Journal Pre-proof

Statistical Analysis Plan (SAP): What is it and how to develop it?

Marc Fradera, Xavier Serra-Aracil

Please cite this article as: Fradera M, Serra-Aracil X, Statistical Analysis Plan (SAP): What is it and how to develop it?, *Cirugí́a Española (English Edition)* (2024), doi: https://doi.org/10.1016/j.cireng.2024.07.007

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Plan de análisis estadístico: ¿Qué es y cómo elaborarlo?

Statistical Analysis Plan (SAP): What is it and how to develop it?

Marc Fradera[1], Xavier Serra-Aracil[2]

[1] Responsable de la Unitat de Suport a la Recerca (USR). Institut d'Investigació i Innovació Parc Taulí I3PT-CERCA.

[2] Consorci Corporació Sanitària Parc Taulí CCSPT, Sabadell. Institut d'Investigació i Innovació Parc Taulí I3PT-CERCA. Profesor Titular de Universidad. Departamento de Cirugía, Universitat Autònoma de Barcelona.

Correspondencia:

Marc Fradera

Responsable de la Unitat de Suport a la Recerca (USR). Institut d'Investigació i Innovació Parc Taulí I3PT-CERCA.

Parc Taulí, Plaça de la Torre de l'Aigua s/n, 08208 Sabadell (Barcelona). España

Correo electrónico: mfradera@tauli.cat

ID ORCID: 0000-0001-9290-6616

In the field of biomedical research, regardless of the study type and design, statistics plays an essential part. It provides us with methods to organise, summarise and analyse data in order to extract valid conclusions from it and facilitate decision-making. Due to the significance of statistics in ensuring accuracy and validity of conclusions derived from data analysis, good planning of the statistical methods and procedures to be employed in the different research project stages is vital. For some time now, the most prestigious scientific journals have requested the attachment of a Statistical Analysis Plan. This provides a detailed description of the proposed statistical methods for project data analysis and accompanies the manuscript for publication assessment.[1,2] Quoting the North American writer Rican Alan Lakein, "Planning is bringing the future into the present so you can do something about it now." Throughout this document we will explain what a Statistical Analysis Plan (known as SAP) is, why it is important, and how to prepare it.

1

What is an SAP?

SAP, or statistical analysis plan, is a document which is separate from the study protocol and which describes and details the intended statistical methods that will be used to analyse the data collected in a research study. To sum up, it describes what variables and results will be collected and what statistical methods will be used to analyse them.[3]

Generally, the study protocol specifies the study design; eligibility criteria; primary and secondary objectives; statistical methods to analyse the main variables; statistical power, and duly justified sample size. Although it is true that the study protocol already contains the main characteristics of the statistical analysis, the SAP is usually a much more complete document. It contains exhaustive technical details on the clinical analysis planned for the main variables; the management of secondary variables; control and/or confounding variables; the confidence intervals that will be used to present the results; the management of missing data, and other relevant specifications.[3,4]

Why is it important to prepare an SAP?

In modern research and open science, transparency and reproducibility are two basic concepts in good research practices to guarantee that the said statistical methods and procedures are accessible and reproducible for other researchers.[5] Having an SAP, first of all, increases the transparency of the analysis. By establishing a detailed plan that describes how the data will be analysed before beginning the study, clarity is provided about the statistical methods and procedures that will be used. This allows researchers and the scientific community to understand in a transparent and concise manner how the results were obtained, facilitating the reproducibility of statistical analyses by personnel outside the research team. This is essential to validate and guarantee the reliability of the findings, contributing to the confidence in and credibility of biomedical research.[6,7] Furthermore, another notable advantage in developing a statistical analysis plan is the efficiency resulting from the necessary communication between the statistician and the researcher when preparing the document.[1] The fact that both the research team and the statistician actively participate in the preparation of the statistical planning that will be carried out will save time when making statistical and

2

methodological decisions during the data analysis process. Although preparing the SAP does require considerable time, it will undoubtedly be a worthwhile investment.

