How to make a research plan

The absolute essentials

CBB Biostatistics seminar
14.12.2023
Raphaela Mayerhofer

Vocabulary

- Study protocol
- Protocol
- Study plan
- Analysis plan
- Statistical analysis plan
- …

“It is my great pleasure to be here and make you feel quite sad about how bad things are.”

Douglas G. Altman, REWARD-EQUATOR conference, Edinburgh 28-30 Sep 2015.

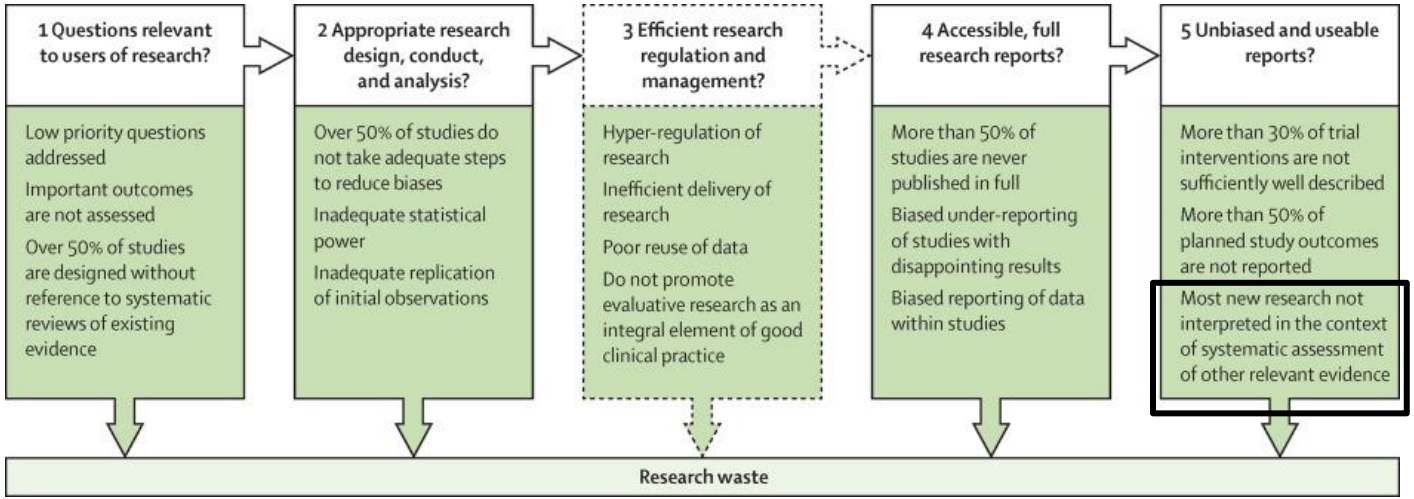
Research waste

- Research that could be answered with existing evidence
- Research questions that are not relevant to patients or clinicians
- Poorly designed studies
- Poor or biased reporting

➢ Poor evidence is expensive and leads to poor clinical decisions (1)

1. Heneghan C, Mahtani KR, Goldacre B, Godlee F, Macdonald H, Jarvies D. Evidence based medicine manifesto for better healthcare. BMJ. 2017 Jun 20;j2973.

What does research waste have to do with protocols?



Moher D, Glasziou P, Chalmers I, Nasser M, Bossuyt PMM, Korevaar DA, et al. Increasing value and reducing waste in biomedical research: who's listening? The Lancet. 2016 Apr;387(10027):1573–86.

→ Inadequate planning (1)

1. Yordanov Y, Dechartres A, Atal I, Tran VT, Boutron I, Crequit P, et al. Avoidable waste of research related to outcome planning and reporting in clinical trials. BMC Med. 2018 Dec;16(1):87.

How having a plan can help

- Forces you to check what already exists
- Check if your aim is relevant
- Identify knowledge gaps, missing equipment, …
- Get independent methodological support
- If no protocol, difficulty in evaluating what was done
- Protect against cognitive biases: making a plan before the data exists and while the outcome is unknown is a form of blinding

What is a statistical analysis plan?

- A map
  - → Part of a grant application
  - → Clinical trial protocol
  - → Master's or PhD thesis project
  - → As part of a larger project
- Set up ***before*** the study begins

What's in a study plan?

