

Quantitative research - Data analysis - Analysis plan

Set-up & Conduct- Process & Analyse data

VERSION

4.0

Aim

To promote structured targeted data analysis.

Requirements

An analysis plan should be created and finalized prior to the data analyses.

Documentation

The analysis plan (Guidelines per study type are provided below)

Responsibilities

- Executing researcher: To create the analysis plan prior to the data analyses, containing a description of the research question and what the various steps in the analysis are going to be. This should also be signed and dated by the PI.
- Project leaders: To inform the executing researcher about setting up the analysis plan before analyses are undertaken.
- Research assistant: N.a.

How To

An analysis plan should be created and finalized (signed and dated by PI) prior to the data analyses. The analysis plan contains a description of the research question and what the various steps in the analysis are going to be. It also contains an exploration of literature (what is already known? What will this study add?) to make sure your research question is relevant (see Glasziou et al. Lancet 2014 on avoiding research waste). The analysis plan is intended as a starting point for the analysis. It ensures that the analysis can be undertaken in a targeted manner, and promotes research integrity.

If you will perform an exploratory study you can adjust your analysis based on the data you find; this may be useful if not much is known about the research subject, but it is considered as relatively low level evidence and it should be clearly mentioned in your report that the presented study is exploratory. If you want to perform an hypothesis-testing study (be it interventional or using observational data) you need to pre-specify the analyses you intend to do prior to performing the analysis, including the population, subgroups, stratifications and statistical tests. If deviations from the analysis plan are made during the study this should be documented in the analysis plan and stated in the report (i.e. post-hoc tests). If you intend to do hypothesis-free research with multiple testing you should pre-specify your threshold for statistical significance according to the number of analyses you will perform. Lastly, if you intend to perform an RCT, the analysis plan is practically set in stone. (Also see [ICH E9 - statistical principles for clinical trials](#))

If needed, an exploratory analysis may be part of the analysis plan, to inform the setting up of the final analysis (see initial data analysis). For instance, you may want to know distributions of values in order to create meaningful categories, or determine whether data are normally distributed. The findings and decisions made during these preliminary exploratory analyses should be clearly documented, preferably in a version two of the analysis plan, and made reproducible by providing the data analysis syntax (in SPSS, SAS, STATA, R) (see chapter: Documentation of data analysis).

The concrete research question needs to be formulated firstly within the analysis plan following the literature review; this is the question intended to be answered by the analyses. Concrete research questions may be defined using the acronym PICO: Population, Intervention, Comparison, Outcomes. An example of a concrete question could be: “Does frequent bending at work lead to an elevated risk of lower back pain occurring in employees?” (Population = Employees; Intervention = Frequent bending; Comparison = Infrequent bending; Outcome = Occurrence of back pain). Concrete research questions are essential for determining the analyses required.

The analysis plan should then describe the primary and secondary outcomes, the determinants and data needed, and which statistical techniques are to be used to analyse the data. The following issues need to be considered in this process and described where applicable:

- In case of a trial: is the trial a superiority, non-inferiority or equivalence trial.
 1. Superiority: treatment A is better than the control.
 2. Non-inferiority: treatment A is not worse than treatment B.
 3. Equivalence: testing similarity using a tolerance range.

In other studies: what is the study design (case control, longitudinal cohort etc).

- Which (subgroup of the) population is to be included in the analyses? Which groups will you compare?;
- What are the primary and secondary endpoints? Which data from which endpoint (T1, T2, etc.) will be used?;
- Which (dependent and independent) variables are to be used in the analyses and how are the variables to be analysed (e.g. continuous or in categories)?;
- Which variables are to be investigated as potential confounders or effect modifiers (and why) and how are these variables to be analysed? There are different ways of dealing with confounders. We distinguish the following:
 - 1) correct for all potential confounders (and do not concern about the question whether or not a variable is a ‘real’ confounder). Mostly, confounders are split up in little groups (demographic factors, clinical parameters, etc.). As a result you get corrected model 1, corrected model 2, etc. However, pay attention to collinearity and overcorrection if confounders coincide too much with primary determinants.
 - 2) if the sample size is not big enough relative to the number of potential confounders, you may consider to only correct for those confounders that are relevant for the association between determinant and outcome. To select the relevant confounders, mostly a forward selection procedure is performed. In this case the confounders are added to the model one by one (the confounder that is associated strongest first). Subsequently, consider to what extent the effect of the variable of interest is changed. Then first choose the strongest confounder in the model. Subsequently, repeat this procedure until no confounder has a relevant effect (<10% change in regression coefficient). Alternatively, you can select the confounders that univariately change the point estimate of the association with >10%.
 - 3) Another option is to set up a Directed Acyclic Graph (DAG), to determine which confounders should be added to the model. Please see <http://www.dagitty.net/> for more information.
- How to deal with missing values? (see Chapter on handling missing data);

Amsterdam Public Health



- Which analyses are to be carried out in which order (e.g. univariable analyses, multivariable analyses, analysis of confounders, analysis of interaction effects, analysis of sub-populations, etc.)?; Which sensitivity analyses will be performed?
- Do the data meet the criteria for the specific statistical technique?

