Newcastle Clinical Trials Unit

Statistical Analysis Plan for the RITPBC Trial

B-Cell Depleting Therapy (Rituximab) as a Treatment for Fatigue in Primary Biliary Cirrhosis

SAP Version number: 1.0
SAP date: 08/12/2016

ISRCTN Number: 03978701
EudraCT Number: 2012-000145-12
Clinicaltrials.gov number: NCT02376335
REC Reference: 12/NE/0095
NUTH Reference: 5997

Trial Statistician: Andrew Bryant

Signature: **Date**:

Reviewed and Approved by:

Signature: Denise Howel (Senior statistician) **Date**:

Signature: Prof. Dave Jones (Chief investigator) **Date**:

CONTENTS

1. INTRODUCTION

This statistical analysis plan (SAP) provides guidelines and presentation for the analysis of the RITPBC trial. This plan, along with all other documents relating to the analysis of this trial, will be stored in the 'Statistical Documentation' of the Trial Master File (TMF) and the final signed SAP will be stored in section 16 of the TMF (16. Statistics / 16.1 Final signed Statistical Analysis Plan).

1.1 Trial Summary

Introduction
Primary Biliary Cirrhosis (PBC) is a liver disease that predominantly affects females, can present for the first time at any age, and which develops over many years. It is caused by the immune system attacking the body's own tissues. People with PBC frequently experience profound fatigue or tiredness which they liken to their "batteries running down", and although people still want to undertake normal activities they simply lack the energy to be able to do them. This reduces quality of life, makes it difficult for people to work, and can end up with them becoming isolated in the community. At present we have no treatment for fatigue in PBC. Finding a treatment for fatigue in PBC is one of the highest research priorities identified by patient groups.

We have shown that PBC patients with fatigue have an abnormality in the way they generate energy within their muscles. This appears to be associated with the presence of an antibody in the blood which is directed against an important protein which normal cells in the body use to generate energy. In recent years new drug treatments have been developed which allow us to safely suppress the part of the immune system which produces antibodies of the type that seem to cause energy production problems in PBC. As yet, however, the extent to which these medicines can improve fatigue through removal of antibodies in PBC has not been tested.

Background
The aim of this study is to undertake a clinical trial to examine the effects of this treatment ("Rituximab") on severe fatigue in PBC to help us understand whether this will be a potentially useful treatment. This will give us information about how energy generation changes in patients with PBC with and without the treatment and will also help us to develop new treatments for fatigue in other diseases. The study has the potential to improve the quality of life of many patients with PBC, for whom there is currently no hope of improvement.

We will perform a randomised controlled trial of Rituximab therapy in PBC compared to placebo with the primary end point of fatigue severity. The study will be performed in a specialised PBC clinical centre.

Trial hypothesis
The B-cell-directed immunotherapeutic agent Rituximab will improve fatigue in PBC (an important and disabling symptom) through its effect on B-cells producing antibodies which inhibit the function of pyruvate dehydrogenase (PDH) an important energy generating enzyme and/or inflammatory cytokines.

Primary objective
To compare the efficacy of B-cell depleting therapy in Primary Biliary Cirrhosis patients followed up for 12 months.

Secondary objectives

- To prospectively evaluate the efficacy and influence of Rituximab upon muscle bioenergetics in Primary Biliary Cirrhosis.

- To examine the effects of Rituximab on the immune function of B-cells in PBC, and explore the links between those changes and impact on fatigue.
- To identify whether improvements in fatigue in Primary Biliary Cirrhosis associate with changes in muscle bioenergetics and /or physical activity levels

Study design
A phase II, single-centre, randomised controlled, double blinded trial comparing Rituximab with placebo in fatigued Primary Biliary Cirrhosis patients over 12 months.

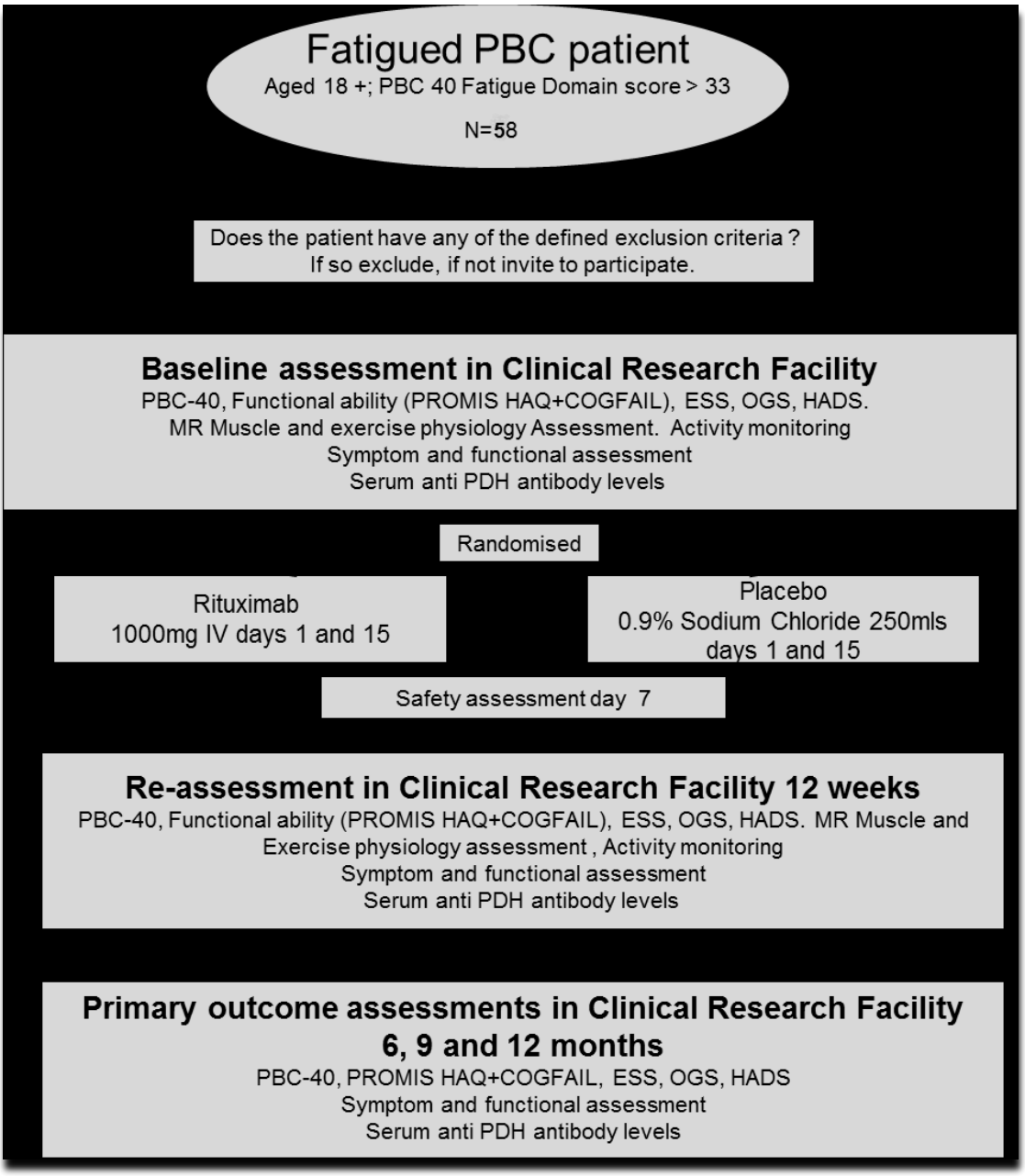
Patient population
Participants will be patients with definite or probable Primary Biliary Cirrhosis established using recognised epidemiological criteria.

Study intervention
The investigational medicinal product used in the clinical trial is Rituximab, 1000mg IV. Patients randomised to receive placebo will receive a control normal saline infusion.

Primary Outcome
The primary outcome variable is fatigue severity in PBC patients, assessed using the fatigue domain of the PBC-40, a fully validated, psychometrically robust, and disease specific quality of life measure.

Target Recruitment
The initial recruitment target was 78 participants but this was revised down to 58 participants when the extension was obtained. Recruitment took place at a single site over a period of 3 years (Recruitment opened 01/10/2012 and closed 01/10/2015).

1.2 Trial Flowchart



2. TIMING AND REPORTING OF INTERIM AND FINAL ANALYSES

The end of the study is defined as last patient, last visit (12 month follow-up visit).

There are no planned interim analyses for efficacy. However if the DMC requires interim analysis for safety then this will be performed, although there are no formal stopping rules implemented within the trial. Final analyses will be carried out when all participants have been followed up.

Previous DMC meetings were held on:

- 14th February 2013
- 9th September 2013
- 25th March 2014
- 23rd September 2014
- 16th March 2015
- 14th September 2015
- 14th June 2016

For all the DMC meetings after the initial one, an open and a closed report are presented. In the closed report, rates of attrition are presented along with summary statistics on PBC-40 score (including fatigue domain score), and PROMIS HAQ and HADS (including anxiety and depression domains) scores. Laboratory outcomes (e.g. haemoglobin, white blood count, alkaline phosphatise, etc) are also presented as well as B-cell data. Outcomes are reported at baseline then quarterly up to 12 months. The closed report has data split by trial arm (labelled as A and B) so interventions were masked in the analyses. The final DMC is scheduled for 31st January 2017.

The final analysis timeline based on planned FU will be December 2016 through to early 2017 (end of study at 12 month FU of last patient was 12th September 2016 with analyses and report to be completed thereafter).

After completion of database reports and checks, the data will be released to the trial statistician no later than an agreed date that will be specified in advance of the data analysis.

3. RECRUITMENT AND RANDOMISATION

3.1 Recruitment

The first RITPBC patient was recruited on 29th October 2012 and the last patient was recruited on 30th September 2015. Randomisation of this last patient did not take place until 18th November 2015. The trial was granted an extension to the initial planned 24 month recruitment period due to slow initial recruitment. A total of 57 participants were recruited in the study.

All 57 participants were recruited from a single centre at Newcastle.

3.2 Randomisation

Randomisation has been conducted by the Newcastle Clinical Trials Unit (NCTU) web based system on a 1:1 ratio and random-permuted blocks with random block length. The treatment allocation will be kept blind from the subjects and the study assessors and investigators until study completion. The randomisation system has generated a treatment arm for each participant that links to the corresponding allocated study drug (blinded). The participant study ID has been clearly documented by the investigator on the trial prescription to ensure the study pharmacist dispenses the correct study medication.
Patients in the study are randomised to receive either:

- Rituximab therapy on days 1 and 15 - study drug (n=29)
- Placebo (0.9% Sodium Chloride 250ml) on days 1 and 15 - control (n=29)

3.3 Ineligible Patients

If any patients are randomised and are subsequently found to be ineligible or withdraw from treatment, all efforts will be made to report the reason for withdrawal as thoroughly as possible. Should a patient withdraw from study drug only, efforts will be made to continue to obtain follow-up data, with the permission of the patient. All statistical analyses will be carried out on an intention to treat basis, retaining patients in their randomised treatment groups and including protocol violator and ineligible patients. Ineligible patients are classed as those randomised patients who are found to subsequently not adhere to the eligibility criteria of the trial. The number of ineligible patients and reasons for ineligibility will be reported and a sensitivity analysis may be conducted and reported if the number of ineligible patients is excessive. Protocol violators will be reported as part of treatment compliance.