- Introduction
  - Justification for study
  - Objectives
  - Research questions and hypotheses

- Procedures and methods
  - Study design
  - Study population
  - Variables/interventions
  - Data management
  - Statistical analysis
  - Management of adverse or unexpected events

Introduction

- Justification for study
  - → What's the health problem?
  - → Who's affected?
  - → How does your study address the problem?

Introduction

- Objectives
  - → Clear and concise statements of what the study will demonstrate, test, evaluate, confirm, or compare
  - → PICO
    - **P**atient or population of interest, or problem
    - **I**ntervention or treatment that is performed on P
    - **C**omparison or control treatment
    - **O**utcome that is to be observed
  - → SMART
    - **S**pecific
    - **M**easurable
    - **A**chievable
    - **R**elevant
    - **T**ime-based (realistic time frame)

Introduction

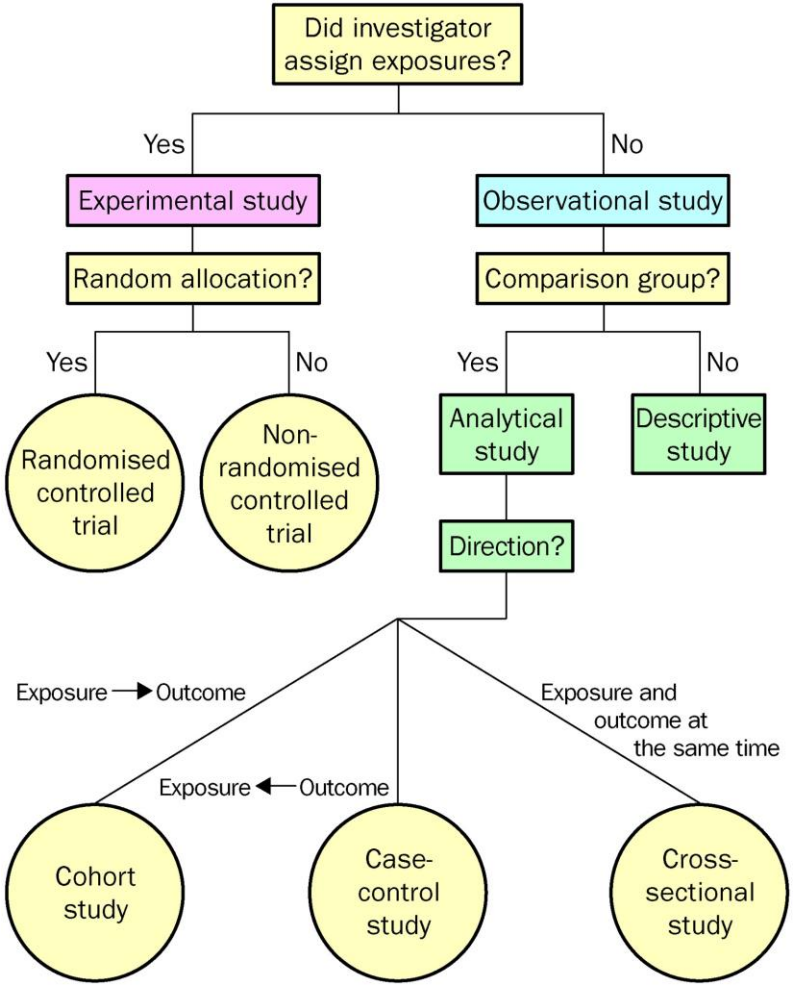
- Research questions and hypotheses
  - → Explicitly state question(s) to be investigated and the hypothesis(es) to be tested in the study
  - → For each specific outcome you're interested in!
  - → **Forms the basis for selecting the appropriate statistical tests!**

- If you don't know the question, you don't know what the answer means

- How to formulate research questions and hypotheses
  - → What is already known about the problem/disease/condition?
  - → What do you still need to find out?
  - → PICO
  - → SMART

Procedures and methods

- Study design
  - → How study design addresses research questions and objectives
    - Experimental/observational
    - Prospective/retrospective

