Standard Operating Procedure CCTU/SOP023

Statistical Analysis Plans

1. Scope

This Standard Operating Procedure will be followed by Cambridge Clinical Trial Unit (CCTU) statisticians working on Cambridge Sponsored CTIMPs or clinical trials managed by the CCTU.

2. Purpose

The purpose of this SOP is to describe the procedure for the preparation of a statistical analysis plan (SAP) for a trial that will comply with the study protocol, GCP guidelines and other statutory and regulatory requirements.

3. Definitions and Abbreviations

The headings below contain the definitions of terms and meaning of abbreviations used within the document.

3.1. Definitions

Term	Definition
Cambridge Sponsored	Sponsored by Cambridge University Hospitals NHS Foundation Trust (CUH); or the University of Cambridge (UoC); or jointly by CUH and UoC OR Sponsored by: Cambridge University Hospitals NHS Foundation Trust (CUH) or CUH jointly with the University of Cambridge or Cambridgeshire & Peterborough NHS Foundation Trust (CPFT) or CPFT jointly with the University of Cambridge
Trial or Study	The terms 'trial' and 'study' can have subtly different meanings, however in this document there is no distinction and 'study' is used throughout. It means, in the most general sense, an activity to collect and analyse data in a systematic fashion.

3.2. Abbreviations

Abbreviation	Meaning
CCTU	Cambridge Clinical Trials Unit
SOP	Standard Operating Procedure
SAP	Statistical Analysis Plan
CRF	Case Report Form
CI	Chief Investigator
CACE	Complier Average Causal Effect
PI	Principal Investigator

4. Undertaken by

- Each study should have a designated Study Statistician, with appropriate qualifications and experience who assumes ultimate responsibility for all statistical aspects of the study
- The designated Study Statistician should be named in the study protocol.
- Some tasks may be delegated to other statisticians involved in the study; the designated Study Statistician must exercise appropriate judgment concerning delegated tasks and should check that these tasks are performed appropriately and accurately
- The Study Statistician will interact closely with other members of the study team, including but not limited to the CI,PI, Project Leader, Clinical Trials Coordinators, Data Managers and Programmers

5. Items Required

- CCTU/TPL007 SAP Template
- CCTU/GD038 Storage and Access to Confidential Information
- CCTU/SOP022 Statistical Principles
- CCTU/TPL064 Query Management Log
- Protocol study specific
- CRF study specific

6. Summary of Significant Changes

Amending FRM079 to TPL064.

Addition of version details of software.

7. Method

The following sections provide a description of the processes to be followed when implementing this document's procedures.

7.1. Development of the Statistical Analysis Plan (SAP)

- The SAP contains the pre-specified statistical analyses to be performed upon completion of the study and detail of any planned interim analyses
- Use CCTU/TPL007 the SAP Template to document the process. Dependant on the study and the stage of the analysis, delete sections as appropriate

7.1.1. No Interim Analyses

- The SAP will be developed during the progress of the study using Template CCTU/TPL 007. This should be complete before the final database lock for use in the final statistical analysis

7.1.2. With Interim Analyses

- The SAP is developed using the SAP Template CCTU/TPL 007 before the data download for any interim analysis

- If multiple interim reports are prepared it must be clear which version of the SAP, data set, and report are linked together even if there are no changes in the SAP
- A copy of a previous version can used to make this clear if needed.
- The SAP for the final analysis should be developed during the progress of the study and must be complete before the final database lock for use in the final statistical analysis.
- If the interim analysis is confidential then the processes outlined in CCTU/GD038 Storage and Access to Confidential Materials must be followed.

7.2. Statistical Analysis Plan Content

- The SAP should be a detailed description of the methods and presentation of data analysis for the study for the main and any interim analyses
- The SAP should provide enough detail for a qualified statistician with no previous experience of the study to perform the required analysis
- Any discrepancies or changes between the analysis plan in the protocol and final SAP should be explained in the SAP
- The SAP should be circulated to other members of the study team for comment (including the CI)
- Any changes to the SAP should be justified and documented in a revision table near the start of the document (see the SAP template)
- Typical contents include:

7.2.1. Authorship

- The statistician, author (if not the study statistician) and Chief Investigator, are required to sign and date the signature box on the front page
- If appropriate an independent reviewer may also sign and date
- All contributors should be listed with details of their contribution even if a signature is not requested. This is appropriate if the SAP is published

7.2.2. Study Background

The background and objectives should be taken directly from the protocol

7.2.3. Populations

The definition of the full analysis population will be documentation if not previously defined in the protocol along with any further populations. This will provide clear guidance on whether to include:

  - Participants who fail to take even one dose of medication
  - Participants with a major ineligibility criterion enrolled in error
  - Participants to exclude in sensitivity analysis due to adverse events, or other reasons, possibly needing expert judgement to judge relatedness

- The process of how participants will be allocated to the populations may be specified in a separate study-specific document refer to CCTU/SOP032 Determination of Analysis Populations.

