Flanders
State of the Art

Statistical Analysis Plan (StAP): Module 1

Teachers: Robrecht Dockx and Raisa Carmen

Team BMK

10/1/2025

Resistent en overal: moeten we ons zorgen maken om de bruine rat?

INBO scientists investigate many different species, topics, processes,...

www.vlaanderen.be/inbo

Vlinderaantallen uitzonderlijk laag: 'Ook een belangrijke indicator voor andere insecten'

RESEARCH INSTITUTE
NATURE AND FOREST

Inbo researchers are asked a lot of questions about nature an Flanders and aim to answer these questions in a scientific way.

raag

vrt nws

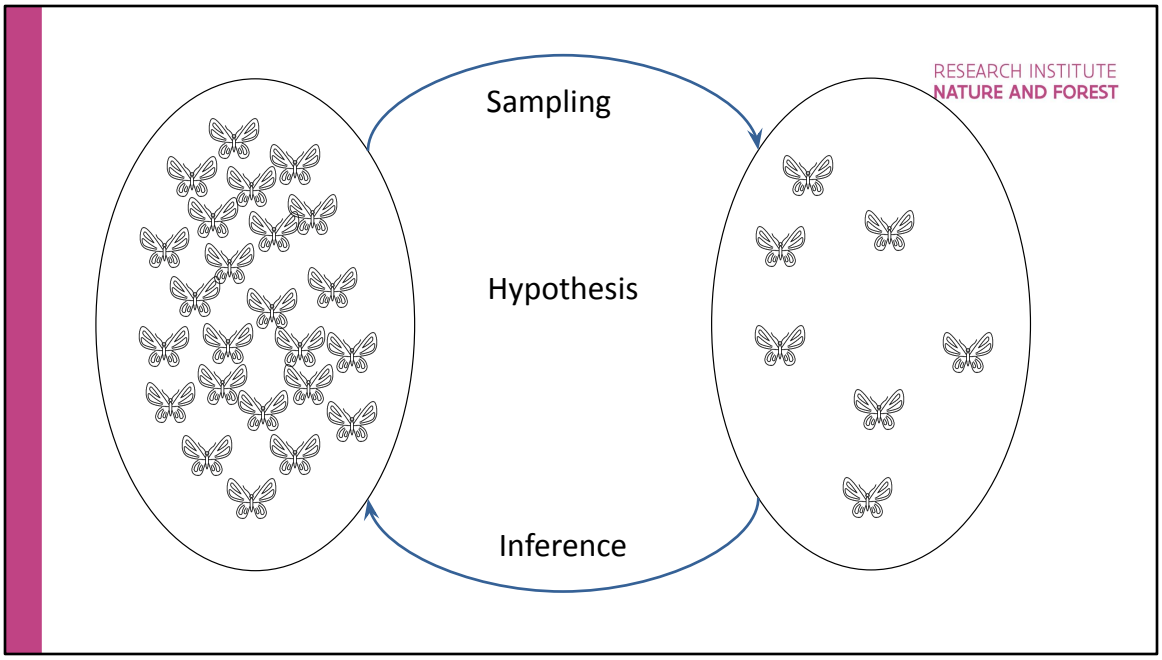
RESEARCH INSTITUTE
NATURE AND FOREST

Gesjoemel bij patrijzentellingen? "Statistisch zeer onwaarschijnlijk wat sommige jagers geteld hebben"

Inbo researchers are asked a lot of questions about nature an Flanders and aim to answer these questions in a scientific way.

Invasieve kreeften overspoelen Belgische wateren: ‘Onder water speelt zich een drama af’

Waarom u helaas steeds minder vogels hoort zingen: ‘Het wordt op steeds meer plaatsen onnatuurlijk stil’



Since we can rarely study the whole population, most INBO research follows the cycle shown above.
To answer a hypothesis about the population, we first take a representative sample of the population.
This sample will be studied or subjected to some tests.
We can then use statistical inference to use what we learn from out sample to make conclusions about the population

Introduction

Types of statistics

Descriptive statistics	Inferential (Inductive) statistics
Describe a certain dataset (sample or full population), for example: •mean •median •minimum/maximum •frequencies •percentages •standard deviation	Analyse a sample to make generalizations or predictions or decisions about a larger population. •Parameter estimates with confidence (credible) intervals •Embedded in probability theory •Requires good quality and thorough data collection

RESEARCH INSTITUTE
NATURE AND FOREST

We will focus on inferential statistics
An interesting historical overview of statistical inference can be found inDavid Freedman's paper from 1999 "From association to causation: some remarks on the history of statistics"

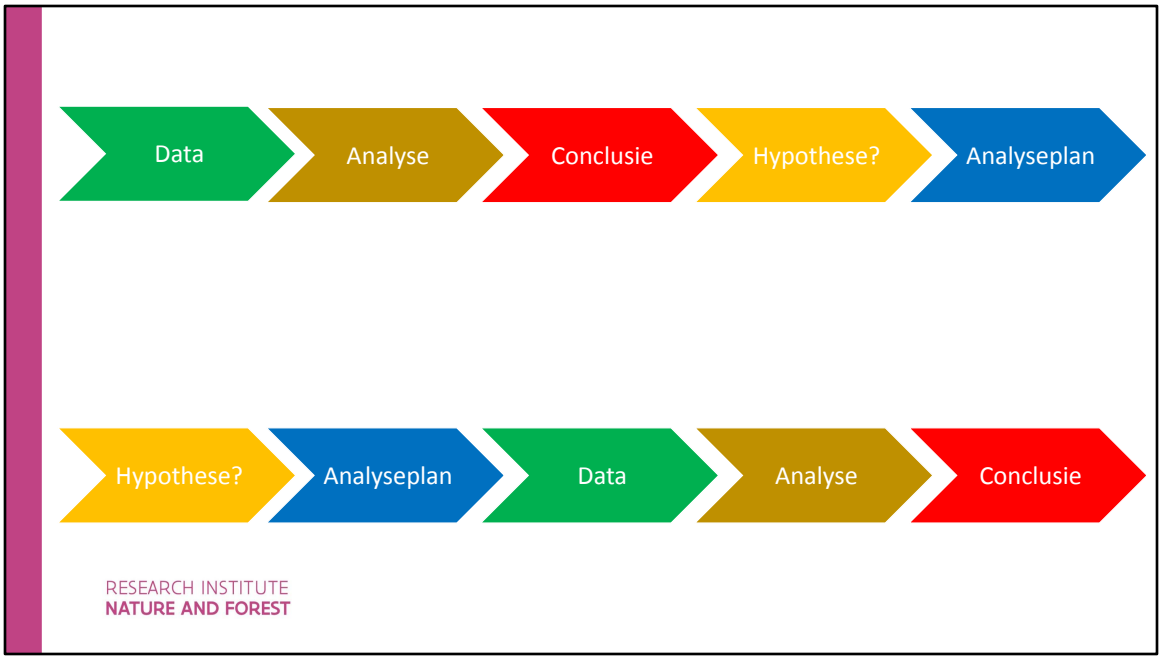
https://projecteuclid.org/journals/statistical-science/volume-14/issue-3/From-association-to-causation--some-remarks-on-the-history/10.1214/ss/1009212409.full
DOI: 10.1214/ss/1009212409

