline at the individual last patient visit within the study.

The evaluation of the primary efficacy variable in this final analysis will be performed for the FAS (intention-to-treat analysis) and for the per-protocol set (per-protocol analysis). The primary analysis will be the intention-to-treat analysis.

The following null hypothesis H_0 will be tested against the alternative hypothesis H_A ($\alpha = 0.025$, one-sided):

H_0: $\pi_A - \pi_B \leq 0$

H_A: $\pi_A - \pi_B > 0$

π_A and π_B denote the response rate in the group of patients receiving ursodeoxycholic acid plus budesonide (treatment group A) and in the group of patients receiving ursodeoxycholic acid plus placebo (treatment group B), respectively.

For confirmatory hypothesis testing the Cochran-Mantel-Haenszel test for comparing two rates with compliance regarding original inclusion criterion no. 6 as stratification factor will be used ($\alpha = 0.025$, one-sided). The compliance regarding original inclusion criterion no. 6 refers to patients meeting the inclusion criterion no. 6 according to the original study protocol effective until amendment no. 3 vs patients randomised in line with inclusion criterion no. 6 as defined in amendment no. 3, who do not meet the original inclusion criterion no. 6.

The original inclusion criterion no. 6 is given as follows:

- Serum alkaline phosphatase ≥ 3 times the upper limit of normal (corresponding to 312 U/L for women and 387 U/L for men) at any time since diagnosis of PBC

or

- ALT or AST ≥ 2 times upper limit of normal (corresponding to 70 U/L for women and 100 U/L for men) at inclusion

or

- Total Bilirubin ≥ 1.0 mg/dL (≥ 17 µmol/L)

or

- Moderate to severe periportal or periseptal lymphocytic interface hepatitis

or

- Periportal and portal fibrosis with numerous septa (Ludwig stage III) without cirrhosis.

After amendment no. 3 inclusion criterion no 6. is defined as:

PBC patients with an incomplete response to ursodeoxycholic acid treatment defined by the failure to achieve s-AP levels < 1.5x upper limit of normal after at least 6 months of treatment with ursodeoxycholic acid and with inflammatory activity according to the mHAI sum score in the histological assessment of the liver.

The study was stopped at the interim analysis since only a minor trend for a treatment effect was observed and due to the increase in patient numbers needed to confirm the treatment effect and to insufficient patient recruitment seen in this study

For the final analysis all patients will be included into stage I (all patients of the first interim analysis + overrun).
The null hypothesis will be tested and the null hypothesis will be rejected, if the test statistic exceeds the critical value 3.641. In this case, superiority of budesonide compared to placebo will have been demonstrated.

For confirmatory hypothesis testing, the inverse normal method of combining p-values of the stratified normal approximation test for two rates will be used, where compliance regarding original inclusion criterion no. 6 will be the stratification factor.

For estimating the treatment effect, the difference between the response rates and the corresponding two-sided 95% multiplicity-adjusted repeated confidence interval for this pairwise difference will be provided.

To demonstrate robustness of the analyses, the following sensitivity analyses of the primary efficacy variable will be performed:

- including all patients of the full analysis set who completed the 3-year treatment period,
- including all patients of the full analysis set who completed < 12 months of the 3-year treatment period (visit Month 9 + remapped visit Month 12 (Month 36/final visit, for patients who discontinued the study after Month 9)),
- including all patients of the full analysis set who completed >= 12 months, but < 36 months of the 3-year treatment period (>= visit 12 (not remapped visit) and <=visit Month 30, + remapped visit Month 36 (Month 36/final visit, for patients who discontinued the study after Month 30)),
- including all patients of the full analysis set with a baseline and a final biopsy,
- including all patients of the full analysis set with a baseline and a final biopsy, who completed the 3-year treatment period.

Additionally the primary efficacy variable will be analysed by Country displaying the response rates without adjustment and treatment effect only.

6.2.2 Secondary efficacy analysis

Secondary endpoints will be analysed for full analysis set (intention-to-treat analysis) and per-protocol set (per-protocol analysis). Continuous variables will be summarized at each scheduled measuring time point using the following standard descriptive summary statistics: number of observations, arithmetic mean, standard deviation, minimum, lower quartile, median, upper quartile, maximum. Categorical data will be described using absolute and relative frequencies. All efficacy variables will be tabulated by treatment group and overall.

