Statistical Analysis Plan Draft

October 2014

CLINICAL TRIAL

Code Number: NCT01451593

A phase II double-blind, randomized, placebo-controlled trial of neuroprotection with phenytoin in acute optic neuritis

APPROVED BY:

<table>
<thead>
<tr>
<th></th>
<th>Signature</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chief Investigator</td>
<td>[Signature]</td>
<td>18/11/14</td>
</tr>
<tr>
<td>Trial Statistician</td>
<td>Daniel Mhmann</td>
<td>20/11/14</td>
</tr>
</tbody>
</table>
CONTENTS

1 Introduction and summary of trial design and sample size
2 Summary of demographic and baseline data
3 Analysis of outcome measures
4 Adverse events
5 Further statistical issues and analyses
1. Introduction and summary of trial design

Trial design and details are given in full in the protocol document Trial Protocol Version 3.0 draft.doc, 14/5/13, but relevant issues are summarised here.

We aim to recruit 90 patients to be randomised into the following two arms, with 45 patients per arm:

- Active treatment (phenytoin)
- Placebo

Randomisation was by minimisation, with five binary minimisation variables (see section 2 below).

The planned duration of the trial is six months, with observation time points at 0, 1, 3 and 6 months. Treatment is given during the first three months only. The following data will be collected:

- Measures of visual function: months 0m, 6m [also 1m, 3m for some measures]
- Optical Coherence Tomography (OCT): 0m, 6m
- Magnetic Resonance Imaging (MRI): 0m, 6m
- Visually Evoked Potentials (VEP): 0m, 6m
- Blood and urine: 0m, 1m, 3m, 6m
- Adherence: 1m, 3m tablet return data and 1 and 3m blood phenytoin levels
- Adverse events

The planned primary outcome measure, on which the sample size calculation (below) was based, is retinal nerve fibre layer thickness (RNFL) measured by OCT, as detailed in the Trial Protocol.

Secondary outcome measures will include other OCT measures, visual function measures, MRI measures and VEP measures.

1.1 Calculation of sample size

Data from a longitudinal study of OCT findings obtained in 22 patients with acute demyelinating optic neuritis who were followed serially from initial presentation for 12-18 months at Moorfields Eye Hospital and the Institute of Neurology (Henderson 2010) was used to calculate the sample size to obtain a power of 80% to detect a treatment effect of 50% at 5% significance level (two-tailed), based on an analysis improving efficiency (compared to a simple comparison of means) by comparing mean follow-up affected eye RNFL adjusted for baseline unaffected eye RNFL (a conventional ANCOVA adjusting for baseline affected eye RNFL is avoided because of affected eye inflammation at baseline). For this calculation, adapting standard ANCOVA methods (Senn 1997), 100% treatment effect was taken as the mean difference between six-month affected eye RNFL and baseline unaffected eye RNFL, -19.2μm; six-month affected eye RNFL standard deviation was 18.32; baseline unaffected vs six-month affected RNFL Pearson correlation coefficient, 0.63. This calculation, before allowance for dropout, gave a requirement of 35 participants per arm to demonstrate a 50% treatment effect under the stated power and significance levels. We consider that a 50% treatment effect is required for the result to be persuasive, given that this is a surrogate marker of potential clinical benefit. This outcome is plausible if one extrapolates from animal evidence of the reduction of axonal loss by phenytoin in EAE (Lo 2003) and assumes a robust relationship between axonal loss and atrophy in the RNFL.
Neuroprotection with phenytoin in optic neuritis

The actual sample size of 45 participants per arm allows for a loss to follow-up of 5%, estimated from a drop out rate of 3/68 observed in a previous optic neuritis trial (Kapoor 1998), and non-adherence of 10%, applying factors of 1/(1-dropout fraction) and 1/(1-non-adherence fraction) to n=35, to give n=45; we also take into account the fact that the sample size was calculated using data from a cohort with clinically isolated optic neuritis, whereas we will include some participants with established MS, in whom there is a possibility of prior changes to the RNFL. However, these changes are likely to be minor as the majority of the participants are likely to have isolated optic neuritis, and even those with MS are likely to have early relapsing disease.