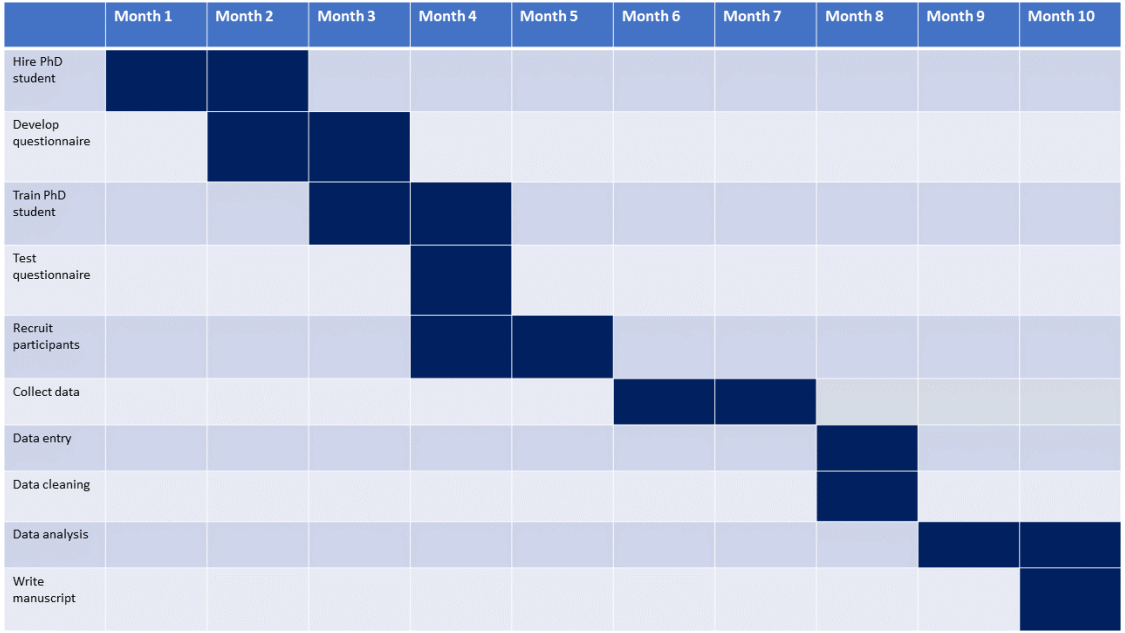


Grimes DA, Schulz KF. An overview of clinical research: the lay of the land. The Lancet. 2002 Jan;359(9300):57–61.

Procedures and methods

- Study design

  → Study timeline



	Month 1	Month 2	Month 3	Month 4	Month 5	Month 6	Month 7	Month 8	Month 9	Month 10
Hire PhD student	■	■								
Develop questionnaire		■	■							
Train PhD student			■	■						
Test questionnaire				■						
Recruit participants				■	■					
Collect data						■	■			
Data entry								■		
Data cleaning								■		
Data analysis									■	■
Write manuscript										■

Procedures and methods

- Study population
  - → Description and source of study population and the target study area
  - → Case definition
  - → Participant inclusion/exclusion criteria
  - → Sampling
    - Simple
    - Stratified
    - Systematic
    - Clustered
  - → Sample size and statistical power
    - Know how many patients are available
    - Calculate sample size and statistical power
      - → Hypotheses
      - → Meaningful effect size
      - → Statistical test

Procedures and methods

- Variables/interventions
  - → Intervention or treatment
    - Drug
    - Lifestyle
    - Surgery

Procedures and methods

- Variables/interventions
  - → Variables
    - Every variable collected in the study must be clearly stated!
    - **Forms the basis for selecting the appropriate statistical tests!**
    - Types of variables
      - → Predictor (causing the effect measured) or outcome (effect being measured) variable
      - → Numeric or character
      - → Continuous (ratio [weight] or interval [temperature])
      - → Categorical (ordinal [cancer stage] or nominal [male/female])

Variable	Type	Description
Age	numeric, continuous	
Sex	character, categorical	Written as "male" or "female" in patient files.
Weight	numeric, continuous	
Comorbidity	numeric, categorical	Charlson Comorbidity Index
Hemoglobin at fibrosis diagnosis	numeric, continuous	

Procedures and methods

- Variables/interventions
  - → Study instruments
    - Questionnaires
    - Laboratory instruments
    - Analytic tests

Procedures and methods

- Data management
  - → Owner of the dataset
  - → Data collection
    - Data collection forms
    - Online/letters/interview
    - Retrieved from external database: how to get access?
  - → Data entry
    - Electronic/by hand
    - Language, date format, missing values
    - Variable dictionary

→ Version control
- Original vs manipulated
- Logbook

→ Storage
- Software (Excel, SQL,…)
- P drive
- MFT
- Store and share files at KI