3.4 Blinding

Assignment to either active or placebo arm will be blinded to both the participant and investigators/assessor (double-blind). A code-break list will be kept in pharmacy; this list should be accessed only in an emergency (preferably with authorisation from the Chief Investigator or Medical Monitor) and the Chief Investigator immediately informed. If the code is broken, details including the participant number, the person who broke the code, why and when the code was broken shall be recorded and maintained in the site file. Code breaks will not be routinely opened for participants who complete study treatment.

At the final visit, the integrity of the blind will be assessed by asking both the participants and their treatment assessor: “Do you think you were receiving Rituximab or the dummy solution? Why do you think this?” The treatment assessor will be asked to record their answer on a separate CRF, and prior to asking the participant to avoid bias.

4. DATA QUALITY

4.1 Forms Returned

Data are collected using case report forms (CRF). Completion rates for each CRF will be reported. CRF’s are completed and collated in the following order:

i) Registration & randomisation form - completed prior to treatment allocation
ii) On Study form - baseline assessment
iii) Treatment forms - one form for each of two infused doses of Rituximab (1000 mg IV) or placebo (0.9% Sodium Chloride 250mls) at visit 2 (day 1) and visit 4 (day 15).
iv) End of Treatment form
v) Follow-up form - one every three calendar months from date of randomisation
vi) Questionnaires for outcomes - baseline, 12 weeks, 6, 9 and 12 months
vii) Serious Adverse Event forms

4.2 Follow-Up

The number of alive patients lost to follow-up will be reported. Follow-up will be compared across treatment groups at baseline, 12 weeks, 6, 9 and 12 months. We will create histograms to check how close these follow up times are to 3, 6, 9 and 12 months and report major departures from these times.

5. STUDY POPULATION

5.1 Defining Populations for Analysis

Analysis will use the **intention-to-treat** population, which includes all participants in the group to which they were randomised, regardless of the intervention that they received. The number of participants who did not receive the intervention to which they were randomised will be reported. We will use a pragmatic ITT approach where patient outcomes are reported at their visit nearest the scheduled appointment date as we do not expect departures from study visits. However, if necessary, we will carry out a sensitivity analysis on the primary outcome using compliant ITT (+/- 1 wk (11-13 weeks) at visit 16) and report the characteristics of participants outside of this.

The number of protocol violators and reasons for violation will be reported if possible. If the proportion of protocol violators is high (say ≥20% non-compliance or other protocol violation) a sensitivity analysis based on actual treatment received may be carried out to investigate the robustness of the conclusions of the study (See section 9 for further details).

5.2 Baseline Patient Characteristics

Demographic and clinical characteristics at baseline will be reported across the two groups descriptively. No significance testing will be carried out due to the randomised nature of the study.

The following data will be recorded (see dummy Table 2):

a. Gender (Male or female)

b. Age in years (DOB and date of randomisation can be used to deduce this)

c. Ethnicity

d. Smoking history (never, ex, current)

e. Alcohol consumption (units per week)

f. BMI (height and weight can be used to calculate this; weight in kg/height in metres^2)

g. Patient location (managed by Newcastle centre for at least one year or not)

h. UDCA use

    a. Yes/no

    b. If yes, responder or non-responder (People on UDCA are responders if they have an Alk Phos of <200 AND a bilirubin of less than 20 when they are enrolled

i. UK PBC risk score at 10 years[a] (outcome). The UK-PBC Risk Score is the projected risk (expressed in percentage) of a PBC patient developing liver related complications (defined as liver failure requiring liver transplantation or liver related death) at 10 years.

Note:
(a) This score is applicable irrespective of being on Ursodeoxycholic acid or not
(b) The 10 year prediction is from time of blood results, not time of diagnosis

6. TREATMENT RECEIVED

Patients will be randomised to receive two infused doses of Rituximab (1000 mg IV) or placebo (0.9% Sodium Chloride 250mls) at visit 2 (day 1) and visit 4 (day 15). In line with recommendations, a conditioning regimen consisting of Paracetamol 1g PO, Chlorphenamine 4mg PO, and Methylprednisolone 100mg IV will be given 30 minutes before each infusion of Rituximab/placebo. Intervention will be given in the setting of a dedicated immunotherapy clinical trials unit.

The proposed clinical trial is single centred, and will be performed in the Newcastle upon Tyne Hospitals NHS Foundation Trust (NUTH). Recruitment was planned to be principally from the large clinical cohort under follow-up at this centre (>500, the largest clinical PBC service in the UK), but recruitment from outside the region was necessary due to slow accrual. Additional patients were signposted via the Northumberland Tyne and Wear and County Durham and Tees Valley CLRNs (consisting of an additional 10+ clinical centres managing PBC patients) and participation identification centres established via the CLRN Hepatology Speciality Groups and the Autoimmune Study Group. In addition CLRNs across the North of England were approached for additional patients. It was also necessary to recruit patients from outside of the North East using the UK PBC Trial platform which required a substantial amendment to the protocol.

A complete breakdown of patient visits and reported outcomes is outlined in the Table 1 (see section 8.2). Where feasible, study visits coincide with routine clinical follow-up to enhance the likelihood of good compliance.

7. SAFETY ANALYSIS

Safety will be assessed in terms of numbers of adverse events (AEs) and adverse reactions in the study groups.

The number of each grade of AE, of each category of AE, will be listed by treatment group, stating those deemed to be treatment related.

The number of treatment related serious adverse events (SAEs), including treatment related deaths, are reported divided by their relationship as ‘definitely’, ‘probably’ and ‘possibly’ related to treatment. The proportions of patients with SAEs will be compared descriptively across treatments and differences assessed for clinical significance.

Study drug must be discontinued if:

- the participant develops elevated serum Alanine Transaminase (ALT)/Aspartate Transaminase (AST) 4 times above normal limits for each local laboratory
- the participant decides she/he no longer wishes to continue
- cessation of study drug is recommended by the investigator

8. OUTCOMES

8.1 Definition and Calculation of Efficacy Outcome Measures

Full details about how questionnaire scores (as well as other outcomes) are calculated (as well as dealing with missing data and scale recoding) for each primary and secondary outcome are given in Appendix 1.

All efficacy outcomes are measured on a numeric scale

Primary outcome measure:

To address the effect of Rituximab on fatigue in PBC the following primary outcome measure was chosen:

The primary outcome variable is fatigue severity in PBC patients, assessed using the fatigue domain of the PBC-40[b], a fully validated, psychometrically robust, disease specific quality of life measure, between baseline and 12 week assessment (PBC-40 fatigue domain score >33 at outset).

Secondary outcome measures:

Symptom severity scores other than fatigue will be calculated from the following questionnaires:

1. Other domains of the PBC-40 questionnaire[b], namely:
    a. **Itch**
    b. **Cognitive**
    c. **Social**
    d. **Emotional**
    e. **Other symptoms**

A total PBC-40 score will not be computed as it has no meaningful interpretation.

Clinical symptom and functional capability scales

2. Epworth Sleepiness Scale **(ESS) score[c,d]** to assess daytime somnolence
3. Orthostatic Grading Scale **(OGS) score[e]** to assess vasomotor autonomic symptoms
4. Patient-Reported Outcomes Measurement Information System Health Assessment Questionnaire **(PROMIS HAQ) score[f,g]** to assess functional status
5. Cognitive Failure questionnaire **(COGFAIL) score[h]**.
6. Hospital Anxiety and Depression Scale **(HADS) score[i]** to assess depressive and anxiety-related symptoms

To further address the effect of Rituximab on fatigue in PBC the following secondary outcome measures were chosen along with safety parameters:

7. **Average perceived fatigue score[l]** will be calculated from participant held fatigue diaries. The diaries measured fatigue using a scale of 1 to 6, where 1 represents no fatigue and 6 extreme fatigue. Participants were asked to complete the diaries six times during the study. They completed the diaries for a period of a week at the beginning of each month at visits: baseline, 1, 3, 6, 9 and 12 months. They returned the diaries at the final visit. The average (mean) score (and SD) will be computed if patient's completed at least 5 days in the week out of the requested 6 times and the overall mean score will range between 1 and 6 inclusive (and be calculated using all days reported by the participant in the diary).
8. **Wrist acceleration per 5 second epoch was calculated with metric Euclidian Norm Minus One (ENMO)**; physical activity assessed using seven day physical activity monitoring[j,k] (previously shown to be impaired in fatigued PBC patients with degree of impairment shown to associate with perceived fatigue severity). Participants wore the

accelerometer device for 7 days (minimum of 5 days on body including at least one weekend day and only days with at least 22 hours of valid data were retained for further analysis). The first and last hour of the measurement were excluded as they are expected to be influenced by the monitor distribution and collection procedure. Monitor non-wear was detected as described previously and imputed by the average accelerometer data on similar time points on different days of the measurement. ENMO was used to summarise the average magnitude of dynamic wrist acceleration over the measurement period. The output from metric ENMO is in m*g* (1m*g* = 0.001*g* = 0.001 x 9.8 m/s^2 = 0.001 x gravity). The following accelerometer measures were assessed pre versus post intervention:

a. **Average magnitude of wrist acceleration per 5 second epoch (millig)**
b. **Average acceleration (millig) during the most active (M5) 5 hr period of each day**

Laboratory parameters

9. **Full blood count (FBC)**
    a. **Haemoglobin (g/L)** (safety parameter)
    b. **White blood count (WBC) (10^9/L)** (safety parameter)
    c. **Platelet count (10^9/L)** (safety parameter)
10. **Liver function test (LFT)**
    a. **Prothrombin time (PT) (seconds)** (safety parameter)
    b. **Bilirubin (micromol/L)** (outcome and safety parameter)
    c. **Alkaline phosphatase (ALP) (U/L)** (outcome and safety parameter)
    d. **Alanine aminotransferase test (ALT) (U/L)** (outcome and safety parameter).
    e. **Aspartate aminotransferase (AST) (U/L)** (outcome and safety parameter).
    f. **Albumin (G/L)** (outcome and safety parameter)
    g. **Gamma-glutamyl transferase test (GGT) (U/L)** (outcome and safety parameter)
    h. **Activated partial thromboplastin time (APTT) (seconds)** (safety parameter)
    i. **C-reactive protein (CRP)** (outcome and safety parameter)
11. **Lipid profile**
    a. **Cholesterol (Low-density lipoprotein** (**LDL)) (mmol/L)** (safety parameter)
    b. **High-density lipoprotein** (**HDL) cholesterol (mmol/L)** (safety parameter)
    c. **Triglyceride (mmol/L)** (safety parameter)
12. **Urea and electrolytes**
    a. **Urea (mmol/L)** (outcome and safety parameter)
    b. **Creatinine (mmol/L)** (outcome and safety parameter)
    c. **Sodium (mmol/L)** (safety parameter)
    d. **Potassium (mmol/L)** (safety parameter)

