Flanders
State of the Art

Statistical Analysis Plan (StAP): Module 1

Teachers: Robrecht Dockx and Raisa Carmen

Team BMK

10/1/2025

Part 2: Details of the StAP

Statistical analysis plan (StAP): module 1

- Overview -

PART I

A) Hypothesis

B) Statistical testing

C) Sampling and treatment allocation

D) Design and conduct of the study

E) Discussion

PART II

F) Details of the planned statistical analysis

- Estimand
- Missing data
- Outlier detection
- Sample size
- Multiple testing

PART III

G) Some StAP examples

Day 2

RESEARCH INSTITUTE
NATURE AND FOREST

Detailing your hypothesis

Estimand

Estimand

- Clinical trials
- Create clarity to what will be estimated
- An estimand is a precise description of the treatment effect reflecting the clinical question posed by a given clinical trial objective.
- Estimate is a numerical value computed by an estimator
- Estimator is a method of analysis to compute an estimate of the estimand using data

RESEARCH INSTITUTE
NATURE AND FOREST

The concept of an estimand originates from clinical trials but is widely applicable.

Estimand

- An estimand is a precise description of the treatment effect reflecting the clinical question posed by a given clinical trial objective.

- 5 components
  - The population
  - The treatments or intervention
  - The outcome variable
  - Accounting for intercurrent events
  - Variable summary

RESEARCH INSTITUTE
NATURE AND FOREST

Intercurrent events

- **Events** occurring **after treatment initiation** that affect either the **interpretation** or the **existence** of the measurements associated with the research question of interest
- Examples in observational studies
- Examples in intervention studies

RESEARCH INSTITUTE
NATURE AND FOREST

Examples in observational studies:
In a monitoring scheme on the state of a certain habitat type (eg heather or bog), you typically select a random spatial sample of locations. Over time, when you return to those same locations, the habitat may have changed eg: the heather has turned into a forest or the bog is drier that it used to be.

Examples in intervention studies:
In a study on growth of different clones of poplar trees, one of the fields where the poplars were planted was flooded due to heavy rain. this caused a lot of the young trees to die and we now have a much smaller sample than expected.

Intercurrent events

- **Events** occurring **after treatment initiation** that affect either the **interpretation** or the **existence** of the measurements associated with the research question of interest

- Adressing intercurrent events
  - Treatment policy strategy
  - Hypothetical strategy
  - Composite variable strategy
  - While on treatment
  - Principle stratum strategy

RESEARCH INSTITUTE
NATURE AND FOREST

Treatment policy strategy: the occurence of the intercurrent event is taken to be part of the treatment condition. Variable value is considered despite the intercurrent events. Poplar clones which are planted in polluted soil are more likely to die such that we can no longer follow and compare their growth to poplars in non-polluted soil. We will measure their height at the end of the study, even if they died.

Hypothetical strategy: a hypothetical scenario is envisaged in which the intercurrent event would not (or would) occur, and participant outcomes corresponding to this hypothetical scenario are used. so in the poplar example above, we could only focus on the growth of the poplars that did not die, this would yield a measure of the growth of poplars on polluted soil, **if they survive**. Or we can impute data for the growth of the trees that died and then do our analysis.

Composite variable strategy: This incorporates the intercurrent event into the endpoint. In the poplar example, we could we could assign growth value of zero to all trees that died. Or we could fit a model that first predicts whether the tree would die or not and, if not, how much it'd grow.

While on treatment: We use the outcome before the intercurrent event occurs. Poplar clones which are planted in polluted soil are more likely to die such that we can no longer follow and compare their growth to poplars in non-polluted soil. We will use their growth just before they died in our analysis.

Principal stratum strategy: We restrict the population to those who would not experience an intercurrent event. Trees who die are not included in the estimand population.

Source: https://www.lexjansen.com/pharmasug/2024/DS/PharmaSUG-2024-DS-374.pdf
https://www.bmj.com/content/384/bmj-2023-076316

Estimands

Observational studies

The estimand describes the expected average difference in eDNA concentration of the African clawed frog **between the standard eradication program and the fish predation eradication program.** The target population consists of **ponds within the Douve area**. The outcome variable is the **$\log_{10}$-transformed eDNA concentration**, and the **change is calculated relative to the baseline measurement (t_0).** Potential intercurrent events, such as heavy **rainfall or drought** are considered.

RESEARCH INSTITUTE
NATURE AND FOREST

Estimands

Intervention studies

The estimand describes the **expected average difference** in eDNA concentration of the African clawed frog between **ponds with introduced fish predators and ponds with no intervention (negative control)**. The target population consists of **artificial experimental ponds (vijverbakken)** with comparable ecological conditions and a known starting population of African clawed frogs. The outcome variable is the **$\log_{10}$-transformed eDNA concentration**, and the change is calculated relative to the baseline measurement (t_0) and averaged across post-intervention timepoints (t_1, t_2, t_3). Potential intercurrent events such as **minor temperature fluctuations or small water losses are being considered**

RESEARCH INSTITUTE
NATURE AND FOREST

Missing data

Missing data

datetime	species
1/1/’25 1:05	owl
1/1/’25 2:45	cat
1/1/’25 3:11	fox
1/1/’25 4:28	mouse
1/1/’25 5:56	rat
1/1/’25 6:32	no animal
1/1/’25 7:44	cat
1/1/’25 8:15	rat
1/1/’25 9:00	cat
1/1/’25 9:23	hedgehog
1/1/’25 11:51	great tit
1/1/’25 13:07	dove
1/1/’25 13:49	finch
1/1/’25 14:28	dove
1/1/’25 14:35	dove
1/1/’25 15:09	cat
1/1/’25 15:56	great tit

datetime	species
1/1/’25 1:05	owl
1/1/’25 2:45	cat
1/1/’25 3:11	[redacted]
1/1/’25 4:28	mouse
1/1/’25 5:56	rat
1/1/’25 6:32	no animal
1/1/’25 7:44	[redacted]
1/1/’25 8:15	rat
1/1/’25 9:00	cat
1/1/’25 9:23	hedgehog
1/1/’25 11:51	great tit
1/1/’25 13:07	dove
1/1/’25 13:49	[redacted]
1/1/’25 14:28	[redacted]
1/1/’25 14:35	dove
1/1/’25 15:09	cat
1/1/’25 15:56	great tit

