Level M Projects
Writing a Statistical Analysis Plan

1 Introduction

A Statistical Analysis Plan (SAP) describes the planned statistical analysis of a studies objectives. In other words, it describes what variables and outcomes will be collected and which statistical methods will be used to analyse them. In clinical trials a SAP is an essential document because very often the individuals collecting the data will not be those who conduct the analyses. Regulatory bodies all over the world expect that the SAP will meet requirements in pre-specifying inferential analyses and other important statistical techniques. It is expected that the Analysis Plan will provide explicit guidance to be followed by the statistician.

A SAP is used to help with reproducibility of studies results and hence is an important document for all types of statistical analyses, not just clinical trials. An individual not involved in a study should be able to reproduce the results in a study when given the final data, study protocol, documentation of the data collection procedures, and the SAP. Setting up an analysis plan also avoids the temptation of cherry-picking results, with its primary goal being to avoid any investigator (or statistician) bias.

2 Timing of the Statistical Analysis Plan Write-up

The SAP should be written prior to any data analysis being conducted.

The International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) gives the following advice:

"The statistical analysis plan (see Glossary) may be written as a separate document to be completed after finalizing the protocol. In this document, a more technical and detailed elaboration of the principal features stated in the protocol may be included (see section 7.1). The plan may include detailed procedures for executing the statistical analysis of the primary and secondary variables and other data. The plan should be reviewed and possibly updated as a result of the blind review of the data (see 7.1 for definition) and should be finalized before breaking the blind. Formal records should be kept of when the statistical analysis plan was finalized as well as when the blind was subsequently broken."

Before any actual analysis is undertaken, therefore, the SAP should be finalised.

3 Main Elements of SAP

The SAP defines the statistical analysis that will be performed and all of the output required to be included in the final report. It should contain the statistical methods to be used - how subject data will be summarised (descriptive statistics such as mean and standard deviation or counts with percents), the statistical tests and models to be used (Fisher's exact test, two-tailed t-tests, generalised

linear model, for example), and so on. Also to be included are data-handling rules for reporting purposes - when and how impute missing/partial data, what combination of values qualifies a subject for efficacy analysis. Finally, a complete table of contents with all Tables/Lists/Figures (TLFs) to be produced and representations of what each TLF will look like but without the actual numbers.

Detailed below is a list of items generally included in a SAP.

- Definition of the populations
- Research objectives and data
  - Specification of the **primary** and **secondary objectives**.
  - A **description of each variable** to be considered in the analysis, its **definition** and **role** (e.g. descriptive (baseline) variables and primary and secondary outcomes). In particular details of any rules, references or programs for calculation of derived variables.
- Baseline Characteristics
  - A summary and description of the study population.
  - A detailed list of the statistics used to assess similarity, and how these will be reported (e.g. means/standard deviations, medians/inter-quartile ranges, proportions). Standard errors, confidence intervals and P-values should not be used.
- Efficacy
  - A description of the statistical methods that are to be used with a statement of the significance level that will be used.
  - Where necessary, a description of the regression models that are to be fitted with interaction terms that would be considered. Information on assumptions that are made and how they will be checked plus a statement about any distributional assumptions that are made.
  - An adjustment for multiple comparisons if there is more than one primary outcome.
  - If appropriate, details of any secondary analyses,
  - A description of confidence intervals and of the methodology to be applied for hierarchical designs that produce repeated/clustered data. These may include summary statistics methods or the use of multi-level/GEE models.
  - A description of the methodology to be used to handle missing data, outliers, non-compliance and withdrawals. An Intention To Treat (ITT) analysis will probably be specified. Other approaches (e.g. per protocol analysis) would need to be justified. Characteristics of those included in the main analysis should be compared with those not in the analysis. A secondary analysis might include a sensitivity analysis that imputes missing data.
  - Any planned subgroup analyses.
  - Detail of software to be used for statistical analysis.

4 SAP template

An example template for a SAP is detailed below. This is just an example, you may require additional sections depending on your study objectives.

- Population
- Primary Objective
- Secondary Objectives
- Data Collection methods
- Variables under consideration
- Missing data procedures
- Numerical and graphical summaries to be presented
- Models to be fitted

5 SAP examples

Several example SAPs are provided on the Level M projects moodle page.