The **B**u**l**lous Pemph**i**goid **S**teroids and **Te**t**r**acyclines **(BLISTER)** Study

EudraCT number: 2007-006658-24
ISRCTN: 13704604

BLISTER STATISTICAL ANALYSIS PLAN
Version 1.0

Version	Date	Comments
0.1	13/01/2011	First draft
0.2	10/05/2011	Updated following SAP meeting on 18/04/2011
0.3	17/05/2011	Updated following TMG meeting on 16/05/2011
0.4	13/06/2011	Minor changes following TMG meeting on 13/06/2011
0.5	04/07/2011	Missing data section updated
0.51	08/03/2012	Minor updates before sending to IDMC
0.6	23/05/2012	Major change to analysis of primary safety outcome
0.7	31/05/2012	Minor change prior to sending to TMG
0.8	18/06/2012	Minor changes following TMG meeting on 18/06/2012
1.0	03/07/2012	Final version

Contents

1. INTRODUCTION

This document details the planned statistical analyses for an investigator-blind, randomised controlled trial that compares the safety and effectiveness of doxycycline (200 mg/day) with prednisolone (0.5 mg/kg/day) for initial treatment of bullous pemphigoid (BP).

Full details of the background to the trial and its design are presented in the trial protocol.

The analyses described in this document will be performed by the designated statistician at the MRC Clinical Trials Unit. All data will be analysed using STATA.

The trial statisticians responsible for writing this document in discussion with the chief investigator and other principal investigators and conducting the final analyses are:
Daniel Bratton, MRC CTU
Sunita Rehal, MRC CTU
Andrew Nunn, MRC CTU.

2. DESIGN

2.1 Summary

This study is a prospective, 2-arm, single-blind, parallel group, multi-centre randomised controlled trial. A total of 256 participants will be recruited to the study (approximately 210 in the UK). It is anticipated that there will be approximately 45 centres in the UK and 7 in Germany.

At baseline (week 0), participants will be randomised to receive either prednisolone (0.5 mg/kg/day) or doxycycline (200mg/day) for 6 weeks. This is the set dose in the single (investigator) blinded phase. From the week 6 visit onwards, the clinician will be allowed to adjust the dosage to meet the clinical needs of the participant. All participants will be followed up for a total of one year from the date of randomisation. Study follow-up visits will take place at 3, 6, 13, 26, 39 and 52 weeks of follow-up.

2.2 Inclusion Criteria

- Aged at least eighteen years old.
- Able to provide written informed consent.
- Clinical features consistent with bullous pemphigoid.

- Have **either** a direct (skin) **or** indirect (serum) immunofluorescence (linear IgG/C3 at epidermal basement membrane zone) positive for bullous pemphigoid
- A total of at least three significant blisters or recent erosions spread over two or more body sites which have appeared in the week prior to study enrolment. Significant blisters are defined as intact blisters containing fluid which are at least 5mm in diameter. However, if the participant has popped a blister, or the blister is at a site that makes it susceptible to bursting such as the sole of the foot, it can be considered part of the blister count, providing there is a flexible (but not dry) roof present over a moist base.
- Free of blisters and any treatment for previous episodes of bullous pemphigoid for at least one year.

2.3 Exclusion Criteria

- Received any of the study medications or other recognised systemic medications for the treatment of the current episode of bullous pemphigoid prior to study entry. Prior topical treatment is permitted.
- Recent administration of a live virus vaccine.
- Mainly or entirely mucosal pemphigoid.
- Known allergy to any member of the tetracycline family.
- Presence of any condition or use of any medication which precludes the use of either of the study drugs.
- Women of childbearing potential who are not taking adequate contraception or who are pregnant, plan to become pregnant during the study duration or lactating.
- Has any other condition which would, in the Investigator's opinion, deem the participant unsuitable for participation in the study.
- Taking part in any other intervention study.

