Cancer Group Statistical Analysis Template Instruction for Use

This is statistical analysis plan template designed to assist with the formation of all analyses (interim and final) involved with a cancer trial. This template should be considered as a guideline/suggested example template and not an instruction for writing a detailed statistical analysis plan.

Sections highlighted in yellow recommended to be completed as it appears in this text.

Sections highlighted in blue are examples and can be used directly in the statistical analysis plan.

Sections highlighted in green were considered standard for cancer trials in the unit but should be amended if appropriate to reflect current working instructions and guidance for writing statistical analysis plans. Trial specific details can also be added as appropriate.

TRIAL Statistical Analysis Plan			
Version Number and Date: Draft/Version x.x DD-MMM-YYYY			
Author	**Position**	**Signature**	**Date**
Reviewed by	Experienced Statistician, affiliation		
	Experienced Statistician, affiliation		
Approved by			

Revision History

Version	Author	Date	Reason for Revision
Draft 0.1			**Protocol version 1.0**
Draft x.x			**First draft**
Final 1.0			**Accepted all changes in Draft 0.x**
Draft 1.1			
Final 2.0			**Accepted all changes in Draft 1.x**

MRC Clinical Trials Unit at UCL

Contacts
Names of MRC CTU contacts

Email: trial@ctu.mrc.ac.uk

This template and all preceding versions will be stored in the Statistical Analysis Master File for this trial, either on paper (currently in cupboards D153 and D154) or electronically if the trial uses an eTMF.

Contents

1. INTRODUCTION

1.1 Background and Rationale

Give a synopsis of trial background and rationale including a brief description of research question and brief justification for undertaking the trial. Include
the trial aims/objectives/hypotheses, treatment/randomisation arms and the patient eligibility criteria. Include a trial flowchart.

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

1.2 Outcome measures

List the primary and secondary outcome measures under the headings below. Give definition of terms in **Section 2.2**.

1.2.1 Primary outcome measure(s)

List of primary outcome measures.

1.2.2 Secondary outcome measures

List of secondary outcome measures.

2. STUDY METHODS

2.1 Trial Design

Brief description of trial design including type of trial (e.g., parallel group, multiarm, crossover, factorial) and allocation ratio and may include brief description of interventions.

Include details on superiority, equivalence, or noninferiority hypothesis testing framework, including which comparisons will be presented on this basis.

2.2 Definition of terms

Give definition of any terms that require explanation.

E.g.

Term	Definition
Overall survival (OS)	Time from randomisation until death from any cause
Censoring	

If a more detailed definition is necessary, include this in the appendix.

2.3 Randomisation

Randomization details, e.g., whether any minimization or stratification occurred (including stratifying factors used or the location of that information if it is not held within the SAP)

2.4 Sample Size Calculations

Give justification for sample size calculations including all relevant sources of information on which the calculation is based on e.g. number of expected events, expected difference in treatment effect. Give details of assumptions used (e.g. proportional hazards).

2.5 Interim Analysis

Give details of any interim analyses to be carried out, the associated time points, any planned adjustment of the significance level(s), as well as details of guidelines for stopping the trial early

If none are planned, then state so.

2.6 Final Analysis

Give details of the timing of final analysis, e.g., all outcomes analyzed collectively, or timing stratified by planned length of follow-up? Also include details of time points at which the outcomes are measured including visit "windows".

3. STATISTICAL PRINCIPLES

Give details of analysis principles. *Sufficient detail should be given to allow independent programming of the primary endpoint from the cleaned analysis files by a statistician who did not perform the initial analysis.* The following information should be included:

- Definition of analysis populations (intention-to-treat, per-protocol, safety etc.)
- Significance levels of tests and level of confidence intervals (CI)
- The use of one-sided or two-sided tests
- Analysis methods, include any formulae
- Subgroup analyses, if applicable
- Definition of adherence to the intervention and how this is assessed including extent of exposure and presented
- Dose intensity analysis, if applicable
- Adjustments for multiple testing and detailing of controlling the type 1 error, if applicable
- Definition and presentation of protocol deviations
- Description of missing data, the assumptions of the missing data mechanisms and how this will be handled during the analysis

For Survival Analysis, it is suggested that:

- Time-to-event data to be presented as Kaplan Meier plots
- Comparisons made by Log-rank χ^2 tests, Log-rank hazard ratios with CI
- Number of events and median follow-up time be reported by comparison groups (follow-up time as median FU in patients last live and by Reversed Kaplan Meier survival plots)

Below is a short example you may use and expand on:

All analyses will be performed on an intention-to-treat basis. All statistical tests will be at the 2-sided p-value of 0.05. There will be no formal adjustment of p-values because of the interim analyses performed.

