## Statistical Analysis Plan (SAP) instructions:

Please complete all of the relevant sections on the following pages. Depending on your study type, not all sections may be applicable, so fill in what is relevant and remove any sections that do not apply. Within each section are questions and comments to help guide your responses – delete these and replace them with your answers.

If your study is a clinical trial, refer to the [Australian Clinical Trials Alliance](https://clinicaltrialsalliance.org.au/resource/statistical-analysis-plan/)for additional guidance on what to include in your SAP. For further information on SAPs, see the resources below:

[USyd: Write a research project plan](https://researcher-hub.sydney.edu.au/research-skills/plan-your-research/write-a-research-project-plan.html)

[A template for the authoring of statistical analysis plans](https://www.sciencedirect.com/science/article/pii/S2451865423000467)

[Guidelines for the Content of Statistical Analysis Plans in Clinical Trials](https://jamanetwork.com/journals/jama/fullarticle/2666509)



Statistical Analysis Plan

Study title: Insert study title

**Study protocol number:** If applicable, insert the study protocol number, ethics approval number or trial registration number.

**Investigators:** Insert chief investigator’s name, company, work address, city, state, postcode, country. Insert names of additional investigators.

**Statistical analysis team members:** Insert name of the primary person responsible for the statistical analysis, their company, work address, city, state, postcode, country. Insert the names of additional statistical analysts.

**Date of plan:** Insert date of plan.

**Statistical Analysis Plan version:** Insert version number, e.g. V1.0 and update in the footer below.

# Statistical Analysis Plan: Insert study title

## Background to the study

### Summary

Use plain language to briefly summarise the study including the research question and design of the study (no more than 100 words).

### Study research question

What is the research question(s) you are trying to answer?

### Study objectives or hypotheses

What are the specific objectives and/or hypotheses in this study? You may have multiple primary and secondary objectives. Is the study exploratory (hypothesis-generating) or confirmatory (hypothesis-testing)?

### Study purpose and rationale

What is the background to the study, the knowledge gap it fills and implications of the results?

## Data details

### Study design

What type of study is it? E.g. randomised control trial, crossover trial, observational cohort study etc. Are there any [reporting guidelines](https://www.equator-network.org/) to follow for your chosen study type? Consider specifying the study duration, any follow-up periods and study setting (e.g. hospital, field, community-based). Refer to the [Experimental Design](https://sydney-informatics-hub.github.io/stats-resources/foundational_statistics.html#experimental-design) workshop for more information.

### Study population

What is the inclusion and exclusion criteria of the subjects in the study?

### Sample size justification

What is the chosen sample size to be used (both total size and per group, if applicable)? What is the rationale for the sample size – have you performed a sample size calculation or pilot study? If a sample size calculation was performed, what were the relevant input values and assumptions used (e.g. smallest effect size, variance, alpha, power)? What methodology (e.g. simulations, hypothesis test, precision approach) was used? Did you adjust for dropout rate and/or multiple testing? Sufficient detail should be provided so an independent person could replicate the sample size calculation. Refer to[Power and Sample Size](https://sydney-informatics-hub.github.io/stats-resources/foundational_statistics.html#power-and-sample-size) for more information.

### Intervention or treatment

Is there an intervention or treatment to which the subjects are exposed? If yes, is the study design independent (i.e. different study subjects in each group) or clustered/within-subject (i.e. the same subjects receive multiple treatments or are grouped in clusters).

### Randomisation, stratification and blinding

Depending on the study design, describe the randomisation procedure including any stratification variables, and any details of who has been blinded during the study and at what stages of the study.

### Ethical and data governance

Describe any ethical approvals obtained for the study, including reference numbers and the approving ethics committee. Mention any data access or permissions required, e.g. data sharing agreements and the [security classification of the data](https://sydneyuni.service-now.com/sm?id=kb_article_view&sysparm_article=KB0015117). Describe how subject privacy and data de-identification will be maintained. See the [Research data management](https://intranet.sydney.edu.au/research-support/managing-research/research-data-management.html) website for more details.

### Databases and datasets used

Where will the data be stored? E.g. Microsoft Excel or REDCAp. What datasets will be used? It can be helpful to include a data dictionary here. See template below and refer to the [Research Essentials](https://sydney-informatics-hub.github.io/stats-resources/foundational_statistics.html#research-essentials) workshop for more information on data types and setting up your dataset.