The following secondary endpoints will be analysed in this study.

<u>Rates</u>

The following rates will be evaluated by analogy to the primary efficacy variable at the patient's last individual visit.

- Rate of patients presenting with cirrhosis or esophageal varices and/or ascites at the end of treatment or patients registered on the liver transplant waiting list or patients with liver related death during 3 years of treatment,
- Rate of patients with improvement of liver histology with respect to stage (according to Ludwig[2]),
- Rate of patients with improvement of liver histology with respect to grade (according to Desmet et al.[3]) and stage (according to Ludwig2).
- Rate of patients with improvement of liver histology with respect to grade (according to Desmet et al.[3]).

To compare the treatment groups a stratified Cochran-Mantel-Haenzel-Test ($\alpha = 0.025$, one-sided) will be applied. Resulting p-values will be interpreted in the exploratory sense.

Normalisation of serum levels of AP and ALT

- Frequencies of normal, increased and decreased AP and ALT values will be computed at each scheduled measuring time point.
- Shift tables displaying changes with respect to the normal range between baseline and each scheduled measuring time point after baseline will be provided.

Normalisation of serum levels of AP or reduction of baseline AP levels by at least 40%

- Rate of patients with improvement or reduction of at least 40% will be calculated and displayed for each scheduled measuring time point,

Improvement of serum levels of AP, bilirubin, AST and GGT

- Standard descriptive summary statistics will be calculated for AP, bilirubin, AST and GGT values at each scheduled measuring time point.
- Standard descriptive summary statistics will be calculated for the absolute and relative change from baseline to each scheduled measuring time point after baseline.

Histology

- Data of the histological assessment according to Ludwig[2] will be summarized by means of standard descriptive statistics at baseline and final/ withdrawal visit.
- The assessment of inflammatory activity (mHAI-score and grading according to Desmet[3] et al.) will be summarized by means of standard descriptive statistics at baseline and final/ withdrawal visit.
- Tables presenting changes from baseline will be provided.

Course of pruritus (measured by VAS)

- The course of pruritus will be displayed summarising assessments at each scheduled measuring time point.
- Tables presenting changes from baseline will be provided.

Course of fatigue/Quality of Life

- The fatigue/Quality of Life data measured by PBC-40 will be summarised by means of standard descriptive statistics at baseline, at V7, V9 and final/ withdrawal.
- In addition, descriptive summary statistics will be calculated for the absolute changes from baseline.

Course of Mayo Risk Score adapted

- The course of the Mayo Risk Score adapted will be summarized by means of standard descriptive statistics at each scheduled measuring time point. For the analysis the recalculated score will be used.
- Tables presenting changes from baseline will be provided.
- Sensitivity analysis will be conducted by assuming that patients with no varices at baseline and esophageal varices at final visit will be having on one side varices at V7 (visit at 12 months) and V9 (visit at 24 months) (point for presence of varices=1) and on another side having no varices at V7 (visit at 12 months) and V9 (visit at 24 months) (point for presence of varices =0).

Assessment of Efficacy

- For assessment of efficacy by investigator/patient, frequencies for the respective categories (very good, good, satisfactory, poor) at each scheduled measuring time point will be provided.

Further analyses concerning laboratory data

- Rate of patients with fall in platelet counts < 135.000/mm³ will be calculated and displayed for each scheduled measuring time point,
- Rate of patients with fall in the prothrombin ratio (INR) of more than 25% and to a value of prothrombin time < 70% will be calculated and displayed for each scheduled measuring time point,
- The course of Fibrosis marker (hyaluronic acid) will be displayed summarising assessments at each scheduled measuring time point.
- Rate of patients with doubling in the serum concentration of hyaluronic acid to more than 80 ug/L or decrease to a value of more than 100 ug/L will be calculated and displayed for each scheduled measuring time point,
- Rate of patients with serum bilirubin levels more than 50 µmol/L (3.0 mg/dL) will be calculated and displayed for each scheduled measuring time point,
- Rate of patients with fall in serum albumin count of 10% and to a value of less than 34 g/L will be calculated and displayed for each scheduled measuring time point,

  Rate of patients with AP<=1.67 ULN AND >=15% decrease in AP AND total bilirubin within the normal limits at 1year, 2 years, 3 years (AP and total bilirubin measured at the same visit). This secondary endpoint will be analysed for the FAS, PP- and modified-PP- sets.