→ Disposal

→ Research data management at KI

→ REDCap

Procedures and methods

- Data management
  - → Example database

patient_id	sex	date_dx	fibrosis	hb_dx	hb_fup	liver_tx	date_tx
1	1	2013-10-08	0	128	145	0	
2	1	2014-11-24	1	155	164	0	
3	2	2016-03-03	3	137	147	1	2018-04-20

Procedures and methods

- Data management
  - → Example variable dictionary

Variable name	Possible values	Coding key	Unit	Explanation
patient_id	1-150	See document key.xlsx stored on KI computer		Patient ID. Unique number per patient, not the personal number
sex	1 / 2	0=female, 1=male		
date_dx	Between 2013-01-01 and 2017-12-31.			Date of diagnosis with fibrosis
fibrosis	0-4			Stage of fibrosis
hb_dx	90-170		g/L	Hemoglobin level at diagnosis
hb_fup	90-170		g/L	Hemoglobin level at last follow-up
liver_tx	0 / 1	0=no, 1=yes		Whether or not the pat received a liver transplantation

Procedures and methods

- Data management
  - → Example logbook

File name	Date created	Description
Studie1_v1.xlsx	2022-05-10	All data thus far collected from journals 2022-05-01 until 2022-05-10. I will add more data later.
Studie1_v2.xlsx	2022-05-12	Added hb_dx values for patient 2. Sent database to statistician.
Studie1_v3.xlsx	2022-05-13 09:00	More data on patients added.
Studie1_v4.xlsx	2022-05-13 12:00	Patient 4 removed for reasons of … .

Procedures and methods

- Statistical analysis
  - → Data analysis plan, including statistical methodology
    - **Depending on research question, hypothesis and type of variable**
    - Descriptive vs inferential (causal)
    - Understand assumptions for different tests
      - → Parametric/non-parametric
      - → Paired tests for paired data
      - → Linearity in linear regression
    - Level of statistical significance (p-value)
    - How to handle outliers and missing data

Procedures and methods

- Statistical analysis
  - → Data analysis plan, including statistical methodology