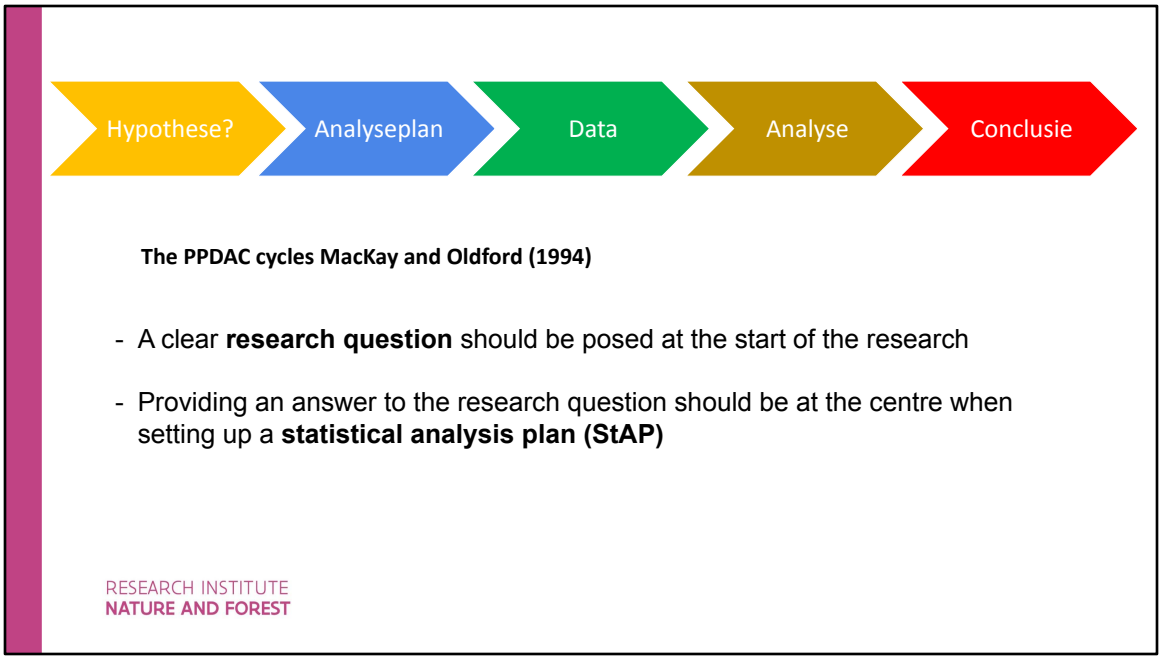
RESEARCH INSTITUTE
NATURE AND FOREST

RESEARCH INSTITUTE
NATURE AND FOREST

Possible consequences

- Problem can't be solved
- Over or underpowered studies
- Data snooping
- No link between the design and the analysis of the study
- Type-III error
- Ethical consequences


Data
Analyse
Conclusie
Hypothese?
Analyseplan
Hypothese?
Analyseplan
Data
Analyse
Conclusie
RESEARCH INSTITUTE
NATURE AND FOREST



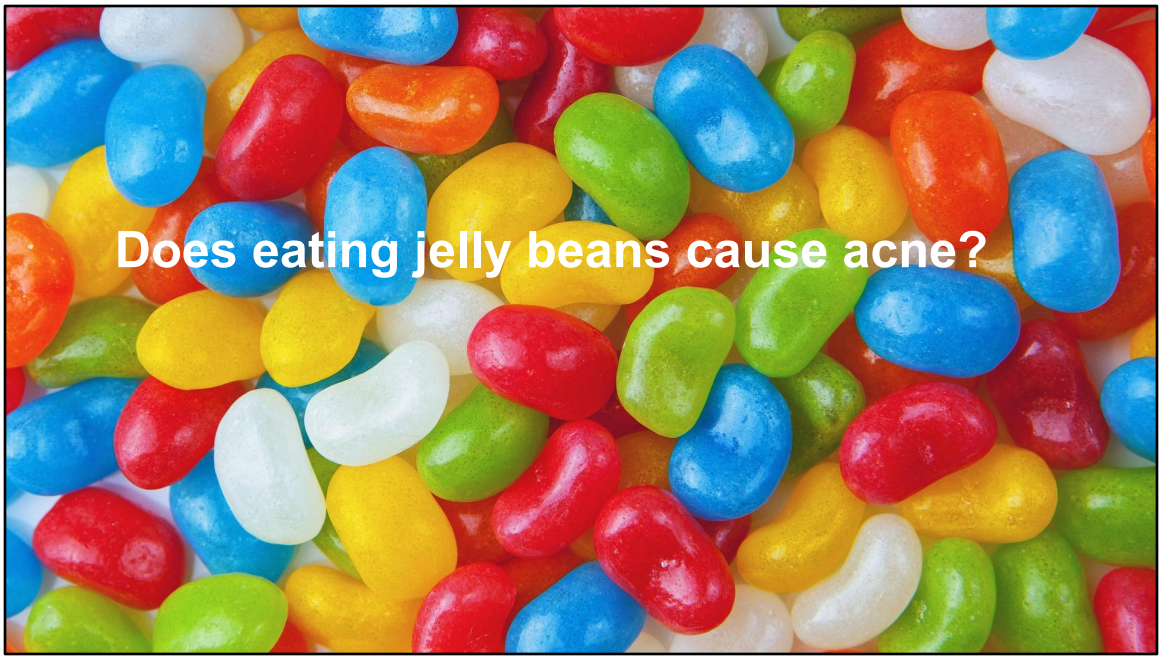
PPDAC = Problem, plan, data, analysis, conclusion

Hypothesis

- Should be set up by the researcher
- Based on past research
- Depicts how two or more variables will interact with one another
- Should be testable

Pitfalls

- Stating the hypothesis after the results are known (**data snooping)**
- Hypothesis testing based on a null hypothesis that is known a priori to be false
- Claiming a non significant result as evidence of no effect or no relationship
- Not reporting non-significant results

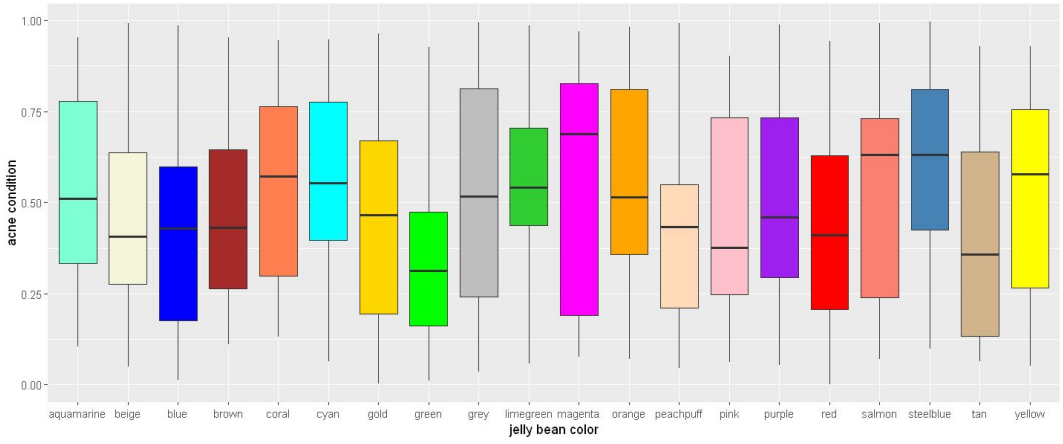


An example of data snooping

- 500 participants
- Acne condition: uniformly distributed between 0 (No acne) and 1 (A lot of acne)
- Jb eating: 0 / 1 variable (No/ Yes)
- Jb colour: 20 colours
- suppose 90% eats jelly beans
- *Acne condition ~ Jb eating*
- P-value 0,67

RESEARCH INSTITUTE
NATURE AND FOREST

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



RESEARCH INSTITUTE
NATURE AND FOREST

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

- Which jelly bean color causes acne?
- Jb_color: 20 different colors!
- t-test for each color
- *Acne condition ~ Jb color*
- **Green color is significant!**
- **p-value = 0.016**

RESEARCH INSTITUTE
NATURE AND FOREST

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

Data snooping involves building you hypothesis based on the data on which you will test the hypothesis. You risk finding patterns that exist by coincidence and concluding there is really something going on which isn't true.
There is a second problem in this example: multiple testing. 20 hypotheses are tested in the data. You should not do this without correcting for multiple testing to avoid inflation of the type 1 error. We'll discuss multiple testing in more detail in the second part of module1

-What if we'd formulated our hypothesis on the color of the jelly bean before we looked at the data?

You're no longer data snooping but you still need to correct for multiple testing. A StAP will also make you think about whether your hypotheses are worth testing. Is it logical that the color would determine the effect on acne? Is there any previous research perhaps on synthetical colorings and acne? You should motivate why you want to test this.