A statistician may need to be consulted regarding the choice of statistical techniques.

It is recommended to already design the empty tables to be included in the article prior to the start of data analysis. This is often very helpful in deciding which analyses are exactly required in order to analyse the data in a targeted manner.

You may consider to make your study protocol including the (statistical) analysis plan public, either by placing in on a publicly accessible website (Concept Paper/Design paper) or by uploading it in an appropriate studies register (for human trials: NTR/EUDRACT/ClinicalTrials.gov, for non-/preclinicaltrials: preclinicaltrials.eu).

Check the reporting guidelines when writing an analysis plan. These will help increase the quality of your research and guide you

<u>Randomised trials</u>	<u>CONSORT</u>	<u>Extensions</u>
<u>Observational studies</u>	<u>STROBE</u>	<u>Extensions</u>
<u>Systematic reviews</u>	<u>PRISMA</u>	<u>Extensions</u>
<u>Study protocols</u>	<u>SPIRIT</u>	<u>PRISMA-P</u>
<u>Diagnostic/prognostic studies</u>	<u>STARD</u>	<u>TRIPOD</u>
<u>Case reports</u>	<u>CARE</u>	<u>Extensions</u>
<u>Clinical practice guidelines</u>	<u>AGREE</u>	<u>RIGHT</u>
<u>Qualitative research</u>	<u>SRQR</u>	<u>COREQ</u>
<u>Animal pre-clinical studies</u>	<u>ARRIVE</u>	
<u>Quality improvement studies</u>	<u>SQUIRE</u>	
<u>Economic evaluations</u>	<u>CHEERS</u>	

References

- Swaen et al: Responsible Epidemiologic Research Practice: a guideline developed by a working group of the Netherlands Epidemiological Society. J Clin Epidemiol. 2018 Aug;100:111-119.
- Glasziou et al: Reducing waste from incomplete or unusable reports of biomedical research. Lancet. 2014 Jan 18;383(9913):267-76.
- [Example of analysis plan \(see under Download\)](#)
- For interim analyse in medical research please also see section 8.1 of the most recent version of the NFU Guideline Quality assurance of research involving human subjects
- ICH E9 - statistical principles for clinical trials
- ICH 6 - Good clinical practice

Audit questions

1. Has an analysis plan been created prior to the start of analysis?
2. Has a concrete research question been formulated in the analysis plan?
3. Have the points described under 'How To' been considered?
4. Has a stepwise description of the analyses to be applied been provided in the analysis?

Amsterdam Public Health



Amsterdam UMC
University Medical Centers



VRIJE
UNIVERSITEIT
AMSTERDAM



UNIVERSITY
OF AMSTERDAM

LINKS

	Link
Wiki statistiek	https://wikistatistiek.amc.nl/index.php/Wiki_Statistiek
Dagitty	http://www.dagitty.net/
ICH E9 - statistical principles for clinical trials	https://www.ema.europa.eu/en/ich-e9-statistical-principles-clinical-trials
ICH 6 - Good clinical practice	https://www.ema.europa.eu/en/ich-e6-r2-good-clinical-practice
NFU Guideline Quality assurance of research involving human subjects	https://www.nfu.nl/sites/default/files/2021-01/21.00024_Guideline_Quality_assurance_of_research_involving_human_subjects_dec20_0.pdf

DOCUMENT HISTORY

Version	Status	Date	Name
4.0	Revision	08DEC2020	Dr. Marieke Blom & Laura van Dongen, Elize Vlainic
3.0	Adaption of the example of analysis plan and minor revisions	23JAN2017	EMGO
2.1	Minor revision	13OCT2016	EMGO
2.0	Revision format	12MAY2015	EMGO
1.2	English translation	01JAN2010	EMGO
1.1	Text in guideline has been re-written with more emphasis on a flexible approach	21JAN2008	EMGO

DOCUMENT APPROVAL

Role	Name	Date
Project Leader	Dr. Seta Jahfari	12MAY2021