1.2 Significance level
Unless otherwise stated, for all statistical tests we will use the conventional two-sided significance level of 5% as the threshold for reporting results as statistically significant.

1.3 Samples for analysis
The primary sample for analysis will be the intention-to-treat (ITT) sample, with secondary per-protocol (PP) samples as described below.

1.3.1 ITT sample
This will comprise all randomized patients with usable data, irrespective of adherence and treatment actually taken and including any patients for which allocation was unblinded during the trial, and any patients with bilateral neuritis or a second acute episode. Details of usable data will be given in the relevant analyses.

1.3.1 PP samples
Since there is no clear prior information on what may be a plausibly effective dose of phenytoin, the per protocol definition will depend solely on adherence, rather than additionally on a minimum prescribed dosage. The primary PP sample will be based on tablet adherence during 0-1 month, and a secondary PP sample will be based on blood phenytoin levels at 1 month.

**Primary PP**
Patients will be included in the per protocol sample (irrespective of whether they are in Active or Placebo) if they satisfy the following two criteria if their adherence percentage at month 1 is greater than a certain threshold (see below):

*Adherence percentage*
If $C = \text{total number of tablets a patient consumed over first month period}$, and $E = \text{total number of tablets a patient should have taken over that period}$, then

adherence percentage = $100 \times \frac{C}{E}$.

The threshold for the adherence percentage will be set at the highest of 90%, 80%,…,30% which includes ≥70% of the ITT sample, and this will then form the primary PP sample.

The primary PP analysis will then compare Active vs Placebo restricted to this PP sample.

**Secondary PP**
This will be a comparison of Active patients who have >0 blood phenytoin level at 1m, vs all Placebo. There is always the danger that a PP sample no longer compares like with like: this danger is more applicable to the secondary PP sample than the primary PP sample, since there is no Placebo subsample which corresponds to the >0 blood phenytoin subsample in Active.
Neuroprotection with phenytoin in optic neuritis

Both PP analyses above will additionally exclude the following categories:
- Patients who are bilaterally affected
- Patients with a second acute episode in the affected eye

2 Descriptive summaries of demographic and baseline data
Demographic and other baseline characteristics will be tabulated by Active and Placebo treatment groups. Descriptive statistics for continuous, ordinal and discrete variables will include as appropriate the mean, standard deviation, median, range and number of observations. Categorical variables will be presented as numbers of patients and percentage of those observed.

A trial profile flow diagram will be provided showing the numbers of patients going through the trial stages, and the number satisfying binary criteria for adherence with protocol.

Tabulated baseline variables will include:

2.1 Minimisation variables (all binary)
- Time from onset (≤ 7 days, > 7 days)
- Centre (London, Sheffield)
- Prior MS diagnosis (yes, no)
- Prescribed disease modifiying treatment (DMD) (yes, no)
- Prescribed steroid (yes, no)

The above five variables were all used with the above binary coding in the minimisation randomisation procedure, and therefore we anticipate these will be relatively evenly balanced between treatment arms at baseline.

Three of these five minimisation variables, ‘Centre’, ‘Prior MS’ and ‘Prescribed DMD’, will be used as pre-specified covariates in the comparisons of Active vs Placebo detailed below, unless the variable has less than 10 subjects in either binary category. In this case the variable will be not be used in the analyses, and this will be clearly indicated in the trial report.

It is thought that the precise number of days between onset and baseline assessment may affect outcome baseline values. Therefore the precise number of days between onset and baseline assessment will be used as a pre-specified covariate in Active vs Placebo comparisons, instead of the binary ‘Time from onset’.