**Table 1.** Data dictionary containing all of the proposed outcome and explanatory variables in the XXXX dataset for the study on XXXX.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Variable name** | **Definition** | **Variable type** | **Categories or units** | **Variable and category coding in dataset** | **Outcome or explanatory variable** |
| Age | Age of participant in years | Discrete, Numerical | Years | age | Explanatory variable |
| Sex | Sex of participant | Binary, categorical | Male, Female | sex; male = 1, female = 2 | Explanatory variable |

### Outcome variables

What is the primary outcome(s)? What is the secondary outcome(s)? List them all below, including the timepoints where the outcome(s) is measured, their unit(s) and variable type(s). It can be helpful to reference the data dictionary above. Mention any calculations or transformations needed to generate the outcome(s), e.g. change from baseline, calculation of a total score. Mention how the outcome(s) will be measured.

### Explanatory/Predictor variables

For each outcome variable, list the relevant explanatory variables, including any intervention/treatment variables, known or potential confounders and control or design variables that you may wish to adjust for. If there are confounders, specify how they will be identified (e.g. through previous literature or Directed Acyclic Graphs (DAGs)). Are there any interaction terms or effect modifiers? It can be helpful to reference the data dictionary.

## Data analysis

### General principles

Include information here about the level of statistical significance (e.g. P < 0.05), any adjustment for multiple comparisons, and the type of confidence intervals to be reported. Mention the statistical analysis software to be used, alongside any relevant packages/modules within the given software, as well as software version numbers. This includes software used for the sample size calculation. Will the code/scripts be made available publicly or be available upon request?

### Data wrangling and cleaning

Mention any data cleaning rules you may use, e.g. categorical variables should have no less than 10 counts per category, as well as how will you handle outliers, missing values and perform any data transformations. Mention information about the creation of new variables, if any.

### Descriptive statistics

You can describe your data through data visualisations or tables. Consider how you will you describe baseline characteristics of participants as well as individual variables, and crude relationships between outcomes and explanatory variables. If there is an intervention/treatment, summary statistics should be provided for each study arm, as well as overall.

E.g. For categorical variables, counts and percentages will be presented in table format. For numerical variables, numerical summary statistics of centre (e.g. mean and median) and spread (maximum, minimum, quartiles, SD) will be presented in table format. For numerical outcomes split by categorical explanatory variables, grouped boxplots will be used. Refer to the [Research Essentials](https://sydney-informatics-hub.github.io/stats-resources/foundational_statistics.html#research-essentials) workshop for more information on exploratory data analysis.

### Inferential analysis

You may choose to split this section up into the following headings: 1. Univariable analysis, 2. Multivariable analysis, 3. Covariate or subgroup analysis, and 4. Sensitivity analysis. You may also wish to have separate sections for each of your primary and secondary outcome variables.

#### Univariable analysis

#### Multivariable analysis

#### Covariate or subgroup analyses

#### Sensitivity analysis

This section will describe the inferential statistical analyses you will use, including any statistical modelling and any sensitivity analyses to confirm the robustness of your final models. Mention the statistical tests used within each of the above headings, how you will check and handle statistical assumptions, your model building strategy, any covariate adjustment and testing of interactions. In the case of statistical assumptions being violated, what alternate approaches will you use? Also mention how you will handle any correlated/clustered or longitudinal repeated measures data and how you will treat missing data (e.g. imputation methods, or complete-case analysis). For more information on statistical modelling, see our [Statistical Modelling](https://sydney-informatics-hub.github.io/stats-resources/statistical_modelling.html) workshop series.

## Data presentation

### Analysis dissemination strategy

With whom will the findings be shared? E.g. Clinical teams, study investigators, trial participants, funders, policy makers, the broader research community or lay people. How will you disseminate the findings and in which format? E.g. Do you have a target journal article or conference in mind?

### Dummy tables

Consider including dummy tables that will be used to present your outcomes, subgroup analyses and/or sensitivity analyses, with placeholder headers and footnotes. List all potential table captions here.

### Dummy figures

Consider including a mock data visualisation that will be used to present your outcomes, subgroup analyses and/or sensitivity analyses. List all potential figure captions here.

## References

Include any relevant references that you cite in the above statistical analysis plan. Follow a specific citation style of your choice. Include references for statistical methods, reporting guidelines and any software packages you use.