

Create a statistical analysis plan describing the study as indicated by Sven (A template will be available on Fronter).

SAP – Statistical Analysis Plan

Hepatitis F in Bengalia

- Purpose/Objective: To present the prevalence of Hepatitis F and characteristics of patients with Hepatitis F in the Bengalian disease registry in 2017.
- Hypotheses: None, since it is a descriptive and explorative study on a topic in a country with no previous scientific publication.
- Data source: The Bengalian national registry of Hepatitis F, complete extraction for 2017. Extraction of depersonalized data have to be performed in Bengalia due to data protection laws.
- Ethical considerations: permission from Regional Ethics Board in Norway reg. 145/166 dated 17.06.2017.
- Population: All persons diagnosed with and/or living with Hepatitis F in Bengalia in 2017 and who was in contact with the health sector due to the disease.
- Working team: first author, 3 co-authors and principal investigator. First author is responsible for reviewing literature, deriving variables, tabulate variables for tables, conduct analyses and drafting the article. The 3 co-authors are responsible for reviewing draft of article. Principal investigator is responsible for reviewing article and checking data.
- Primary variables: sex, date of birth, in which region they live, level of education, civil status, height, weight, years since diagnosis, comorbidities, type of treatment, how many years of treatment, type of health coverage (private health insurance or universal health care), level of work.
- Derived variables: age, body mass index, number of years with diagnosis, socio-economic status according to scale.
- Statistical methods: Chi-square, t-test and Wilcoxon rank-sum test to test for differences in mean, median and categorical variables between sex and age groups.
- Schedule: Analysis start; February 2018.

First draft result 1. April 2018.

First manuscript draft 15. May 2018.

- Target journal: Journal of international Hepatitis studies, impact factor 2.565. Open access, maximum 3500 words + references and 4 tables.
- Statistical software: Stata/MP version 15.0
- Storage of data: all do-files for cleaning data, creating variables and running analyses are made available in GitHub in an open access repository. Provide log and output of what is being done continuously in the process with a time-stamp. All variables should be given short and preferably self-explanatory names. Derived variables need to be added to the codebook of the study.
- The original extracted datafile is stored at an SQL-server available to those involved in the study.
- Skeleton tables, example:

Characteristics	Males	Females
Mean age		
Education level		
Civil status		
Work level		
Etc....		

Question 5

Create a good DM plan for the project.

- Write a statistical analysis plan (SAP) to create an overview of the project, which is agreed upon by all involved in the study.
- Make clear who will have access to data in the study.
- The original dataset is stored on an SQL-server, which can be easily transferred and opened in most common statistical programs. The main program used in the study is STATA/MP version 15.0.
- The original dataset is stored using UTF-8.
- Create a common work directory for do-files, log entries and other files related to the study stored in the cloud at the university. Make separate folders for the different stages of the study with self-explanatory names.

- Make a codebook of all the original and derived variables used in the study, list if the variable is mandatory, derived, and how missing and unknown observations are coded.
- When the original dataset is imported into Stata, first write a do-file with names and labels for all variables. The names must be short and as self-descriptive as possible.
- Second, write do-files for cleaning the dataset. Make the lists in chronological order, naming them in such a way that it is self-explanatory.
- For each do-file created and stored, including when performing analyses, save output and write a log-entry to explain the content of the do-file with a time-stamp, save all in the common work directory with the same name-ending for easy identification.
- Do not store different versions of the dataset in STATA, instead run the required do-files in chronological order to obtain the cleaned and labelled data file ready for analysis.
- If an update of software is necessary during the study, all analyses have to be run again.
- Perform backup of all material once a week.
- If anyone involved in the project change jobs or for some reason will not be involved in the project any longer, they are obliged to notify the rest of the group and will no longer have access to the common working directory and original dataset.
- When the article is published, the data content of the study have to be archived at the registry office or the university with the possibility of being made available to other researchers/reviewers on request. Files to be included in the archive are:
 - Final statistical research plan
 - Copy of original dataset
 - Codebook for the study
 - Do-files used during the study for cleaning data, create variables and run analyses, in chronological order with timestamps for reproducibility purposes.
 - Logs with explanations of what was done and why
 - Output and results included in the article
 - Copy of the article
 - Ethical permission to use the data material