Time-to-event data will be presented using Kaplan-Meier plots. In all time-to-event analyses, patients that have not experienced the event in question (e.g. progression) will be censored on the date last seen. Time-to-event data will be tested using a log-rank χ^2 test. The number of events observed and the log-rank expected number of events will be presented. Log-rank hazard ratios will be calculated to test the difference between the treatment arms as this methods makes no assumptions about proportionality of the hazards on each arm. All hazard ratios (HR) will be presented with a 2-sided 95% confidence interval.
To assess whether any treatment is more or less effective in well-defined subgroups, $\chi 2$ tests for heterogeneity or, when appropriate, trend will be performed. Forest plots will be presented to visually summarize the consistency of an effect over the subgroups.
Median follow-up time will be calculated. Firstly, using a Kaplan-Meier approach, taking date last seen (if alive) to be an event and death as the time of censorship. The median time will be read for each group from the life tables. Secondly, in patients who were last alive, the median time to last follow-up will be calculated using standard summary statistics.

Additional details could include:
How graphs will be labelled and scaled
Important time points in your analysis
Any stratification in your analysis
Adjustment for any factors in your model
Model Assumptions
Details of non-standard statistical techniques used
How certain variables would be derived from other variables

Suggested approach for the analysis of a survival outcome when the proportional hazards assumption is violated is given below. It closely follows that presented in Royston and Parmar, 2011:

1. The treatment effect should be tested using a logrank test and conclusions regarding the null hypothesis drawn accordingly. The logrank test is known to have good power under PH.

2. Regardless of whether the log-rank test in step 1 is 'significant', the treatment effect should be tested for non-PH. The best-known approach is the Grambsch–Therneau test based on scaled Schoenfeld residuals from a Cox model. Since a bare P-value from a hypothesis test is not particularly informative, we also recommend producing a graphical diagnostic.

3. If there is no evidence of non-PH, the primary summary of the treatment effect is the HR andits confidence interval.

4. If there is evidence of non-PH, the primary measure of the treatment effect switches to the difference in restricted mean survival time at $t*$ and its confidence interval. Either the flexible parametric or the pseudovalues method may be used. The value of $t*$ should be used as specified apriori in the trial protocol. A reasonable choice for $t*$ should be clinically motivated. In the absence of such a choice, a default $t*$ may be taken to be slightly below the maximum expected follow-up time. Although a single estimate of the HR is no longer meaningful, it is of scientific interest in a secondary analysis to estimate and plot the HR as a function of time, as mentioned at step 1.

4. DATA

4.1 CRF Forms and variables

Full details of data collection and timings are described in the trial protocol.
A copy of the CRFs and Quality of Life (QL) questionnaires (if applicable) is presented in the protocol and the Trial Master File. Details of the variables are presented as in the metadata which forms part of the Trial Master File.

4.2 Management of datasets

Below is the standard policy for management of data in the Cancer Group as given in the Statistical standard operating procedure. You may add additional trial specific data management details to this.

At the time of analysis:

For trials databases stored in Macro/OpenClinica:

- The Statistician will file out from Macro/OpenClinica a dataset of all data stored in the database. This will act as the frozen dataset. It is the responsibility of the Statistician to accurately record the date of freezing and ensure all data is retrieved.
- New data can continue to be entered onto Macro/OpenClinica database.
- If any outstanding data queries are resolved during the analysis that relate to data in the frozen dataset (e.g. problems that are found during analysis or amended CRFs that are returned to CTU), the main Macro/OpenClinica database should be changed under the oversight of the Trial Manager.

For trial databases stored in CACTUS:

- Data is downloaded at regular intervals (these are agreed with DMS and stored in a designated location in the SMF). The Statistician will designate which dataset will act as the frozen dataset. It is the responsibility of the Statistician to accurately record the date of freezing and ensure all data is retrieved.
- If any outstanding data queries are resolved during the analysis that relate to data in the frozen dataset (e.g. problems that are found during analysis or amended CRFs that are returned to CTU), the CACTUS database should be changed under the oversight of the Trial Manager.

4.3 Data completion schedule

Describe the time points for completion of clinical record forms and quality of life questionnaires. You may include a timetable.

4.4 Data verification

Below is the standard policy for verification of data collected in clinical trials carried out by the Cancer group. You may add any additional trial specific data verification procedures.

Data verification, consistency and range checks will have been performed at the data entry stage by the MRC CTU, as well as checks for missing data (copies can be found in the Trial Master File). Additional range, consistency and missing data checks will be performed, as appropriate, when the analysis is performed (and when the datasets for analysis are

constructed). All variables will be examined for unusual, outlying, unlabelled or inconsistent values.