2.4 Randomisation

The randomisation will be based on a computer generated pseudo-random code using random permuted blocks of randomly varying size, created by the Nottingham Clinical Trials Unit (CTU) in accordance with their standard operating procedure (SOP) and held on a secure server. Access to the sequence will be confined to the Nottingham CTU Data Manager.
Participants will be allocated in a 1:1 ratio to the doxycycline and prednisolone treatment arms. Randomisation will be stratified by disease severity. This is defined by the number of blisters present at baseline:

- Mild = less than10
- Moderate = between 10 and 30 blisters
- Severe = more than 30 blisters

3. STUDY TREATMENT

At randomisation participants will be allocated to receive either doxycycline (200mg/day) or prednisolone (0.5mg/kg/day). During the first 6 weeks of follow-up investigators are blinded to treatment allocation and the treatment regimen is fixed. Between randomisation and the week 3 visit participants are permitted to use up to 100g of topical mometasone furoate or equivalent "potent" topical steroid as per normal clinical practice. Between the week 3 and week 6 visits participants are prohibited from using topical steroids.

At the week 6 visit investigators are unblinded to treatment allocation and are allowed to alter the dose of the allocated treatment as necessary for the rest of follow-up. Furthermore from this visit onwards participants are permitted to use up to 30g/week of topical mometasone furoate or equivalent "potent" topical steroid.

4. OUTCOME MEASURES

4.1 Primary outcome measures

Absolute difference between the two treatment arms in the:

- proportion of participants classed as a treatment success (3 or fewer significant blisters present on examination) at 6 weeks (a non-inferiority comparison)
- proportion of participants experiencing grade 3 (severe), 4 (life-threatening) and 5 (fatal) adverse events which are possibly, probably or definitely related to bullous pemphigoid medication in the 52 weeks following randomisation (a superiority comparison)

4.2 Secondary outcome measures

For the secondary and tertiary endpoints a participant will be classed as a treatment success at each visit if they have 3 or fewer significant blisters present on examination **and** have not had their treatment modified (change of medication or dose of randomised medication increased) on account of a poor response prior to the visit. A participant will be deemed to be a treatment failure if they are not a treatment success or if their cause of death prior to the visit is possibly, probably or definitely related to BP or its treatment as judged by an independent adjudicator.

The following secondary endpoints will be non-inferiority comparisons:
Absolute difference between the two treatment arms in the:

- proportion of participants classed as a treatment success at 6 weeks
- proportion of participants classed as a treatment success at 13 weeks
- proportion of participants classed as a treatment success at 52 weeks
- proportion of participants who have a further episode of bullous pemphigoid defined as a change or escalation of treatment due to

worsening of disease during their participation in the study after previously being classed as a treatment success (either 3 or fewer significant blisters present on prior examination or previously classed as a treatment success on the treatment log)

The following secondary endpoints will be superiority comparisons:
Absolute difference between the two treatment arms in the:

- proportion of participants classed as a treatment success at 6 weeks and are alive at 52 weeks
- proportion of participants reporting any adverse events which are possibly, probably or definitely related to BP medication in the 52 weeks following randomisation
- quality of life (Euroqol EQ5D and Dermatology Life Quality Index DLQI questionnaires completed at 6, 13, 26, 39 and 52 weeks)
- cost-effectiveness over 12 months

4.3 Tertiary outcome measures

The following tertiary endpoints will be non-inferiority comparisons:
Absolute difference between the two treatment arms in the:

- proportion of participants who, on examination at 6 weeks, are completely blister free
- proportion of participants classed as a treatment success at 3 weeks

The following tertiary endpoints will be superiority comparisons:
Difference between the two treatment arms in:

- mortality over the 52 week follow-up period
- amount of potent and super-potent topical corticosteroids used during the 52 weeks of follow-up (total number of grams/week)

5. SAMPLE SIZE CALCULATIONS

A total of 256 participants will be recruited to the study (128 per arm). This should be sufficient to detect a clinically relevant absolute difference of 20% in grade 3, 4 and 5 side effects within one year of randomisation (primary safety objective). This is based on 60% incidence with prednisolone versus 40% with doxycycline at 80% power and a two-sided 5% significance level allowing for a 20% loss to follow-up by one year.

The effectiveness outcome at 6 weeks will be expressed as a two sided 90% confidence interval for the absolute difference in success rates (in terms of blister count) between the prednisolone (control) arm and doxycycline (intervention) arm. It has been assumed that the point estimate for this difference will be 25%, based on an expected response rate of 95% in the control arm and 70% in the intervention. For the upper bound of the 90%

confidence interval for this difference not to exceed an absolute rate difference of **37%** (see figure) with 80% power, a total of 111 evaluable participants per group would be required.

The primary effectiveness outcome measure is at 6 weeks. The attrition rate in the initial 6 weeks is likely to be low (5%). It should be noted that no spontaneous resolution of this condition would be expected at 6 weeks.