Since also taking steroid medication at or near the time of baseline outcome assessment may affect outcome values, for the purposes of pre-specified covariate adjustment in the Active vs Placebo outcome comparisons, the binary ‘Prescribed steroid’ variable will be replaced by a three-category variable specific to the outcome being compared:

- Days between steroid prescription and baseline outcome assessment
  - > 30
  - 6, ≤ 30
  - 1, ≤ 5

2.2 Other important variables potentially relevant to analyses

Dates:
- Date of onset
- Start date (and end date) for steroid
Neuroprotection with phenytoin in optic neuritis

Dates of MRI scan
Dates of OCT, VEP, visual assessment and phenytoin dispensing

Date of London OCT machine failure
Date of repaired OCT return

Using the above variables, the following will be reported:
Days between onset and MRI scan
Days between onset and OCT/visual assessment
Days between onset and first phenytoin (if different from above)
Three-category variable ‘Days between steroid prescription and baseline assessment’ specific to MRI and to OCT or visual assessment.

MRI scanner if different from Centre
OCT assessor initials:
London – baseline assessor RR, 6m assessor SM
Sheffield – various

Visual function assessor initials for London and Sheffield
Brain MRI

- Binary Brain MRI performed
- Binary Dissem in space
- Binary MS criteria satisfied
- Binary other pathology

Socio-demographic
- Age
- Gender

2.3 Outcome variables observed at baseline and later
Summaries of baseline values for these measures will be reported.

Primary outcome measures
OCT: average RNFL thickness [0, 6 months]. Available for affected and unaffected eye.

Secondary outcome and related measures
OCT, available for affected and unaffected eye:
- Average macular volume [0,6m]
- Average papillomacular bundle thickness [0,6m]
- Four sector-specific RNFLs: temporal, nasal, inferior and superior

Also other OCT measures
Some further OCT measures may be available for analysis, including the following five ‘Pittsburgh’ thickness measures:
- RNFL, GCL + IPL, OPL + INL, ONL

Optic nerve MRI measures [0, 6]. Available for affected and unaffected eye:
- Optic nerve/lesion length (from T2).
- Optic nerve/lesion area (from T1).
- Optic nerve/lesion MTR mean
  - Distance of lesion from globe (three categories: orbital/cranial/canalicular)
  - Binary lesion indistinct/fragmented at 6m
Neuroprotection with phenytoin in optic neuritis

**VEP measures** [0, 6]. Available for affected and unaffected eye. For large and small checked separately:
- Amplitude
- Latency

**Measures of visual function** [0,(1,3) 6]. Available for affected and unaffected eye.
- Logmar visual acuity [0,6, but 1,3 available by conversion of Snellen]
- Snellen visual acuity [0,1,3,6]
- Contrast letter sensitivity: Sloan 1.25% and 2.5% [0,6]
- Colour vision: Farnsworth-Munsell 100 and root error score [0,6]

**Blood and urine** [0, 1, 3, 6]
- Neurofilament concentrations: these will be reported and analysed separately, not for the main trial report under this analysis plan.

### 2.4 Adherence variables

Phenytoin levels [1,3]

**Capsule adherence-related data** [0 to 1m, 1-3m]
- Total dosage prescribed for the period
- Number of 100mg, of 50mg and of 25mg capsules dispensed (and date dispensed)
- Number of 100mg, of 50mg and of 25mg capsules returned (and date returned)

### 3 Analysis of outcome measures and other measures collected during follow-up

Where the analysis method allows this, Active vs Placebo comparisons for a given outcome will be adjusted for all of the following binary minimisation variables which have at least 10 patients in their smallest category:

- Centre, Prior MS, Prescribed DMD

and in addition the following two variables which were described in 2.1 above: i) the precise number of days between onset and the baseline assessment of the outcome being compared; ii) the three-category 'steroid at time of baseline assessment' variable. No other covariates will be used unless pre-specified below. This set of up to five variables is abbreviated in model schematics below as **PSC**.