datetime	species
1/1/’25 1:05	[redacted]
1/1/’25 2:45	[redacted]
1/1/’25 3:11	fox
1/1/’25 4:28	mouse
1/1/’25 5:56	rat
1/1/’25 6:32	[redacted]
1/1/’25 7:44	[redacted]
1/1/’25 8:15	[redacted]
1/1/’25 9:00	cat
1/1/’25 9:23	hedgehog
1/1/’25 11:51	great tit
1/1/’25 13:07	dove
1/1/’25 13:49	finch
1/1/’25 14:28	dove
1/1/’25 14:35	[redacted]
1/1/’25 15:09	cat
1/1/’25 15:56	great tit

datetime	species
1/1/’25 1:05	owl
1/1/’25 2:45	cat
1/1/’25 3:11	fox
1/1/’25 4:28	mouse
1/1/’25 5:56	[redacted]
1/1/’25 6:32	no animal
1/1/’25 7:44	cat
1/1/’25 8:15	rat
1/1/’25 9:00	cat
1/1/’25 9:23	hedgehog
1/1/’25 11:51	great tit
1/1/’25 13:07	[redacted]
1/1/’25 13:49	finch
1/1/’25 14:28	[redacted]
1/1/’25 14:35	[redacted]
1/1/’25 15:09	cat
1/1/’25 15:56	great tit

The camera fails at random times, leading to a distorted picture which the AI software cannot identify.

The camera is much more likely to make unclear pictures at night than during the day.

The AI software is a lot worse in recognizing doves than any of the other animals.

Research Question: which animals are in my garden and when?

Movement sensor triggers a camera to make a picture. Next, AI software is used to identify the animal
Left= the full dataset. Note that the movement sensor may sometimes trigger a picture when there’s no animal. This is not the type of missing data we’re explain here. We can safely remove that record at 6:32 AM.

Other 3 tables have datasets with missing data. In reality, we wouldn’t know what we’re missing if we have a dataset with missing data.
There’s a different type of missingness at play in each of these examples which we’ll get into in the next slides.

Disclaimer: of course, we would be able to manually check unidentified pictures and label them ourselves but let’s imagine that wouldn’t be possible (especially in the last table)

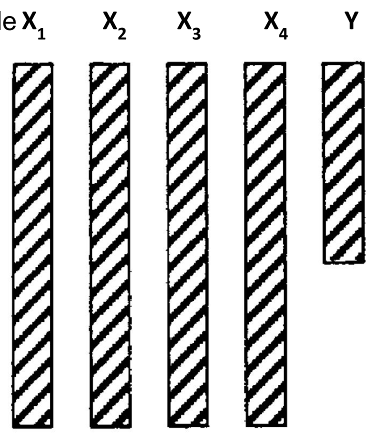
Source photo: https://www.vlaanderen.be/inbo/projecten/camerai-evinbo

Details of the planned statistical analysis –

Missing data

•**Univariate missingness**: confined to a single variable

- eg: lack of germination of a seed
- Because the data were incorrectly recorded
- Missing plot problem



RESEARCH INSTITUTE
NATURE AND FOREST

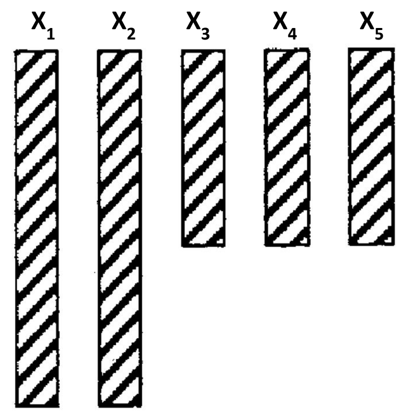
example: suppose our independent variables are:
-the soil
-the temperature
-the species of the seed
-...
the dependent variable is how long it takes for a seed to germinate an get its first two leaves. If a seed does not germinate, we cannot measure growth or how long it takes to germinate so Y will be missing for those.

Details of the planned statistical analysis –

Missing data

•**Multivariate missingness**: confined to ≥ 2 patterns

- Incomplete variables
- Fully observed
- Item nonresponse
- Surveys/questionnaires



RESEARCH INSTITUTE
NATURE AND FOREST

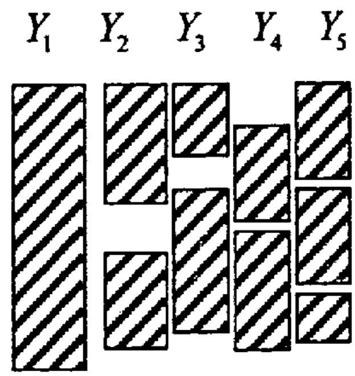
This pattern is typical for surveys where people stop filling in the survey and the next variables are no longer recorded. Often, people might stop filling in the survey at the same questions.

Details of the planned statistical analysis –

Missing data

•**Longitudinal missingness**:

- Drop out
- Do not return
- Monotone missing
- Rarely monotone



RESEARCH INSTITUTE
NATURE AND FOREST

In longitudinal studies, missingness often occurs.
monotone missingness means that, once a subject drops out, they never come back.
However, missingness is rarely purely monotone. They might come back again later.

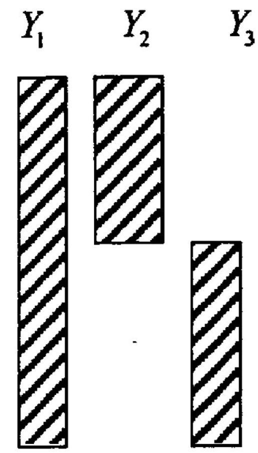
eg: fire salamanders can be identified by the pattern on their belly (or DNA). They may be recaptured many times over the years and age, weight and size are recorded. However, some individuals may be missed in some years/capture campaigns. (= non-monotone missingness).
https://www.inbo.be/inbo/publicaties/populatieonderzoek-en-beheersvisie-voor-de-vuursalamander-in-vlaanderen

eg: the health of old trees with cultural value may be tracked over time. Once the tree dies, though, they are no longer tracked. (=monotone missingness)

Details of the planned statistical analysis –

Missing data

- **File matching missingness**: confined to 2 variables
  - 2 variables never observed together
  - Parameters of association
  - Combining data from two data sources



RESEARCH INSTITUTE
NATURE AND FOREST

Typical example when two datasets are merged and they do not contain the same variables.
eg: table one contains the following variables: location (x & y), and nitrogen level in the air for urban environments
table two contains the following variables: location (x & y), and nitrogen level in the soil for rural environments
We can merge the two datasets and thus obtain more information on nitrogen in the environment but each row will either have nitrogen in the soil or in the soil.