- What if I'm not sure what I want to test?

Split up your data set in two parts: one for hypothesis building and one of hypothesis testing. You should not look at the data for hypothesis testing before you've finished your data exploration and hypothesis building on the

first part of your data

- What if I come up with a new scientific hypothesis based on existing data?

Ideally, you'd set up a new experiment to gather new data, similar to the existing dataset, and you test your hypothesis only in this new dataset.

Data snooping

Data snooping refers to the practice of looking at the data to decide how to analyze it in a way that is not planned in advance. In other words, **performing statistical inference after looking at the data.**

- It can be the result of ignorance.
- It can be the result of unethical and misleading intentions.
- Will inflate the Type I error

If the choice of statistical analysis method becomes **data-dependent**, then the frequentist properties of estimators and hypothesis tests change.

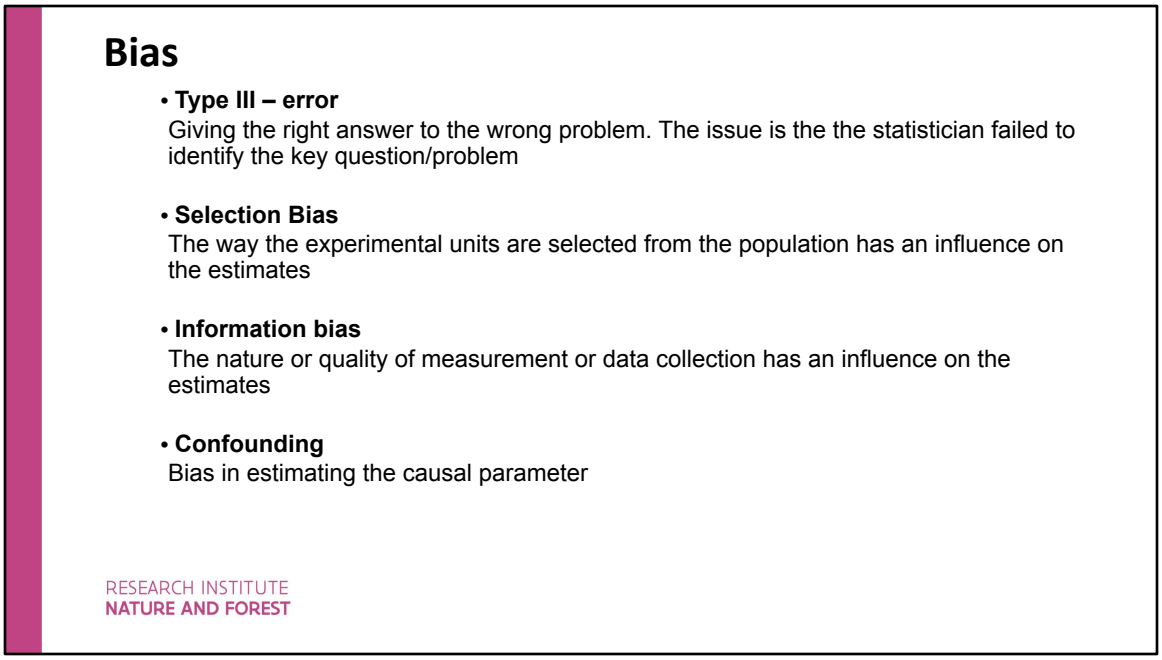
Set up a StAP to prevent data snooping

Statistical analysis plan (StAP)

- Information about the research
- Should be sufficient to reconstruct the setup and analysis
- Assess the appropriateness of the used methods
- Can be altered but should be well documented
- **Helps avoid bias**
- The StAP should be finalised **before** the **data collection** starts

RESEARCH INSTITUTE
NATURE AND FOREST

The “methods used” here should be understood in a broad sens, covering not only the statistical model but also how to deal with intercurrent events, missing data, outliers, etc.



Selection bias

You will get biased results if your sample is not representative for the population you are interested in.

Information bias

- The study does not have a double-blind design—i.e., the researchers know whether a participant is assigned to the control or the experimental group.
- The researchers use different methods to assess outcomes in each group. For example, using medical records for one group and self-report questionnaires for the other when studying disease status.
- The independent variable (e.g., exposure to toxic substances) and/or the dependent variable (e.g., risk of lung cancer) are recorded inaccurately. This can be due to errors in recording an individual's history, different disease definitions, or different diagnostic criteria among experts.
- Instruments for objective measurements (e.g., weight) is not correctly calibrated, results are registered incorrectly, or data are switched during the data entry or data cleaning phase.

Gains

- Money
- Time
- Energy
- Animals

RESEARCH INSTITUTE
NATURE AND FOREST

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

Day 1

PART II

F) Details of the planned statistical analysis

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

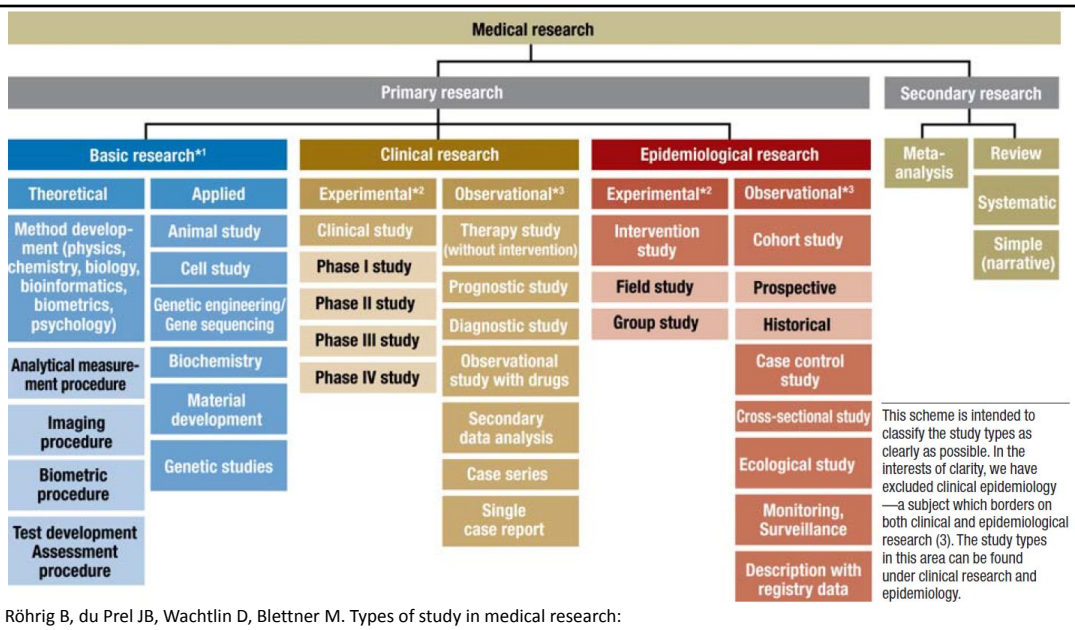
PART III

G) Some StAP examples

Day 2

RESEARCH INSTITUTE
NATURE AND FOREST

Part 1



www.vlaanderen.be/inbo

Röhrig B, du Prel JB, Wachtlin D, Blettner M. Types of study in medical research: part 3 of a series on evaluation of scientific publications. Dtsch Arztebl Int. (2009).

The principles for a StAP are best/first developed for clinical trials but can be applied in various branches of research.

A) Hypothesis

Hypothesis

- Should be set up by the researcher
- Based on past research
- Depicts how two or more variables will interact with one another
- Should be testable

The B@seball project

Background

In schools across Belgium, we investigate the effect of biodiversity in and around the school on the physical and mental health of the students. We also look at the opportunities to reduce health inequalities among children through biodiversity in the school environment.

Hypothesis:

Is there a difference between children in green and grey schools

RESEARCH INSTITUTE
NATURE AND FOREST

•Define the **population** and set well-defined **inclusion-exclusion criteria**.