To conclude, the development of an SAP in biomedical research projects is essential to guarantee the transparency, reproducibility, objectivity and validity of the analyses, while promoting effective communication between researchers and statisticians. This contributes significantly to the quality and credibility of the project in question.

When should the SAP be made?

The SAP must be prepared either at the same time or shortly after completing the protocol.[8] In the case of experimental studies, if necessary, the SAP can be updated before unblinding the study to guarantee the transparency, accuracy and validity of the analyses.[9] In prospective observational studies, the SAP should be completed before the inclusion of the first patient.[10] When the study is retrospective observational, it is also advisable to have a SAP and, in this case, its version must be final before closing the database to begin analysis.[3] One aspect to consider, whatever the type of study, is to detail each update with its dated version.

How is an SAP prepared?

As previously commented upon, given the level of detail and specificity required by the document, it is essential that the research staff and the person or persons who will be responsible for the statistical analysis collaborate in the preparation of the SAP. Several scientific publications serve as guides for the preparation of SAPs,[3, 9, 10] and a recent publication even provides an extensive and complete template for the scientific community, with the different sections to include and how they should be completed.[8]

Sections to be included in an SAP.

The Statistical Analysis Plan must address different sections to provide detail on the statistical management that will ensue and thus enable other researchers to replicate the analysis with similar data sets. Some consensus exists among the scientific community on the main sections and sections that the SAP should contain, thanks to the work of Gamble et al in 2017[9]:

3

1. <u>Administrative information</u>: This includes the title of the project, the SAP version and the protocol, the different revisions that have been made, the signatures of those people involved in the preparation of the document, and their roles.

2. <u>Introduction</u>: This section contextualises the project with its scientific justification and the research questions intended to be answered.

3. <u>Design and methods</u>: A detailed description of the study methodology. This section also includes the statistical justification for the sample size calculation, as well as the randomisation procedures to be carried out, if applicable. Additionally, this section specifies the proposed interim analyses and the criteria for stopping the study early based on the results obtained in said analyses.

4. <u>Statistical assumptions</u>: This section details both the confidence intervals and the level of statistical significance that will be assumed. Another part of this section is the definition of adherence, protocol deviations and the population that will be analysed (by intention to treat, by protocol, etc.)

5. <u>Study population</u>: Here the eligibility criteria of the sample are explained, as well as the follow-up time, management of loss to follow-up, etc. The baseline characteristics that will be collected for each of the study participants are also detailed.

6. <u>Data analysis</u>: This is the broadest and most detailed section, since it is where what is stated in the protocol must be expanded. All primary and secondary variables to be analysed must be well documented, together with the units of measurement for each of them. All statistical analyses planned for these variables, the management of missing data and the statistical software that will be used to develop the analysis are also detailed.

There are some small differences between the content of an SAP depending on the type of study. Table 1 specifies the different sections that should be included in randomised clinical trials, in prospective and retrospective observational trials.

References

Referencias

4

1. New England Journal of Medicine. New Manuscripts, https://www.nejm.org/author-center/new-manuscripts [consultada el 14 de mayo de 2024]

2. JAMA Surgery. Instructions for Authors, https://jamanetwork.com/journals/jamasurgery/pages/instructions-for-authors#SecProtocols [consultada el 14 de mayo de 2024]

3. Yuan I, Topjian AA, Kurth CD, Kirschen MP, Ward CG, Zhang B, Mensinger JL. Guide to the statistical analysis plan. Paediatr Anaesth. 2019 Mar;29(3):237-242. DOI: 10.1111/pan.13576.

4. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH). Statistical Principles for Clinical Trials (E9). 1998. https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-9-statistical-principles-clinical-trials-step-5_en.pdf [consultada el 14 de mayo de 2024]

5. Nosek BA, Alter G, Banks GC, Borsboom D, Bowman SD, Breckler SJ, et al. Promoting an Open Research Culture. Science. 2015;348(6242):1422-5. DOI: 10.1126/science.aab2374.