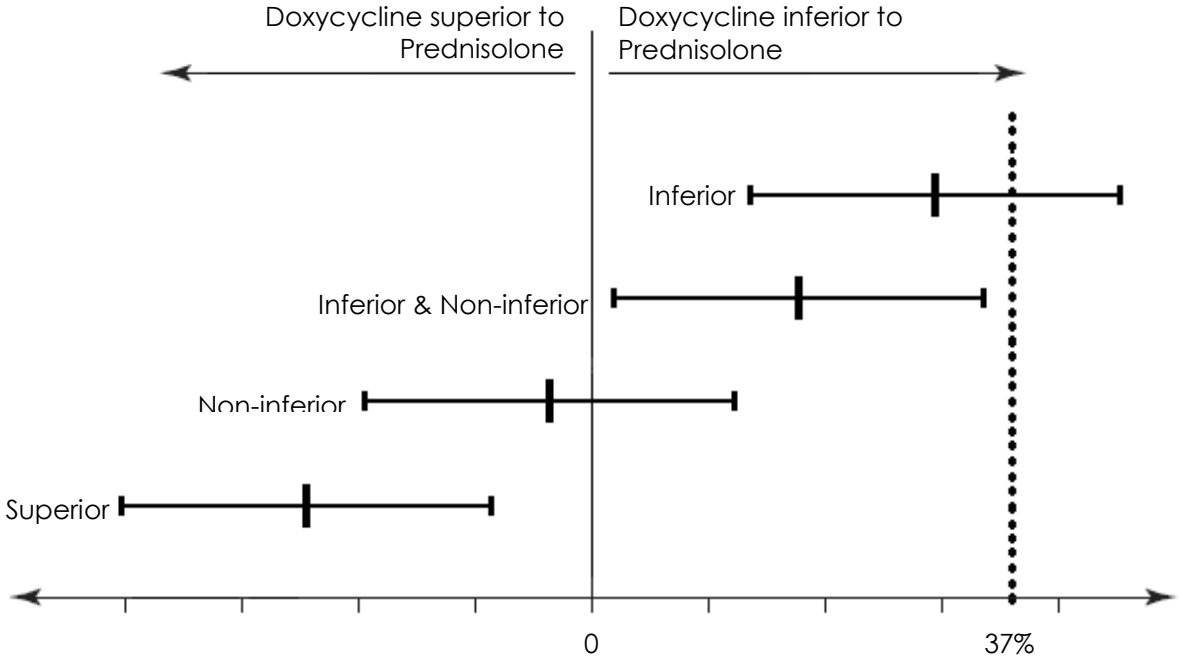

Figure: Four possible outcomes of a non-inferiority trial and their interpretations for the doxycycline arm

6. ANALYSIS PRINCIPLES

6.1 Intention-to-treat and per-protocol definitions

All superiority analyses will be on an intention-to-treat (ITT) basis and all non-inferiority analyses will be performed on **both** the ITT and per-protocol (PP) populations (1).

The ITT population will consist of those participants who fulfil the eligibility criteria, have been randomised to receive either study drug and have data on the outcome of interest.

For each non-inferiority outcome, the doxycycline arm will be considered non-inferior to the prednisolone arm if the upper bound of the confidence

interval for the difference in proportions is less than the agreed non-inferiority margin of an absolute difference of 37% in **both** the ITT and PP analyses

For the non-inferiority outcomes at weeks 3 and 6, participants will be **excluded** from the per-protocol population if, for reasons other than treatment success or failure (determined by the investigator assessing the patient), they have:

- Increased the dose of their allocated treatment
- Changed treatment or added a new treatment to their allocated treatment
- Used topical steroids between visit weeks 3 & 6 (week 6 outcome only)
- Missed more than three consecutive days of treatment
- Committed other protocol deviations deemed to be violations by an independent adjudicator blinded to treatment allocation

For non-inferiority outcomes after week 6, the per-protocol populations will consist of those participants who were included in the per-protocol analysis of the 6 week primary effectiveness outcome and:

- Do not miss more than 3 consecutive weeks of allocated treatment between 6 and 52 weeks (regardless of whether dose has been increased or decreased) unless they have stopped for good clinical response
- Have used no more than 30g of topical steroids per week after week 6
- Have not added systemic steroids to doxycycline (if allocated) or doxycycline or another immunosuppressant to prednisolone (if allocated) unless for poor clinical response.
- Do not commit other protocol deviations deemed to be violations by an independent ad hoc adjudication panel

6.2 Stratification factor

Randomisation will be stratified by disease severity. This is defined by the number of blisters present at baseline:

- Mild = less than 10 blisters
- Moderate = between 10 and 30 blisters
- Severe = more than 30 blisters

All analyses will be adjusted for baseline disease severity to optimise power and reduce bias.

