SAP and haven package

Subtitle

Haoran Teng
MCW

June 27, 2022

Overview

1. SAP
2. haven

SAP

- Biomedical research has been struck with the problem of study findings that are not reproducible
- with the advent of large databases and the development of powerful statistical software to perform data analyses, it has become easier to find associations and form conclusions from data without following the scientific method of forming an a-priori hypothesis.
- This approach may yield statistically significant associations without clinical relevance, false positive findings, or biased results due to "fishing" for the desired results.

SAP

- To improve reproducibility, transparency, and validity among clinical trials, the National Institute of Health (NIH) updated its grant application requirements for clinical trials, which among other items, mandates trial registration, posting of results within ClinicalTrials.gov, and submission of the original statistical analysis plan(SAP) along with the research protocol.
- describes what variables and outcomes will be collected and which statistical methods will be used to analyze them.
- According to Dr F. Collins, director of the NIH, the goal is to dissuade "P-value hacking, where people sort of shop around for a statistical test to give them the P value that they love".

SAP

- there is an argument to also include a SAP for observational studies (hypothesis-driven and/or data-driven), as findings from many observational studies have failed to be replicated
- Anesthesia and Analgesia require indication of "whether the primary outcome(s) were defined and established a-priori at initiation of the study design or were created post hoc during data exploration ("data mining") and accompanying statistical analysis. . . "

SAP

TABLE 1 Components of a statistical analysis plan (SAP) for randomized control trials and observational studies[12]

		Study type		
	SAP items	**Randomized control**	**Prospective observational**	**Retrospective observational**
1	Study title that matches the study protocol	X	X	X
2	Trial registration number	X		
3	SAP version number with dates	X	X	X
4	Version of protocol referenced	X	X	X
5	SAP revision history	X	X	X
6	Reason for each SAP revision	X	X	X
7	Timing of SAP revisions relative to interim analyses.	X	X	
8	SAP contributors with roles and responsibilities	X	X	X
9	Person writing SAP	X	X	X
10	Senior statistician	X	X	X
11	Principal investigator	X	X	X
12	Background and rationale of study	X	X	X
13	Objectives and hypotheses	X	X	X
14	Study type	X	X	X
15	Randomization details	X		
16	Sample size calculation, if applicable	X	X	X
17	Superiority, equivalence, or noninferiority hypothesis testing framework.	X		
18	Interim analysis, timing of analysis, and person performing interim analysis, if applicable	X	X	
19	Adjustment of the significance level due to interim analysis	X	X	
20	Guidelines for stopping study early	X	X	

SAP

21	Timing of final analysis	X	X	X
22	Timing and time interval for assessing each outcome; visit windows	X	X	
23	Level of statistical significance (P values) and whether one- or two-sided	X	X	X
24	Plan and rationale for adjustment for multiplicity, if applicable, including how type 1 error will be controlled	X	X	X
25	Confidence intervals to be reported and whether one- or two-sided	X	X	X
26	Definition of intervention adherence and how it will be presented	X		
27	Definition and summary of protocol deviations	X	X	
28	Definition of population being analyzed	X	X	X
29	Reporting of screening data to describe representation of study population, if applicable	X	X	X
30	Inclusion and exclusion criteria	X	X	X
31	Recruitment strategy	X	X	
32	Level and timing of withdrawal	X	X	
33	Presentation of withdrawal and follow-up data	X	X	
34	Baseline patient characteristics and how they will be descriptively summarized	X	X	X
35	Definitions of outcomes and sequence of measurement	X	X	X
36	Specific measurements and units	X	X	X
37	Calculations or transformations used to derive outcome	X	X	X
38	Analysis method used	X	X	X
39	Presentation of treatment effects	X		
40	Covariates and adjustments	X	X	X
41	Methods to check for distributional assumptions	X	X	X
42	Alternative methods if distributional assumptions are false	X	X	X

SAP

43	Sensitivity analysis for each outcome if applicable	X	X	X
44	Subgroup definition and analysis, if applicable	X	X	X
45	Method for handling missing data	X	X	X
46	Additional statistical analysis, if applicable	X	X	X
47	Details on summarizing safety data	X	X	
48	Statistical packages used for analysis	X	X	X
49	Reference to standard operating procedure or additional documents	X	X	X

Items 35-37 apply to each primary and secondary outcome.

Components of a statistical analysis plan (SAP) for randomized control trials and observational studies[yuan, 2019].

haven

- Haven enables R to read and write various data formats used by other statistical packages
- Haven is part of the tidyverse. The easiest way to get haven is to install the whole tidyverse: install.packages("tidyverse") Alternatively, install just haven: install.packages("haven")

haven

- SAS: *read_sas*() reads .sas7bdat + .sas7bcat files and *read_xpt*() reads SAS transport files (version 5 and version 8).
- SPSS: *read_sav*() reads .sav files and *read_por*() reads the older .por files.*write_sav*() writes .sav files.
- Stata: *read_dta*() reads .dta files (up to version 15). *write_dta*() writes .dta files (versions 8-15).

haven

- The output objects are tibbles
- Translate value labels into a new labelled() class, which preserves the original semantics and can easily be coerced to factors with *as_factor*()

References


Lan Yuan (2019)
Guide to the statistical analysis plan
pediatric anesthesia 10.1111/pan.13576

The End