Bioenergetics in PBC

13. Bioenergetics outcomes will be continuous variables and will include:
    a. **Minimum** muscle calculated **pH post-exercise** (outcome)
    b. **pH recovery half-time** (outcome)
    c. **pH Fall with Exercise** (outcome)
    d. **Area under curve (AUC) for pH** (outcome)
    e. **Anaerobic threshold** (outcome)
    f. **We will additionally record resting PH level in the two arms for comparability but this will not be included in core analyses**

Immunoglobin Levels

14. **Serum IgG levels will be continuous variables and will include:**
    a. **Total Immunoglobulin G (IgG)** level (outcome and safety parameter)

b. **Total Immunoglobulin M (IgM)** level (outcome and safety parameter)
c. **Total Immunoglobulin (Ig)** level (calculated by adding IgG, IgM and IgA to form total Ig)

Sub-fractions of AMA

15. Sub-fractions of AMA antibody outcomes:
    a. **AMA titre will be categorical variable** (e.g. reciprocal of 40, 80, 160, 320, 640, >640 categories)
    b. **Sub-fraction of AMA (Anti-PDC titre) antibody** will be continuous variable (Anti-PDC titre) antibody will be continuous variable – individual titres (levels of anti-PDC activity) are generated for each serum sample at baseline and then 3 months (visit 16) to 12 months (visit 19) are expressed as a percentage of the visit 1 titre (e.g. baseline (visit 1) is always 100%).

B-cell data

16. **CD19 B-cell^{m}** depletion expressed as a **%** (outcome)

Safety outcomes

17. **Serious adverse events (SAEs)** will be listed and not categorised a priori since we expect so few**.**

Outcomes measures for future analyses (described in section 10 in Future analyses):

Cytokines

18. Cytokines will include the following outcomes:
    a. **Tumour necrosis factor-alpha** (**TNF-alpha) (pg/mL)**
    b. **Interleukin1-beta (IL1-beta) (pg/mL)**
    c. **Interleukin6 (IL6) (pg/mL)**
    d. **IFNy (gamma-interferon) (pg/mL)**
    e. **Granulocyte-macrophage colony-stimulating factor** (**GMCSF) (pg/mL)**

8.2 Data collection and outcome assessments

Table 1: RITPBC Schedule of Events:

	Visit 0 Screening	Visit 1	Visit 2	Visit 3	Visit 4	Visits 5-15	Visit 16	Visit 17	Visit 18	Visit 19 Final
	-2 Weeks	Baseline and Randomisation	Day 1 1st infusion	Day 7 Safety Visit	Day 15 2nd infusion	Treatment +1-11 weeks	Treatment +12 weeks	Treatment + 6 months	Treatment + 9 months	Treatment + 12 months
Physical examination[1]	X		X	X	X	TELEPHONE REVIEW/ CONCOMITANT MEDICATIONS/ THERAPIES CHECK AND AE REVIEW	X	X	X	X
Obtain informed consent	X									
Pregnancy test		X								
Hep B/C[5]	X									
PBC- 40		X					X	X	X	X
PROMIS HAQ questionnaire		X					X	X	X	X
COGFAIL questionnaire		X					X	X	X	X
HADS questionnaire		X					X	X	X	X
Issued Fatigue diary[2]		X								
Return Fatigue diary										X
ESS/OGS questionnaires		X					X	X	X	X
FBC, LFT, U&Es and CRP[3]		X	X	X	X		X	X	X	X
Random lipid profile[4]		X		X			X	X	X	X
Autoantibodies & Immunoglobulins		X					X			
Coagulation studies		X					X			
Adverse events			X	X	X		X	X	X	X
Concomitant medication	X	X	X	X	X		X	X	X	X
Activity monitors		X					X			
Blood for serum, RNA and white cells		X					X	X	X	X

Muscle MRI		X					X			
Anaerobic threshold		X					X			
Randomisation		X								
Rituximab / placebo infusion			X		X					

1.Physical examination includes vital signs (height, weight, blood pressure, heart rate, temperature and respiratory rate)
2. Patients are issued with Fatigue diary at Baseline and are asked to complete it for a one week period at the beginning of each month for each visit at baseline, 1, 3, 6, 9 and 12 months. They return it at visit 19.
3. LFT includes AST and GGT
4. Random lipid profile includes total cholesterol, HDL and triglycerides
5. Hepatitis B serology test including HBsAg (Hepatitis B surface antigen) and HBc (Hepatitis B core antibody)

9. DATA ANALYSIS

9.1 Sample size calculation and clinically important difference

The primary outcome is the PBC-40 fatigue domain score (range 11–55) after 12 weeks of intervention. The SD of fatigue scores is 8 units (based on the PBC-40 derivation studies utilising >1000 patients[m]), with a correlation of 0.6 between baseline and follow-up time points based on previous studies.

The study was originally powered to detect a mean change in PBC-40 fatigue domain score of 5 units (equating to an average of 0.5 point change per question; a difference in PBC-40 score demonstrated to be associated with significantly higher levels of social function and, therefore, deemed to be clinically significant) with a power of 90% and a 5% significance level. This equated to 35 participants in each group providing data on the primary outcome (PBC-40 fatigue score at 12 weeks): incorporating an assumption of a 10% attrition gave a total of 78. The number of participants lost to follow-up, or withdrawing consent prior to initial treatment was expected to be minimal.

However, since recruitment rates were lower than planned, even after recruiting from outside the North East region, the funder (NIHR EME) agreed to a 6-month extension to the study, a reduction in the power of the trial to 80% and a revised target sample size (with other estimates of clinically important difference, standard deviation and attrition rate remaining the same) of 58 participants (29 per arm). These changes to the design of the trial were submitted as a substantial amendment to both the REC and MHRA and were accepted during October 2015.

9.2 Missing data

Where applicable, details about crude imputation (dealing with missing data) in questionnaires for each primary and secondary outcome are given in Appendix 1.

Data with missing observations due to loss to follow-up will be examined to determine the extent of missingness. We do not propose the use of any other imputation techniques.

We will report the number (percentage) of missing data for all variables. An allowance for loss to follow-up has been included in the sample size calculation (see above).

9.3 Descriptive Analyses

Descriptive analysis of the study population will be summarised for each randomised group and will include:

- Baseline patient characteristics (section 5.2)
- Adverse event safety outcomes
- All primary and secondary outcomes (section 8.1)

A full description of outcomes that will be reported in the trial is given in section 8.1 and Appendix 1.

9.3.1 Patient characteristics at baseline

Baseline (BL) characteristics of the study population will be summarised separately within each randomised group (see section 5.2, items a-i, for full list of variables). Characteristics such as

gender, ethnicity, smoking history and patient location will be summarised by reporting the number (%) in each category, whereas age in years, alcohol consumption, BMI, resting pH and UK PBC risk score (at 10 years prediction) will be reported as mean (SD) and range or median (IQR) as appropriate (see dummy Table 2).

Dummy Table 2: Baseline characteristics by intervention arm

	Rituximab (n=)				**Placebo (n=)**			
Categorical variables	**N**		**(%)**		**n**		**(%)**	
Sex:								
Male								
Ethnicity:								
White								
Non-white								
Smoking status:								
Never								
Past								
Current								
Managed by Newcastle centre for at least year:								
Yes								
UDCA use:								
Yes								
If yes: Responder								
Continuous variables	**n**	**Mean (SD)**	**Median (IQR)**	**Range**	**n**	**Mean (SD)**	**Median (IQR)**	**Range**
Age in years								
Alcohol consumption (units per week)								
BMI								
UK PBC risk score at 10 years								

A comparison of the distribution of baseline patient characteristics (see above and section 5.2, items a-i) between completion (defined as completing at least 50% of questions which would result in having an overall calculated score) and non-completion of the primary outcome (PBC-40 Fatigue domain) at 3 months will also be carried out to see if any systematic differences between the two groups (see dummy Table 3).

Dummy Table 3: Characteristics of those who did and did not complete PBC 40 fatigue domain questionnaire by intervention arm

	Completion* (n=)				Non-completion$ (n=)			
Categorical variables	**N**		**(%)**		**n**		**(%)**	
Sex:								
Male								
Ethnicity:								
White								
Non-white								
Smoking status:								
Never								
Past								
Current								
Managed by Newcastle centre for at least year:								
Yes								
UDCA use:								
Yes								
If yes: Responder								
Continuous variables	**n**	**Mean (SD)**	**Median (IQR)**	**Range**	**n**	**Mean (SD)**	**Median (IQR)**	**Range**
Age in years								
Alcohol consumption (units per week)								
BMI								
UK PBC risk score at 10 years								

* Defined as completing at least 50% of PBC 40 fatigue domain questions which would result in having an overall calculated score

& Defined as completing less than 50% of PBC 40 fatigue domain questions which would result in not having an overall calculated score

9.3.2 Outcome measures

9.3.2.1 Primary outcome

PBC-40 fatigue domain

For PBC-40 fatigue domain, the mean (SD) and range (or median and IQR as appropriate) will be recorded at each reported time point in each arm. This outcome will be analysed at BL, 12 weeks and 6, 9 and 12 months and results will be tabulated (see dummy Table 4).

Dummy Table 4: PBC-40 fatigue domain score at different timepoints by intervention arm

PBC-40 fatigue domain score	Rituximab				Placebo			
Timepoint (scale range: 11-55)	**n**	**Mean (SD)**	**Median (IQR)**	**Range**	**n**	**Mean (SD)**	**Median (IQR)**	**Range**
Baseline								
3 months								
6 months								
9 months								
12 months								

9.3.2.2 Secondary outcomes

Other PBC-40 domains

For the other PBC-40 domains (listed in section 8.1, items 1a-e), the mean (SD) and range (or median and IQR as appropriate) will be recorded at each reported time point in each arm. These outcomes will be analysed at BL, 12 weeks and 6, 9 and 12 months and results will be tabulated (see dummy Tables 5a-b).

Clinical symptom and functional capability questionnaires

For all questionnaire scores (see section 8.1, items 2-6), the mean (SD) and range (or median and IQR as appropriate) will be recorded at each reported time point in each arm. These outcomes will be analysed at BL, 12 weeks and 6, 9 and 12 months and results will be tabulated (see dummy Tables 5a-b).