6.3 Other covariates

In addition to the stratification factor, age and Karnofsky score will also be adjusted for as continuous variables in all analyses. All adjustment factors will be entered into models as fixed effects variables.

6.4 Other principles

- No adjustments for multiple testing will be made
- Differences between the two treatment arms will be calculated by subtracting the favourable outcome rate in the doxycycline arm from that in the prednisolone arm
- Non-inferiority will be assessed at the one-sided 5% significance level using a one-sided 95% confidence interval or equivalently the upper bound of a two-sided 90% confidence interval
- All non-inferiority analyses will use an absolute difference of 37% as the non-inferiority margin
- Superiority comparisons will be made at the two-sided 5% significance level using two-sided 95% confidence intervals and the p-value for the treatment difference (under the null hypothesis that there is no difference) on the ITT population only
- Participants with no follow-up data on a particular outcome will be excluded from the analysis of that outcome

7. ANALYSIS DETAILS

7.1 Participant flowchart

Participant throughput, from those screened for entry through those who are eligible (meet all inclusion criteria and no exclusion criteria) for the trial will be reported. The throughput of participants from those eligible to those randomised and to those that are included in the ITT and PP analyses of the primary endpoints will be summarised in a flowchart.

The number of participants who are excluded (failure to satisfy inclusion and exclusion criteria, refusal to participate), discontinued from allocated treatment, and discontinued from follow-up will be reported.

7.2 Baseline characteristics

Participants in the two treatment arms will be described separately by treatment arm with respect to all characteristics recorded at baseline.

Categorical variables will be summarised by number and percentage in each category and continuous variables will be summarised by mean and standard deviation or median, 25th and 75th percentiles as appropriate. No formal statistical tests will be performed since any differences between treatment groups at baseline should be the result of chance rather than bias.

7.3 Primary endpoint analyses

Primary effectiveness outcome

The number and percentage of treatment successes (3 or fewer blisters present at 6 week examination) in the ITT and PP populations will be summarised by treatment arm. Participants who have died or withdrawn

from follow-up before the week 6 visit will be excluded from the analysis. The adjusted absolute difference between the two treatment arms in the proportion of treatment successes at 6 weeks will be estimated using a binomial regression model with the identity link function (2).

Primary safety outcome

Only adverse events (AEs) of grade 3 or above which are possibly, probably or definitely related to BP medication and have occurred or worsened to grade 3 or more during the 52 weeks following randomisation will be considered in the analysis of the primary safety outcome. The relatedness of the grade 3 and 4 AEs to BP treatment is judged by the investigator examining the patient and verified by the chief investigator through the serious adverse reaction reporting process. The relatedness of grade 5 (fatal) adverse events to BP treatment will be judged by an independent adjudicator.

Participants with AE data available for at least one study visit will be included in the analysis. Missing data at other visits will be imputed as follows:

- For each study visit (weeks 3, 6, 13, 26, 39, 52) a binary variable indicating whether a participant has experienced a new or worsened grade 3, 4 or 5 adverse event (AE) at that visit will be generated
- Missing data in these variables will be imputed using multiple imputation using chained equations (`ice` command in Stata)
- Imputation models will use logistic regression to impute missing outcomes and will include as predictor variables: AE indicator variables for other study visits, age, Karnofsky score, baseline disease severity and variables indicating the number of grade 1, 2, 3 and 4 AEs present at baseline (note grade 5 is death and so is not included)
- Imputations will be done separately within treatment groups
- Fifty imputations will be generated
- For each imputed dataset, the adjusted absolute difference between the two treatment arms in the proportion of participants experiencing adverse events of interest during 52-week follow-up will be estimated using a binomial regression model with the identity link function. Treatment estimates from each imputed dataset will then be combined using Rubin's rules.

7.4 Secondary endpoint analyses

See section 4.2 for definition of treatment success for the secondary and tertiary outcomes.