In the schematics for any models given below, the primary analysis sample is the ITT; however, the secondary PP sample analyses will follow the same model but with the binary Active indicator replaced by a corresponding indicator appropriate for comparing protocol-adherent patients. The 'Secondary Analyses' referred to below indicate analysis methods, rather than samples (and will be repeated for the PP samples).

If for any outcome variable whose measurement is not automated two different assessors have had to be used at the same time point, and this does not coincide with the Centre, a secondary analysis will be performed with an assessor indicator entered as an additional covariate. Such an additional covariate will be identified before analyses and clearly indicated in the trial report.

Where appropriate, normality and homoscedasticity of residuals, or other model assumptions, will be checked. If departures from assumptions are observed in an analysis, it will be repeated using a non-parametric bootstrap or other appropriate non-parametric method.
Neuroprotection with phenytoin in optic neuritis

3.1 Primary outcome, RNFL.
RNFL data will be available from OCT measurements at 0 (baseline) and 6 months.

3.1.1 Primary analysis
The primary analysis for this outcome will use the following simplified schematic for a multiple linear regression:

\[
\text{RNFL}_{\text{affected}, \text{6m}} = \alpha + \beta \text{Active} + \text{RNFL}_{\text{unaffected}, \text{0m}} + \text{PSC}
\]

In this model the coefficient \( \beta \) is the estimated difference in means, Active – Placebo, adjusting for the covariates, and this estimate, its 95% confidence interval and p-value will form the primary trial result.

The rationale of this model is that the minimisation covariates will tend to reduce residual error if they contribute significantly, but without affecting efficiency if they do not (since they will be balanced across the two arms, ie independent of the Active covariate); and the baseline fellow RNFL, as well as tending to reduce residual error, will additionally allow the interpretation of the treatment effect as a comparison of thinning of the RNFL relative to the healthy value in the unaffected eye. The baseline value for the affected eye may be distorted by the inflammatory pathology in the acute stage, and thus is not usable as in a conventional baseline adjusted comparison (ANCOVA). The two covariates replacing the minimisation binary variables, ‘Time from onset’ and ‘Prescribed steroid’ are expected also to be sufficiently well balanced across Active and Placebo not to materially reduce power.

3.1.2 Secondary analyses
The purpose of the secondary analyses i)-iv) outlined below is a) to confirm the robustness of the result of Model (1) in 3.1.1 above; b) to investigate possible non-therapeutic effects of phenytoin. Model (1) above implemented for ITT with no exclusions will give the result to be reported; if this result is materially different or compromised in the light of the secondary analyses, an appropriate caution or comment will be reported.

i) A secondary analysis will examine whether there has been a change in the unaffected eye; examining

\[
\text{change}_{\text{unaffected}} = \text{RNFL}_{\text{unaffected}, \text{6m}} - \text{RNFL}_{\text{unaffected}, \text{0m}}
\]

and whether this change is different in Active vs Placebo:

\[
\text{change}_{\text{unaffected}} = \alpha + \beta \text{Active} + \text{RNFL}_{\text{unaffected}, \text{0m}} + \text{PSC}
\]

If there is evidence of a difference, an additional comparison will be carried out designed to make allowance for osmotic shrinkage in the Active unaffected eye, using as outcome the difference between affected and unaffected RNFL at 6 months:

\[
(\text{RNFL}_{\text{affected}, \text{6m}} - \text{RNFL}_{\text{unaffected}, \text{6m}}) = \alpha + \beta \text{Active} + \text{PSC}
\]

ii) Model (1) will be repeated for the ITT sample minus any bilateral or second episode patients.

iii) Model (1) will be repeated with an additional covariate indicating (London) patients whose 6m OCT measurements were performed on either an interim OCT machine or the repaired original machine.
3.2 **Secondary outcomes**

ITT and PP analyses will be conducted on the secondary outcome measures itemised in section 2.3 above: secondary OCT, optic nerve MRI, VEP and visual function. Unless otherwise specified below, the analyses for these secondary outcomes will again use take the form of a multiple regression with an Active/Placebo indicator with the same set of pre-specified covariates, \( PSC \).