7.2.4. Flow of Participants

- A flow diagram following the Consort structure
- http://www.consort-statement.org/ and reporting requirements of regulatory agencies includes the numbers of:
    - Screened, eligible and consenting participants
    - Participants assigned
    - Participants receiving the allocated treatment
    - Participants completing the study protocol
    - Participants analysed for the primary outcome
    - A breakdown for each treatment group

A comparison of screening data between the participants included and excluded should be produced if any relevant data is recorded before consent to check how representative the study population is.

7.2.5. Research Hypotheses and Data

- Specification of the primary and secondary hypotheses
- A description of each variable to be considered in the analysis, its definition and role. For example:
    - Descriptive (baseline) variables
    - Primary and secondary outcomes
    - Details of any rules, references, or programs for calculation of derived variables

7.2.6. Endpoints

- Further clarification on the interpretation of observations that are not provided in the protocol may be provided
- Describe the schedule of all the observations or assessments of primary or key secondary endpoints relevant to the statistical analysis
- Include details of time-windows used to group observations, how composite or derived variables (e.g. the mean over weeks 4-8) are calculated
- Any further details required to avoid ambiguity
- If appropriate identify and document in the SAP the short text used to identify specific variables of interest in the database

7.2.7. Baseline Characteristics

- A summary and description of the study population and, if the study is randomised, a check that the baseline characteristics are balanced between treatment arms
- A detailed list of the variables used to assess similarity, and how these will be reported (e.g. means/standard deviations, medians/inter-quartile ranges, proportions)
- If appropriate, a comparison of the characteristics of participants included in and excluded from the study
- Standard errors, confidence intervals and P-values should not be used

7.2.8. Treatment Allocation

- The methods used to allocate treatment are taken from the protocol
- In the case of stratified randomisation or minimisation all the baseline variables used must be listed
- If a comparison between treatments needs to be kept confidential before the final analysis is produced, then do not include details of block size or detailed categories for the stratification factors.
- It is possible to use blinded data, combined with knowledge of which block a participant belongs to, to provide estimates of treatment effects. A reference to a confidential document describing the treatment allocation procedure will suffice

7.2.9. Treatment Received

- A summary of treatments received over the scheduled treatment regimen
- A summary of patterns of missing data and compliance to the protocol

7.2.10. Efficacy

- An evaluation of the efficacy of the treatment tested. Both p-values and confidence intervals should be provided with estimates of effect size for the primary endpoint(s)
- An adjustment for multiple comparisons to be specified if there is more than one primary endpoint or more than one primary comparison
- If appropriate, a secondary analysis for treatment efficacy that is adjusted for predefined highly prognostic baseline characteristics and variables used in stratified randomisation
- An evaluation of efficacy with respect to secondary outcomes
    - Confidence intervals should be provided with estimates of effect size
    - p-values should either be omitted or interpreted with caution as the study may not be appropriately powered
    - Such analyses should be considered as hypothesis generating rather than providing firm conclusions
- A definition of the populations to be used in the final analysis
    - A "Full Analysis" population will probably be specified
    - Other approaches (e.g. per protocol analysis) need to be justified
- A description of the methodology to be used to manage missing data, outliers, non-compliance and withdrawals
    - Characteristics of those included in the main analysis should be compared with those not in the analysis
    - Secondary analysis should include a sensitivity analysis that inputs missing data according to a range of assumptions
    - In the context of non-compliance, generally an Intention-To-Treat (ITT) method will be used to group participants according to the treatment arm they were allocated to, rather than according to what treatment(s) were actually applied. Other methods may also be applied (e.g. CACE) if assumptions and interpretation are made clear
- Any planned subgroup analyses as specified in the protocol

- The addition of further subgroup analyses not specified in the protocol requires further justification
- Regression models with interaction terms are typically used for analysis
- Details of software to be used for the statistical analysis
- Details of any further analysis, with the assumptions they may require e.g. CACE analysis, sensitivity analysis to missing data assumptions
- Alternative methods to use if the assumptions made in the primary analysis do not hold, and the method used to check these assumptions

