UNIVERSITY OF BIRMINGHAM

OPT: The Outpatient Polyp Treatment Trial

Statistical Analysis Plan (SAP)

Version Number: 1.1

Effective Date: 23/2/2013 (version 1.0 23/12/2008 – minor changes documented below)

Prepared By: Lee Middleton

1. Introduction

This document gives a detailed statistical analysis plan for the OPT Randomised Controlled Trial, and should be read in conjunction with the current trial protocol.

2. Changes from original SAP (v1.0)

Some minor additions to the original SAP have been added to this version. These comprise of further sensitivity analysis and further description of analysis to minor end points. These are highlighted below in *italics*. No other fundamental changes have been made to the SAP since original approval by the independent Data Monitoring Committee.

3. Study Design

Setting

OPT is a non-inferiority trial designed to determine reliably whether Outpatient polyp removal under local anaesthetic is no worse (or not worse than a pre-specified margin) than Inpatient surgery for women with uterine polyps, and to determine the relative cost-effectiveness of each strategy.

Interventions

Outpatient treatment (a.k.a 'Outpatient' treatment) versus standard Inpatient treatment.

Sample Size

The sample size is chosen to give good statistical power to preclude any clinically important inferiority of Outpatient polypectomy compared to Inpatient treatment.

Outpatient treatment is more convenient for women in that no inpatient stay is required and is also likely to cost substantially less. We believe, therefore, that Outpatient will be the treatment of choice even if 25% less women (in relative terms) have alleviated symptoms at 6 months, i.e. the margin of non-inferiority is set at 0.75. Making the assumption that Inpatient treatment will be 90% successful (as judged by the primary outcome) and Outpatient 80% successful, a sample size of approximately 200 in each arm (400 in total) will be needed to rule out a success rate of less than 67.5% in the Outpatient arm with 90% power, i.e. not more than 25% worse (0.675/0.90=0.75). This calculation was based on a conservative two-sided test at the 5% level (equivalent to a one-sided test at the 2.5% level). To also allow for a 15%

loss to follow-up, the target sample size is inflated to 240 patients in each group (i.e. 480 patients in total).

Primary Outcome

The primary end-point is based on the woman's assessment of their own bleeding and is formed of a dichotomous (yes/no) response. This question differs depending on the referral reason (see table below), but in all cases responding 'yes' will be defined as a success.

Bleeding problem	**Patient type**	**Assessment of bleeding question**
Heavy menstrual	Pre-menopausal & post-menopausal on sequential HRT preparations	Has your bleeding returned to an acceptable level (y/n)?
Intermenstrual	Pre-menopausal & post-menopausal on sequential HRT preparations	Has your intermenstrual bleeding stopped (y/n)?
Bleeding not expected	Post-menopausal taking no HRT or 'no bleed' preparations	Has your bleeding stopped (y/n)?

Primary time-point

Data collected at six months follow-up will be considered the primary time point for analysis. Data will also be collected at one and two years follow-up.

4. General considerations

Levels of confidence and p-values

All results will be presented as point estimates and 95% confidence intervals along with associated p-values from two-sided tests. Analysis will be performed in SAS v9.2.

Analysis population

Analysis will be performed intention to treat (ITT) in the first instance, although as recommended in the CONSORT statement[1] and by Jones[2] a 'per protocol' (PP) analysis for the primary outcome will also be performed as some protection for any theoretical increase in the risk of type I error (erroneously concluding non-inferiority). The ITT analysis will include all randomised patients in the groups they were allocated, regardless of whether the women received this, or indeed any, treatment. The PP analysis will include only those women who received their allocated treatment at the time of their initial operation.

Missing Data

In the first instance, analysis will be completed on received data only with every effort made to follow up participants even after protocol treatment violation to minimise any potential bias.

In addition to this primary method the following analysis will also be completed as a sensitivity analysis (*not included on the original version of the SAP*): To examine the possible impact of missing data on the results, analysis using a multiple imputation approach will be performed on the primary outcome measure. Missing responses will be simulated using a Markov chain Monte Carlo method (MCMC) that assumes an arbitrary missing data pattern and a multivariate normal distribution. Variables including treatment group, the three subgroup variables (listed below) and a variable for each time-point will be included in the model and used to generate 20 simulated data-sets. Analysis will be then be performed (as per the primary analysis proposed) on each set with the results combined using Rubin's rule to obtain a single set of results (treatment effect estimate and confidence intervals).

Late responses

Questionnaires at each time point will be excluded and treated as missing data if they are returned after the subsequent questionnaire has been sent to the patient (e.g. a six month form returned after nine months will be included. A six month formed returned after one year will be excluded and treated as missing). If a late form, which would otherwise be excluded, is the only form available for the later time point it will be included at the subsequent time point (one year in this example). However, if a separate form is returned in time this will be included and the late six month form discarded. All forms will be assumed to have been completed on the completion date

written on the form. If this date is missing, forms will be assumed to have been completed on the date they were received.