Perceived fatigue score

For fatigue score (listed in section 8.1, item 7), the mean (SD) and range (or median and IQR as appropriate) will be recorded from participant fatigue diaries at each reported time point in each arm. This outcome will be analysed at BL, 4 and 12 weeks and 6, 9 and 12 months and results will be tabulated (see dummy Tables 5a-b).

Dummy Table 5a: Questionnaire scores at baseline and 3 months by intervention arm

	Rituximab								Placebo							
	Baseline				3 months				Baseline				3 months			
Questionnaire (*possible scores*)	**n**	**Mean (SD)**	**Median (IQR)**	**Range**	**n**	**Mean (SD)**	**Median (IQR)**	**Range**	**n**	**Mean (SD)**	**Median (IQR)**	**Range**	**n**	**Mean (SD)**	**Median (IQR)**	**Range**
PBC-40 domain: **Itch (*0-15*)** **Cognitive (*6-30*)** **Social (*8-50*)** **Emotional (3-15)** **Other symptoms (6-35)**																
ESS (*0-24*)																
OGS (*0-20*)																
PROMIS-HAQ (0-100)																
COGFAIL (0-100)																
HADS (0-42)																
Fatigue diary* score (1-6)																

* A footnote will note the descriptive statistics for fatigue diary score at 1 month as this was additional timepoint in which diaries were recorded

Dummy Table 5b: Questionnaire scores at 6, 9 and 12 months by intervention arm

	Rituximab									Placebo								
	6 months			9 months			12 months			6 months			9 months			12 months		
Questionnaire (*possible scores*)	n	Mean (SD)	Median (IQR) [Range]	n	Mean (SD)	Median (IQR) [Range]	n	Mean (SD)	Median (IQR) [Range]	n	Mean (SD)	Median (IQR) [Range]	n	Mean (SD)	Median (IQR) [Range]	n	Mean (SD)	Median (IQR) [Range]
PBC-40 domain: Itch (*0-15*) Cognitive (*6-30*) Social (*8-50*) Emotional (3-15) Other symptoms (6-35)																		
ESS (*0-24*)																		
OGS (*0-20*)																		
PROMIS-HAQ (0-100)																		
COGFAIL (0-100)																		
HADS (0-42)																		
Fatigue diary score (1-6)																		

ENMO outcomes

For the overall average magnitude of wrist acceleration per 5 second epoch (millig) and average acceleration (millig) during the most active (M5) 5 hr period of each day (listed in section 8.1, items 8a-b), the mean (SD) and range (or median and IQR as appropriate) will be recorded at each reported time point (pre and post) in each arm. This outcome will be analysed at BL (pre) and 3 months after treatment (post) and also as a difference in values from pre to post in each group and results will be tabulated (see dummy Table 6).

Dummy Table 6: ENMO outcomes at baseline and 3 months by intervention arm

ENMO outcome	Rituximab						Placebo					
Timepoint	**n**	**Mean**	**SD**	**Median**	**IQR**	**Range**	**n**	**Mean**	**SD**	**Median**	**IQR**	**Range**
Baseline (pre)												
Average ENMO												
ENMO best 5 hrs												
3 months (post)												
Average ENMO												
ENMO best 5 hrs												

Laboratory parameters

For all laboratory bloods data (See section 8.1, items 9-12), the mean (SD) and range (or median and IQR as appropriate) will be recorded at each reported time point in each arm. These outcomes will be analysed at BL, 12 weeks and 6, 9 and 12 months and results will be tabulated (see dummy Table 7a-d).

Dummy Table 7a: Full blood count and liver function test laboratory parameters at baseline and 3 months by intervention arm

	Rituximab								Placebo							
	Baseline				3 months				Baseline				3 months			
Laboratory parameter	n	Mean (SD)	Median (IQR)	Range	n	Mean (SD)	Median (IQR)	Range	n	Mean (SD)	Median (IQR)	Range	n	Mean (SD)	Median (IQR)	Range
Full blood count																
Haemoglobin (g/L)																
WBC (10^9/L)																
Platelet count (10^9/L)																
Liver function test																
PT (secs)																
Bilirubin (mmol/L)																
Alkaline phosphatase (U/L)																
ALT (U/L)																
AST (U/L)																
Albumin (G/L)																
GGT (U/L)																
APTT (secs)																
CRP																

Dummy Table 7b: Full blood count and liver function test laboratory parameters at 6, 9 and 12 months by intervention arm

	Rituximab									Placebo								
	6 months			9 months			12 months			6 months			9 months			12 months		
Laboratory parameter	**n**	**Mean (SD)**	**Median (IQR) [Range]**	**n**	**Mean (SD)**	**Median (IQR) [Range]**	**n**	**Mean (SD)**	**Median (IQR) [Range]**	**n**	**Mean (SD)**	**Median (IQR) [Range]**	**n**	**Mean (SD)**	**Median (IQR) [Range]**	**n**	**Mean (SD)**	**Median (IQR) [Range]**
Full blood count																		
Haemoglobin (g/L)																		
WBC (10^9/L)																		
Platelet count (10^9/L)																		
Liver function test																		
Bilirubin (mmol/L)																		
Alkaline phosphatase (U/L)																		
ALT (U/L)																		
AST (U/L)																		
Albumin (G/L)																		
GGT (U/L)																		
CRP																		

Dummy Table 7c: Lipid profile and urea and electrolytes laboratory parameters at baseline and 3 months by intervention arm

	Rituximab								Placebo							
	Baseline				3 months				Baseline				3 months			
Laboratory parameter	n	Mean (SD)	Median (IQR)	Range	n	Mean (SD)	Median (IQR)	Range	n	Mean (SD)	Median (IQR)	Range	n	Mean (SD)	Median (IQR)	Range
Lipid profile																
LDL (mmol/L)																
HDL (mmol/L)																
Triglyceride (mmol/L)																
Urea and electrolytes																
Urea (mmol/L)																
Creatinine (mmol/L)																
Sodium (mmol/L)																
Potassuim (mmol/L)																

Dummy Table 7d: Lipid profile and urea and electrolytes laboratory parameters at 6, 9 and 12 months by intervention arm

	Rituximab									Placebo								
	6 months			9 months			12 months			6 months			9 months			12 months		
Laboratory parameter	**n**	**Mean (SD)**	**Median (IQR) [Range]**	**n**	**Mean (SD)**	**Median (IQR) [Range]**	**n**	**Mean (SD)**	**Median (IQR) [Range]**	**n**	**Mean (SD)**	**Median (IQR) [Range]**	**n**	**Mean (SD)**	**Median (IQR) [Range]**	**n**	**Mean (SD)**	**Median (IQR) [Range]**
Lipid profile																		
LDL (mmol/L)																		
HDL (mmol/L)																		
Triglyceride (mmol/L)																		
Urea and electrolytes																		
Urea (mmol/L)																		
Creatinine (mmol/L)																		
Sodium (mmol/L)																		
Potassuim (mmol/L)																		

Bioenergetics outcomes

For all bioenergetics outcomes (See section 8.1, items 13a-e), the mean (SD) and range (or median and IQR as appropriate) will be recorded at each reported time point in each arm. These outcomes will be analysed at BL and 12 weeks and results will be tabulated (see dummy Table 8).

Dummy Table 8: Bioenergetics outcomes at baseline and 3 months by intervention arm

	Rituximab				Placebo			
Bioenergetics outcomes	**n**	**Mean (SD)**	**Median (IQR)**	**Range**	**n**	**Mean (SD)**	**Median (IQR)**	**Range**
Min pH post exercise:								
Baseline								
3 months								
pH recovery HT								
Baseline								
3 months								
pH fall with exercise								
Baseline								
3 months								
AUC for pH								
Baseline								
3 months								
Anaerobic threshold								
Baseline								
3 months								

Immunoglobin levels

For serum Ig (IgG, IgM and total Ig level) levels (See section 8.1, items 14a-c), the mean (SD) and range (or median and IQR as appropriate) will be recorded at each reported time point in each arm. These outcomes will be analysed at BL and 12 weeks and results will be tabulated (see dummy Table 9).

Sub-fraction of AMA (titre)

AMA titre data (See section 8.1, items 15a) will also be reported categorically at each time point in each arm. These outcomes will be analysed at BL, 12 weeks and 6, 9 and 12 months and results will be tabulated using following categories: reciprocal of 40, 80, 160, 320 and 640 or greater (see dummy Table 10).

We will also report a more specific measure of the autoantibody which reacts against a certain protein (PDC), this is the Anti-PDC antibody level (See section 8.1, items 15b). Sub-fraction of AMA (Anti-PDC titre) antibody will be continuous variable – individual titres (levels of anti-PDC activity) are generated for each serum sample at baseline (visit 1) and then at 3 months (visit 16) to 12 months (visit 19) and are expressed as a percentage of the visit 1 titre (e.g. visit 1 is always 100%). The mean (SD) and range (or median and IQR as appropriate) will be recorded at 3, 6, 9 and 12 months relative to baseline in each arm and results will be tabulated (see dummy Table 10).

Dummy Table 9: Immunoglobin level outcomes up to 12 months by intervention arm

	Rituximab				Placebo			
Ig outcomes	**n**	**Mean (SD)**	**Median (IQR)**	**Range**	**n**	**Mean (SD)**	**Median (IQR)**	**Range**
Serum Ig levels								
Total IgG: **Baseline** **3 months** **Total IgM:** **Baseline** **3 months** **Total Ig level:** **Baseline** **3 months**								

Dummy Table 10: AMA titre categories and anti-PDC antibody level up to 12 months by intervention arm up to 12 months by intervention arm

	Rituximab		Placebo	
AMA titre	**n**	**(%)**	**n**	**(%)**
At baseline: **40:** **80:** **160:** **320:** **>640:**				
At 3 months: **40:** **80:** **160:** **320:** **>640:**				
At 6 months: **40:** **80:** **160:** **320:** **>640:**				
At 9 months: **40:** **80:** **160:** **320:** **>640:**				
At 12 months: **40:** **80:** **160:** **320:** **>640:**				

Anti-PDC titre (% from baseline):	n	Mean (SD)	Median (IQR)	Range	n	Mean (SD)	Median (IQR)	Range
Anti-PDC titre (% from baseline): **3 months** **6 months** **9 months** **12 months**								

B-cell data

B-cell data (See section 8.1, item 16) will also report the mean (SD) and range (or median and IQR as appropriate) at each reported time point in each arm. These outcomes will be analysed at BL, 12 weeks and 6, 9 and 12 months and results will be tabulated (see dummy Table 11a-b).

