Writing SAPs

Yue Jiang

Duke University

A disclaimer

The following material was used during a live lecture. Without the accompanying oral comments and discussion, the text is incomplete as a record of the presentation. A full recording may be found via Zoom on the course Sakai site.

What is an SAP?

A **statistical analysis plan** is a tool often used in clinical trials that establishes the research aims, study design and variables, statistical methods and models, and rationale for choosing such methods. It often includes analyses regarding sample size and power, as well as logistical details regarding any randomization procedure, data entry, quality assurance, and database management.

Importantly, SAPs are written *prior* to initiating of the clinical trial in question.

Why is having an SAP important?

What is an SAP?

> We will use t-tests and chi-square tests to assess continuous and categorical variables, respectively.

is basically saying

> We will use beakers to hold solutions and burettes to perform titration.

SAPs should provide *strategies* and *rationales* instead of simply listing procedures. Explain *why* each statistical tool is being used and how it specifically addresses one of the research aims (don't be afraid to restate them!).

Discuss rationale, appropriateness, advantages, and limitations, including mention of competing reasonable methods that were not selected (and why).

What is an SAP?

Strong SAPs will discuss main analyses (used to drive storyline of paper), but also include auxiliary analyses to support the main results:

- Sensitivity analyses help evaluate robustness of the main results to assumptions or reasonable differences in choice of methods, inclusion of outliers, etc.
- Goodness of fit analyses help place main results in context of the overall data.

Be *comprehensive* in anticipating data issues. Mention strategies for dealing with incomplete data or missing data, questionable data values, violation of distributional or modeling assumptions, and multiple comparisons.

What is an SAP?

Provide detailed "template" mock-ups of all figures and tables to be included in the main manuscript. You may consider including a detailed **data dictionary** providing variables you intend to collect, the timing of variable collection (if multiple visits are specified), and the expected format and units of such variables.

You may also refer to existing guidance for certain types of trials:

- CONSORT: Consolidated Standards of Reporting Trials
- STROBE: Strengthening the Reporting of Observational Studies in Epidemiology
- TREND: Transparent Reporting of Evaluations wiht Nonrandomized Designs

What to include in an SAP?

JAMA | Special Communication

Guidelines for the Content of Statistical Analysis Plans in Clinical Trials

Carrol Gamble, PhD; Ashma Krishan, BSc; Deborah Stocken, PhD; Steff Lewis, PhD; Edmund Juszczak, MSc; Caroline Doré, BSc; Paula R. Williamson, PhD; Douglas G. Altman, DSc; Alan Montgomery, PhD; Pilar Lim, PhD; Jesse Berlin, ScD; Stephen Senn, PhD; Simon Day, PhD; Yolanda Barbachano, PhD; Elizabeth Loder, MD, MPH

What to include in an SAP?

Table. SAP Guidance Document: Recommended Items to Address in a Clinical Trial SAP[a]

Section/Item	Index	Description
Section 1: Administrative Information		
Title and trial registration	1a	Descriptive title that matches the protocol, with SAP either as a forerunner or subtitle, and trial acronym (if applicable)
	1b	Trial registration number
SAP version	2	SAP version number with dates
Protocol version	3	Reference to version of protocol being used
SAP revisions	4a	SAP revision history
	4b	Justification for each SAP revision
	4c	Timing of SAP revisions in relation to interim analyses, etc
Roles and responsibility	5	Names, affiliations, and roles of SAP contributors
Signatures of:	6a	Person writing the SAP
	6b	Senior statistician responsible
	6c	Chief investigator/clinical lead

What to include in an SAP?

Section 2: Introduction		
Background and rationale	7	Synopsis of trial background and rationale including a brief description of research question and brief justification for undertaking the trial
Objectives	8	Description of specific objectives or hypotheses

What to include in an SAP?