Timing of interim analysis

Interim analyses of primary and major secondary end-points will be conducted on behalf of an independent DMC and will occur at least at yearly intervals following commencement of recruitment (or more frequently if deemed necessary). This will include a full safety report. Content of reporting (e.g. presentation of demographic or output tables and plots) will be agreed prior to recruitment with the DMC members along with stopping/modifying criteria (see terms of reference for more information). Pragmatic stopping criteria will be applied where an overwhelmingly convincing difference, likely to change clinical practice, would need to be seen in the primary outcome for the DMC to recommend the study stops early or needs to be modified (see DMC charter for more details).

Timing of final (main) analysis for dissemination

The final analysis will be performed when all recruited patients have reached the two year follow-up stage (or earlier should the Data Monitoring Committee (DMC) recommend that the trial stop earlier if, for example, one particular treatment is overwhelmingly beneficial – see DMC terms of reference). It will include all scores from completed questionnaires up to and including this time. Six month follow-up will be considered the primary analysis time however.

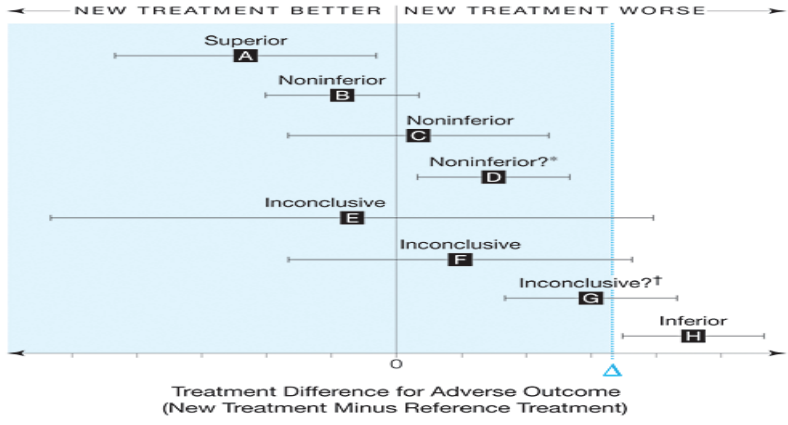
Timing of other planned analyses

There are no further analyses planned. The last follow-up time is at two years.

5. Proposed analyses

Primary endpoint

Unadjusted risk ratios and 95% confidence intervals will be calculated for the primary outcome. These will be generated through the use of a log-binomial regression model. A chi-squared test will be used to examine statistical significance. We will only conclude non-inferiority with Outpatient treatment if the lower band of the 95% confidence interval is not less than the 25% (in relative terms) margin of non-inferiority. See figure below.



Secondary endpoints

Secondary endpoints measured on a continuous scale (scores from MMAS, Euroqol questionnaires, and VAS scores) will be analysed at each time point using a linear model (analysis of covariance) adjusting for baseline score. A repeated measures analysis including all assessment time-points will also be performed for these end-points. Models here included parameters allowing for group, time and baseline score and in the first instance assume a constant treatment effect over time. Time by treatment interaction will be explored though by including this parameter in the linear model. Furthermore, paired t-tests at each time point were used to investigate change scores within groups (*this latter analysis was not included in the original SAP but was considered to be informative*).

Standard tests will be used for other outcome measures: Cochran-Armitage test for trend for ordinal responses, t-tests for continuous data and chi-squared tests for binary and categorical responses.

Planned subgroup analysis

Subgroup analysis will be limited to the primary outcome and to the stratification variables pre-specified in the protocol which comprise of the following (*due to the very small numbers sampled taking HRT and with a history of use of Tamoxifen these two originally proposed subgroup variables will be ignored*):

- Type of bleeding (post-menopausal/heavy menstrual/inter-menstrual)
- Location of uterine polyp (fundal versus non-fundal)

- Type of uterine polyp (endometrial ‘glandulocystic’ versus fibrous)

Standard tests for interaction will be used to explore the effects of these subgroups prior to any examination of effect sizes within and between subgroups, i.e. by testing the statistical significance of interaction parameters (treatment by subgroup) included in the log-binomial regression model. Effect sizes will only be examined if interaction effects are shown to be statistically important (the value of $p<0.05$ will be used here).

Sensitivity analysis

As a sensitivity analysis estimates adjusted for the variables listed above (subgroups) for primary and important secondary endpoints will be generated by adding them to the corresponding linear models (*this is in addition to the original SAP*). Further sensitivity analysis on the primary outcome will be analysis excluding those women who have gone on to receive a further related procedure (e.g. further polyp removal/hysterectomy/ablation) and also an exploration of primary outcome results without forms that were received more than 3 months after their due date. Other sensitivity analysis (*not included in the original SAP)* will include analysis of the primary outcome without those women who had a LNG-IUS system fitted at the time of original operation and also an analysis of bleeding scores following a log-transformation to stabilise the variance.

Safety data

Tables of frequencies of Serious Adverse Events (SAEs) by treatment group will be reported. It is not anticipated these events will be formally analysed – the low anticipated frequencies of events mean we would have low power to detect any differences through any hypothesis testing.

Health Economic analyses

As indicated in the protocol there will also be an economic analysis. The details of this analysis are documented separately.

Any deviations from this plan will be described in the final report.

[1]Reporting of Noninferiority and equivalence trials. An extention of the CONSORT statement. Piaggio et al. *JAMA.* 2006; 295:1152-1160.

[2]Trials to assess equivalence: the importance of rigorous methods. Jones et al. *BMJ.* 1996; 313:36-9.