Statistical Analysis Plan (SAP)

Please complete all relevant sections. Instructions are in red and should be deleted when completing the SAP. The purpose of this template is to provide a general layout, but sections can be reformatted as you wish.

The creation of this template for the Duke BERD Methods Core was made possible by Grant Number UL1TR002553 from the National Center for Advancing Translational Sciences (NCATS) of the National Institutes of Health (NIH), and NIH Roadmap for Medical Research.

Title

CRU/Department/Division/Center

IRB Number

Investigators:

Primary Investigator

Collaborative Lead

Co-authors (if known)

Analysis Biostatistician(s)

Biostatistics Supervisor

Lead Biostatistician

Subject Matter Expert

Original Creation Date

Version Date

Project Folder Location

Project Goal(s) *Manuscript, abstract, presentation, etc.*

Submission Deadline(s) *Deadline(s) of above goal(s)*

Effort Estimate (optional)

Investigator Agreement

☐ All statistical analyses included in an abstract or manuscript should reflect the work of the biostatistician(s) listed on this SAP. No changes or additional analyses should be made to the results or findings without discussing with the project biostatistician(s).

☐ All biostatisticians on this SAP should be given sufficient time to review the full presentation, abstract, manuscript, or grant and be included as co-authors on any abstract or manuscript resulting from the analyses.

☐ If substantial additional analysis is necessary or the aims of the project change, a new SAP will need to be developed.

☐ Publications resulting from this SAP are supported in part by the Duke CTSA and must cite grant number UL1TR002553 and be submitted to PubMed Central.

☐I have reviewed the SAP and understand that any changes must be documented.

	Acknowledged by: Click or tap here to enter text. *Date:* Click or tap to enter a date.
Activity Log	*This section should be used to track any amendments or changes made to the SAP. It should include detailed information about the changes and the date.* *This is where you should put any information about whether this is an entirely new SAP and an addendum was not appropriate. Here you should state why you had to write an entirely new SAP. Otherwise, include the changes as a 'change log' in the addendum section below. State reason and details if this is a new SAP for a project that has undergone changes (i.e. dataset changed because the original one was not appropriate and a new SAP was created).* *OPTIONAL: This section can also be used to track your project milestones/meetings/status if you find it helpful.*

Acronyms	*XXX*	*Definition here*
	YYY	*Definition here*
	XXX	*Definition here*

1 Study Overview

Background/Introduction: *This section should be written by the author of the SAP so other biostatisticians can easily follow and to ensure everyone understands the background and goals of the project. If pieces are copied from the protocol or other documents, cite the document. References are not necessary here.*

1.1 Study Aims

[insert aims (this is in scientific terms)]

1.2 Study Hypotheses

[insert hypotheses (this is in statistical terms)]
If the analysis is descriptive, simply say so and skip the rest of this section.

1.2.1 Primary Hypotheses

[insert]
You should only have 1-2 primary hypotheses and the rest should be secondary.

1.2.2 Secondary Hypotheses

[insert]

2 Study Population

2.1 Inclusion Criteria

- [insert]

2.2 Exclusion Criteria

- [insert]

2.3 Data Acquisition

Fill in all relevant information:

Study design	
Data source/how the data were collected	
Contact information for team member responsible for data collection/acquisition	
Date or version (if downloaded, provide date)	
Data transfer method and date	
Where dataset is stored	

Notes: *Include any additional details that give information about how this data set was acquired.*

Description:
[insert]

3 Outcomes, Exposures, and Additional Variables of Interest

A. *If you choose not to use the table format below, make sure all the information outlined below is still included.*

3.1 Primary Outcome(s)

You should only have 1-2 primary outcomes and the rest should be secondary.

Outcome	**Description**	**Variables and Source**	**Specifications**
			e.g., how is it coded?

3.2 Secondary Outcome(s)

Outcome	**Description**	**Variables and Source**	**Specifications**
			e.g., how is it coded?

3.3 Additional Variables of Interest

This is optional but generally this would be all your covariates of interest. If there are any variables that need special calculations etc. include them here or in a data dictionary that is an appendix to this SAP which you can reference here.