6. Munafò M, Nosek B, Bishop D, et al. A manifesto for reproducible science. Nat Hum Behav. 2017;1:0021. DOI: 10.1038/s41562-016-0021.

7. Landis S, Amara S, Asadullah K, et al. A call for transparent reporting to optimize the predictive value of preclinical research. Nature. 2012;490:187–91. DOI: 10.1038/nature11556.

8. Stevens G, Dolley S, Mogg R, Connor JT. A template for the authoring of statistical analysis plans. Contemp Clin Trials Commun. 2023 9;34:101100. DOI: 10.1016/j.conctc.2023.101100.

9. Gamble C, Krishan A, Stocken D, et al. Guidelines for the Content of Statistical Analysis Plans in Clinical Trials. JAMA. 2017;318(23):2337–2343. DOI:10.1001/jama.2017.18556

10. Hiemstra, B., Keus, F., Wetterslev, J. et al. DEBATE-statistical analysis plans for observational studies. BMC Med Res Methodol. 2019;19,233. DOI: https://doi.org/10.1186/s12874-019-0879-5

Table 1. Sections and components that the Statistical Plan Analysis must contain for randomised clinical trials, prospective observational studies and retrospective observational studies.

		Study type		
	Sections and components that the SAP must contain	**Randomised clinical trial**	**Prospective observational**	**Retrospective observational**
1	Project title	X	X	X
2	Study registration number	X		

3	SAP version number and date	X	X	X
4	Study protocol version	X	X	X
5	SAP review history	X	X	X
6	Reasons for SAP reviews	X	X	X
7	Time of SAP reviews in relation to the interim analyses	X	X	
8	SAP collaborators, with responsibilities and roles	X	X	X
9	Name of the person who wrote the SAP	X	X	X
10	Name of the senior statistician	X	X	X
11	Name of the principal researcher	X	X	X
12	Study background and justification	X	X	X
13	Hypothesis and objectives	X	X	X
14	Study type	X	X	X
15	Randomisation details	X		
16	Estimation and justification of sample size, if applicable	X	X	X
17	Focus of the superiority hypothesis tests, equivalence or non inferiority	X		
18	Interim analysis, time of analysis and person performing the analysis, if applicable	X	X	
19	Adjustment of the significance level due to interim analysis	X	X	
20	Indications for early termination of the study	X	X	
21	Final analysis moment	X	X	X

6

22	Schedule of visits and time interval to evaluate each result	X	X	
23	Statistical significance levels (p values) and whether they are unilateral or bilateral	X	X	X
24	Plan and justification for multiplicity adjustment, if applicable, including how type 1 error is controlled	X	X	X
25	Confidence intervals to be reported and whether they are unilateral or bilateral	X	X	X
26	Definition of adherence to the intervention and how it will be presented	X		
27	Definition and summary of protocol deviations	X	X	
28	Definition of the population analysed	X	X	X
29	Report screening data to describe representation of the 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 early withdrawal of patients from the study	X	X	
33	Presentation of early withdrawal and follow-up data	X	X	
34	Baseline characteristics of the patients and how they will be descriptively summarised.	X	X	X
Points 35-37 apply to each of the primary and secondary results.				
35	Definitions of results and sequence of measurement	X	X	X
36	Specific measurements and units for each variable	X	X	X
37	Estimations and transformations used to obtain the result	X	X	X

7

38	Methods of analysis used	X	X	X
39	Presentation of treatment or intervention effects	X		
40	Covariates and adjustments	X	X	X
41	Methods for confirming distribution assumptions	X	X	X
42	Alternative methods used if distribution assumptions are not met	X	X	X
43	Sensitivity analysis for each outcome, if applicable	X	X	X
44	Subgroup definition and analysis, if applicable	X	X	X
45	Methods for missing data management	X	X	X
46	Additional statistical analysis, if applicable	X	X	X
47	Safety data summary details	X	X	
48	Statistical packages used for analysis	X	X	X
49	Reference to standard operating procedure or additional documents	X	X	X

Adaptation of Yuan et al (2019), under licence permits.

8