Section 3: Study Methods		
Trial design	9	Brief description of trial design including type of trial (eg, parallel group, multiarm, crossover, factorial) and allocation ratio and may include brief description of interventions
Randomization	10	Randomization details, eg, whether any minimization or stratification occurred (including stratifying factors used or the location of that information if it is not held within the SAP)
Sample size	11	Full sample size calculation or reference to sample size calculation in protocol (instead of replication in SAP)
Framework	12	Superiority, equivalence, or noninferiority hypothesis testing framework, including which comparisons will be presented on this basis
Statistical interim analyses and stopping guidance	13a	Information on interim analyses specifying what interim analyses will be carried out and listing of time points
	13b	Any planned adjustment of the significance level due to interim analysis
	13c	Details of guidelines for stopping the trial early
Timing of final analysis	14	Timing of final analysis, eg, all outcomes analyzed collectively or timing stratified by planned length of follow-up
Timing of outcome assessments	15	Time points at which the outcomes are measured including visit "windows"

What to include in an SAP?

Section 4: Statistical Principles		
Confidence intervals and *P* values	16	Level of statistical significance
	17	Description and rationale for any adjustment for multiplicity and, if so, detailing how the type 1 error is to be controlled
	18	Confidence intervals to be reported
Adherence and protocol deviations	19a	Definition of adherence to the intervention and how this is assessed including extent of exposure
	19b	Description of how adherence to the intervention will be presented
	19c	Definition of protocol deviations for the trial
	19d	Description of which protocol deviations will be summarized
Analysis populations	20	Definition of analysis populations, eg, intention to treat, per protocol, complete case, safety

What to include in an SAP?

Section 5: Trial Population		
Screening data	21	Reporting of screening data (if collected) to describe representativeness of trial sample
Eligibility	22	Summary of eligibility criteria
Recruitment	23	Information to be included in the CONSORT flow diagram
Withdrawal/follow-up	24a	Level of withdrawal, eg, from intervention and/or from follow-up
	24b	Timing of withdrawal/lost to follow-up data
	24c	Reasons and details of how withdrawal/lost to follow-up data will be presented
Baseline patient characteristics	25a	List of baseline characteristics to be summarized
	25b	Details of how baseline characteristics will be descriptively summarized

What to include in an SAP?

Section/Item	Index	Description
Section 6: Analysis		
Outcome definitions		List and describe each primary and secondary outcome including details of:
	26a	specification of outcomes and timings. If applicable include the order of importance of primary or key secondary end points (eg, order in which they will be tested)
	26b	specific measurement and units (eg, glucose control, hbA_{1c} [mmol/mol or %])
	26c	any calculation or transformation used to derive the outcome (eg, change from baseline, QoL score, time to event, logarithm, etc)
Analysis methods	27a	what analysis method will be used and how the treatment effects will be presented
	27b	any adjustment for covariates
	27c	methods used for assumptions to be checked for statistical methods
	27d	details of alternative methods to be used if distributional assumptions do not hold, eg, normality, proportional hazards, etc
	27e	any planned sensitivity analyses for each outcome where applicable
	27f	any planned subgroup analyses for each outcome including how subgroups are defined
Missing data	28	Reporting and assumptions/statistical methods to handle missing data (eg, multiple imputation)
Additional analyses	29	Details of any additional statistical analyses required, eg, complier-average causal effect[10] analysis
Harms	30	Sufficient detail on summarizing safety data, eg, information on severity, expectedness, and causality; details of how adverse events are coded or categorized; how adverse event data will be analyzed, ie, grade 3/4 only, incidence case analysis, intervention emergent analysis
Statistical software	31	Details of statistical packages to be used to carry out analyses
References	32a	References to be provided for nonstandard statistical methods
	32b	Reference to Data Management Plan
	32c	Reference to the Trial Master File and Statistical Master File
	32d	Reference to other standard operating procedures or documents to be adhered to

Power/sample size analysis

What is the definition of statistical power and why should we care?

Power is the probability of rejecting the null hypothesis when it is false: P(reject H_0 | H_0 is false) - often calculated for *specific* alternatives.

Check out an interactive visualization of some factors that are related to power.

Power/sample size analysis

Why care about choosing a sample size/power?

- To show that under certain required conditions, a hypothesis test has a good chance of showing the anticipated difference, if it really exists
- To be more confident that a null result is not simply a sample of excessive variability
- To show a funding agency that the study has a reasonable chance of reaching a useful result
- To show that necessary resources (human, animal, financial, time, etc.) will be minimized

Note that for multiple specific hypotheses of interest, each with their own tests and estimates of interest, you may come to different conclusions when evaluating each!