Variable	**Description**	**Variables and Source**	**Specifications**
			e.g., how is it coded?

4 Statistical Analysis Plan

[insert]

Every analysis step should refer back to an aim/hypothesis/objective. You can include just one deadline for the project, but if there are multiple major stopping points, you should include individual deadlines for each milestone. Good practice is to number each item here and reference them in your code. If you have to wait for data, you should include a disclaimer that X analysis cannot be completed until the data is collected and deadlines cannot be determined until you receive the data.

For more details on what to include, see the SAP Checklist document.

4.1 Demographic and Clinical Characteristics ("Table 1")

[insert]

4.2 Analyses Plan for Aim 1

[insert]

4.3 Analyses Plan for Aim 2

[insert]

ETC

5 Limitations

All design and analysis limitations (this will grow as you do the study and should be included in the report)

6 Addendum for Additional Analyses

[insert if applicable]

All post-hoc analyses should be described here. If minor changes are made to the main analysis (e.g. adding a covariate to the model), this can be changed in the main analysis section above and a note should be added to the activity log.

7 Appendix

This section can be used to include table shells, example figures, or anything else that does not belong in the body of the SAP but you feel should be included.

8 References

If needed, this section can included citations to statistical methods, programming software (including R packages), and/or relevant clinical literature.

Statistical Analysis Plan Checklist

Below you will find a checklist of recommended items to include in a statistical analysis plan. Some of these are specific to clinical trials (based on this JAMA paper) and some are other are specific to observational studies (based on STROBE/RECORD guidelines), so every item will not be necessary for every project. The biostatistician should start with the SAP template above and add in necessary information from the checklist. Item numbers that are starred (*) are not explicitly included in the SAP template and should be added by the author if relevant to the project. This checklist was developed using the CONSORT 2010 Checklist.

Section/Topic	Item #	Description	Included (Yes/No/NA)
Administrative Information			
Study Information	1a	Descriptive title that matches the protocol, with SAP either as a forerunner or subtitle	
	1b	Trial registration number, protocol version number, and/or IRB number.	
	1c	CRU/Department/Division/Center/other collaborative unit that the study falls under	
Roles and responsibility	2a	Listing of principal investigators, clinical leads, and co-authors (if known)	
	2b	Name and affiliation of SAP author(s)	
	2c	Names, affiliations, and roles of other SAP contributors (e.g. senior statistician)	
SAP Information	3	SAP version number, with date of current version and original creation date	
Project Information	4a	Project folder location	
	4b	Project goals (e.g. manuscript, abstract, presentation, etc.)	
	4c	Project deadlines (of listed goals)	
	4d	Effort estimate	
Investigator Agreement			
Investigator Agreement	5	Confirmation that BERD Method Core's collaborative process has been reviewed, that all statistical analyses included in an abstract or manuscript should reflect the SAP, no changes should be made to the SAP without discussing with the SAP author, all biostatisticians on the SAP are co-authors on the manuscript, and that publications resulting from the SAP must cite grant number UL1TR002553 and be submitted to PubMed Central	
Signatures	6	Signatures of SAP author, senior statistician, and principal investigator(s)	
Activity Log			
SAP revisions	7a	SAP revision history with dates	
	7b	Justification for each SAP revision	
	7c*	Timing of SAP revision in relation to any interim analyses or submissions	

[1]Turner L, Shamseer L, Altman DG, et al. Consolidated standards of reporting trials (CONSORT) and the completeness of reporting of randomised controlled trials (RCTs) published in medical journals. *Cochrane Database Syst Rev*. 2012;11:MR000030. Published 2012 Nov 14. doi:10.1002/14651858.MR000030.pub2

