**Chapter 11 Statistical analysis plans**

Statistical analysis plans (SAP) are also known as data analysis plans (DAP) or reporting analysis plans (RAP). A statistical analysis plan describes the study variables and the plan for analyzing a data **before** conducting the analysis; this is essentially the strategy for connecting the study objective to the data analysis that will answer the research question. SAPs have been used in biomedical research and in clinical trials for many years; statistical analysis plans for clinical trials are registered and made publicly available in repositories such as [ClinicalTrials.gov](http://www.clinicaltrials.gov). In fact, the National Institutes of Health (NIH) in the United States established policies for reporting NIH-funded clinical trials in 2016, requiring researchers to report full protocol and statistical analysis plan, along with levels of specification for outcome measures, information about adverse events and collection method, and baseline information and characteristics associated with primary outcome measures (1). Pre-registering SAPs can prevent “*P-value hacking”*, which can occur when researchers “*shop around for a statistical test to give them the P-value that they love*” (2). By registering pre-specified SAPs, researchers can help improve the study reproducibility and reduce bias (3).

In observational studies, SAPs are much less adopted compared to clinical trials (4); however, the discussion around its use and value have been growing. In this chapter, we discuss the use of SAPs for observational studies, and propose some key components of SAP for observational studies.

**11.1. The value of statistical analysis plans in observational studies**

Many observational studies are based on large datasets, or “big data,” which is defined as heterogeneous datasets linked to a single dataset, with a large number of observations and variables, and that is either real-time or frequently updated (5). With these big data and powerful statistical software and methods, finding statistically significant associations without pre-established study objectives, research questions and hypotheses has become easier (2). These types of analyses can produce statistically significant findings without implications to clinical relevance or justification. SAPs can be useful in ensuring that the analytical methods are planned ahead of time in relation to the research question and objectives, and that this procedure is transparent.

As the findings from observational studies may have an impact on public health policies, guidelines and decision-making, it is critical to ensure that these studies are of high standard, that analyses are pre-specified based on relevance to public health, and that they are replicable. When there is no pre-establisehd SAP specifying the primary outcome variable, outcome reporting bias can occur (6). Many efforts have been made to reduce reporting bias in observational studies, such as STROBE guidelines (7). The use of SAPs has also been suggested, and that only the variables that researchers pre-specified as variables of interest be made available to them to limit post hoc analyses (8,9). Some even argue that SAPs should be required even before obtaining data, during the application stage of data access (10,11). In fact, to obtain access to big data, it is often required to submit a data request form that contains some key elements of a SAP (12,13).

SAPs also have an important role in identifying potential biases, such as selection bias (based on the inclusion/exclusion criteria) or measurement bias, and can help researchers plan how to minimize and address these biases.

**11.2. Guide on writing an SAP for observational studies**

Based on the guidelines for SAP for clinical trials (14) and literature suggesting its’ adaptation for observational studies (2,8,11), we suggest the following four key components for writing SAP for observational studies in health sciences research:

* **Study objectives and hypotheses**
  + Broad research area, study background and rationale
  + Research question (e.g. using PICOT framework (15) and FINER criteria, see Chapter 8)
  + Hypothesis and aims
* **Study population**
  + Study design (e.g. cross-sectional, prospective cohort)
  + Study sample and inclusion/exclusion criteria
  + Study period (time points under consideration in the data source)
  + Baseline characteristics of study population
* **Study variables: definitions, types, how they are measured**
  + Outcome variables
  + Explanatory/exposure variables
  + Covariates (e.g. mediators, colliders, confounders)
  + Derived variables
* **Statistical analysis methods**
  + Defined level for statistical significance
  + Plans for handling missing data, correlation, bias and confounding, and repetitive analyses
  + Details on model building and variable selection
  + Details on additional methods if model assumptions do not hold (e.g. normality, proportional hazards)
  + Strategies for interaction or subgroup analysis and sensitivity analyses

Finally, SAPs can be useful in observational studies because they encourage detailed and rigorous planning of the study rather than disorganized and spontaneous data analysis. They can also optimize the resources to focus on the right methods for the research questions, and ensure methodological transparency and replicability of findings. We propose the following two broad questions that can be used to determine whether the SAP is appropriate:

1. Does the SAP help in answering the research question or achieving the original study objective?
2. Are the planned analyses appropriate in the context of the research question?