Given the thorough nature of our follow-up procedure we expect the issue of missing data to be relatively minimal. We anticipate high compliance with initial data collection as this is close to the time of patient registration. If any data is missing imputation will not be done.

Any problems with trial data will be queried with the Trial Managers, Data Managers, or statisticians, as appropriate. If possible, data queries will be resolved, although it is accepted that due to administrative reasons and data availability a small number of problems will continue to exist. This will be minimised.

4.5 Data coding

State where a list of variable coding can be found if this is not included in the appendix. Give details of coding of free-text variables if this is applicable to your analyses.

5. TRIAL POPULATION

Describe what and how data collected in the trial will be presented. Below are some suggested headings and example of ways of presenting or assessing data:

5.1 Recruitment and follow-up patterns

- Summary of eligibility criteria (how many patients were eligible and how many were excluded due to violating each inclusion/exclusion criteria)
- Example of the CONSORT diagram
- Screening data and failures
- Recruitment by year and centre
- The number of forms returned
- For surviving patients, the time since last follow-up form received e.g. by centre
- Withdrawal and lost of follow-up summaries (including timings of those taken place)

6. ANALYSIS

The results of the analyses will be reported following the principle of the ICH E3 guidelines on the Structure and Content of Clinical Study Reports.

6.1 Baseline Characteristics

List all baseline characteristics/stratification factors and describe how they will be reported (e.g. by treatment arm etc).

6.2 Trial treatment

- Time from randomisation to start of treatment
- Number of treatments administered
- Treatment compliance/tolerance
- Frequency and reason for dose modification and delays, treatment termination
- Additional treatments given

6.3 Analysis of the Primary Outcome(s)

In this section list which analysis will be used (described under Statistical Principles), the population that will be used and how results will be presented.

Include any calculation or transformation used to derive the outcome(s) (eg, change from baseline, QoL score, time to event, logarithm, etc).

6.3.1 Subgroup Analysis

6.3.2 Sensitivity Analysis

6.4 Analysis of the Secondary Outcomes

As per Section 6.3.

6.4.1 Subgroup Analyses

6.4.2 Sensitivity Analyses

6.5 Prognostic factors

Give details of the investigation of prognostic factors, if applicable.

6.6 Quality of life analysis

- A reminder of questionnaires used
- Assessment of baseline QL

- Construction of QL subscales/ scores
- QL hypotheses and time-points

Expert advice should be sought before planning and conducting any quality of life analyses and refer to "Clinical Trials in Cancer" – D. Girling, M. Parmar, S. Stenning, R. Stephens and L. Stewart.

6.7 Health Economics

Health economics (HE) analyses are generally not performed at the CTU. Give details of HE analyses that will be carried out by outside members, E.g. what data will be released, to whom. Refer to the HE statistical analysis plan if one is available.

6.8 Additional Analyses

Use this section to describe any analyses that have not been described yet.

6.9 Statistical Software

Details of statistical packages to be used to carry out analyses.

7. HARMS

This section should include information on the safety data in addition to what was described in Section 6.4, given safety/toxicity is a secondary outcome. Any additional summaries not described above can be included here as follows.

Include sufficient detail on summarizing safety data, e.g., information on severity, expectedness, and causality; details of how adverse events/serious adverse events are coded or categorized; the time period that these will be collected and any how changes over time will be examined;

Describe how adverse event data will be analyzed, i.e., grade 3/4 only, incidence case analysis, intervention emergent analysis. Include any description of assessment of baseline symptoms, if applicable.

8. ADDITIONAL INFORMATION IN THE CLINICAL PAPER

Include any additional information that will be included in the clinical paper, e.g. outline of reviews by IDMC (timing and recommendation)

Include in the appendix any additional information that is relevant to the analyses such as:

- Additional sample size issues
- Changes in trial objectives e.g. change in comparisons groups
- Surveys carried out to aid hypothesis generation
- Detailed definition of an event
- How the Statistical Analysis Plan for the final data related to the SAP for interim analyses

9. REFERENCES

9.1 Referenced for non-standard statistical methods

Royston, P. and Parmar, M.K.B. (2011), The use of restricted mean survival time to estimate the treatment effect in randomized clinical trials when the proportional hazards assumption is in doubt. Statist. Med., 30: 2409-2421. https://doi.org/10.1002/sim.4274

9.2 References to trial related documents (DMP, TMF, SMF)

9.3 Reference to SOPs or other policy documents

The development of the statistical analysis plan follows all relevant CTU SOPs. If any additional documents and policies have been followed, list them here and provide details in the Appendix.

APPENDIX 1:

APPENDIX 2:

APPENDIX 3:

Acknowledgement

This Statistical Analysis Template is constructed with the aid of an original document by Matthew Sydes.