Dummy Table 11a: B-cell outcome at baseline and 3 months by intervention arm

	Rituximab								Placebo							
	Baseline				3 months				Baseline				3 months			
Outcome	n	Mean (SD)	Median (IQR)	Range	n	Mean (SD)	Median (IQR)	Range	n	Mean (SD)	Median (IQR)	Range	n	Mean (SD)	Median (IQR)	Range
CD19 B-cell depletion (%)																

Dummy Table 11b: B-cell outcome at 6, 9 and 12 months by intervention arm

	Rituximab									Placebo								
	6 months			9 months			12 months			6 months			9 months			12 months		
Outcome	n	Mean (SD)	Median (IQR) [Range]	n	Mean (SD)	Median (IQR) [Range]	n	Mean (SD)	Median (IQR) [Range]	n	Mean (SD)	Median (IQR) [Range]	n	Mean (SD)	Median (IQR) [Range]	n	Mean (SD)	Median (IQR) [Range]
CD19 B-cell depletion (%)																		

Serious adverse events (SAEs)

For SAEs (see section 8.1 item 17), we will collect information on the number of patients in each intervention arm who experienced an event and the number of patients assessed: we will report percentages in each arm and these will be tabulated (see dummy Table 12).

Dummy Table 12: Serious adverse events at 6, 9 and 12 months by intervention arm

Patient	SAE onset date	Description	Start date of study drug	Last date of study drug	Seriousness	Causality to study intervention	Outcome of event	Randomisation group
ID Number	dd/mm/yyy	text	dd/mm/yyyy	dd/mm/yyyy	1/2/3/4	**Y**(expected / unexpected) /**N**	text	Rituximab/Placebo

9.4 Additional Descriptive Analyses

Liver Disease Activity
Although not powered to demonstrate biologically significant effects on severity of underlying liver disease, the study will provide important pointers to any effect which would inform the design of future studies.

We will analyse change in serum alkaline phosphatase level (see section 8.1, item 10c) and classify a drop in baseline alkaline phosphatase (ALP) of >15% or normalisation (within normal range) as being clinically significant. We will report the number (%) of clinically significant results in each trial arm.

The analyses will be descriptive in nature and will compare change in ALP at 12 weeks from baseline by trial arms as well as change at 12 months from baseline.

9.5 Association

9.5.1 Bioenergetics in PBC

To address whether fatigue score is linked to energetics in PBC:

- Correlation between change in PBC-40 fatigue domain score and change in bioenergetics outcomes (separately for all five outcomes listed in section 8.1, items 13a-e)

Pearson product-moment correlation will be used throughout.

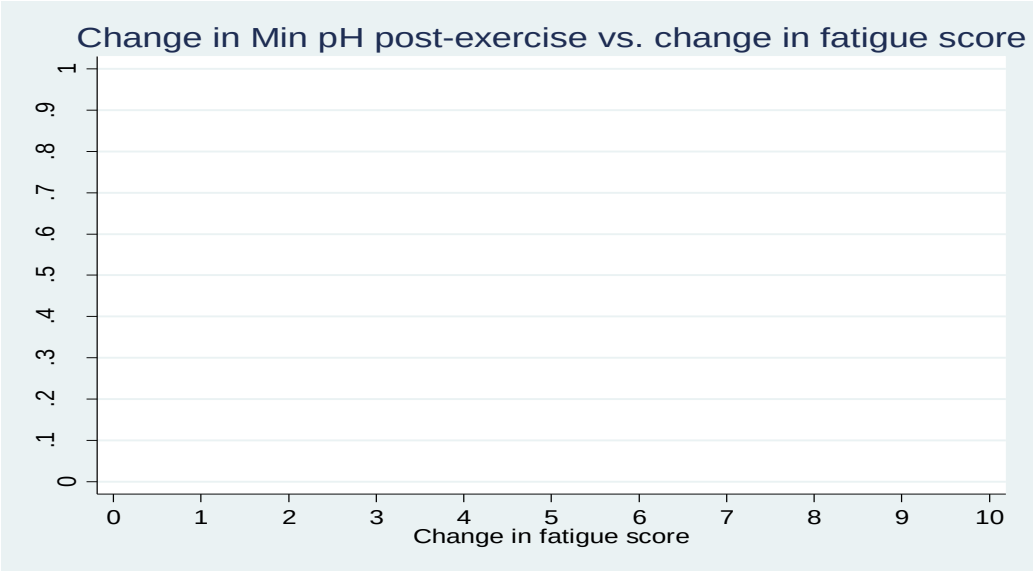
The changes will be measured as a difference from 12 week measurement to baseline so correlations on 5 variables for each trial arm will be reported and tabulated (see dummy Table 13).

Dummy Table 13: Correlations (Pearson) between changes in PBC-40 fatigue domain score and changes in bioenergetics outcomes by intervention arm

Change in PBC-40 fatigue score	**Rituximab**			**Placebo**			**Overall**		
Correlation with:	n	corr	95 % CI	n	corr	95% CI	n	corr	95% CI
Change at 3m from baseline:									
Min pH post-exercise									
pH recovery half-time									
pH fall with exercise									
Area under curve (AUC) for pH									
Anaerobic threshold									

Change in PBC-40 fatigue domain score and change in bioenergetics outcomes from 12 week measurement to baseline will also be plotted on a scatter plot.

See below for dummy template of scatter plot:

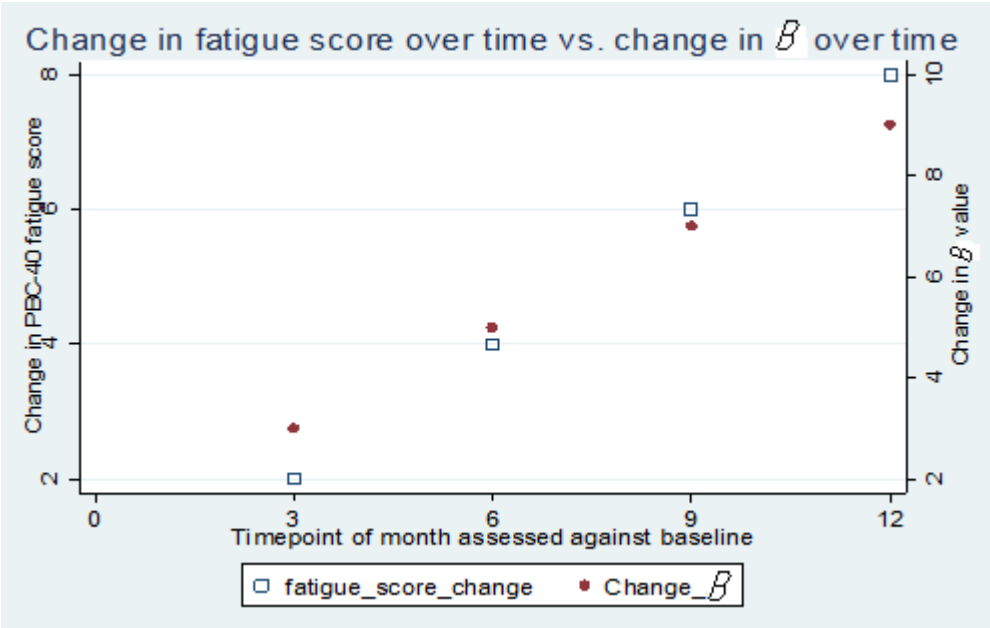


9.5.2 Biological change

An analysis that describes trend over time, namely, change in PBC40 fatigue score over time plotted against change in CD19 B-cell depletion values (listed in section 8.1, item 15) over time will be carried out

The changes will be reported as a difference from baseline at 12 weeks, 6, 9 and 12 month measurements.

See below for example dummy template of trend plot:



9.6 Inferential analyses

The primary hypothesis to be tested is H_0: There is no difference in mean PBC40 fatigue scores between Rituximab and placebo as a treatment for fatigue in patients with PBC.

Where applicable, a two-sided significance level of $p<0.05$ will be used throughout.

A full description of outcomes that will be reported in the trial is given in Section 8.1 and Appendix 1.

9.6.1 Primary efficacy endpoint

9.6.1.1 At 12 weeks

PBC-40 fatigue domain
The primary endpoint, PBC-40 fatigue domain scores (see section 8.1 for details) will be compared at 12 weeks between intervention and placebo group using a forward selection stepwise approach in multiple linear regression. Covariates for selection in the model will include age in years, UK PBC risk model score at 10 years and patient location (managed by Newcastle centre for at least one year or not) at baseline and selected allowing 10% level of significance. Baseline PBC-40 fatigue score will be included in the final model along with covariates chosen in the stepwise procedure. Transformation of covariates will be performed if appropriate (also applies to secondary analyses below). The results will be reported as a difference in means with a 95% confidence interval. Bootstrap estimation will be used if the error distribution is non-Normal (see dummy Table 14).

Dummy Table 14: Multiple linear regression showing difference in mean PBC-40 fatigue domain score at 12 weeks between intervention arms

		Rituximab		Placebo		Difference
PBC-40 fatigue domain	**Observed range**	**n**	**Mean[a] (SD)**	**n**	**Mean[a] (SD)**	**Adj. diff in means[b] (I-C) (95%CI)**
Fatigue score (11-55)						

[a] univariate analysis reporting mean and SD at 12 weeks without adjustment
[b] multivariate analysis reporting difference in means between groups at 12 weeks with adjustment for baseline score and covariates selected from stepwise procedure

9.6.1.2 Up to 12 months

Analyses will be restricted to repeated measures analysis of variance (ANOVA) as specified in the protocol because we expected data to be sparse. Therefore results will be based on unadjusted estimates. More advanced methods such as multilevel mixed effects models were deemed inappropriate.

Repeated measures analysis

The time course of the comparison between intervention and control groups over the 12 month follow-up period will be assessed for primary outcome (PBC-40 fatigue domain) outlined in section 8.1 using repeated measures ANOVA using time points at baseline, 12 weeks, 6, 9 and 12 months (see dummy Table 15).

Dummy Table 15: Repeated measures ANOVA for fatigue score up to 12 months by intervention arm

	Rituximab										Placebo											
PBC-40 questionnaire	**Baseline**		**3 months**		**6 months**		**9 months**		**12 months**		**Baseline**		**3 months**		**6 months**		**9 months**		**12 months**		**Test statistics**	
Fatigue domain	**n**	**Mean (SD)**	**n**	**Mean (SD)**	**n**	**Mean (SD)**	**n**	**Mean (SD)**	**n**	**Mean (SD)**	**n**	**Mean (SD)**	**n**	**Mean (SD)**	**n**	**Mean (SD)**	**n**	**Mean (SD)**	**n**	**Mean (SD)**	**F test**	**P value**
																					Between:	
Fatigue score (11-55)																					**Trial arms**	
																					Trial*timepoints	

9.6.2 Secondary efficacy endpoints

9.6.2.1 At 12 weeks

Secondary endpoints (see section 8.1 for details) will be compared at 12 weeks between intervention and placebo group using a forward selection stepwise approach in multiple linear regression as outlined in section 9.5.1.1 above for primary endpoint. Covariates for selection in each model will also include age in years, UK PBC risk model score at 10 years and patient location (managed by Newcastle centre for at least one year or not) at baseline and selected allowing 10% level of significance. Baseline score for each outcome in question will be included in the final model along with covariates chosen in the stepwise procedure.