Details of the planned statistical analysis –

Missing data

Missing data can hide true values that are meaningful for analysis

RESEARCH INSTITUTE
NATURE AND FOREST

Details of the planned statistical analysis –

Missing data

- Mechanisms that lead to missingness
  - MCAR

If missingness does not depend on the values of the data Y, missing or observed, the data are called missing completely at random (MCAR)

Example: A handful of measurements are missing because a machine in the laboratory was broken.

Example: The wildlife camera fails at random times, leading to a distorted picture which the AI software cannot identify.

RESEARCH INSTITUTE
NATURE AND FOREST

Missing data

datetime	species
1/1/'25 1:05	owl
1/1/'25 2:45	cat
1/1/'25 3:11	fox
1/1/'25 4:28	mouse
1/1/'25 5:56	rat
1/1/'25 6:32	no animal
1/1/'25 7:44	cat
1/1/'25 8:15	rat
1/1/'25 9:00	cat
1/1/'25 9:23	hedgehog
1/1/'25 11:51	great tit
1/1/'25 13:07	dove
1/1/'25 13:49	finch
1/1/'25 14:28	dove
1/1/'25 14:35	dove
1/1/'25 15:09	cat
1/1/'25 15:56	great tit

datetime	species
1/1/'25 1:05	owl
1/1/'25 2:45	cat
1/1/'25 3:11	[redacted]
1/1/'25 4:28	mouse
1/1/'25 5:56	rat
1/1/'25 6:32	no animal
1/1/'25 7:44	[redacted]
1/1/'25 8:15	rat
1/1/'25 9:00	cat
1/1/'25 9:23	hedgehog
1/1/'25 11:51	great tit
1/1/'25 13:07	dove
1/1/'25 13:49	[redacted]
1/1/'25 14:28	[redacted]
1/1/'25 14:35	dove
1/1/'25 15:09	cat
1/1/'25 15:56	great tit

The camera fails at random times, leading to a distorted picture which the AI software cannot identify.

There doesn't seem to be a real pattern to the missingness.

datetime	species	missing (R)
1/1/'25 1:05	owl	1
1/1/'25 2:45	cat	1
1/1/'25 3:11	[redacted]	0
1/1/'25 4:28	mouse	1
1/1/'25 5:56	rat	1
1/1/'25 6:32	no animal	1
1/1/'25 7:44	[redacted]	0
1/1/'25 8:15	rat	1
1/1/'25 9:00	cat	1
1/1/'25 9:23	hedgehog	1
1/1/'25 11:51	great tit	1
1/1/'25 13:07	dove	1
1/1/'25 13:49	[redacted]	0
1/1/'25 14:28	[redacted]	0
1/1/'25 14:35	dove	1
1/1/'25 15:09	cat	1
1/1/'25 15:56	great tit	1

RESEARCH INSTITUTE
NATURE AND FOREST

We can add a missingness indicator, typically called “R”, which is 0 if the species is missing and 1 otherwise.

Details of the planned statistical analysis –

Missing data

•Mechanisms that lead to missingness

- MCAR

A complete case analysis is allowed under the MCAR assumption

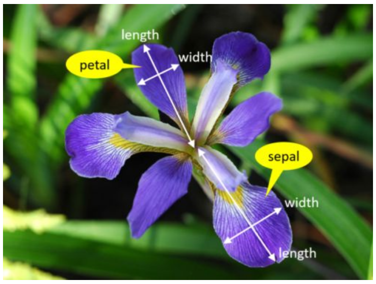
$$E[Y|X, R = 1] = E[Y|X] \text{ and } E[Y|R=1] = E[Y]$$

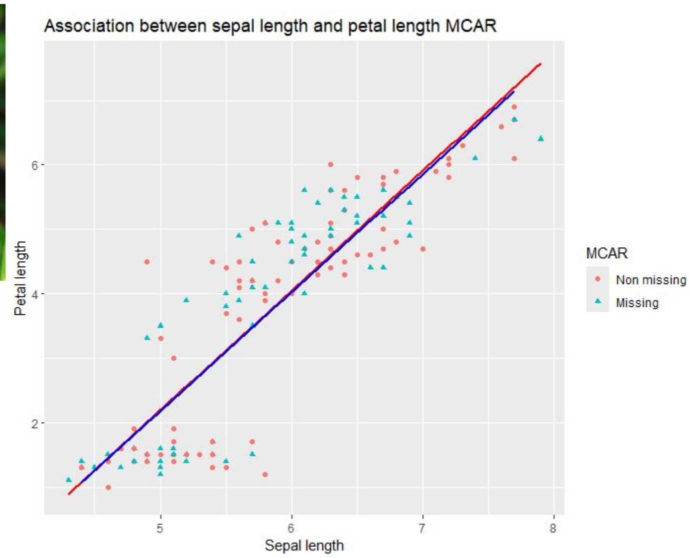
R is the missingness process.
Variable *j* is missing for respondent *i* if *Rij*=0

RESEARCH INSTITUTE
NATURE AND FOREST

E[Y|X, R=1] = The expected value of Y given X, only for the data without missingness (ie if we would do a complete case analysis).
E[Y|X] = The expected value of Y given X.

if E[Y|X, R=1] = E[Y|X] and E[Y|R=1] = E[Y], a complete case analysis will give us unbiased results.





dataset	E[Y]
Full	3,758
MCAR	3.757951

RESEARCH INSTITUTE
NATURE AND FOREST

The iris dataset is a well-known dataset in R.

In this simulated dataset, observations are missing completely at random
If we'd calculate the average petal length on the full dataset (which is unknown in reality) or only on the non-missing data points (= complete case analysis), we get (almost) the same results.

Details of the planned statistical analysis – Missing data

- Mechanisms that lead to missingness
  - MAR

The probability of a value being missing is independent of the unobserved variables conditional on the observed variables

Example: individuals with higher stress levels are less likely to report their blood pressure. Missingness in blood pressure is related to an observed variable (stress)

Example: The camera is much more likely to make unclear pictures at night than during the day.

RESEARCH INSTITUTE
NATURE AND FOREST

“independent of the unobserved variables conditional on the observed variables” means that the observed variables can explain the missingness very well already.