The analyses described below will be repeated for the ITT sample minus any bilateral or second episode patients, and an appropriate caution or comment will be reported if these exclusions materially affect the reported result (with no exclusions).

### 3.2.1 OCT secondary outcomes

The analysis will follow Model 1 with the further OCT outcomes substituted for RNFL, and with the same covariates \( PSC \). Secondary analysis will add a covariate indicating patients whose 6m OCT measurements were performed on either an interim machine or the repaired machine.

### 3.2.2 Optic nerve MRI

The analysis of the lesional optic nerve MRI measures will follow Model 1, with the MRI measure substituted for RNFL, and with the following addition to the pre-specified covariates \( PSC \): the three-category globe-lesion distance variable (described in 2.3 Optic Nerve MRI section).

Lesions which at 6 months have become indistinct or fragmented will be categorised as such and omitted from the comparison above for length. The lesion MTR will still be analysed; and the proportion of such lesions will be compared in a separate analysis (see below).

**Secondary analyses**

i) proportion of indistinct or fragmented lesions at 6m will be compared between Active and Placebo using an exact chi-square test.

ii) Changes between baseline and 6m in the optic nerve measures for the unaffected (contralateral) region will be compared between Active and Placebo using Model 2 above, with the optic nerve measure substituted for RNFL in that model, and the same ‘globe-lesion distance’ addition to \( PSC \) described above. If there is a evidence of a difference consistent with osmotic shrinkage, then allowance for this will be made using Model 3 above, again with the appropriate substitution and same covariates.

iii) Changes between baseline and 6m in the lesion measures will be compared between Active and Placebo, using a regression of the change on the trial group indicator and the same covariates.

### 3.2.3 VEP

These measures will be analysed using Model 1 with the VEP variables substituted for RNFL, and with the same covariates \( PSC \).

However, the following changes to the variables will be carried out in order to avoid potential biasing of results by exclusion of any patients with vision too poor to give a valid reading:

Patients with absent amplitudes and latencies **because of poor vision**, will be assigned an amplitude value of 0 and a latency of 200.

### 3.2.4 Visual function

These measures will be analysed using Model 1, above, with the visual function measures substituted for RNFL; and with the same covariates \( PSC \).

Secondary analyses for Snellen/logmar will use visual function data from all available time points to model trajectories of visual recovery in the affected eye in the Active and Placebo
Neuroprotection with phenytoin in optic neuritis

groups. For this analysis a linear mixed model will be used, with either piecewise linear or polynomial terms in time, depending on the trajectory shape.

4 Adverse events
Adverse events and serious adverse events will be separately tabulated by trial arm. Active vs Placebo comparisons will separately compare serious and other adverse events between trial arms.

The null hypotheses that there is not a greater proportion or rate of adverse events in Active compared to Placebo will be tested with one-tailed tests and P<0.05 significance threshold, using univariable comparisons of proportions or rates, as appropriate.

4.1 Adverse events and blood phenytoin level
Appropriate regression techniques (logistic or poisson) will be used to estimate any association between phenytoin level as explanatory variable and adverse event-related rates or proportions as outcomes.

5 Further statistical issues and analyses

5.1 OCT reproducibility
Multi-technician OCT data will be analysed for reproducibility across technicians. OCT longitudinal stability across the interim or repaired London machine will be investigated using the unaffected eye measurements in Placebo.

5.2 Blood phenytoin and outcome measures
As a secondary analysis, multiple linear regression will be used to investigate any associations between serum phenytoin level and primary and secondary outcome measures.

5.3 Time from onset of visual loss
As a secondary analysis, regression will be used to investigate any associations between primary and secondary outcomes and the time interval between onset and first phenytoin prescription.

5.4 Sensitivity and other planned analyses
If there is evidence that data missing from the six-month outcome may be potentially biasing the treatment effect estimate, a sensitivity analysis will be carried out to investigate the extent of this potential problem.