Other PBC-40 domains

Remaining PBC-40 domain scores (listed in section 8.1, items 1a-e) will be compared at 12 weeks between intervention and placebo group using multiple linear regression adjusted for baseline domain score and covariates identified from stepwise procedure. The results will be reported as a difference in means with a 95% confidence interval. Bootstrap estimation will be used if the distribution is non-Normal (see dummy Table 16).

Clinical symptom and functional capability scales

Questionnaire scores (listed in section 8.1, items 2-6) at 12 weeks will be analysed separately as above using multiple linear regression adjusted for baseline questionnaire score and covariates identified from stepwise procedure. The results will be reported as a difference in means with a 95% confidence interval. Bootstrap estimation will be used if the distribution is non-Normal (see dummy Table 16).

Participant diaries

Fatigue diary score (from fatigue diaries listed in section 8.1, item 7) will be compared at 12 weeks between intervention and placebo group using multiple linear regression adjusted for baseline fatigue diary score and covariates identified from stepwise procedure. The results will be reported as a difference in means with a 95% confidence interval. Bootstrap estimation will be used if the distribution is non-Normal (see dummy Table 16).

Dummy Table 16: Multiple linear regression showing difference in mean questionnaire scores at 12 weeks between intervention arms

		Rituximab		Placebo		Difference
Questionnaire	**Observed range**	**n**	**Mean[a] (SD)**	**n**	**Mean[a] (SD)**	**Adj. diff in means (I-C) (95%CI)[b]**
PBC-40 domain: **Itch (0-15)** **Cognitive (6-30)** **Social (8-50)** **Emotional (3-15)** **Other symptoms (6-35)**						
ESS (0-24)						
OGS (0-20)						
PROMIS-HAQ (0-100)						
COGFAIL (0-100)						
HADS (0-42)						
Fatigue diary score (1-6)						

[a] univariate analysis reporting mean and SD at 12 weeks without adjustment

[b] multivariate analysis reporting difference in means between groups at 12 weeks with adjustment for baseline score and covariates selected from stepwise procedure

ENMO outcomes

ENMO outcomes (listed in section 8.1, item 8a-b) will be compared at 12 weeks between intervention and placebo group using multiple linear regression adjusted for baseline (pre) measurement and covariates identified from stepwise procedure. The results will be reported as a difference in means with a 95% confidence interval. Bootstrap estimation will be used if the distribution is non-Normal (see dummy Table 17).

Dummy Table 17: Multiple linear regression showing difference in mean ENMO measurements at 12 weeks between intervention arms

		Rituximab		Placebo		Difference
ENMO outcome	**Observed range**	**n**	**Mean[a] (SD)**	**n**	**Mean[a] (SD)**	**Adj. diff in means (I-C) (95%CI)[b]**
Average ENMO						
ENMO best 5 hrs						

[a] univariate analysis reporting mean and SD at 12 weeks without adjustment

[b] multivariate analysis reporting difference in means between groups at 12 weeks with adjustment for baseline measurement and covariates selected from stepwise procedure

Bioenergetics in PBC

Bioenergetics outcomes (namely, minimum pH post-exercise, pH recovery half-time, pH fall with exercise, AUC for pH & Anaerobic threshold listed in section 8.1, items 13a-e) will be compared (separately) at 12 weeks between intervention and placebo group using multiple linear regression adjusted for baseline value (for bioenergetics outcome in question) and covariates identified from stepwise procedure. The results will be reported as a difference in means with a 95% confidence interval. Bootstrap estimation will be used if the distribution is non-Normal (see dummy Table 18).

Dummy Table 18: Multiple linear regression showing difference in mean bioenergetics outcomes at 12 weeks between intervention arms

		Rituximab		Placebo		Difference
Outcome	**Observed range**	**n**	**Mean[a] (SD)**	**n**	**Mean[a] (SD)**	**Adj. diff in means[b] (I-C) (95%CI)**
pH post-exercise						
pH recovery half-time						
pH fall with exercise						
Area under curve (AUC) for pH						
Anaerobic threshold						

[a] univariate analysis reporting mean and SD at 12 weeks without adjustment

[b] multivariate analysis reporting difference in means between groups at 12 weeks with adjustment for baseline value and covariates selected from stepwise procedure

Biological outcomes

Biological outcomes (namely, serum immunoglobin level outcomes (IgG and IgM) listed in section 8.1, items 14a-b) will be compared (separately) at 12 weeks between intervention and placebo group using multiple linear regression adjusted for baseline value (for bioenergetics outcome in question) and covariates identified from stepwise procedure. The results will be reported as a difference in means with a 95% confidence interval. Bootstrap estimation will be used if the distribution is non-Normal (see dummy Table 19).

Dummy Table 19: Multiple linear regression showing difference in mean serum Ig outcomes at 12 weeks between intervention arms

		Rituximab		Placebo		Difference
Serum Ig outcome	**Observed range**	**n**	**Mean[a] (SD)**	**n**	**Mean[a] (SD)**	**Adj. diff in means[b] (I-C) (95%CI)**
IgG						
IgM						

[a] univariate analysis reporting mean and SD at 12 weeks without adjustment

[b] multivariate analysis reporting difference in means between groups at 12 weeks with adjustment for baseline value and covariates selected from stepwise procedure

9.6.2.2 Up to 12 months

Repeated measures analyses

The time course of the comparison between intervention and control groups over the 12 month follow-up period will be assessed for remaining PBC-40 domains, questionnaire scores from clinical symptom and functional capability scales, diary fatigue scores, bioenergetics values and liver function test (LFT) outcomes (namely, bilirubin, alkaline phosphatase) using repeated measures ANOVA (as described in section 9.5.1.2) using timepoints at baseline, 12 weeks, 6, 9 and 12 months.

Additionally, fatigue diaries were completed at 4 weeks so the repeated measures analysis for this outcome will include an additional timepoint.

More specifically repeated measures analyses will include the following secondary outcomes listed in section 8.1:

- PBC-40 domains (items 1a-e) (see dummy Table 20)
- questionnaire scores from clinical symptom and functional capability scales (items 2-6) (see dummy Table 21)
- diary fatigue scores (item 7) (see dummy Table 22)
- LFT outcomes (items 10b-c, namely bilirubin & alkaline phosphatase) (see dummy Table 23)

Dummy Table 20: Repeated measures ANOVA for PBC-40 questionnaire scores up to 12 months by intervention arm

	Rituximab										Placebo											
	Baseline		3 months		6 months		9 months		12 months		Baseline		3 months		6 months		9 months		12 months		Test statistics	
PBC-40 domain:	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	F test	P value
																					Between:	
Itch (*0-15*)																					Trial arms	
																					Trial*timepoints	
Cognitive (*6-30*)																					Trial arms	
																					Trial*timepoints	
Social (*8-50*)																					Trial arms	
																					Trial*timepoints	
Emotional (3-15)																					Trial arms	
																					Trial*timepoints	
Other symptoms (6-35)																					Trial arms	
																					Trial*timepoints	

Dummy Table 21: Repeated measures ANOVA for PBC-40 questionnaire scores up to 12 months by intervention arm

	Rituximab										Placebo											
	Baseline		3 months		6 months		9 months		12 months		Baseline		3 months		6 months		9 months		12 months		Test statistics	
Questionnaire	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	F test	P value
																					Between:	
ESS (*0-24*)																					Trial arms	
																					Trial*timepoints	
OGS (*0-20*)																					Trial arms	
																					Trial*timepoints	
PROMIS-HAQ (0-100)																					Trial arms	
																					Trial*timepoints	
COGFAIL (0-100)																					Trial arms	
																					Trial*timepoints	
HADS (0-42)																					Trial arms	
																					Trial*timepoints	

Dummy Table 22: Repeated measures ANOVA for fatigue diary scores up to 12 months by intervention arm

	Rituximab										Placebo											
	Baseline		3 months		6 months		9 months		12 months		Baseline		3 months		6 months		9 months		12 months		Test statistics*	
Fatigue diary score	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	F test	P value
																					Between:	
Diary score (1-6)																					Trial arms	
																					Trial*timepoints	

* Additionally fatigue diary score was recorded at 1 month so this timepoint will be included in the repeated measures ANOVA

Dummy Table 23: Repeated measures ANOVA for liver function test (LFT) outcomes up to 12 months by intervention arm

	Rituximab										Placebo											
	Baseline		3 months		6 months		9 months		12 months		Baseline		3 months		6 months		9 months		12 months		Test statistics	
LFT outcome	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	F test	P value
																					Between:	
Bilirubin																					Trial arms	
																					Trial*timepoints	
Alkaline phosphatase																					Trial arms	
																					Trial*timepoints	

9.7 Statistical Software

Trial data are input by individual site staff into a MACRO database held and maintained by the Newcastle Clinical Trials Unit.

Data will be downloaded directly from MACRO into the statistical software package StataIC (version 12^{o} but any subsequent update will be recorded at time of final download). Statistical analyses will be carried out by the Trial Statistician at NCTU downloading snapshots of the data at time-points agreed by the TMG.

10. FUTURE ANALYSES

10.1 Descriptive analyses

Cytokines outcomes

For cytokines outcomes (See section 8.1, items 18a-e), the mean (SD) and range (or median and IQR as appropriate) will be recorded at each reported time point in each arm. These outcomes will be analysed at BL, 12 weeks and 6, 9 and 12 months and results will be tabulated (see dummy Table 24a-b).

Dummy Table 24a: Cytokine outcomes at baseline and 3 months by intervention arm

	Rituximab								Placebo							
	Baseline				3 months				Baseline				3 months			
Cytokines	**n**	**Mean (SD)**	**Median (IQR)**	**Range**	**n**	**Mean (SD)**	**Median (IQR)**	**Range**	**n**	**Mean (SD)**	**Median (IQR)**	**Range**	**n**	**Mean (SD)**	**Median (IQR)**	**Range**
TNF-alpha																
IL1-beta																
IL6																
IFN-gamma																
GMCSF																

Dummy Table 24b: B-cell and cytokine outcomes at 6, 9 and 12 months by intervention arm

	Rituximab									Placebo								
	6 months			9 months			12 months			6 months			9 months			12 months		
Cytokines	**n**	**Mean (SD)**	**Median (IQR) [Range]**	**n**	**Mean (SD)**	**Median (IQR) [Range]**	**n**	**Mean (SD)**	**Median (IQR) [Range]**	**n**	**Mean (SD)**	**Median (IQR) [Range]**	**n**	**Mean (SD)**	**Median (IQR) [Range]**	**n**	**Mean (SD)**	**Median (IQR) [Range]**
TNF-alpha																		
IL1-beta																		
IL6																		
IFN-gamma																		
GMCSF																		

10.2 Association

To address whether biological change with Rituximab is antibody (Ab) dependent the following secondary outcome measures were chosen:

An analysis that examines the correlation between the changes in PBC-40 fatigue domain score and change in cytokines will be carried out. The five cytokines that will be used in the correlation analyses are shown in section 8.1 (items 18a-e).