  Definition of time frames:

  1 year: all data documented at visit Month 12, without using remapped visit Month 12 (Month 36/final visit, for patients who discontinued the study after Month 9)

  2 years: all data documented at visit Month 24, without using remapped visit Month 24 (Month 36/final visit, for patients who discontinued the study after Month 18)

  3 years: all data documented at visit Month 36, without using remapped visit Month 36 (Month 36/final visit, for patients who discontinued the study after Month 30)

  1 year (LOCF): all data documented at visit Month 12, missing values will be replaced by using LOCF

  2 years (LOCF): all data documented at visit Month 24, missing values will be replaced by using LOCF

  3 years (LOCF): all data documented at visit Month 36, missing values will be replaced by using LOCF

GLOBE score:

- Summary at baseline/final visit,
- Summary of absolute change from baseline to final visit,

- Subgroup analyses (GLOBE Score <=0.30 (responder) versus GLOBE score >0.30 (non-responder)) will be conducted for the primary efficacy variable (rate of patients with response defined as improvement in liver histology with respect to inflammation and no progression of fibrosis compared to baseline at the individual last patient visit within the study).

To compare the treatment groups a mixed effects model for repeated measures will be applied with treatment group, visit and (treatment group x visit)-interaction as independent factors and the respective baseline value as covariate. This applies to the course of pruritus and to the ALT and AP values. Resulting p-values will be interpreted in the exploratory sense.

6.3 Safety analysis

The safety evaluations will include analyses of AEs, vital signs, laboratory measurements, bone metabolism, bone mineral density, suppression of adrenal function and assessment of tolerability by investigator and patient.

The analysis of safety will be based on the safety analysis set.

6.3.1 Adverse events

All adverse events with onset on or after treatment day 1 will be analysed as treatment emergent adverse events (TEAE).

Adverse events which occurred more than 30 days after last intake of study medication will be analysed as follow-up adverse events.

If the date of first intake of study medication is unknown, the adverse event will be counted as treatment emergent adverse event, if its onset is on/after visit 1. If the start day of an adverse event is unknown and the start month is in the range between the month of first intake of study medication and the month of last intake of study medication it also will be counted as treatment emergent adverse event.

For the final analysis the following analyses concerning adverse events (AE) will be performed for the Safety Analysis Set:

Table:

- Overview of adverse events (this includes number and frequency of patients with at least one AE, Number and frequency of patients with at least one treatment-emergent AE, Number and frequency of patients with at least one follow-up AE, Number and frequency of patients with at least one drug-related TEAE (causality = certain, probable/likely or possible) to Budesonide/ Placebo, Number and frequency of patients with at least one drug-related TEAE (causality = certain, probable/likely or possible) to UDCA, Number and frequency of patients with at least one treatment emergent serious adverse event (SAE), Number and frequency of patients with at least one AE leading to premature discontinuation of study medication, Number and frequency of patients with at least one TEAE leading to premature study termination
- Overview of the absolute and relative frequencies of patients with at least one TEAE by seriousness, severity, causality, measures and outcome
- Summary of treatment-emergent AEs by system organ class and preferred term and, in addition, by seriousness, severity, causality, measures and outcome.
- Summary of follow-up AEs by system organ class and preferred term and, in addition, by seriousness, severity, causality, measures and outcome.
- Summary of drug-related treatment-emergent AEs by system organ class and preferred term
- Summary of treatment emergent serious adverse events by system organ class and preferred term
- Summary of treatment emergent adverse events leading to premature discontinuation of study drug by system organ class and preferred term

Listings:

- Death - AEs with outcome=’fatal’ including investigator's term, system organ class, preferred term, start date, stop date, time interval between last administration of study medication and onset of AE, frequency, seriousness, severity, causality, measures, challenge and outcome;
- SAEs (including AEs with outcome = `fatal`) including investigator's term, start date, stop date, time interval between last administration of study medication and onset of AE, frequency, seriousness, severity, causality, measures, challenge and outcome;
- Unexpected adverse drug reactions (regarding Budesonide) including investigator's term, system organ class, preferred term, start date, stop date, time interval between last administration of study medication and onset of AE, frequency, seriousness, severity, causality, measures, challenge and outcome;
- AEs leading to premature discontinuation of study drug or premature study termination due to AEs including investigator's term, system organ class, preferred term, start date, stop date, time interval between last administration of study medication and onset of AE, frequency, seriousness, severity, causality, measures, challenge, outcome, cause of termination and a “flag” which indicates whether this AE was expected or not.