Missing data

datetime	species
1/1/'25 1:05	owl
1/1/'25 2:45	cat
1/1/'25 3:11	fox
1/1/'25 4:28	mouse
1/1/'25 5:56	rat
1/1/'25 6:32	no animal
1/1/'25 7:44	cat
1/1/'25 8:15	rat
1/1/'25 9:00	cat
1/1/'25 9:23	hedgehog
1/1/'25 11:51	great tit
1/1/'25 13:07	dove
1/1/'25 13:49	finch
1/1/'25 14:28	dove
1/1/'25 14:35	dove
1/1/'25 15:09	cat
1/1/'25 15:56	great tit

datetime	species
1/1/'25 1:05	owl
1/1/'25 2:45	cat
1/1/'25 3:11	[redacted]
1/1/'25 4:28	mouse
1/1/'25 5:56	rat
1/1/'25 6:32	no animal
1/1/'25 7:44	[redacted]
1/1/'25 8:15	rat
1/1/'25 9:00	cat
1/1/'25 9:23	hedgehog
1/1/'25 11:51	great tit
1/1/'25 13:07	dove
1/1/'25 13:49	[redacted]
1/1/'25 14:28	[redacted]
1/1/'25 14:35	dove
1/1/'25 15:09	cat
1/1/'25 15:56	great tit

datetime	species
1/1/'25 1:05	[redacted]
1/1/'25 2:45	[redacted]
1/1/'25 3:11	fox
1/1/'25 4:28	mouse
1/1/'25 5:56	rat
1/1/'25 6:32	[redacted]
1/1/'25 7:44	[redacted]
1/1/'25 8:15	[redacted]
1/1/'25 9:00	cat
1/1/'25 9:23	hedgehog
1/1/'25 11:51	great tit
1/1/'25 13:07	dove
1/1/'25 13:49	finch
1/1/'25 14:28	dove
1/1/'25 14:35	[redacted]
1/1/'25 15:09	cat
1/1/'25 15:56	great tit

The camera fails at random times, leading to a distorted picture which the AI software cannot identify.

The camera is much more likely to make unclear pictures at night than during the day.

Here, the time when the picture is taken can explain the missingness quite well.
We could fit a model that predicts missingness based on the time-of-day and that would explain missingness very well:
R_i ~ 1 + time-of-day_i

Details of the planned statistical analysis –

Missing data

- Mechanisms that lead to missingness
  - MAR

Try imputation instead!

Be carefull with a complete case analysis

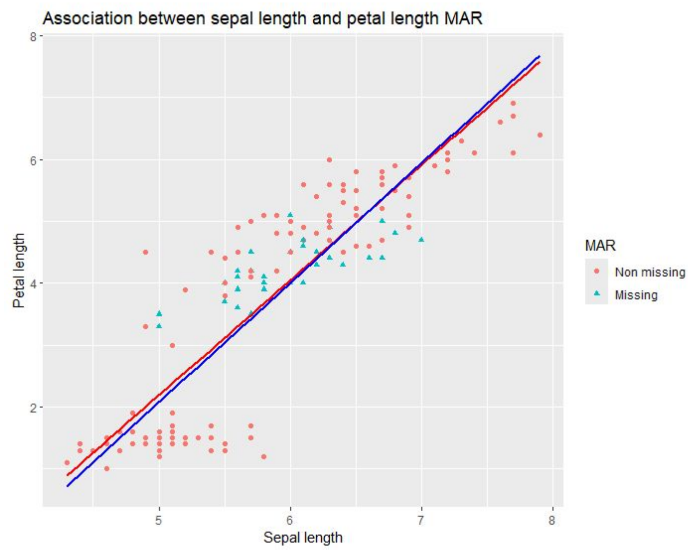
$$E[Y|X, R = 1] = E[Y|X] \text{ and } E[Y|R=1] \neq E[Y]$$

RESEARCH INSTITUTE
NATURE AND FOREST

E[Y|X, R=1] = E[Y|X] meaning that we can predict which animals would occur if we take into account the time-of-day. In other words, for each hour of the day, we can still get a good idea of which animals appear in front of the camera.

E[Y|R=1] != E[Y] means that, if we don't take into account the observed variables (time of day), we do not get a reliable idea of which animals appear in front of the camera. We would overestimate how many day-time animals are in our garden and underestimate how many night-time animals are in our garden.

datas et	E[Y]	$\square_0$	$\square_1$
Full	3,758	-7.10	1.86
MCAR	3.758	-7.12	1.86
MAR	3.68	-7.58	1.93



RESEARCH INSTITUTE
NATURE AND FOREST

Here, we simulated a missingness pattern under MAR; the missingness is more likely for average petal lengths.

Suppose we fit a model where we aim to predict the petal length (Y) based on the sepal length (X)
The blue graph shows the relationship between sepal and petal length with MAR. We see that the model still fits pretty welland the estimated beta’s are similar.
However, the mean petal length (E[Y]) differs quite substantially from the real average petal length.

Details of the planned statistical analysis –

Missing data

•Mechanisms that lead to missingness

- NMAR

If the mechanism depends on Y_i the mechanism is NMAR since it depends on Y_i that are missing, assuming that there are some.

Example: individuals with very severe depression might be less likely to participate or answer questions about their depression, due to shame or stigma. The missingness is related to the unobserved severity level

Example: The AI software is a lot worse in recognizing doves than any of the other animals.

RESEARCH INSTITUTE
NATURE AND FOREST

We wouldn't know that we are mainly missing doves in our dataset

Missing data

datetime	species
1/1/'25 1:05	owl
1/1/'25 2:45	cat
1/1/'25 3:11	fox
1/1/'25 4:28	mouse
1/1/'25 5:56	rat
1/1/'25 6:32	no animal
1/1/'25 7:44	cat
1/1/'25 8:15	rat
1/1/'25 9:00	cat
1/1/'25 9:23	hedgehog
1/1/'25 11:51	great tit
1/1/'25 13:07	dove
1/1/'25 13:49	finch
1/1/'25 14:28	dove
1/1/'25 14:35	dove
1/1/'25 15:09	cat
1/1/'25 15:56	great tit

datetime	species
1/1/'25 1:05	owl
1/1/'25 2:45	cat
1/1/'25 3:11	[redacted]
1/1/'25 4:28	mouse
1/1/'25 5:56	rat
1/1/'25 6:32	no animal
1/1/'25 7:44	[redacted]
1/1/'25 8:15	rat
1/1/'25 9:00	cat
1/1/'25 9:23	hedgehog
1/1/'25 11:51	great tit
1/1/'25 13:07	dove
1/1/'25 13:49	[redacted]
1/1/'25 14:28	[redacted]
1/1/'25 14:35	dove
1/1/'25 15:09	cat
1/1/'25 15:56	great tit