Study Overview			
Background and introduction	8	Synopsis of scientific background and rationale for the study	
Aims and Hypotheses	9a	List of all scientific aims/objectives of the study, with specifications of primary, secondary, etc.	
	9b	List of all statistical hypotheses (corresponding to the scientific aims), with specifications of primary, secondary, etc.	
Variables of Interest	10a	List of all outcome/endpoint variables, with a description of their coding/units, timing, and source, corresponding to the statistical hypotheses. If any variables are defined using ICD or CPT codes, list them out.	
	10b	List of all exposure variables, with a description of their coding/units, timing, and source, corresponding to the statistical hypotheses. If any variables are defined using ICD or CPT codes, list them out.	
	10c	List of any additional variables of interest (e.g. covariates, potential confounders, effect modifiers, etc.) in the analysis	
	10d*	Location of data dictionary (or provided as an appendix)	
	10e*	Report category boundaries if continuous variables are collapsed into categories, and describe any other relevant data transformations	
Causal Graph	11*	May be helpful to include a DAG or other graph/diagram that describes the way the variables of interest are presumed to relate to each other	
Study Methods			
Study Plan and Design	12a	Description of the study design (e.g. parallel group randomized trial, case-control study, cohort study, etc.)	
	12b*	Study setting, location, and relevant dates (e.g. periods of enrolment, exposure, follow-up, and collection)	
	12c*	Description of intervention or exposure groups, with allocation ratios, and details of any matching criteria	
	12d*	Details on randomization (e.g. stratification factors) and blinding procedures	
	12e	List of eligibility and/or inclusion/exclusion criteria	
	12f*	Description of screening/enrolment/recruitment processes	
	12g*	Description of patient flow (e.g. CONSORT diagram)	
	12h*	Description of analysis population (e.g. intention to treat, per protocol, etc.)	
	12i*	Definitions of adherence/compliance, protocol deviations, loss-to-follow-up, adverse events, etc.	
	12j*	Time points at which outcomes are measured	

	12k*	Timing of final analyses (are all outcomes analysed collectively, or will short-term outcomes be analysed separately from long-term outcomes, etc.)	
Sample Size	13a*	Sample size calculation or justification (either provided in full or summarized, with link to original source)	
	13b*	Description of pre-planned subgroup analyses, power for these analyses, and planned multiple comparison adjustment procedures	
Interim Analyses	14a*	Description of what interim analyses will be conducted at which time points, and what methods used to adjust significance levels due to the interim analysis	
	14b*	Details of any guidelines (e.g. safety, futility) for stopping the study early	
	14c*	Details of any changes to trial design due to interim analyses (e.g. enrolling more patients)	
Data	15a	Description of data collection/acquisition process, with contact information for team member responsible	
	15b	Description of data flow/transfer from primary data collection through to creation of final analysis dataset	
	15c	Data transfer method and date	
	15d	Folder location where datasets are stored	
	15e*	Description of any additional data management, quality control, or processing undertaken	
	15f*	If any data are extracted from a database, a description of the database and the query used for the extraction, and whether/how it was merged with any data from outside that database. If the study involved linkage of databases, consider use of a flow diagram to demonstrate the data linkage process, including the number of individuals with linked data at each stage.	
	15f*	Description of any other data sources incorporated in the analysis	
Missing Data	16a*	Description of sources and magnitudes of missing data	
	16b*	Description of how missing data patterns will be presented/summarized (may be helpful to have a table shell or draft CONSORT-style diagram)	
	16c*	Description of contingency plans for handling missing data in analysis	
Simulations	17a*	If conducting a simulation, a description of the purpose of the simulation and its design (e.g. fully factorial, partially factorial, grid search, etc.)	
	17b*	Define the fixed and variable factors or parameters in the simulation, the estimands/targets of the simulation, and the performance measures to be estimated (with justifications of their relevance to the estimands/targets)	
	17c*	Description of the tabular and graphical presentations of simulation results and their interpretation	