**References**

1. Zarin DA, Tse T, Williams RJ, Carr S. Trial Reporting in ClinicalTrials.gov — The Final Rule. N Engl J Med [Internet]. 2016 Nov 17 [cited 2021 Nov 21];375(20):1998–2004. Available from: /pmc/articles/PMC5225905/

2. Yuan I, Topjian AA, Kurth CD, Kirschen MP, Ward CG, Zhang B, et al. Guide to the statistical analysis plan. Paediatr Anaesth. 2019;29(3):237–42.

3. Kahan BC, Ahmad T, Forbes G, Cro S. Public availability and adherence to prespecified statistical analysis approaches was low in published randomized trials. J Clin Epidemiol [Internet]. 2020 Dec 1 [cited 2021 Nov 21];128:29–34. Available from: http://www.jclinepi.com/article/S0895435620301979/fulltext

4. Thor M, Oh JH, Apte AP, Deasy JO. Registering Study Analysis Plans (SAPs) Before Dissecting Your Data—Updating and Standardizing Outcome Modeling. Front Oncol. 2020 Jun 24;10:978.

5. Ehrenstein V, Nielsen H, Pedersen AB, Johnsen SP, Pedersen L. Clinical epidemiology in the era of big data: New opportunities, familiar challenges. Clin Epidemiol [Internet]. 2017 Apr 27 [cited 2021 Nov 22];9:245–50. Available from: /pmc/articles/PMC5413488/

6. Cafri G, Paxton EW. Mitigating Reporting Bias in Observational Studies Using Covariate Balancing Methods. Obs Stud [Internet]. 2018 [cited 2021 Nov 22];4(1):292–6. Available from: https://doi.org/10.1353/obs.2018.0009https://muse.jhu.edu/article/793374/summaryObservationalStudies

7. Von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: Guidelines for reporting observational studies. PLoS Med. 2007 Oct;4(10):1623–7.

8. Thomas L, Peterson ED. The value of statistical analysis plans in observational research: Defining high-quality research from the start [Internet]. JAMA - Journal of the American Medical Association American Medical Association; Aug 22, 2012 p. 773–4. Available from: https://jamanetwork.com/journals/jama/fullarticle/1352120

9. Williams RJ, Tse T, Harlan WR, Zarin DA. Registration of observational studies: Is it time? [Internet]. Vol. 182, CMAJ. Canadian Medical Association; 2010 [cited 2021 Nov 21]. p. 1638–42. Available from: /pmc/articles/PMC2952011/

10. Trinh QD, Sun M. Statistical analysis plans in observational research [Internet]. Vol. 309, JAMA - Journal of the American Medical Association. American Medical Association; 2013 [cited 2021 Nov 22]. p. 32. Available from: https://jamanetwork.com/journals/jama/fullarticle/1555125

11. Hiemstra B, Keus F, Wetterslev J, Gluud C, Van Der Horst ICC. DEBATE-statistical analysis plans for observational studies. BMC Med Res Methodol [Internet]. 2019 Dec 9 [cited 2021 Nov 21];19(1):1–10. Available from: https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/s12874-019-0879-5

12. DARS: how to make an application [Internet]. NHS Digital. 2021 [cited 2021 Nov 22]. Available from: https://digital.nhs.uk/services/data-access-request-service-dars/dars-guidance/dars-how-to-make-an-application

13. The PopData/HDP Data Access Request (DAR) process [Internet]. Population Data BC. 2021 [cited 2021 Nov 22]. Available from: https://www.popdata.bc.ca/data\_access/DAR\_process

14. Gamble C, Krishan A, Stocken D, Lewis S, Juszczak E, Doré C, et al. Guidelines for the content of statistical analysis plans in clinical trials. JAMA - J Am Med Assoc [Internet]. 2017 Dec 19 [cited 2021 Nov 21];318(23):2337–43. Available from: https://jamanetwork.com/journals/jama/fullarticle/2666509

15. Thabane L, Thomas T, Ye C, Paul J. Posing the research question: not so simple. Can J Anesth Can d’anesthésie 2008 561 [Internet]. 2008 Dec 24 [cited 2021 Oct 31];56(1):71–9. Available from: https://link.springer.com/article/10.1007/s12630-008-9007-4