•Define the **outcome variable**

Predictor variable or independent variable

Outcome variable or dependent variable

RESEARCH INSTITUTE
NATURE AND FOREST

Types of variables

Categorical variables:

Entities are divided into distinct categories

- Binary variable: 2 categories
  e.g., dead or alive, male or female
- Count data: infinite categories
  e.g., Number of dead trees in a forest, number of animals seen on camtrap data
- Nominal variable: +2 categories
  e.g., yellow, green, red, blue, …
- Ordinal variable: logical order in categories
  e.g., fail, pass, cum laude, magna cum laude, summa cum laude

Continuous variables:

Entities get a distinct score

- Interval variable: equal intervals represent equal differences but no true zero (no ratios possible)
  e.g., Fahrenheit temperature
  freezing – boiling of water = 180° apart
- Ratio variable: you can count with the scores
  e.g., 16 on test score = 2 * 8

RESEARCH INSTITUTE
NATURE AND FOREST

- Define the **population** and set well-defined **inclusion-exclusion criteria**.

- Define the **outcome variable**

- Primary endpoint or outcome
  - Relate directly to the main research question.
  - The study is powered on the primary outcome measure.
  - Form the basis for conclusions and influence decision-making.

- Secondary endpoint or outcome
  - Can provide supporting evidence for the main conclusions.
  - Can guide future research.
  - The study is not powered on the secondary outcome variables
  - Be careful drawing conclusions.

RESEARCH INSTITUTE
NATURE AND FOREST

The B@seball project

In schools across Belgium, we investigate the effect of biodiversity in and around the school on the physical and mental health of the students. We also look at the opportunities to reduce health inequalities among children through biodiversity in the school environment.

Aims

To be able to detect a 5% difference in allergy sensitivity between children in green and grey schools. Secondly we are interested in the children's ability to focus in class, activity levels, interest in outdoor play,…

RESEARCH INSTITUTE
NATURE AND FOREST

The B@seball project

In schools across Belgium, we investigate the effect of biodiversity in and around the school on the physical and mental health of the students. We also look at the opportunities to reduce health inequalities among children through biodiversity in the school environment.

Aims

To be able to detect a **5% difference in allergy sensitivity** between children in green and grey schools. Secondly, we are interested in the children's ability to focus in class, activity levels, interest in outdoor play,…

RESEARCH INSTITUTE
NATURE AND FOREST

This is the primary outcome

The B@seball project

In schools across Belgium, we investigate the effect of biodiversity in and around the school on the physical and mental health of the students. We also look at the opportunities to reduce health inequalities among children through biodiversity in the school environment.

Aims

To be able to detect a **5% difference in allergy sensitivity** between children in green and grey schools. Secondly, we are interested in the children's **ability to focus in class, activity levels, interest in outdoor play**,…

RESEARCH INSTITUTE
NATURE AND FOREST

These are secondary outcomes

The B@seball project

In schools across Belgium, we investigate the effect of biodiversity in and around the school on the physical and mental health of the students. We also look at the opportunities to reduce health inequalities among children through biodiversity in the school environment.

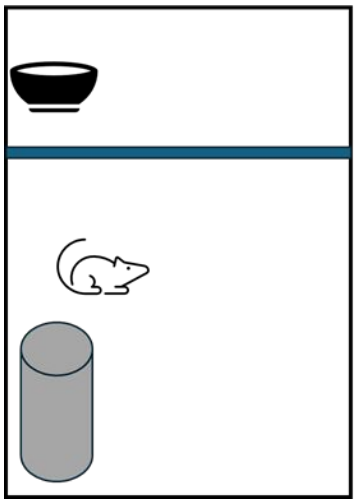
Aims

To be able to detect a 5% difference in allergy sensitivity **(concentration IgG)** between children in green and grey schools. Secondly, we are interested in the children's ability to focus in class, activity levels **(amount of play hours)**, interest in outdoor play,…

RESEARCH INSTITUTE
NATURE AND FOREST

These are examples of the primary and secondary outcome variables
In reality, they ideally wanted to test allergy sensitivity with skin prik tests but reverted used the ISAAC survey instead since it is less invasive.

Another example where an insulation board manufacturer wants to know whether rats can / would damage his insulation boards or not.



RESEARCH INSTITUTE
NATURE AND FOREST

What the experiment will look like

- The **primary** and **secondary endpoints**

Probability that rats gnawed through insulation board after X days and proportion of damage-free insulation boards after X days.

- **Multiple primary outcome or composite endpoints**

Instead of examining probability of gnawing through a board and amount of damage separately, consider damage-free "survival" of insulation boards.

- **Surrogate endpoint**

Faster to measure grams of materials gnawed off after X days than probability of gnawed-through plates after being installed in a building for 5 years.

- **Categorized endpoint**

Gnawed through after X days: binary outcome or time-to-event?

- **Baseline and longitudinal covariates**

Get insulation material type, age and sex at baseline.

RESEARCH INSTITUTE
NATURE AND FOREST

- The **primary** and **secondary outcomes**

Growth of different poplar clones after five years as primary outcome. Straightness of the trunk and disease-free growth as secondary outcomes.

- **Multiple primary outcomes or composite outcome**

Growth measured as height and trunk circumference or estimate volume of the tree trunk.

- **Surrogate outcomes**

Measure trunk circumference at 1m above the ground instead of counting growth rings.

- **Categorized outcomes**

Growth in centimeters or whether or not the growth is larger than a certain threshold (0/1 variable).

- **Baseline and longitudinal outcomes**

Get type of clone, information on the plot (soil, location, incline) at baseline. Correct for rainfall, temperature,… during the year(s).

RESEARCH INSTITUTE
NATURE AND FOREST

An alternative example with poplar trees

B) Statistical testing

Intermezzo - definitions

- **Type-I error** or alpha is the probability to reject the null hypothesis despite the fact the the null hypothesis is true (false positives)
- **Type-II error** is the probability of not rejecting the null hypothesis despite the fact the the null hypothesis is false (false negatives)
- **P-value** is the probability to observe data as extreme or more extreme then observed given that the null hypothesis is true
- **Power** or probability of correctly rejecting the null hypothesis

Illustration

RESEARCH INSTITUTE
NATURE AND FOREST

The illustration shows type1 error, type 2 error and the power visually.
Try the followingthings and see what happens:

- Play around with the difference between H0 and Ha: the smaller the effect you want to detect, the lower your power
- Play around with the variance in the H0 and Ha distributions: the higher the variance, the lower your power
- A one-sided test will ceteris paribus yield more power if Ha is in the same direction as your test. However, if Ha is in the different direction of your test, your type 2 error increases a lot and power quickly goes to zero. We therefore warn against using a one-sided test (see further on)
- You can increase the power if you accept a higher type 1 error or if you increase your sample size.

If the data is gathered according to the protocol and correctly digitized by the partridge counters, we will use the data to calculated the number of partridges in the game management unit.
If the data is very likely NOT gathered according to the protocol and correctly digitized by the partridge counters, we will NOT use the data to calculated the number of partridges in the game management unit since this would lead to an overestimation of the number of partridges.

Null hypothesis (H_0)

Hunters' count data is gathered according to the protocol and correctly digitized.

Alternative hypothesis (H_A)

Hunters' count data is not gathered according to the protocol and correctly digitized

Type - I error (α)

0.01

RESEARCH INSTITUTE
NATURE AND FOREST

We choose a very low alpha since we absolutely want to avoid rejecting the count data if it was gathered according to the protocol.
i.e. the probability that we reject H0 while it is true (= alpha) is very low.

CHALLENGE

TRUE or FALSE

If p-value < 0.01, we can conclude that the data is not gathered according to the protocol.

If p-value >=0.01, we can conclude that the data is gathered according to the protocol.

RESEARCH INSTITUTE
NATURE AND FOREST

The first statement is TRUE
The second statement is FALSE

We will use the data to calculate partridge densities if we do not have (strong enough) evidence that it was not gathered according to the protocol.
This is not proof that the data was properly gathered and digitized

H_0: $\mu_A = \mu_B$ H_A: $\mu_A \neq \mu_B$

H_0: $\mu_A \geq \mu_B$ H_A: $\mu_A < \mu_B$

H_0: $\mu_A \leq \mu_B$ H_A: $\mu_A > \mu_B$

- Can only reject the H_0 when there is enough evidence and conclude H_A to be true
- If not enough evidence to reject H_0, H_0 not necessarily true
- In other words: ''Absence of evidence is not evidence of absence''
- 1-sided tests only when strong evidence

RESEARCH INSTITUTE
NATURE AND FOREST

You need very strong arguments to support a one-sided test. You need to argument and proof that:

- You are sure the effect is one-sized
- You are not interested if the effect turns out to be in the other direction than you expected.