datetime	species
1/1/'25 1:05	[redacted]
1/1/'25 2:45	[redacted]
1/1/'25 3:11	fox
1/1/'25 4:28	mouse
1/1/'25 5:56	rat
1/1/'25 6:32	[redacted]
1/1/'25 7:44	[redacted]
1/1/'25 8:15	[redacted]
1/1/'25 9:00	cat
1/1/'25 9:23	hedgehog
1/1/'25 11:51	great tit
1/1/'25 13:07	dove
1/1/'25 13:49	finch
1/1/'25 14:28	dove
1/1/'25 14:35	[redacted]
1/1/'25 15:09	cat
1/1/'25 15:56	great tit

datetime	species
1/1/'25 1:05	owl
1/1/'25 2:45	cat
1/1/'25 3:11	fox
1/1/'25 4:28	mouse
1/1/'25 5:56	[redacted]
1/1/'25 6:32	no animal
1/1/'25 7:44	cat
1/1/'25 8:15	rat
1/1/'25 9:00	cat
1/1/'25 9:23	hedgehog
1/1/'25 11:51	great tit
1/1/'25 13:07	[redacted]
1/1/'25 13:49	finch
1/1/'25 14:28	[redacted]
1/1/'25 14:35	[redacted]
1/1/'25 15:09	cat
1/1/'25 15:56	great tit

The camera fails at random times, leading to a distorted picture which the AI software cannot identify.

The camera is much more likely to make unclear pictures at night than during the day.

The AI software is a lot worse in recognizing doves than any of the other animals.

We might think that we can still explain the missingness based on the time of day since doves are mainly active duringthe day so missingness will occur mostly during the day.
However, in reality we are missing mostly doves and we wouldn't know this.

(expert judgement may help since an expert may notice that we observed a lot less doves than expected. In reality, we might also be able to look at the pictures that the AI software couldn't identify, conclude the AI algorithm doesn't recognize doves, and improve the AI algorithm with a bigger training set with doves)

Details of the planned statistical analysis –

Missing data

- Mechanisms that lead to missingness

  - NMAR

Be carefull with a complete case analysis

We won't have enough information for imputation either!

$$E[Y|X, R = 1] \neq E[Y|X] \text{ and } E[Y|R=1] \neq E[Y]$$

Or

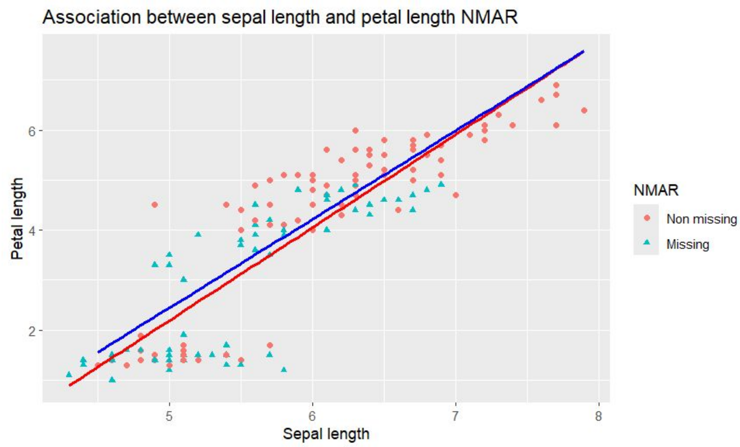
$$E[Y|X, R = 1] = E[Y|X] \text{ and } E[Y|R=1] \neq E[Y]$$

RESEARCH INSTITUTE
NATURE AND FOREST

Our model may or may not give us reliable estimates but E[Y] will always be biased.

In the camera trap example, we will always underestimate the number of doves in our garden.

datase t	E[Y]	□$_0$	□$_1$
Full	3,758	-7.10	1.86
MCAR	3.758	-7.12	1.86
MAR	3.68	-7.58	1.93
NMAR	4.22	-6.67	1.80

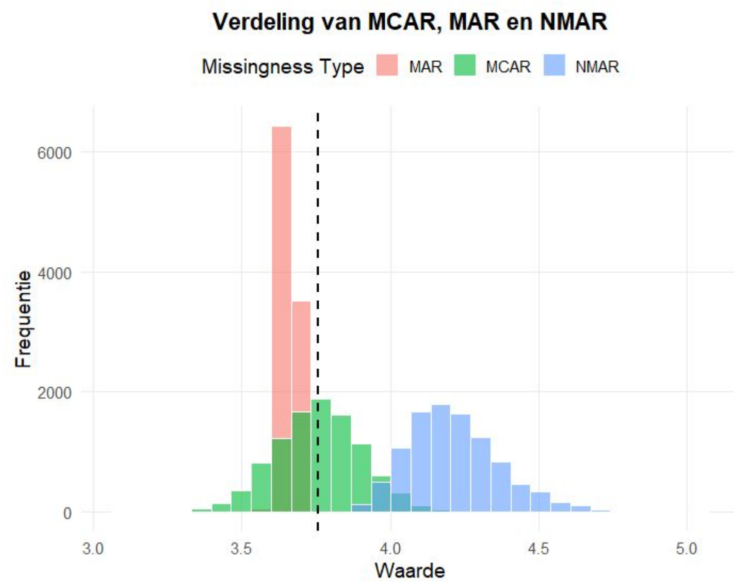


RESEARCH INSTITUTE
NATURE AND FOREST

Here, MNAR is simulated such that it depends on a non-observed variable (eg the species)

Both E[Y] and the beta’s are highly impacted in complete-case analysis.

dataset	E[Y]
Full	3,758
MCAR	3.757
MAR	3.68
NMAR	4.22



RESEARCH INSTITUTE
NATURE AND FOREST

Missing data

datetime	species
1/1/'25 1:05	owl
1/1/'25 2:45	cat
1/1/'25 3:11	fox
1/1/'25 4:28	mouse
1/1/'25 5:56	rat
1/1/'25 6:32	no animal
1/1/'25 7:44	cat
1/1/'25 8:15	rat
1/1/'25 9:00	cat
1/1/'25 9:23	hedgehog
1/1/'25 11:51	great tit
1/1/'25 13:07	dove
1/1/'25 13:49	finch
1/1/'25 14:28	dove
1/1/'25 14:35	dove
1/1/'25 15:09	cat
1/1/'25 15:56	great tit