The changes will be measured as a difference from 12 week measurement to baseline and also 12 month measurement to baseline so correlations on 5 variables (at the two timepoints) for each trial arm will be reported and tabulated (see dummy Table 25).

Dummy Table 25: Correlations (Pearson) between changes in PBC-40 fatigue domain score and changes in cytokine outcomes by intervention arm

Change in PBC-40 fatigue score	**Rituximab**			**Placebo**			**Overall**		
Correlation with:	n	corr	95 % CI	n	corr	95% CI	n	corr	95% CI
Change at 3m from baseline:									
TNF-alpha									
IL1-beta									
IL6									
IFN-gamma									
GMCSF									
Change at 12m from baseline:									
TNF-alpha									
IL1-beta									
IL6									
IFN-gamma									
GMCSF									

The dummy template scatter plot depicted in section 9.5.1 will also be used for change in PBC-40 fatigue score and change in cytokines analyses.

10.3 Biological change

To address whether biological change with Rituximab is Ab dependent the following subgroups analyses will be carried out:

These analyses mimic the change in fatigue domain score and cytokine correlations shown in section 10.2 (see above), but these new subgroup analyses will be restricted to those who responded where 'response' is defined as a decrease in PBC-40 fatigue domain score of 5 or more points. Since the number of 'responders' in this case could be low there could be problems with the correlations outlined below and issues with regression to the mean. We will consider this during the interpretation of the analysis.

An analysis that examines the correlation between the changes in PBC-40 fatigue domain score and change in cytokines will be carried out. The five cytokines that will be used in the correlation analyses are shown in section 8.1 (item 18a-e).

The changes will be measured as a difference from 12 week measurement to baseline and also 12 month measurement to baseline, so correlations on 5 variables (at the two timepoints) for each trial arm will be reported and tabulated (see dummy Table 26).

Dummy Table 26: Correlations (Pearson) between changes in PBC-40 fatigue domain score and changes in cytokine outcomes of responders by intervention arm

Change in PBC-40 fatigue score	**Rituximab**			**Placebo**			**Overall**		
Correlation with:	n	corr	95 % CI	n	corr	95% CI	n	corr	95% CI
Change at 3m from baseline:									
TNF-alpha									
IL1-beta									
IL6									
IFN-gamma									
GMCSF									
Change at 12m from baseline:									
TNF-alpha									
IL1-beta									
IL6									
IFN-gamma									
GMCSF									

11. STORAGE AND ARCHIVING

MACRO, a clinical data management software package will be used for data entry and processing, allowing a full audit trail of any alterations made to the data post entry. Original

questionnaires, CRFs and consent forms will be securely archived at the Newcastle upon Tyne Hospitals NHS Foundation Trust archive facility for fifteen years following publication of the last paper or report from the study.
Data will be handled, computerised and stored in accordance with the Data Protection Act 1998. No participant identifiable data will leave the study site.

12. APPENDIX

Appendix 1

Description of outcome scoring, recoding, calculations and dealing with missing data

Primary outcome

PBC-40 fatigue domain score[b]: Fatigue domain comprises 11 questions with all questions ranging from 1-5 (never, rarely, sometimes, most of the time and always) with overall score ranging from 11-55. Code in Macro was given as 0-4 for each question so recoding to 1-5 was necessary:

Recode: (0=1) (1=2) (2=3) (3=4) (4=5) (e.g. 0 recoded as 1, 1 recoded as 2 … 4 recoded as 5).

If data are missing from fatigue domain (typically missed or duplicated answers) the whole domain should be discarded if <50% of items are completed. If >50% of responses are present then the median value for the completed items in the fatigue domain should be ascribed to the missing item.

Secondary outcomes

PBC-40 domains[b]

PBC-40 symptoms domain score

Symptoms domain comprises 5 questions ranging from 1-5 (never, rarely, sometimes, most of the time and always), 1 question 5-1 and 1 question 0-5 (additionally included a 'did not apply' option scored as zero) with overall score ranging from 6-35. Code in Macro was given as 0-4 in 6 question and 0-5 in one question so recoding to 1-5. 5-1 or 0-5 was necessary:

Recode: (0=1) (1=2) (2=3) (3=4) (4=5) (e.g. 0 recoded as 1, 1 recoded as 2 … 4 recoded as 5) in 5 questions (Q2, 4-6)

Recode: (0=5) (1=4) (2=3) (3=2) (4=1) in one question (Q1)

Recode: (0=1) (1=2) (2=3) (3=4) (4=5) (5=0) in one question (Q3)

PBC-40 itch domain score

Itch domain comprises 3 questions ranging from 0-5 (did not apply, never, rarely, sometimes, most of the time and always) with overall score ranging from 0-15. Code in Macro was given as 0-5 but coded incorrectly so recoding was necessary:

Recode: (0=1) (1=2) (2=3) (3=4) (4=5) (5=0)

PBC-40 cognitive domain score

Cognitive domain comprises 6 questions with all questions ranging from 1-5 (never, rarely, sometimes, most of the time and always) with overall score ranging from 6-30. Code in Macro was given as 0-4 for each question so recoding to 1-5 was necessary:

Recode: (0=1) (1=2) (2=3) (3=4) (4=5) (e.g. 0 recoded as 1, 1 recoded as 2 … 4 recoded as 5).

PBC-40 emotional domain score

Emotional domain comprises 3 questions with all questions ranging from 1-5 (never, rarely, sometimes, most of the time and always) with overall score ranging from 3-15. Code in Macro was given as 0-4 for each question so recoding to 1-5 was necessary:

Recode: (0=1) (1=2) (2=3) (3=4) (4=5) (e.g. 0 recoded as 1, 1 recoded as 2 … 4 recoded as 5).

PBC-40 social domain score

Social domain comprises 2 questions ranging from 1-5 (never, rarely, sometimes, most of the time and always), 6 questions 5-1 and 2 questions 0-5 (additionally included a 'did not apply' option scored as zero) with overall score ranging from 8-50. Code in Macro was given as 0-4 in 6 question and 0-5 in one question so recoding to 1-5. 5-1 or 0-5 was necessary:

Recode: (0=1) (1=2) (2=3) (3=4) (4=5) (e.g. 0 recoded as 1, 1 recoded as 2 … 4 recoded as 5) in 5 questions (Q32, Q40)

Recode: (0=5) (1=4) (2=3) (3=2) (4=1) in one question (Q34-39)

Recode: (0=1) (1=2) (2=3) (3=4) (4=5) (5=0) in one question (Q29, Q31)

If data are missing from a domain (typically missed or duplicated answers) the whole domain should be discarded if <50% of items are 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.

Clinical symptom and functional capability scales

PROMIS-HAQ[f,g]
PROMIS-HAQ questionnaire comprises 20 questions about ability to carry out daily activities (used to assess function) with all questions ranging from 0-4 (without difficulty, with a little difficulty, with some difficulty, with much difficulty, unable to do) with overall score ranging from 0-80. This is then converted to a score out of 100 with higher scores indicating worse functional ability. The PROMIS HAQ measures the functional and physical ability of the participants (covering, dressing, arising, eating, walking, hygiene, reach, grip and activity). Code in Macro was given as 1-5 for each question so recoding to 0-4 was necessary:

Recode: (5=4) (4=3) (3=2) (2=1) (1=0) (e.g. 5 recoded as 4, 4 recoded as 3 … 1 recoded as 0).

If ≥80% (at least 16 questions out of 20 answered) of questions have been answered then the median value for each participants questionnaire score will be ascribed to any missing questions.

COGFAIL questionnaire[h]
COGFAIL questionnaire comprises 25 questions about memory and concentration with all questions ranging from 4-0 (very often, quite often, occasionally, very rarely, never) with overall score ranging from 0-100. MICRO data not downloaded yet as not reported in closed report.

If ≥80% (at least 20 questions out of 25 answered) of questions have been answered then the median value for each participants questionnaire score will be ascribed to any missing questions.

Epworth Sleepiness Scale (ESS) questionnaire[c,d]
ESS questionnaire comprises 8 questions about sleep (in particular sleepiness in the daytime (somnolence)) with all questions ranging from 0-3 (would never dose off, slight chance of dozing, moderate chance of dozing, high chance of dozing) with overall score ranging from 0-24. MICRO data not downloaded yet as not reported in closed report.

If ≥80% (at least 7 questions out of 8 answered) of questions have been answered then the median value for each participants questionnaire score will be ascribed to any missing question.

Orthostatic Grading Scale (OGS) questionnaire[e]

OGS questionnaire comprises 5 questions about dizziness and vasomotor autonomic symptoms with all questions ranging from 0-4 (ranging from never/do not etc to always/mostly etc) with overall score ranging from 0-20. MICRO data not downloaded yet as not reported in closed report.

If ≥80% (at least 4 questions out of 5 answered) of questions have been answered then the median value for each participants questionnaire score will be ascribed to any missing question.

Hospital Anxiety and Depression Scale (HADS) questionnaire[i]

HADS questionnaire comprises 14 questions (7 related to anxiety and 7 related to depression) related to anxiety and depression with questions ranging from 0-3 and 3-0 (ranging from not at all/never/hardly at all etc to very often/most of the time etc) with overall score ranging from 0-42. MICRO data did not need to be recoded in the closed report.

If ≥80% (at least 12 questions out of 14 answered) of questions have been answered then the median value for each participants questionnaire score will be ascribed to any missing questions.

Habitual physical activity Euclidian Norm Minus One (ENMO) outcomes[j,k]

Participants completed physical activity monitoring using wrist worn triaxial accelerometers (GENEA, Unilever Discover, Sharnbrook, Bedfordshire, UK). The accelerometer was worn continuously on the right wrist for a period of 5-7 days in free-living conditions. Accelerometer data was processed in R (www.cran.r-project.org) using R-package GGIR. The first and last hour of the measurement were excluded as they are expected to be influenced by the monitor distribution and collection procedure. Monitor non-wear was detected as described previously and imputed by the average accelerometer data on similar time points on different days of the measurement. Patients were included in the analysis if they had worn the monitor for a minimum time period of 5 days (with at least 1 of these days on the week-end). Only days with at least 22 hours of valid data were retained for further analysis. Thereafter, the average magnitude of wrist acceleration per 5 second epoch was calculated with metric Euclidian Norm Minus One (ENMO) as previously described. ENMO was used to summarise the average magnitude of dynamic wrist acceleration over the measurement period. The output from metric ENMO is in m*g* (1m*g* = 0.001*g* = 0.001 x 9.8 m/s^2 = 0.001 x gravity).