Participants who have died during follow-up and whose deaths are possibly, probably or definitely related to the study treatment will be classed as treatment failures for the remainder of the one year follow-up period.

Participants with missing blister count at a particular visit will be classed as treatment failures if they had their treatment modified (changed treatment or dose increased) on account of a poor response before this visit. Participants who withdrew from the trial due to treatment failure will also be classified as such for the remainder of follow-up. Participants whose outcomes cannot be ascertained will be excluded from the analysis of that outcome.

All absolute differences in proportions will be estimated using the treatment coefficient obtained from a binomial regression model with the identity link function.

The following secondary endpoints will be non-inferiority analyses and will use the same non-inferiority margin as the primary effectiveness outcome (absolute difference of 37%):
The absolute difference between the two treatment arms in:

- the proportion of participants classed as a treatment success at 6 weeks
- the proportion of participants classed as a treatment success at 13 weeks
- the proportion of participants classed as a treatment success at 52 weeks
- the proportion of participants who have a further episode of bullous pemphigoid during their participation in the study after previously being classed as a treatment success (see section 4.2 for definition).

The following secondary endpoints will be superiority analyses:
The absolute difference between the two treatment arms in:

- the proportion of participants classed as a treatment success at 6 weeks and are alive at one year
- the proportion of participants reporting any new adverse events, or adverse events which have increased in severity from baseline, which are possibly, probably or definitely related to BP treatment during the 52 weeks following randomisation. The five grading categories will also be compared separately.
- quality of life (Euroqol EQ5D and Dermatology Life Quality Index DLQI) scores will be analysed separately in multilevel models for repeated measures with score as the response variable and including visit week and baseline scores as fixed effects with participant-specific and visit-specific random intercepts (with the latter nested within the former). These models will allow differences between the two treatment arms in the quality of life scores to be estimated at each visit week. An unstructured covariance structure will be used in each model.

7.5 Tertiary endpoint analyses

All absolute differences in proportions will be estimated using the treatment coefficient obtained from a binomial regression model with the identity link function.

The following tertiary endpoints will be non-inferiority analyses and will use the same non-inferiority margin as the primary effectiveness outcome (absolute difference of 37%):
The absolute difference between the two treatment arms in:

- the proportion of participants who, on examination at 6 weeks, are completely blister free
- the proportion of participants who are classed as a treatment success at 3 weeks

The following tertiary endpoints will be superiority analyses:

- All-cause mortality will be illustrated using a Kaplan-Meier plot. To test for a difference between treatment groups Cox regression will be used, however, if the proportional hazards assumption is violated restricted mean survival time will be used (3). Participants withdrawing from the trial will be censored at the date of withdrawal. Deaths will also be summarised in a table by treatment group and relatedness to treatment
- The average amount of potent and super-potent topical corticosteroid (grams/week) used will be summarised by treatment arm and type of topical steroid. Multivariable regression models will be used to compare the average amounts used between the two treatment arms

All primary, secondary and tertiary outcomes will also be summarised by treatment arm using descriptive statistics (mean (SD), median (IQR) for continuous outcomes and number (%) for binary outcomes).

7.6 Cost-effectiveness analysis

Cost-effectiveness will be analysed by Professor James Mason (University of Durham) and a separate analysis plan will be written for this endpoint.

8. ADDITIONAL ANALYSES

8.1 Sensitivity analyses

Primary safety outcome

A sensitivity analysis will be performed on the primary safety outcome where participants with missing AE data at a particular visit are assumed not to have experienced a grade 3, 4 or 5 event at that visit. This will only apply to participants who have at least one visit with AE data observed.

Due to the difficulty and subjectivity in determining the causality of deaths, another sensitivity analysis will be performed in which all deaths, regardless of their relationship to BP medication, are classified as an unfavourable outcome in the analysis.

8.2 Imputation analysis on 6 week blister count

Missing data can lead to a loss of power and biased results. Baseline data is unlikely to be missing but there may be unobserved outcomes (e.g. in those who are lost to follow-up). Since participants with missing blister counts at 6 weeks will be excluded from the primary effectiveness analysis, imputation will be used to determine the sensitivity of the observed results to the missing data.