datetime	species
1/1/'25 1:05	owl
1/1/'25 2:45	cat
1/1/'25 3:11	[redacted]
1/1/'25 4:28	mouse
1/1/'25 5:56	rat
1/1/'25 6:32	no animal
1/1/'25 7:44	[redacted]
1/1/'25 8:15	rat
1/1/'25 9:00	cat
1/1/'25 9:23	hedgehog
1/1/'25 11:51	great tit
1/1/'25 13:07	dove
1/1/'25 13:49	[redacted]
1/1/'25 14:28	[redacted]
1/1/'25 14:35	dove
1/1/'25 15:09	cat
1/1/'25 15:56	great tit

datetime	species
1/1/'25 1:05	[redacted]
1/1/'25 2:45	[redacted]
1/1/'25 3:11	fox
1/1/'25 4:28	mouse
1/1/'25 5:56	rat
1/1/'25 6:32	[redacted]
1/1/'25 7:44	[redacted]
1/1/'25 8:15	[redacted]
1/1/'25 9:00	cat
1/1/'25 9:23	hedgehog
1/1/'25 11:51	great tit
1/1/'25 13:07	dove
1/1/'25 13:49	finch
1/1/'25 14:28	dove
1/1/'25 14:35	[redacted]
1/1/'25 15:09	cat
1/1/'25 15:56	great tit

datetime	species
1/1/'25 1:05	owl
1/1/'25 2:45	cat
1/1/'25 3:11	fox
1/1/'25 4:28	mouse
1/1/'25 5:56	[redacted]
1/1/'25 6:32	no animal
1/1/'25 7:44	cat
1/1/'25 8:15	rat
1/1/'25 9:00	cat
1/1/'25 9:23	hedgehog
1/1/'25 11:51	great tit
1/1/'25 13:07	[redacted]
1/1/'25 13:49	finch
1/1/'25 14:28	[redacted]
1/1/'25 14:35	[redacted]
1/1/'25 15:09	cat
1/1/'25 15:56	great tit

The camera fails at random times, leading to a distorted picture which the AI software cannot identify.

The camera is much more likely to make unclear pictures at night than during the day.

The AI software is a lot worse in recognizing doves than any of the other animals.

Movement sensor triggers a camera to make a picture -> AI software to identify animal
RQ: which animals are in my garden and when?
Left= the full dataset. Note that the movement sensor may sometimes trigger a picture when there's no animal. This is not the type of missing data we're explain here. We can safely remove that record at 6:32 AM.

Other 3 tables have datasets with missing data. In reality, we wouldn't know what we're missing if we have a dataset with missing data.
There's a different type of missingness at play in each of these examples which we'll get into in the next slides.

Disclaimer: of course, we would be able to manually check unidentified pictures and label them ourselves but let's imagine that wouldn't be possible (especially in the last table)

Source photo: https://www.vlaanderen.be/inbo/projecten/camerai-evinbo

Details of the planned statistical analysis –

Missing data

MCAR, MAR and NMAR are untestable assumption.

RESEARCH INSTITUTE
NATURE AND FOREST

We can almost never know for sure which type of missingness we have.

Details of the planned statistical analysis –

Missing data

•Complete case analysis

- Only people/animals that responded
- Missingness in Y **and missingness in X**
- Change of target population
- Inference
- Might need bigger sample size

RESEARCH INSTITUTE
NATURE AND FOREST

complete case analysis is usually inappropriate, since the investigator is usually interested in making inferences about the entire target population, rather than the portion of the target population that would provide responses on all relevant variables in the analysis. Our aim is to describe a collection of techniques that are more generally appropriate than complete-case analysis when missing entries in the data set mask underlying values.

Details of the planned statistical analysis –

Missing data

- Remedial measures
  - Avoid missing data
  - Regression mean imputation
  - Stochastic regression imputation
  - Multiple regression imputation
  - Multiple Imputation via Chained Equations (MICE)
  - Sensitivity analysis

RESEARCH INSTITUTE
NATURE AND FOREST

Challenges: MAR, MCAR, or NMAR?

- People with very low incomes tend to not fill in their income in a survey.
- A field worker accidentally forgets to fill in one of the data fields on his meetnetten-app.
- A bird radar records bird tracks and weather information 24/7. The radar often fails to identify birds when it's raining.
- In a tobacco/smoking study, younger participants report their values less often
- In a study about growth in trees on different soils, trees that die will not be followed up.

RESEARCH INSTITUTE
NATURE AND FOREST

Challenges: MAR, MCAR, or MNAR?

- People with very low incomes tend to not fill in their income in a survey.

MNAR if we cannot discern low incomes from any other recorded variable

- A field worker accidentally forgets to fill in one of the data fields on his meetnetten-app.

MCAR

- A bird radar records bird tracks and weather information 24/7. The radar often fails to identify birds when it's raining.

MAR; we can explain missingness with a recorded variable: whether it is raining or not.

- In a tobacco/smoking study, younger participants report their values less often

MAR if age is a recorded variable: We can explain the missingness by the age.

- In a study about growth in trees on different soils, trees that die will not be followed up.

MAR/MNAR?: If we can explain the death by other recorded variables, it is MAR. If we cannot explain death and it are the trees that grow less well that die, we'd have MCAR?

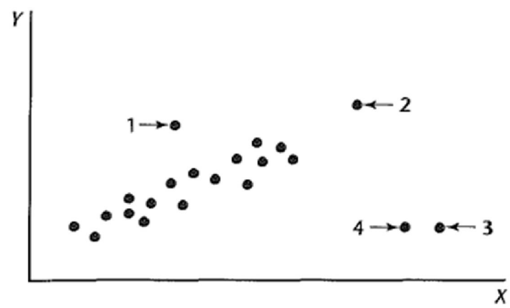
RESEARCH INSTITUTE
NATURE AND FOREST

It's not always straightforward to know which of the missingness patterns we have since we don't know much about the studies.
Some of these solutions are surely debateable.

Outlier detection

Details of the planned statistical analysis –

Outliers and influential observations



RESEARCH INSTITUTE
NATURE AND FOREST

[Make your own dataset, add and remove outliers and see their effect]

1 is not really an outlier for x or y separately (it is within the expected range). But, given x, it is unexpectedly high and it will influence the regression.
3 and 4 are outliers in the X-direction and they will influence the regression.
2 is an outlier in the X and Y-direction but it might fit our model; its influence on the regression will be minimal.

Expert knowledge is needed to assess whether the “outliers” may be true data or not. Is there a second cluster of data around 3 and 4 which is undersampled?

Details of the planned statistical analysis –

Outliers and influential observations

Outliers in Y variable

- Residuals
- Standardized residuals
- Studentized residuals
- Deleted residuals
- Studentized deleted residuals
- Hat matrix and Leverage

Outliers in X variable

- Hat matrix and Leverage

RESEARCH INSTITUTE
NATURE AND FOREST

How can we identify outliers?