6.3.2 Clinical laboratory evaluation

Routine laboratory parameters will be measured within the scope of the baseline examinations (within four weeks), at all Interim Visits and at the Final Visit.

The following routine laboratory parameters will be determined:

- Haematology: Leukocytes, erythrocytes, haemoglobin, haematocrit, mean corpuscular volume, platelets,
- Biochemistry: Creatinine, total protein, serum albumin, serum bilirubin, α1-globulin, α2-globulin, β globulin, γ-globulin, ferritin, glucose (fasting), potassium, sodium, hCG (baseline and final visit, women of childbearing potential only), cystatin C (every 12 months), α1-anti-trypsin (baseline visit only), IgA Anti-transglutaminase (baseline visit only)

- Liver function parameters: AST, ALT, γ-GT, total bilirubin, alkaline phosphatase, cholinesterase, GLDH
- Coagulation: Prothrombin time (PT) (% and INR), PT(sec) and partial thromboplastin time (PTT)
- Urinalysis: pH, glucose, protein, leukocytes and erythrocytes.

Laboratory data will be subjected to both a quantitative analysis (descriptive summary statistics) and qualitative analysis where frequencies of normal, increased and decreased values will be computed. By analysing Cystacin C, the change of the value as well as the change of the lower/upper limit over the time will be taken into consideration (see also 5.1, 'Cystatin C corrected').

The following outputs will be provided for laboratory parameters (Haematology, Biochemistry, Coagulation, and Urinalysis):

- Summary at each scheduled measuring time point (Haematology, Biochemistry ,Urinalysis and Coagulation (INR only)).
- Summary of absolute change from baseline to each scheduled measuring time point after baseline (Haematology, Biochemistry and Coagulation (INR only)).
- Frequencies of normal, increased and decreased values at each scheduled measuring time point (Haematology, Biochemistry, Coagulation, and Urinalysis).
- Shift tables displaying changes with respect to the normal range between baseline and each scheduled measuring time point after baseline (Haematology, Biochemistry, Coagulation, and Urinalysis).
- A listing of all patients with abnormal (increased or decreased) values at any time point.
- The incidence of newly occurring and worsening abnormalities of laboratory parameters according to their direction (increase or decrease) values at any time point.

The following outputs will be provided for Bone metabolism (Bone specific AP), suppression of adrenal function (plasma cortisol), bone mineral density (DXA):

- Summary at each scheduled measuring time point.
- Summary of absolute change from baseline to each scheduled measuring time point after baseline.
- A listing of all patients with abnormal (increased or decreased) values at any time point.
- The incidence of newly occurring and worsening abnormalities of laboratory parameters according to their direction (increase or decrease) values at any time point.

6.3.3 Assessment of tolerability

For assessment of tolerability by investigator/patient, frequencies for the respective categories (very good, good, satisfactory, poor) at each scheduled measuring time point will be provided.

6.3.4 Vital signs

Measurements of vital signs (blood pressure, pulse rate, body weight) at each scheduled time points will be summarized. Further, standard descriptive summary statistics will be computed for the absolute change from baseline to each scheduled time point after baseline.

A patient data listing of all vital signs measurements will be provided.

In addition, a patient data listing of all clinically significant abnormal vital signs will be provided.

- Weight < 30 kg or weight > 150 kg or a decrease /increase of 5 kg during study period was observed or
- Pulse rate < 50 per min or > 100 per min or
- Systolic blood pressure < 90 mmHg or > 180 mmHg or diastolic blood pressure < 60 mmHg or > 100 mmHg or
- Adverse event.