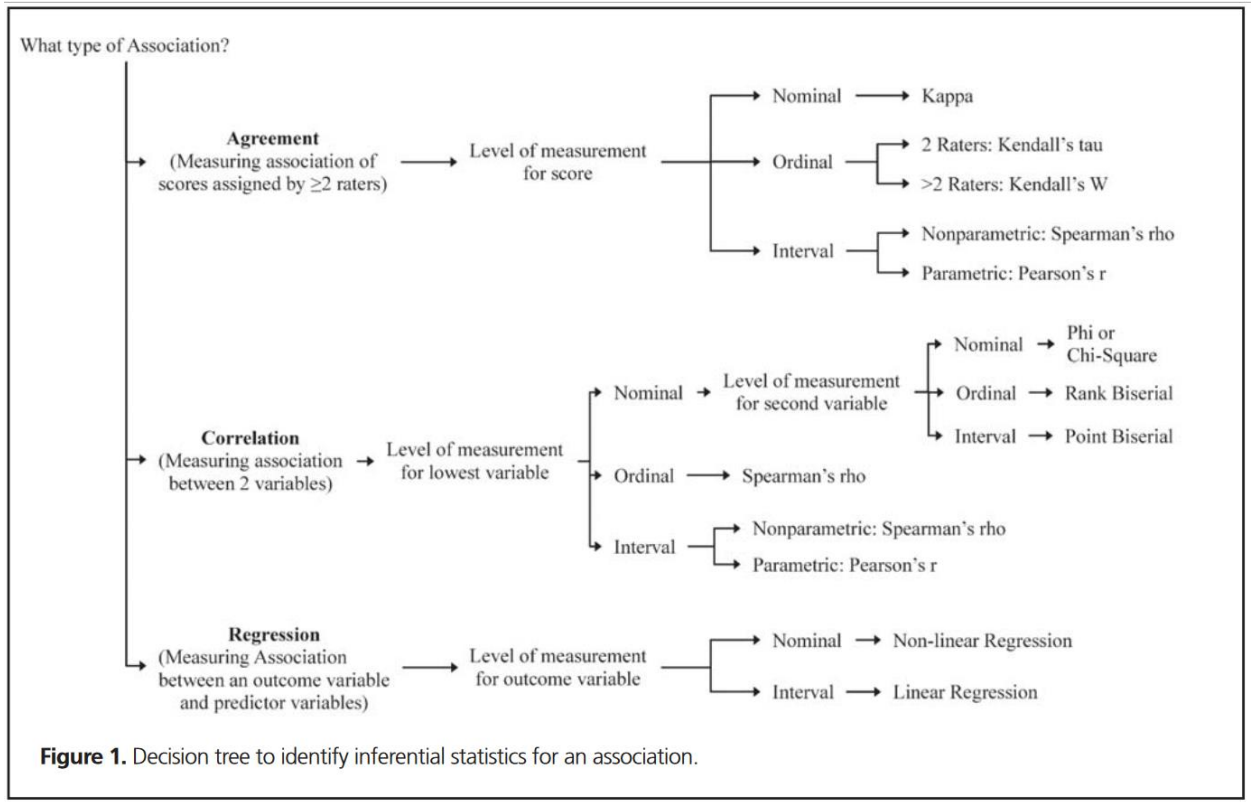


Figure 1. Decision tree to identify inferential statistics for an association.

Simpson SH. Creating a Data Analysis Plan: What to Consider When Choosing Statistics for a Study. CJHP. 2015 Aug 25 ;68(4).

Procedures and methods

- Statistical analysis
  - → Data analysis plan, including statistical methodology

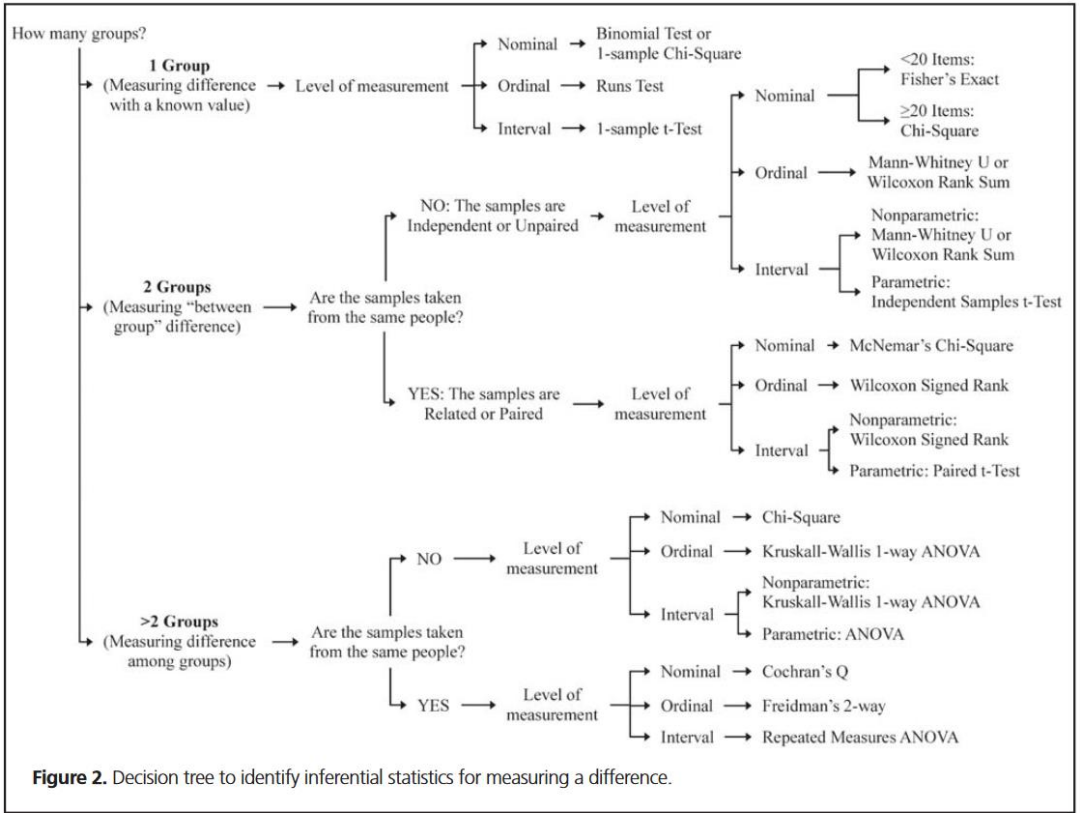


Figure 2. Decision tree to identify inferential statistics for measuring a difference.

Simpson SH. Creating a Data Analysis Plan: What to Consider When Choosing Statistics for a Study. CJHP. 2015 Aug 25 ;68(4).