7.2.11. Interim Analysis

- A statement of when the interim analysis will be performed. This may be outlined in the study protocol and made more explicit in a Data Monitoring Committee Charter document or similar
- Details of which data will be summarised and analysed in the interim analysis (usually primary outcome data, safety parameters and study recruitment) and the methods used
- The SAP template gives further guidance as to what content may be expected or omitted from an interim analysis compared to a final analysis. This can vary from study to study

7.2.12. Safety Analysis

- Describe which coding system will be used
- Describe how repeated events within the same participant will be reported
- When giving rates, clarify which denominator will be used, typically a group of participants
- Check that the format of the reporting is consistent with the reporting requirements in place (EudraCT).

7.2.13. References

- Include a section of references which can include:
    - Non-standard statistical methods
    - Protocol
    - Data Management plan
    - Trial Master File

7.3. Dataset Issues

For each analysis:

- The SAP must be finalised prior to data lock or snapshot for interim analyses
- Prior to data cleaning, data locking and unblinding processes must be followed and documented
- The study database should be frozen by the data manager or the person responsible for the data. Freezing means that a complete audit trail is in place to link the data back to source, and provide proof that no editing has occurred outside of the correct processes for data querying

- The statistician/designee should work on a copy of the study database that has been downloaded by the data manager. The copy should not require manual intervention to obtain, but rather be an automated process where statistical code takes an internal temporary copy of the frozen data each time it is run
- The statistician should refer data queries to the data manager to investigate
    - State the study, site, participant, visit/time point/etc. (whichever is applicable), CRF section, question, etc.
    - State succinctly what the issue is, that is: discrepancy, non-clarity, inconsistency, 'missingness', range issues, etc.
    - State the query. The question should NOT be leading or prescriptive
- Any amendments to the study database should be documented and a new version supplied to the statistician
- The software used for analysis will depend on the preference of the statistician or person performing the statistical analysis, the analyses should be based only on software that has been widely used. The version of the software should be captured in the SAP and in the output of the software
- The SAP may be an appropriate document to record plans to share data or analysis code at some point in the future

7.4. Coding

- To provide high quality code the following points should be followed
- Any outputs should have;
    - The date and time included
    - The name of the code file that produced the analysis
    - The author
    - A log file that captures versions of the software and any external add-on code
- At the start of any code file there should be a set of comments that give
    - the author
    - the date and time of writing
    - references to inputs and outputs
    - reference to any parent code file that runs the child code file
    - the architecture, or overview of how a set of code files are organised

Avoid hard coding. Produce a table-of-tables document as part of or append to the SAP that describes the tables and figures to be produced. This will evolve and become explicit input into the code to provide meta-data such as table titles, table numbers, footnotes etc. Produce dummy tables if the explicit layout of table is non-standard

7.5. Quality Assurance

- A check list of statistical documents required for a study should be maintained, a quality assurance checklist can be completed by the reviewer to ask specific question relating to the statistical components of the study
- There should be a complete audit trail from source data, through CRF and database to the statistical analysis code and the final report documents

- Running the statistical code and producing the reports must be reproducible
- Prior to finalisation the SAP must be reviewed by a statistician (distinct from the author) and approved by the Chief Investigator

7.6. EudraCT Reporting

- The final study should allow simple extraction of the statistics required according to the requirements of the regulatory agencies (EudraCT at the time of writing). The SAP author to verify the current requirements. This can be very specific for example, regarding how age values are categorised, and how the frequency of adverse events is reported

8. Monitoring Compliance with and the Effectiveness of this Document

a. Process for Monitoring Compliance and Effectiveness

As part of routine monitoring visits, audit and inspection

b. Standards/Key Performance Indicators

This process forms part of a quality management system and is reviewed according to CCTU procedures. Standard Operating Procedures are reviewed every two years.

9. References

The Institute of Clinical Research, Abbreviations used in Clinical Trials.

MHRA, Good Clinical Practice "Grey Guide"

International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use, 1998, Statistical Principles for Clinical Trials.

10. Associated Documents

CCTU/SOP032 Determination of Analysis Populations

11. Equality and Diversity Statement

This document complies with the Cambridge University Hospitals NHS Foundation Trust service equality and diversity statement.

12. Disclaimer

It is the user's responsibility to check against the electronic library that this printed out copy is the most recent issue of this document.

Review date	2 years (or earlier in light of new evidence) from approval date
Owning department:	CCTU QA
Supersedes:	CCTU/SOP023 version 4
Local reference:	CCTU/SOP023 version 5