In practice, we almost always use a two sided test, even when our hypothesis is directional. This is because we (usually) still care about a significant finding in the "unexpected" direction.

For readers interested in Bayesian statistics, we recommend the following:

- Andrew Gelman, Gelman A., Francis Tuerlinckx & Tuerlinckx F. (2000). Type S error rates for classical and Bayesian single and multiple comparison procedures. Computational Statistics 15 (3): 373-390. https://doi.org/10.1007/s001800000040.
- https://sites.stat.columbia.edu/gelman/research/published/retropower20.pdf

B) Statistical testing

H_0: $|\mu_A - \mu_B| \geq \delta$ $\quad$ H_A: : $|\mu_A - \mu_B| < \delta$

H_0: $\mu \leq -\delta$ $\quad$ H_A: $\mu > -\delta$

H_0: $\mu \leq \delta$ $\quad$ H_A: $\mu > \delta$

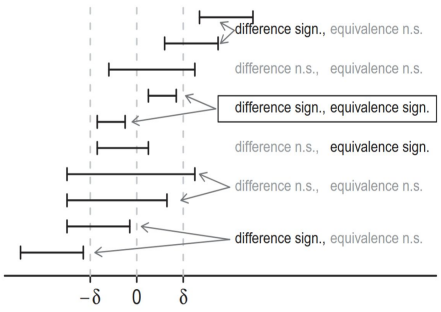
where $\mu = \mu_A - \mu_B$
And $\delta > 0$

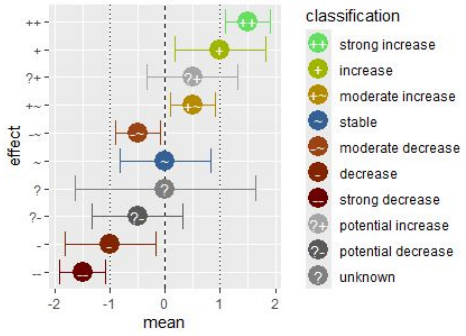
- δ is the minimal meaningful distance and set by experts in the field
- δ is critical in power and sample size calculations

Example: Labo investigates whether a new, cheaper method to measure concentrations in samples yields the same results as their current method. They use a paired t-test and allow for a maximum difference of δ.

RESEARCH INSTITUTE
NATURE AND FOREST

B) Statistical testing





RESEARCH INSTITUTE
NATURE AND FOREST

Meyners, M. (2012). Equivalence tests – A review. *Food Quality and Preference, 26*(2), 231–245.
https://doi.org/10.1016/j.foodqual.2012.05.003
Effectclass package

Equivalence tests should be applied whenever the aim of the study is not to show differences, but to conclude similarity.

```
#Script voor effectclass figuur:
library(effectclass)
library(tidyverse)
ds <- data.frame(mean = c(0, 0.5, -0.5, 1, -1, 1.5, -1.5, 0.5, -0.5, 0),
sd = c(1, 0.5, 0.5, 0.5, 0.5, 0.25, 0.25, 0.25, 0.25, 0.5))
ds$lcl <- qnorm(0.05, ds$mean, ds$sd)
ds$ucl <- qnorm(0.95, ds$mean, ds$sd)
ds$effect <- classification(ds$lcl, ds$ucl, 1)
ds <- ds %>%
mutate(effect =factor(as.factor(effect),
levels = rev(c("++", "+", "?+", "+~", "-~", "~", "?", "?-", "-",  "--"))))

ggplot(ds, aes(x = effect, y = mean, ymin = lcl, ymax = ucl, link_sd = sd)) +
stat_effect(threshold = 1) +
coord_flip()
```

CHALLENGE

Define H_0 and H_a:

- The average yield per acre for all types of corn in a recent year was 161.9 bushels. An economist believes that the average yield per acre is different this year.

- Poplar trees with genotype A grow faster than those with genotype B

RESEARCH INSTITUTE
NATURE AND FOREST

CHALLENGE

Define H_0 and H_a:

- The average yield per acre for all types of corn in a recent year was 161.9 bushels. An economist believes that the average yield per acre is different this year.

$$H_0: \mu_A = \mu_B \qquad H_A: \mu_A \neq \mu_B$$

- Poplar trees with genotype A grow faster than those with genotype B

$$H_0: \mu_A \leq \mu_B \qquad H_A: \mu_A > \mu_B$$

RESEARCH INSTITUTE
NATURE AND FOREST

Note that you would need strong arguments to perform a one-sided test in the second example (eg: we know the genetic sequence of both clones and know that, because of that A will grow at least as fast as B. Additionally, we would not be interested in a significant effect in the other direction).

CHALLENGE

Reformulate the following research questions into testable hypotheses:

- How do windmills affect bird behaviour?
- What are the health benefits of eating an apple a day?
- Which airlines have the most delays?
- What effect does daily use of social media have on the attention span of under-16s?

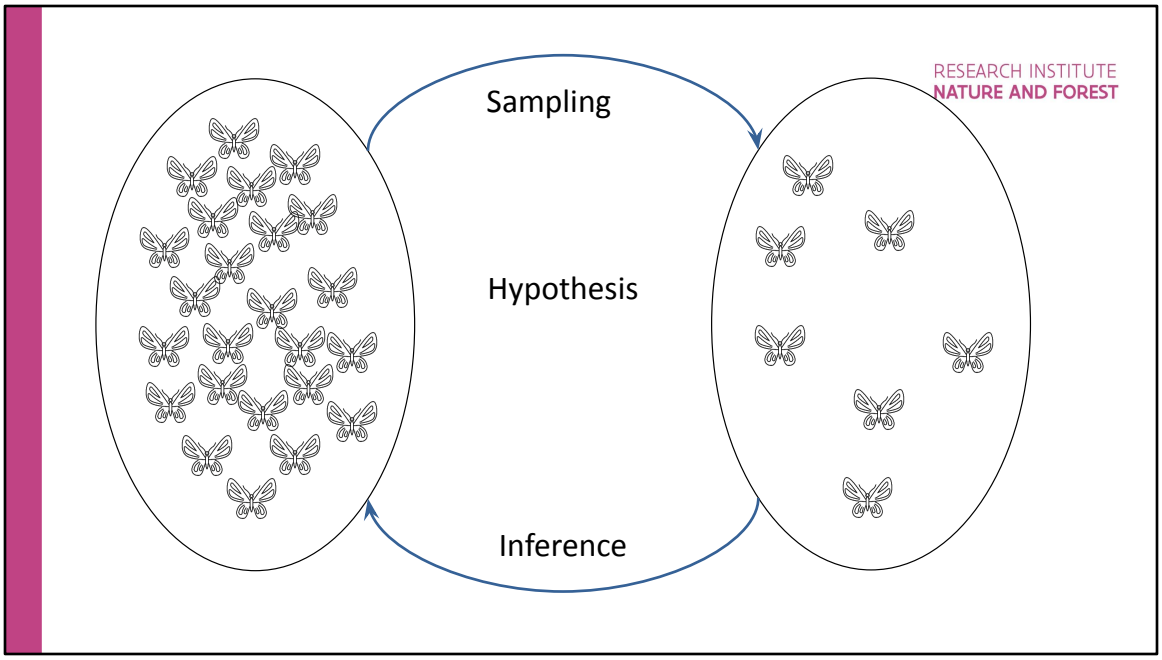
RESEARCH INSTITUTE
NATURE AND FOREST

Reformulate by detailing the population, what is your primary outcome? How will you measure/quantify the primary outcome?
Eg:

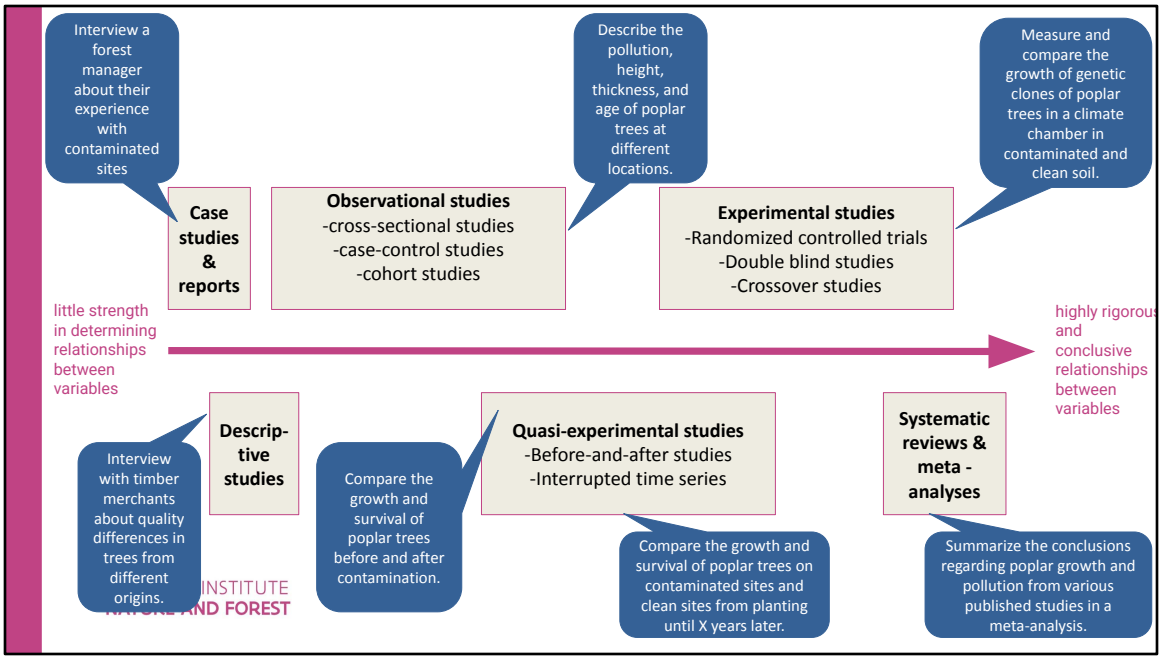
- Population may be breeding grassland birds/ migratory birds/...
  We might measure the number of birds (flux) that visits an area reduces after a windmill is built in that area. / The number of breeding birds is lower in fields with windmill's shadow flicker, compared to similar fields without shadow flicker. / …
- Population may be over -60s / people in elderly homes (easier to monitor their activity and consumption)/
  We might measure frequency of doctor's visits, BMI, blood tests,...
  We might do a retrospective cohort study where we ask the general population about their apple consumption in the last year and relate that to their current health
- Population: Low-cost airlines vs premium airlines, only zaventem airlines
  how to measure delays? total minutes delay, delay over a certain threshold, take into account transfers to other flights? take into account number of passengers?
- Is it ethical to force some -16s to use social media more if you suspect

- negative consequences? Maybe it’s better to do a cohort in stead? Effect we might be interested in might be attention span, activity levels, anxiety, self-image, confidence,...

C) Sampling and treatment allocation



In most studies, we aim to proof causal relationships between variables.



Inspired by:
https://www.fplanque.com/living/science/scientific-studies-research-quality/

Intervention studies	Observational studies
Researcher manipulates one or more variables	Variable of interest is not under influence by the investigator
Two treatments - New treatment vs control - New treatment vs old treatment - New treatment vs placebo	Can compare different treatments or groups
Random sampling	Random sampling
Random treatment allocation	

RESEARCH INSTITUTE
NATURE AND FOREST

Random treatment allocation and random sampling will reduce (eliminate) bias.

Random sampling

- Select a part of the entire population of N units
- Give an answer to how the data (sample of size n) was selected
- Each sample has a known probability P(n) of being selected
- **Inference about the entire population**
- **Unbiased estimates of population quantities**

RESEARCH INSTITUTE
NATURE AND FOREST

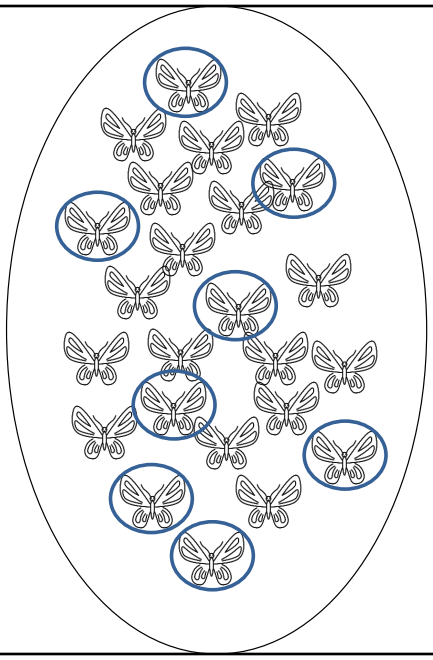
Can either be under the control of the investigator or due to nature
In monitoring schemes, there's often a difference between the population and the sampling frame (steekproefkader). Ideally, they completely overlap.

An example would be the monitoring scheme for forest birds en de farm birds in Flanders. Grid cells are randomly sampled from 5800 grid cells which only represent 55% of Flanders. This means that we can only make inference for these 5800 grid cells and, strictly speaking, we cannot make inference to the whole of Flanders.

Sampling

- **Simple random sampling**
- Stratified Sampling
- Cluster and systematic sampling
- Capture-recapture Sampling
- Spatial sampling
- Adaptive sampling

RESEARCH INSTITUTE
NATURE AND FOREST



can either be with replacement or without replacement. With replacement will give an biased estimate for the population and will be less efficient (higher variance)

Sampling

Simple random sampling

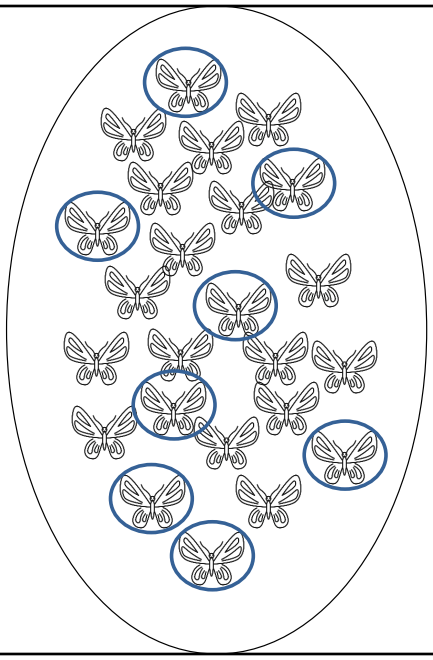
Advantages

- Representative sample if **n** is large enough
- Simple and straightforward statistical analysis

Disadvantages

- Larger sampling variance
- Estimates in specific sub-areas not possible

RESEARCH INSTITUTE
NATURE AND FOREST



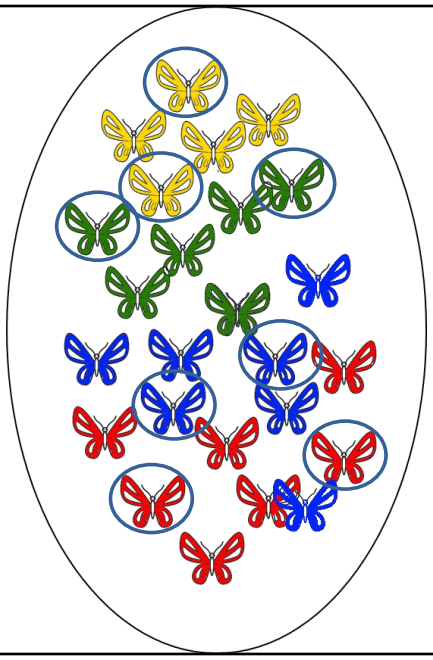
In simple random sampling, the probability that an element of the population is part of the sampling frame is the same for all elements.

example: Water Framework Directive
Make a list of all water bodies in Flanders
Randomly generate N numbers (position in the list of water bodies)
Select the water bodies at those positions in the list as your random sample.