Statistical Analysis Plan

Statistical Significance	18a*	Hypothesis testing framework (e.g. superiority, equivalence, non-inferiority), or description of alternative analytic framework (e.g. evaluation of a posterior in a Bayesian analysis, etc.)	
	18b*	Level of significance for primary hypotheses, including a description and rationale for any multiple comparisons adjustment or Type I error control procedures	
	18c*	Description of any decision-making rules based on confidence intervals, credible intervals, prediction intervals, Bayes' factors, or other alternative inferential methods	
	18d*	Description of how the results of any hypothesis tests (or alternative inferential methods) will be interpreted with respect to both the statistical hypotheses and scientific aims/objectives of the study	
Descriptive Statistics	19a*	List of characteristics (e.g. demographic, clinical) to be summarized descriptively (e.g. "Table 1")	
	19b*	Description of how these characteristics will be summarized descriptively (e.g. means/medians vs. N (%), tabular displays, graphical displays, etc.)	
	19c*	Summarize follow-up time (e.g. average and total amount) and number of events	
Analysis Methods	20a	For each aim/hypothesis (see items 9a/9b), a description of what analysis method will be used and how the results from this method will be reported and interpreted	
	20b*	Description of any transformations, standardizations, covariate or confounder adjustments, weighting, or stratification methods to be used and why.	
	20c*	For each analytic method proposed, a description of the assumptions of that method and what processes will be used to evaluate whether or not those assumptions hold	
	20d*	Details of contingency plans/alternative methods to be used if the assumptions are found not to hold	
	20e*	In the case of non-standard test statistics, formulas provided for the test statistic with a description of the mathematical null hypothesis, how significance is determined, and how the test statistic is interpreted	
	20f*	In the case of regression models, formulas provided for the full model with a description of which parameters are to be used, how they will be interpreted, how confidence intervals will be constructed, etc.	
	20g*	In the case of survey, hierarchical/nested, or clustered data, a description of what methods will be used to adjust for the data structure and why (e.g. if using a GEE, describing which correlation structure and why it was chosen, etc.)	
	20h*	For non-continuous outcomes, clearly explain the effect used (e.g. risk difference, risk ratio, odds ratio, etc.), whether it is relative or absolute, and justify why that was chosen as the effect measure of interest	
	20i*	Documentation of any non-standard methods used (e.g. using alternative degree of freedom calculation methods, using a non-canonical link function, etc.)	

	20j*	Description of any limitations, sources of bias, internal/external validity, and other relevant discussions concerning the interpretation and generalizability of the design or methods used	
Additional Analysis Methods	21a*	Description of any pre-planned sensitivity analyses and how they will be interpreted	
	21b*	Description of pre-planned subgroup analyses, power for these analyses, and planned multiple comparison adjustment procedures	
	21c*	Description of any additional post-hoc calculations or analyses (e.g. evaluating interaction/modification effects, calculating mediation or local average treatment effects, evaluation of AUROC curves, etc.)	
	21d*	If conducting any bootstrap analyses, a description of the sampling algorithm and number of iterations used	
	21e*	If conducting any cross-validation procedures, a description of how the cross-validation is conducted (e.g. leave-one-out, train/validation/test, etc.)	
Exploratory Analyses	22a*	Description and justification for any pre-planned exploratory analyses and what methods will be used to conduct them	
	22b*	Framework for conducting any unplanned exploratory analyses and how they will be integrated into the planned analysis	
Software	23*	List of statistical software (along with version numbers) to be used for each phase of the analysis; in the case of R or Stata, additionally list any requisite installed packages and their version numbers	
Other	24*	Description of any additional planned analyses of the data (e.g. a safety analysis looking at adverse event rates for a Data Safety Monitoring Board, etc.)	
Tables and Figures			
Table Shells	25*	Example tables related to any of the conducted analyses; if possible including any available preliminary data	
Example Figures	26*	Example figures related to any of the conducted analyses; if possible including any available preliminary data.	
References			
References	27a	References for any non-standard statistical methods used	
	27b	References (and locations) for any relevant protocols, standard operating procedures, or other documents cited in the SAP	
Additional Information			
Appendices	28*	If necessary, appendices may be included (e.g. a full data dictionary, a copy of a Case Report Form, etc.)	
Addendums	29*	Any additional analyses conducted that were not included in the SAP should be documented in an addendum, describing the purpose of the additional analysis, when it was conducted, and by whom	