The following accelerometer measures were assessed pre versus post intervention: average acceleration (millig) during the most active (M5) and least active (L5) 5 hr period of each day, time (min/d) spent in moderate-to-vigorous physical activity (MVPA) using a ≥100 mg cut-off with 1 and 5 min bouts.

Participant held fatigue diaries[l]

Average fatigue score will be calculated from participant held fatigue diaries. The diaries measured fatigue using a scale of 1 to 6, where 1 represents no fatigue and 6 extreme fatigue. Participants were asked to complete the diaries six times during the study. They completed the diaries for a period of a week at the beginning of each month at visits: baseline, 1, 3, 6, 9 and 12 months. They returned the diaries at the final visit.

Participants were asked to complete the diaries six times during the study. They completed the diaries for a period of a week at the beginning of each month at visits: baseline, 1, 3, 6, 9 and 12 months. They returned the diaries at the final visit. The average (mean) score (and

SD) will be computed if patient's completed at least 5 days in the week out of the requested 6 times and the overall mean score will range between 1 and 6 inclusive (and be calculated using all days reported by the participant in the diary).

Liver function test (LFT) Alanine aminotransferase test (ALT) and Aspartate aminotransferase (AST) variables

It was not possible to re-run the blood samples from pre-Aug 2015 in the new AST/ALT assays as the samples have been discarded. The clinical team identified data from two sets of non-study samples to compare the old/new assays and a correction factor was needed to convert values generated from the new assay to the old. The prediction from the two datasets was almost perfect as indicated by an R squared value from the linear regression models of >99.9% which showed the model fit was excellent:

ALT=3.7786+1.1994*new value of ALT
AST=0.0935+1.0106*new value of AST

B-cell depletion[m]

B-Cell depletion value is given as the number of CD19+ve cells (B cells) as a percentage of the CD45+ve cells (total lymphocyte population). CD19 is present on all B-cell subsets other than plasma cells and CD45 is a general leucocyte marker.

Anaerobic Threshold

Participants will cycle on a stationary ergometer (Corival, Lode, Nederland) at between 60-90rpm. The test will be terminated voluntarily by the participant or when they were unable to maintain a pedal frequency of 60 revolutions per minute (RMP). Expired air will be collected at rest and during exercise using a breathing mask and analysed online using a gas analysis system (MetaLyzer II, CORTEX, Germany) and heart rate (Polar Electro, Polar, Finland). Peak cardiovascular fitness will be calculated in metabolic equivalents (one MET is equivalent to the oxygen consumption whilst laying quietly or approximately 3.5 ml/kg/min oxygen consumption). Anaerobic threshold will be assessed using the computerised v-slope method and values compared for before and after therapy as the outcome measure.

Anti-PDH Antibody Reactivity

PBC is characterised by autoantibody directed at PDH which is highly effective at inhibiting PDH function *in vitro*. Autoantibody of all isotypes can be quantified using ELISA, with the IgG3 fraction typically predominating. Anti-PDH can also be quantified using a PDH-inhibition-based functional assay; an assay of relevance given the hypothesis being tested in the proposed study. Pilot studies of Rituximab in PBC have demonstrated sustained reduction in anti-PDH antibody of all isotypes. Anti-PDH antibody total and individual isotype levels and antibody functional inhibitory capacity will be studied on day 0 and at the primary end point (12 weeks after therapy). Antibody levels will also be correlated with long term fatigue status during the secondary follow-up period to 12 months. Anti-PDH levels and isotype patterns will be assessed using a well established ELISA developed within our research group. A variety of highly purified native autoantigens (bovine and human PDH) and recombinant (full-length E2 component or the major autoantigenic E2 inner lipoyl domain epitope) will be employed as coating antigens on Immulon 4HBX 96 well microtitre plates (5µg/ml). Detection of bound anti-PDH antibodies will be detected using goat anti-human IgG (including individual IgG1-4 isotypes), IgM and IgA peroxidase conjugated antibody (Sigma). Bound peroxidase activity will be visualised using o-phenylenediamine and measured at 492nm[32]. Anti-PDH inhibitory activity will be assessed using an established assay[13]. In this assay, PDH activity will be measured at 30°C by monitoring the production of nicotinamide adenine dinucleotide (NADH)

at 340 nm. The reaction is initiated by the addition of 2mM pyruvate, to a mixture containing diluted purified bovine PDH, 50mM KPO_4, 0.2mM TPP, 1mM $MgCl_2$, 2.5 mM NAD, 0.13mM coenzyme A, 2.6mM cysteine hydrochloride, in a final volume of 1ml at pH7.4. One unit of enzyme activity catalyses the production of 1µmol of NADH per minute. Before initiating the reaction, serum samples (5µl) are incubated for 30 minutes at 30°C with the PDH containing mixture. Inhibitory capacity of sera is assessed as a percentage of reactivity observed when pre-incubated with phosphate-buffered saline. The specific anti-lipoic acid component of the anti-PDH antibody response will be quantified using the subtractive approach previously described by Bruggraber et al. In the analysis phase impact of Rituximab on fatigue in PBC will be correlated with changes in individual autoantibody isotype responses and with PDH-inhibitory capacity of serum.

Muscle Acidosis

Magnetic resonance data will be acquired prior to first infusion and after 12 weeks follow-up using a 3T Intera Achieva scanner (Philips, Best, NL) with a 14cm diameter ^{31}P surface coil for transmission/reception of signal and the in-built body coil for anatomical imaging. The protocol used for acquisition and analysis has been described fully elsewhere but briefly involved controlled plantar flexion using a purpose-built exercise apparatus developed for operation within the MRI scanner. Participants will perform 2 x 180s bouts of plantar flexion contractions at 25% and then 35% of MVC, with each bout preceded by 60s of rest and followed by 390s of recovery. Phosphorous spectra will be collected at 10s intervals, as previously described. Quantification of spectra will be undertaken using the jMRUI software with metabolite concentrations and metabolic calculations performed as described previously. In particular we will evaluate the minimum pH seen in the exercise and recovery period, the time required post-exercise for pH to return to within 0.01 units of baseline levels (calculated as the sum for each individual for the three bouts to form a total pH recovery time) and the mean “area under the curve” for pH for the 3 exercise episodes which reflects total acid exposure.

Liver Disease Activity

Although not powered to demonstrate biologically significant effects on severity of underlying liver disease the study will provide important pointers to any effect which would inform the design of future studies. Change in serum biochemical end-points has been accepted by The Food and Drug Association (FDA) as an appropriate end-point for clinical trials in PBC. The outcome measure which we will use will be reduction in serum alkaline phosphatase level and attainment of the previously identified positive outcome measure of drop in baseline alkaline phosphatase of >15% or normalisation.

Safety

Safety was **a**ssessed in terms of numbers of adverse events and adverse reactions in the study groups.

References

a) Carbone M, Sharp SJ, Flack S, Paximadas D, Spiess K, Adgey C, Griffiths L, Lim R, Trembling P, Williamson K, Wareham NJ, Aldersley M, Bathgate A, Burroughs AK, Heneghan MA, Neuberger JM, Thorburn D, Hirschfield GM, Cordell HJ, Alexander GJ, Jones DE, Sandford RN, Mells GF; members of the UK-PBC Consortium. The UK-PBC risk scores: Derivation and validation of a scoring system for long-term prediction of end-stage liver disease in primary biliary cholangitis. Hepatology. 2016 Mar; 63(3):930-50. DOI: 10.1002/hep.28017.

b) Jacoby A, Rannard A, Buck D, et al. Development, validation, and evaluation of the PBC-40, a disease specific health related quality of life measure for primary biliary cirrhosis. *Gut*. 2005;54(11):1622-9. doi:10.1136/gut.2005.065862.

c) Johns MW. A New Method for Measuring Daytime Sleepiness. The Epworth Sleepiness Scale. Sleep. 1991; 14(6): 540-5.

d) Johns M. Sleepiness in different situations measured by the Epworth Sleepiness Scale. *Sleep* 1994;17:703–10.

e) Schrezenmaier C, Gehrking JA, Hines SM, Low PA, Benrud-Larson LM, Sandroni P. Evaluation of Orthostatic Hypotension: Relationship of a New Self-Report Instrument to Laboratory-Based Measures. *Mayo Clin Proc*. 2005; 80(3):330-4

f) Bruce B, Fries JF. The Health Assessment Questionnaire (HAQ). *Clin Exp Rheumatol* 2005;23:S14–18.

g) Khanna D, Krishnan E, DeWitt EM, Khanna PP, Spiegel B, Hays RD. Patient-Reported Outcomes Measurement Information System (PROMIS®) -- The future of measuring patient reported outcomes in rheumatology. *Arthritis Care & Research*. 2011;63(Suppl 11):S486-S490. doi:10.1002/acr.20581.

h) Broadbent DE, Cooper PF, FitzGerald P, Parkes KR. The Cognitive Failures Questionnaire (CFQ) and its correlates. *Br J Clin Psychol*. 1982 Feb;21 (Pt 1):1-16.

i) Zigmond A, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatria Scand* 1983;67:361–70.

j) Benjamin E. Garfield, Jane L. Canavan, Cayley J. Smith, Karen A. Ingram, Ria P. Fowler, Amy L. Clark, Michael I. Polkey, William D-C. Stanford Seven-Day Physical Activity Recall questionnaire in COPD. *European Respiratory Journal* Aug 2012, 40 (2) 356-362; DOI: 10.1183/09031936.00113611

k) Physical activity: Jones DEJ. Pathogenesis of primary biliary cirrhosis. Gut 2007;56:1615–24.

l) Jopson L, Newton JL, Palmer J, Floudas A, Isaacs J, Qian J, Wilkinson J, Trenell M, Blamire A, Howel D, Jones DE. RITPBC: B-cell depleting therapy (rituximab) as a treatment for fatigue in primary biliary cirrhosis: study protocol for a randomised controlled trial. *BMJ Open* 2015;5:e007985 doi:10.1136/bmjopen-2015-007985.

m) Zarkhin V, Lovelace PA, Li L, et al. Phenotypic evaluation of B-cell subsets after rituximab for treatment of acute renal allograft rejection in pediatric recipients. *Transplantation* 2011;91: 1010–18.

n) Barcelona Pares A, Caballeria L, Rodes J. Excellent long-term survival in patients with primary biliary cirrhosis and biochemical response to ursodeoxycholic acid. *Gastroenterology* 2006;130:715–20.

o) StataCorp. 2015. *Stata Statistical Software: Release 14*. College Station, TX: StataCorp LP.