Sampling

- Simple random sampling
- **Stratified Sampling**
- Cluster and systematic sampling
- Capture-recapture Sampling
- Spatial sampling
- Adaptive sampling

RESEARCH INSTITUTE
NATURE AND FOREST



In stratified sampling, the population is partitioned into regions or strata, and a sample is selected by some design within each stratum.the selections in different strata are made independently, the variances of estimators for individual
strata can be added together to obtain variances of estimators for the whole population.
Since only the within-stratum variances enter into the variances of estimators, the principle of stratification is to partition the population in such a way that the units within a stratum are as similar as possible. A geographical region may be stratified into
similar areas by means of some known variable such as habitat type, elevation, or soil type.

Sampling

Stratified Sampling

Proportionally Stratified Sampling

- **Definition:**

  The number of elements in a stratum is **proportional to the "size" of the stratum**.

- **Advantages:**
  - All strata are adequately represented in the sample.
  - Smaller variances → smaller sample size needed.
- **Disadvantages:**
  - Prior knowledge of the population is required.

RESEARCH INSTITUTE
NATURE AND FOREST

Disproportionally Stratified Sampling

- **Definition:**

  The number of elements in a stratum is **not proportional to the "size" of the stratum**.

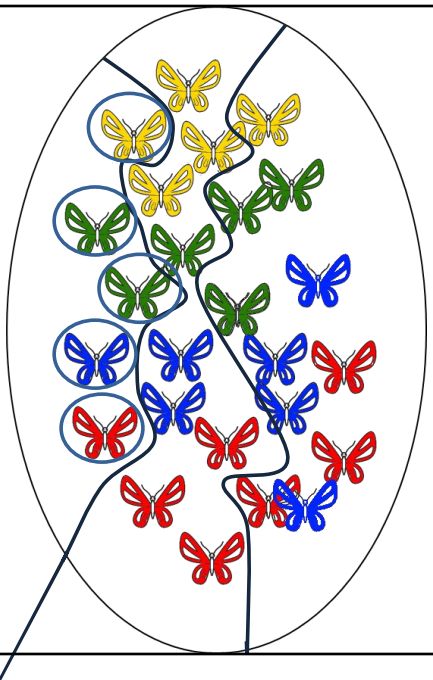
- **Advantages:**
  - Oversampling of smaller strata allows them to be studied individually.
- **Disadvantages:**
  - Analysis is more complex; **weights must be applied** in estimation.

In stratified sampling, the population is partitioned into regions or strata, and a sample is selected by some design within each stratum.the selections in different strata are made independently, the variances of estimators for individual
strata can be added together to obtain variances of estimators for the whole population.
Since only the within-stratum variances enter into the variances of estimators, the principle of stratification is to partition the population in such a way that the units within a stratum are as similar as possible. A geographical region may be stratified into
similar areas by means of some known variable such as habitat type, elevation, or soil type.

Sampling

- Simple random sampling
- Stratified Sampling
- **Cluster and systematic sampling**
- Capture-recapture Sampling
- Spatial sampling
- Adaptive sampling

RESEARCH INSTITUTE
NATURE AND FOREST



Cluster sampling
In cluster sampling, a primary unit consists of a cluster of secondary units, usually in close proximity to each other.

example with pigs: One stable contains multiple pens. Each pen can therefore be seen as a cluster of different animals. These animals do not necessarily have to have the same characteristics. And we are going to select the pens at random.
A random cluster is chosen and all elements in that cluster are in the sampling frame.

Systematic sampling;
In systematic sampling, a single primary unit consists of secondary units spaced in some systematic fashion throughout the population.
example with pigs: 1 stable has 20 pens and we first select one random pen, then we'll select every other pen after that randomly chosen one.

Sampling

Cluster and systematic sampling

- Elements in the population are divided into **groups, called "clusters"** (e.g., transects).
- A **random sample of clusters** (more than one) is then selected.
- **All elements within the selected clusters** are included in the sample.

RESEARCH INSTITUTE
NATURE AND FOREST

Cluster sampling

Sampling

Cluster and systematic sampling

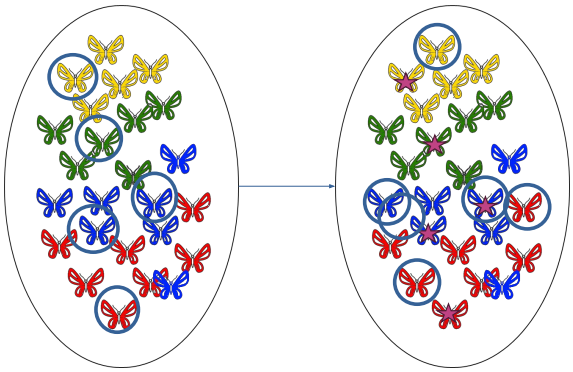
- A special case of **simple random sampling**
- The first element is random
- Every k^{th} element is selected

RESEARCH INSTITUTE
NATURE AND FOREST

Systematic sampling

Sampling

- Simple random sampling
- Stratified Sampling
- Cluster and systematic sampling
- **Capture-recapture Sampling**
- Spatial sampling
- Adaptive sampling



RESEARCH INSTITUTE
NATURE AND FOREST

In capture–recapture sampling to estimate the total number of individuals in a population, an initial sample is obtained and the individuals in that sample are marked or otherwise identified. A second sample is obtained independently, and
it is noted how many of the individuals in that sample are marked. If the second
sample is representative of the population as a whole, the sample proportion of marked individuals should be about the same as the population proportion of
marked individuals. From this relationship, the total number of individuals in the
population can be estimated.

Sampling

Capture-recapture Sampling

- Estimate the total number of individuals (τ)
- Initial sample (X)
- Second sample (x, y)

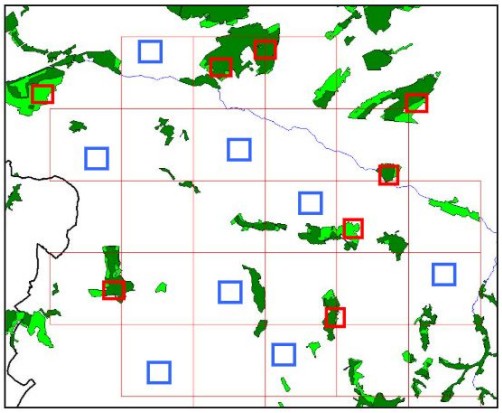
$$\frac{x}{y} = \frac{X}{\tau}$$

RESEARCH INSTITUTE
NATURE AND FOREST

In capture–recapture sampling to estimate the total number of individuals in a population, an initial sample is obtained and the individuals in that sample are marked or otherwise identified. A second sample is obtained independently, and
it is noted how many of the individuals in that sample are marked. If the second
sample is representative of the population as a whole, the sample proportion of marked individuals should be about the same as the population proportion of
marked individuals. From this relationship, the total number of individuals in the
population can be estimated.

Sampling

- Simple random sampling
- Stratified Sampling
- Cluster and systematic sampling
- Capture-recapture Sampling
- **Spatial sampling**
- Adaptive sampling



RESEARCH INSTITUTE
NATURE AND FOREST

Samples inside and outside HRL
Ideally, you should take into account spatial autocorrelation (locations close to each other are similar) and spatial heterogeneity (properties that change in space, e.g., soil, slope, temperature, etc.).
We will discuss this in more detail in the following module(s) because spatial sampling is often carried out at INBO (and in particular in the monitoring networks).
Can be uniform, stratified, random, mesh-based.

Here is a lot of additional information about spatial samples & designs for ecological research.