Procedures and methods

- Statistical analysis
  - → Analysis software
    - R, SPSS, Epi Info, SAS, STATA, Microsoft Excel,…
    - Version number
    - Packages used
    - Year of release
    - Author/manufacturer

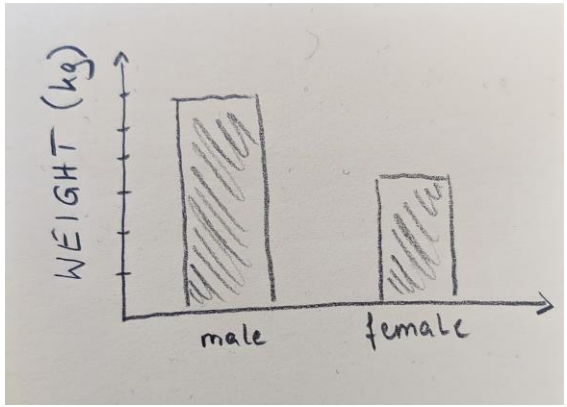
Procedures and methods

- Statistical analysis
  - → Shell tables and figures
    - In anticipation for the results that will be obtained from different levels of analysis

Table 1: Example of a shell table from univariate analysis

Variable		Label
Age		Mean (SD)
Sex		
	Female	N (%)
	Male	N (%)
Fibrosis		N (%)

N=frequency, SD=standard deviation



Procedures and methods

- Statistical analysis
  - → Shell tables and figures
    - In anticipation for the results that will be obtained from different levels of analysis

Table 2: Example of a shell table from bivariate analysis

Variable		Fibrosis	No fibrosis	OR(CI)	p-value
Sex					
	Female	N (%)	N (%)		
	Male	N (%)	N (%)		

OR=odds ratio, CI=95% confidence intervall

Table 3: Example of a shell table from multivariate analysis

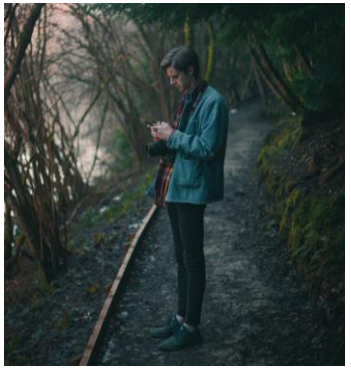
Variable		aOR(CI)	p-value
Age			
Sex			
	Female	Reference	
	Male		
Alcohol			
	No	Reference	
	Yes		
Overweight			
	No	Reference	
	Yes		

aOR = adjusted odds ratio

Procedures and methods

- Management of adverse or unexpected events
  - → How to respond to unexpected findings
  - → How to respond to changes in the study environment
  - → How to identify, manage, and report adverse events

How to get out of freeze mode?



- Nothing is scarier than a blank page

- No part of your project should feel like a black box!!
  - → Essential questions
    - What am I you doing?
    - Why am I doing it?
    - What will it help me to understand/find out?

Resources

- Publications
  - → Simpson SH. **Creating a Data Analysis Plan: What to Consider When Choosing Statistics for a Study**. Can J Hosp Pharm. 2015 Jul-Aug;68(4):311-7. doi: 10.4212/cjhp.v68i4.1471. PMID: 26327705; PMCID: PMC4552232.
  - → Lang TA, Altman DG. **Statistical analyses and methods in the published literature: The SAMPL guidelines**. Medical Writing. 2016 Sep;25(3).
  - → Vetter TR. **Fundamentals of Research Data and Variables: The Devil Is in the Details**. Anesthesia & Analgesia. 2017 Oct;125(4):1375–80.
- https://www.equator-network.org/
- Vetenskapsrådet
  - → https://www.vr.se/english/analysis/reports/our-reports/2017-08-31-good-research-practice.html
  - → https://www.vr.se/english/mandates/open-science/open-access-to-research-data/data-management-plans.html
- KI resources
  - → https://staff.ki.se/research-support-services
  - → https://staff.ki.se/clinical-research
  - → https://ki.se/en/research/centre-for-bioinformatics-and-biostatistics-cbb

“We need less research, better research, and research done for the right reasons.”

Altman DG. The scandal of poor medical research. BMJ. 1994 Jan 29;308(6924):283–4.


KAROLINSKA INSTITUTET
ANNO 1810
Karolinska
Institutet