Calculating power

Suppose $X \sim N(0, 3)$ and $Y \sim N(\_\_, \_\_)$, and that you are interested in testing

$$H_0 : \mu_X = \mu_Y$$
$$H_1 : \mu_X \neq \mu_Y$$

What is the anticipated power at $\alpha = 0.05$ if you have 20 subjects from population X and 20 from Y if $Y \sim N(1, 3)$? How about $Y \sim N(5, 12)$? What if we had 40 subjects from Y or if we specified $\alpha = 0.01$ instead?

- Built-in functions in software
- Formulas do exist for certain types of (simple) analyses!
- In real-life, **simulation** is often used to simulate power across a wide range of potential alternatives and across a wide range of potential data patterns (demonstration given in class)

Power/sample size analysis

However: there is *no* place for power when analyzing results -- it is irrelevant for doing so!

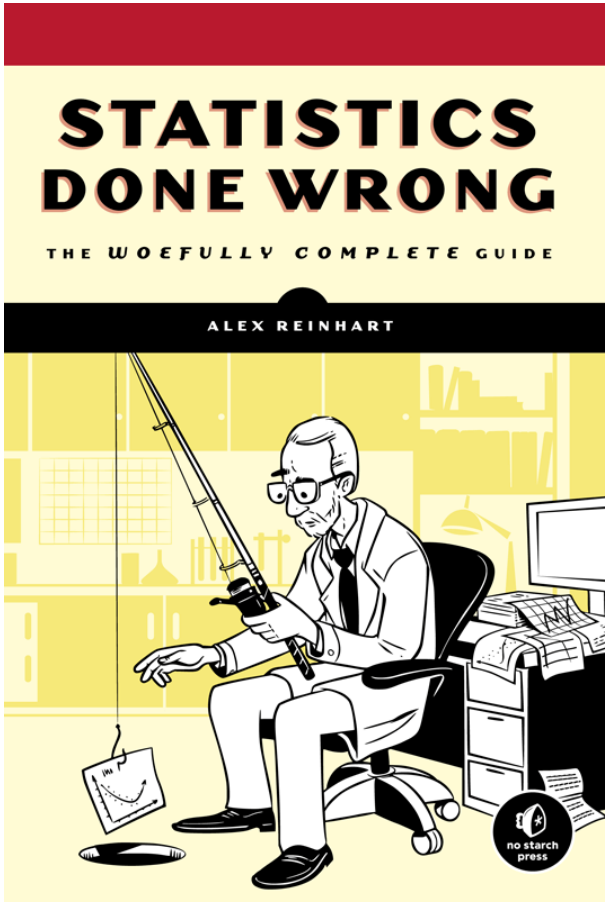
- Power is only useful in the planning stages; observed confidence intervals and point estimates are all that's needed for (frequentist) analysis.
- No additional information can be obtained by performing any kind of power calculation.
- After data are collected, these are just previous conjectures about expected behavior - they provide no assistance in interpreting the study's data.
- Inclusion of pre-study power analyses may lead to misinterpretations regarding study results.

Post hoc power analyses are similarly worthless (similarly for sample size analyses).

Power/sample size analysis

Does this in fact mean that power/sample size calculations are not important?

A cautionary tale



Example adapted from Reinhart, 2015: Statistics Done Wrong

A cautionary tale

Suppose 100 independent researchers conducted 100 pilot studies in which the true magnitude of the effect was small. Because these were pilot studies, each investigator is only able to study a small number of patients.

Because of the small sample size, estimators have high variability -- there is low **precision** in these estimates. Each investigator plans to test the null hypothesis of zero effect (even though we can probably suspect that any tests will have low power).

A cautionary tale

Suppose the true unknown power of each test is 7%. We would expect 7 of the investigators to obtain a statistically significant p-value and conclude that the effect is not-zero.

These 7 investigators have made the *correct* decision - they have not made false discoveries! Remember, the effect truly exists.

What are some potential consequences of this sequence of events?

A cautionary tale

Because of the small sample size, the estimate of the *magnitude* of any effect will be gigantic if the p-value is statistically significant (given the high variability, how else would we have rejected the null hypothesis with such few patients?).

The p-value will be significant here only when the estimate happens to be large.

Under this understanding, what are some other potential consequences? What is a simple way we can mitigate some of these consequences?