Missing-at-random

Under the missing-at-random assumption the missingness of the data is assumed to depend only on the observed data. The plausibility of this assumption will be determined by comparing summary statistics of observed data between the group of participants with and the group without missing data. If the assumption is plausible there should be a difference in some of the summary statistics between these two groups. If this is the case an imputation analysis of the primary 6 week blister count outcome will proceed as follows:

- Participants with at least one blister count available during follow-up will be included in the imputation analysis
- At each study visit (week 3, 6, 13, 26, 39, 52) a binary variable indicating whether a participant has 3 or fewer blister (treatment success) is recorded
- Missing data in these binary variables will be imputed using multiple imputation using chained equations (`ice` command in Stata)
- Imputation models will use logistic regression to impute missing outcomes and will include as predictor variables: treatment success indicator variables for other study visits, age, Karnofsky score and baseline disease severity
- Imputations will be done separately within treatment groups
- Fifty imputations will be generated
- For each imputed dataset, the adjusted difference between the two treatment arms in the proportion of participants classed as a treatment success at 6 weeks will be estimated using a binomial regression model with the identity link function. Treatment estimates from each imputed dataset will then be combined using Rubin's rules.

Missing-not-at-random

The 6 week effectiveness outcome will also be assessed under a range of missing-not-at-random assumption (i.e. the missingness of the data depend on the value of the missing data). Missing outcome data will be imputed under various assumptions around the informative missing odds ratio (IMOR). The IMOR for each arm is the ratio of the odds of treatment successes between

those participants with missing data and those with observed data (4). IMOR=0 imputes all missing data as treatment failure. Varying the IMOR will show how sensitive the primary results are to the assumptions made to the missing data. The imputation analysis will consist of three parts:

i. Impute missing values in the prednisolone arm as treatment failures and vary the IMOR for the doxycycline arm
ii. Impute missing values in the doxycycline arm as treatment failures and vary the IMOR for the prednisolone arm
iii. Vary the IMOR in both arms

Low IMORs which substantially change the conclusions of the trial will show that the observed results are sensitive to the assumptions made upon the missing data.

8.3 Subgroup analyses

Effectiveness outcome at 6 weeks

For some patients, the blister count at 6 weeks may have been performed by an investigator who knew what treatment the patient was on. To determine whether this may have introduced bias into the results of the 6 week effectiveness outcome an interaction test will be performed to compare the treatment effects in patients who were and were not assessed by an investigator who knew the treatment allocation. This analysis will be performed on both the primary and secondary definition of treatment success at 6 weeks.

Subgroup analyses of treatment success (using both definitions) at 6 weeks by the three categories of baseline disease severity (mild, moderate and severe) will be performed. A global test for a treatment interaction will be used to determine whether treatment effects are different in the three categories of disease severity.

Primary safety outcome

The difference between the two treatment arms in the primary safety outcome will be compared by age at baseline. Age will be categorised by its quartiles and the treatment effect in each subgroup will be estimated and compared using a global test for interaction. Should a strong treatment-covariate interaction exist a continuous treatment effect plot will be obtained from a fractional polynomial model (`mfpi` command in STATA) to show the treatment-covariate interaction in more detail (5). The results from the two analyses should be consistent; however, should the two models not agree the results from the subgroup analysis will be used.

9. SIGNATURES OF APPROVAL

Date: 03/07/2012
Version: 1.0

Signatures

Name	Trial Role	Signature	Date
Hywel Williams	Chief Investigator		
Andrew Nunn	Senior Trial Statistician		

10. REFERENCES

1. D'Agostino RB, Sr., Massaro JM, Sullivan LM. Non-inferiority trials: design concepts and issues - the encounters of academic consultants in statistics. Stat Med. 2003 Jan 30;22(2):169-86.
2. Hardin JW, Cleves MA. sbe29: Generalized linear models: Extensions to the binomial family. Stata Technical Bulletin. 1999(50):21-5.
3. Royston P, Parmar MKB. The use of restricted mean survival time to estimate the treatment effect in randomized clinical trials when the proportional hazards assumption is in doubt. Statistics in Medicine. 2011;30(19):2409-21.
4. Higgins JP, White IR, Wood AM. Imputation methods for missing outcome data in meta-analysis of clinical trials. Clin Trials. 2008;5(3):225-39.
5. Royston P, Sauerbrei W. A new approach to modelling interactions between treatment and continuous covariates in clinical trials by using fractional polynomials. Stat Med. 2004 Aug 30;23(16):2509-25.