Not all outliers are influential. We will determine which are influential by seeing what happens to the fitted values when that observation is deleted.

Details of the planned statistical analysis –

Outliers and influential observations

Are the outliers influential observations?

- On a single fitted value: DFFIT
- On the entire model/all fitted values: Cook’s Distance
- On the each of the model parameters ($\beta 0$, $\beta 1$, …): DFBETA’s

RESEARCH INSTITUTE
NATURE AND FOREST

DFFIT and cook’s Distance will often give similar conclusions.
If you care most about the “stability” of the regressor coefficients, you might run DFBETA

Details of the planned statistical analysis –

Outliers and influential observations

Are the outliers influential observations?

- Redo your analysis, excluding a suspicious case
- Sensitivity analysis
- Conclusions should not depend on 1 (or a few) case(s)!
- Never forget your good judgment

Sample size

Sample size calculation

- Determine the number of experimental units
- Balance cost experimental unit and **increase in precision** per additional unit
- More of the same does not lead to more significance
- Non controllable constraints on sample size

RESEARCH INSTITUTE
NATURE AND FOREST

Explore power, sample size, population parameters

Some consideration virtually always has to be given to the number of experimental units to be used and, where subsampling of units is employed, to the number of repeat observations per unit.
The issue is then most commonly to decide whether the resources are sufficient to yield enough precision to justify proceeding at all.

Can we afford a big enough sample size to answer our questions? If not, can we answer a subset of the questions that is still worthwhile with the available budget?

Sample size calculation

- Clearly defined statistical hypothesis
- The design
- Type-I error
- Type-II error
- Smallest relative difference (effectsize)

RESEARCH INSTITUTE
NATURE AND FOREST

voor meetnetten:
https://www.vlaanderen.be/publicaties/ontwerp-en-evaluatie-van-meetnetten-voor-het-milieu-en-natuurbeleid-leidraad-voor-de-meetnetontwerper

Sample size calculation

Assumptions

- Distribution of the data
- Design

Methods

- Exact solutions
- Asymptotc solutions
- Simulations

RESEARCH INSTITUTE
NATURE AND FOREST

We often need to make some assumptions on the distribution of the data (mean/variance/…). It may be possible to set up an interim analysis (see intervention study of group sequential design in part 1 of module1) to get a better estimate of these parameters.
For simple designs, we can get exact or asymptotic solutions. For more complex designs, we often need to refer to simulations.

Multiple testing

Multiple testing

- Often more than one hypothesis of interest
  - Multiple doses
  - Multiple endpoints
  - Multiple populations
  - Multiple looks

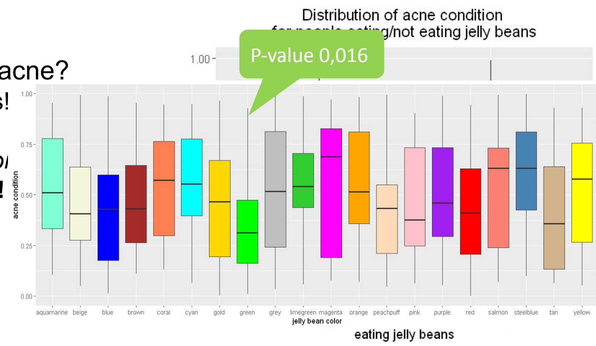
Try to avoid inflation of the type-I error

RESEARCH INSTITUTE
NATURE AND FOREST

If we want to test more than one hypothesis, we need to account for these tests to avoid an inflation of the type 1 error.

Remember the jelly beans example

- Does eating jelly beans cause acne?
  - Part 1: 500 participants
    - Acne_condition: uniformly distributed between 0 and 1
    - Jb_eating: 0 / 1 variable (No/ Yes); suppose 90% eats jelly beans
    - *Acne_condition ~Jb_eating*
    - P-value 0,6688
  - Part 2: which color causes acne?
    - Jb_color: 20 different colors!
    - t-test for each color
      - *Acne_condition ~Jb_color*
    - **Green color is significant!**



RESEARCH INSTITUTE
NATURE AND FOREST

Source: https://datascience.stanford.edu/news/data-snooping

Multiple testing

Control the family-wise error rate (FWER)

Example:

- **We test 20 different independent true null hypotheses, what is the probability the at least one hypothesis will be significant at a 5% level?**

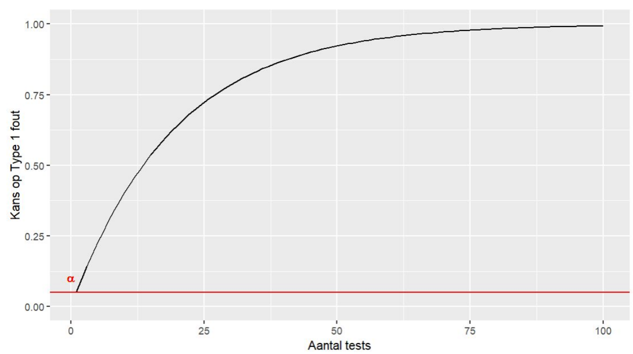
A) 5%-10%
B) 10%-25%
C) 25%-50%
D) >50%

# hypotheses	1	5	10	20
$1 - 0{,}95^m$	5%	23%	40%	64%

RESEARCH INSTITUTE
NATURE AND FOREST

the type I error for the whole study, not just for an individual test

Multiple testing



RESEARCH INSTITUTE
NATURE AND FOREST

See also: https://istats.shinyapps.io/ExploreCoverage/

Code:

```
library(latex2exp)
library(tidyverse)
tibble(nb_tests = seq_len(100)) %>%
    mutate(type1 = 1-0.95^nb_tests) %>%
     ggplot(aes(x=nb_tests, y = type1)) +
    geom_line() +
    scale_y_continuous(limits = c(0,1)) +
    geom_hline(yintercept = 0.05, color = "red") +
    geom_text(label = TeX("$\\alpha$"), color = "red", x = 0, y = 0.1)+
    xlab("Aantal tests") +
    ylab("Kans op Type 1 fout")
```

Methods to account for multiple testing

Control the family-wise error rate:

- Bonferroni method
- The Holm Procedure
- The Hochberg Procedure
- Fixed sequence Procedure
- The Fallback Procedure

Before looking at the data, randomly divide data into two parts:

- One to be for used for discovery purposes (generating hypotheses)
- the other to be used for confirmatory purposes (testing hypotheses).