7 CHANGES FROM THE PLANNED ANALYSIS IN STUDY PROTOCOL

Additional analysis concerning efficacy:

- Absolute and relative change of serum levels of AP, bilirubin, AST and GGT
- Rate of patients with AP<=1.67 ULN AND >=15% decrease in AP AND total bilirubin within the normal limits at 1yr, 2yrs, 3yrs (and 1 year LOCF, 2 years LOCF, 3 years LOCF)
- Rate of patients with improvement of liver histology with respect to grade (according to Desmet et al.[3])

Additional outputs will be provided for Bone metabolism (Bone specific AP), suppression of adrenal function (plasma cortisol), bone mineral density (DXA):

- A listing of all patients with abnormal (increased or decreased) values at any time point.
- The incidence of newly occurring and worsening abnormalities of laboratory parameters according to their direction (increase or decrease) values at any time point.

A new score, the modified GLOBE score, will be calculated at baseline and at final visit (LOCF) (Lammers et al. (4)) and analysed as follow:

- Summary at baseline/final visit,
- Summary of absolute change from baseline to final visit,
- Subgroups analyses will be conducted GLOBE score <=0.30 versus GLOBE score >0.30 for the primary endpoint.

Additional subgroups for the analysis of the primary endpoint were added:

- including all patients of the full analysis set who completed < 12 months of the 3-year treatment period,
- including all patients of the full analysis set who completed >= 12 months, but < 36 months of the 3-year treatment period

The Mayo Risk Score as defined in the protocol was not calculable and replaced by the Mayo Risk Score adapted defined as:

Mayo Risk Score adapted = $0.0295 \times$ (age in years) + $0.5373 \times \log_e$ (total bilirubin in mg/dL) - $0.8389 \times$ (serum albumin in g/dL) + $0.5380 \times \log_e$ (AST in IU/L) + $1.2426 \times$ (points for presence of varices).

Analysis of Mayo Risk Score adapted: Sensitivity analysis will be conducted by assuming that patients with no varices at baseline and esophageal varices at final visit will be having on one side varices at V7 (visit at 12 months) and V9 (visit at 24 months) (point for presence of varices=1) and on another side having no varices at V7 (visit at 12 months) and V9 (visit at 24 months) (point for presence of varices =0).

Results of the liver biopsy at baseline will be shown for randomised patients versus not randomised patients (screened patients who were not randomised).

Definition of a new population (modified Per-Protocol)

Analysis of one particular secondary endpoint (Rate of patients with AP<=1.67 ULN AND >=15% decrease in AP AND total bilirubin within the normal limits at 1yr, 2yrs, 3yrs (and 1 year LOCF, 2 years LOCF, 3 years LOCF)) for the mPP-population

8 REFERENCES

1 Ishak K, Baptista A, Bianchi L, Callea F, De GJ, Gudat F, et al. Histological grading and staging of chronic hepatitis. J Hepatol. 1995;22(6):696-9.

2 Ludwig J. The nomenclature of chronic active hepatitis: An obituary. Gastroenterology. 1993;105(1):274-8.

3 Desmet VJ, Gerber M, Hoofnagle JH, Manns M, Scheuer PJ. Classification of chronic hepatitis: diagnosis, grading and staging. Hepatology. 1994;19(6):1513-20.

4 Lammers WJ, Hirschfield GM, Corpechot C, Nevens F, Lindor KD, Janssen HLA, Floreani A, Ponsioen CY, Mayo MJ, Invernizzi P, Battezzati PM, Parés A, Burroughs AK, Mason AL, Kowdley KV, Kumagi T, Harms MH, Trivedi PJ, Poupon R, Cheung A, Lleo A, Caballeria L, Hansen BE, van Buuren HR, on behalf of the Global PBC Study Group, Development and Validation of a Scoring System to Predict Outcomes of Patients With Primary Biliary Cirrhosis Receiving Ursodeoxycholic Acid Therapy, *Gastroenterology* (2015), Vol 149 (7):1804-1812

5 Norlund L, Fex G, Lanke J, Von Schenk H, Nilsson JE et al. Reference intervals for the glomerular filtration rate and cell-proliferation markers: Serum cystatin C and serum β_2-microglobulin/cystatin C-ratio. *Scan. J. Clin. Lab. Invest*. 54 (6) 463-470 (1997)