Report your results honestly and carefully

RESEARCH INSTITUTE
NATURE AND FOREST

Bron: https://datascience.stanford.edu/news/data-snooping

Bonferroni is the most well known but also the strictest method to control the FWER.

If you are not sure what you'd like to investigate in your dataset, avoid type 1 error inflation by using only part of your data for hypothesis building. Once you've established which hypotheses you'd like to test, test them in the other part of the data which you have never looked at before.

If you tested hypotheses which were not significant, be honest to report those tests as well and correct for them accordingly.

Measurement error

Measurement error

It seems intuitive that producing a result under challenging circumstances makes it all the more impressive. If you learned that a friend had run a mile in 5 minutes, you would be respectful; if you learned she had done it while carrying a heavy backpack, you would be awed. The obvious inference is that she would have been even faster without the backpack.

Eric Loken & Andrew Gelman. "Measurement error and the replication crisis". Science, Feb 2023.

Measurement error

But should the same intuition always be applied to research findings?

Eric Loken & Andrew Gelman. "Measurement error and the replication crisis". Science, Feb 2023.

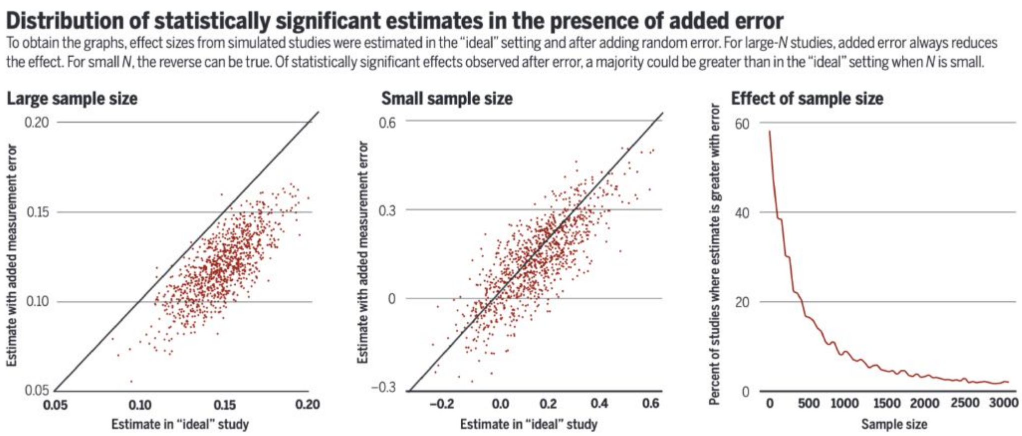
RESEARCH INSTITUTE
NATURE AND FOREST

If significance is found *with* measurement error, can we assume that the results would have been even more significant / impressive *without* measurement error?

Measurement error

Should we assume that if statistical significance is achieved in the presence of measurement error, the associated effects would have been stronger without noise? We caution against the fallacy of assuming that that which does not kill statistical significance makes it stronger

Eric Loken & Andrew Gelman. "Measurement error and the replication crisis". Science, Feb 2023.



Eric Loken & Andrew Gelman. "Measurement error and the replication crisis". Science, Feb 2023.

RESEARCH INSTITUTE
NATURE AND FOREST

NO! Although this idea used to be supported, we now know that measurement error may also exaggerate results.
https://www.science.org/doi/10.1126/science.aal3618
In noisy research settings, poor measurement can contribute to exaggerated estimates of effect size.

References

Fisher, R. A. (1951). *The Design of Experiments* (2nd ed.). Edinburgh: Oliver & Boyd.

Morgenstern, H. (1993). *Principles of Study Design in Environmental Epidemiology*. Oxford: Oxford University Press.

Du Prel, J. B., Röhrig, B., Wachtlin, D., & Blettner, M. (2009). Types of study in medical research: Part 3 of a series on evaluation of scientific publications. *Deutsches Ärzteblatt International, 106*(15), 262–268. https://doi.org/10.3238/arztebl.2009.0262

Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2004). *Applied Linear Statistical Models* (5th ed.). Boston, MA: McGraw-Hill/Irwin.

Thompson, S. K. (2012). *Sampling* (3rd ed.). Hoboken, NJ: John Wiley & Sons.

Montgomery, D. C. (2017). *Design and Analysis of Experiments* (3rd ed.). Hoboken, NJ: John Wiley & Sons.

Chapman & Hall / CRC. *Design and Analysis of Experiments with R*. (Controleer exacte auteur en jaar; vermoedelijk ca. 2015)

INBO. (2008). *Ontwerp en evaluatie van meetnetten voor het milieu- en natuurbeleid*. Brussel: Instituut voor Natuur- en Bosonderzoek. https://www.vlaanderen.be/inbo/publicaties/ontwerp-en-evaluatie-van-meetnetten-voor-het-milieu-en-natuurbeleid

Duchateau, L., Dockx, R., Goethals, K., Vynck, M., Vangroenweghe, F., & Burvenich, C. P. G. (2024). Treatment randomisation at animal or pen level? Statistical analysis should follow the randomisation pattern! *Laboratory Animals*. https://journals.sagepub.com/doi/10.1177/00236772241247274

European Commission / ICH. (2019). *ICH E9 Guideline: Statistical Principles for Clinical Trials*. Geneva: ICH.

RESEARCH INSTITUTE
NATURE AND FOREST

Part 3: some StAP examples

Statistical analysis plan (StAP): module 1

- Overview -

PART I

A) Hypothesis

B) Statistical testing

C) Design of the study

D) Conduct of the study

PART II

E) Details of the planned statistical analysis

- Estimand
- Missing data
- Outlier detection
- Sample size
- Multiple testing

PART III

F) Some StAP examples

RESEARCH INSTITUTE
NATURE AND FOREST

Questions

- Are all questions in the template clear?
- Are all questions answered clearly enough?
  - Would you give more or less details?
- Would it be feasible to fill in such a template for your own project?
- Is any important information missing?

RESEARCH INSTITUTE
NATURE AND FOREST

Questions to keep in mind when going over the examples.
Feel free to add comments and remarks to the examples.

Questions

- Are all questions in the template clear?
- Are all questions answered clearly enough?
  - Would you give more or less details?
- Would it be feasible to fill in such a template for your own project?
- Is any important information missing?
- Procedure:
  - first draft (researcher)
  - second draft with help/feedback of BMK
  - review of the StAP (other BMK team member)
  - publish finalized StAP
- Overlap with DMPs?

RESEARCH INSTITUTE
NATURE AND FOREST

Flanders
State of the Art

Q&A