Sampling

Spatial sampling

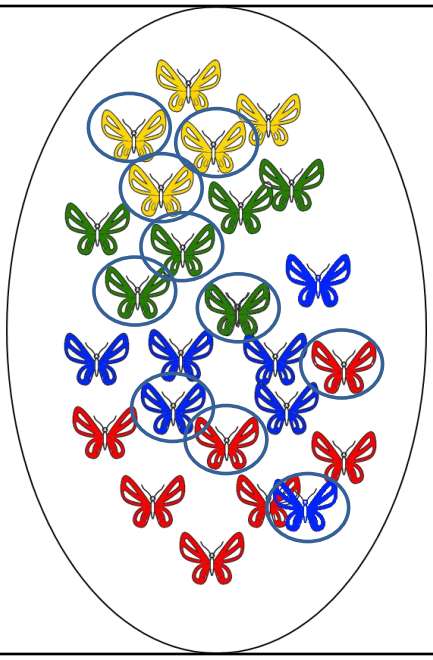
- Sampling in 2/3 dimensions
- Estimate a mean value in an area
- Spatially autocorrelated data
- Spatially heterogeneous data

RESEARCH INSTITUTE
NATURE AND FOREST

Sampling

- Simple random sampling
- Stratified Sampling
- Cluster and systematic sampling
- Capture-recapture Sampling
- Spatial sampling
- **Adaptive sampling**

RESEARCH INSTITUTE
NATURE AND FOREST



adaptive sampling;
Suppose you sample butterflies of each color with the same probability as you aim to get approximately the same amount of butterflies of each color in your sample.
You will sample 10 butterflies in total.
By chance, the first 3 butterflies are all yellow. You temporarily adjust your sampling probability for yellow butterflies down such that the probability you get another yellow goes down.
Especially interesting if you do not know in advance how many butterflies you will get in the end.

https://www.wiley.com/en-us/Adaptive+Sampling-p-9780471558712

Sampling

Adaptive sampling

- Sampling based on previous observations
- Take advantage of unknown population characteristics
- Get more precise estimate
- Biased estimates

RESEARCH INSTITUTE
NATURE AND FOREST

This is often used in medical research to balance, for instance, gender (male/female) or smoking behavior (smoker vs non-smoker). Suppose we want to recruit 100 people to participate in a study on people with heart diseases. The cardiologist will get a notification at random intervals to ask a patient who is with him during consultation to participate in the study. Suppose that, after 50/100 people have been recruited, only 15 of them are male. The computer will then still notify the cardiologist to ask a patient to participate at random intervals but with a higher probability for male patients then for female patients.

At INBO, an example where adaptive sampling is used, is in a survey of a rare, spatially clustered animal or plant population. Suppose want to investigate at what life stage (seedling, blooming, wilting,...) forest plants are at different points in time. We first take a spatial random sample which we will visit repeatedly over time. For common forest plants, we will likely find enough plants in our initial spatial random sample to get a good idea of the plants' life stage at each point in time. However, we might only be able to find some rare plants at a very limited number of locations. Therefor, once we find a rare plant, we can add randomly selected locations close to that location to our sample, in the hopes to find more of this rare plant to get good enough sample

size for this species specifically. Analyzing data that has been collected in this way, however, is not straightforward.
Adaptive sampling can also be used to systematically track down invasive species that typically come in clusters. Once an invasive species is found (often unexpectedly / by accident), we'll randomly investigate other locations in the neighborhood but, once it is found somewhere, the locations close by are more likely to be visited as well.

More information can be found a book by Thompson & Seber (1996) called "Adaptive sampling" (ISBN 978-0-471-55871-2)

Sampling

Two or multistage sampling

- In general, any sampling strategy is possible at each stage
  - Natural link with mixed model
  - Used when a complete list of all members of the population does not exist and is inappropriate
- Example: B@seball study:
  - Step 1: select schools (spatial sample, stratified over SES & type of environment (urban/rural))
  - Step 2: randomly select children within each of the school, stratified by age/gender
- Example: Aphid infestation
  - Step 1: spatial sample of grid cells, stratified over habitat types (that may contain oaks)
  - Step 2: inventorize oak trees in the grid cells that were selected and take a random sample of oaks, stratified over age / size
  - Step 3: Take a random sample of leaves from each tree and check for aphids

RESEARCH INSTITUTE
NATURE AND FOREST

Examples:

- B@seball study: First, schools are spatially sampled in Belgium. Next, n students are randomly selected in each school, stratified by gender.
- LUCAS (Land Use and Coverage Area frame Survey) protocol; First, a cheap screening method is used to characterise all population elements and to define strata. Next, a random sample is taken from each of the strata.
- We cannot know where all oak trees are in Flanders. Therefor, we first use expert judgement to know which habitats may contain oaks. We divide all area of Flanders which may have oak trees into smaller grid cells of equal size. We will only make a list of oak trees within the selected grid cells in stead of in the whole of Flanders.

source image= forestryimages.org

Intermezzo - definitions

- **Experimental Unit** is the item under study upon which something is changed. This could be raw materials, human subjects, or just a point in time. It is the item you randomly assign the treatment to.

- **Observational Unit** is the item under study upon which the variable of interest is observed.

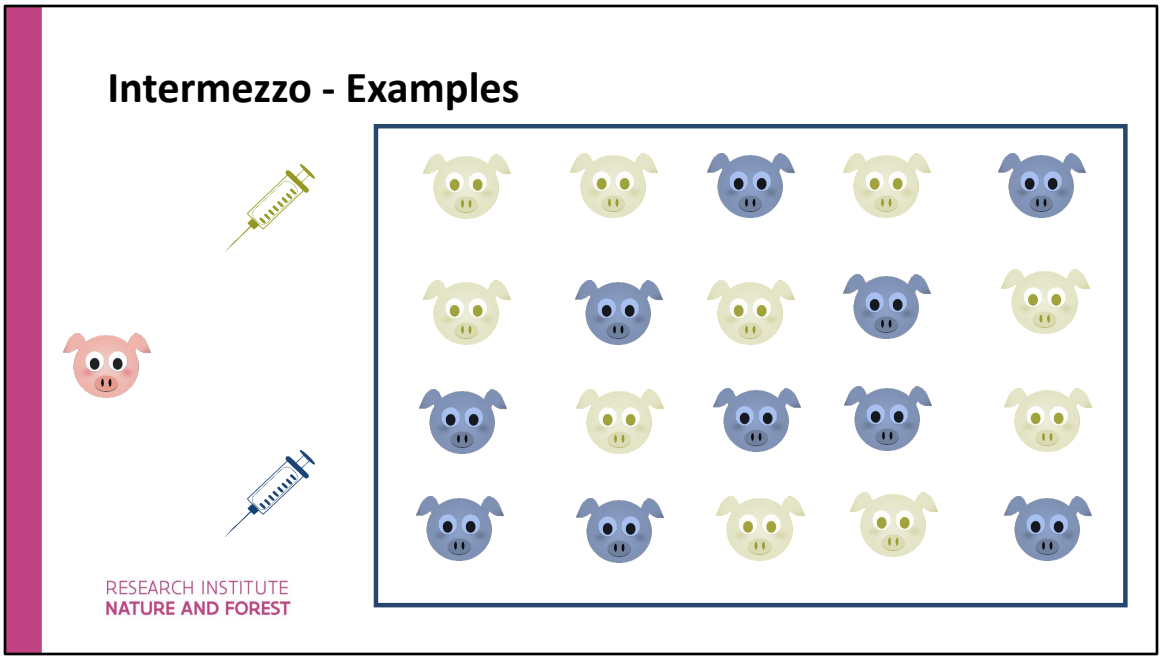
- **Independent Variable** is one of the variables used to predict, control or describe the dependent variable Treatment factor: often can be influenced by the researcher, especially in intervention studies.
  - Control variables: cannot be influenced by researcher but may have an effect on Y (especially important in observational research).

- **Dependent Variable** (or the Response denoted by Y) is the characteristic of the experimental unit that is measured after each experiment or run.

Antibodies ~ Treatment

RESEARCH INSTITUTE
NATURE AND FOREST

suppose we want to uncover the relationship between the treatment a pig receives and the number of antibodies in their blood a couple of days after the injection.

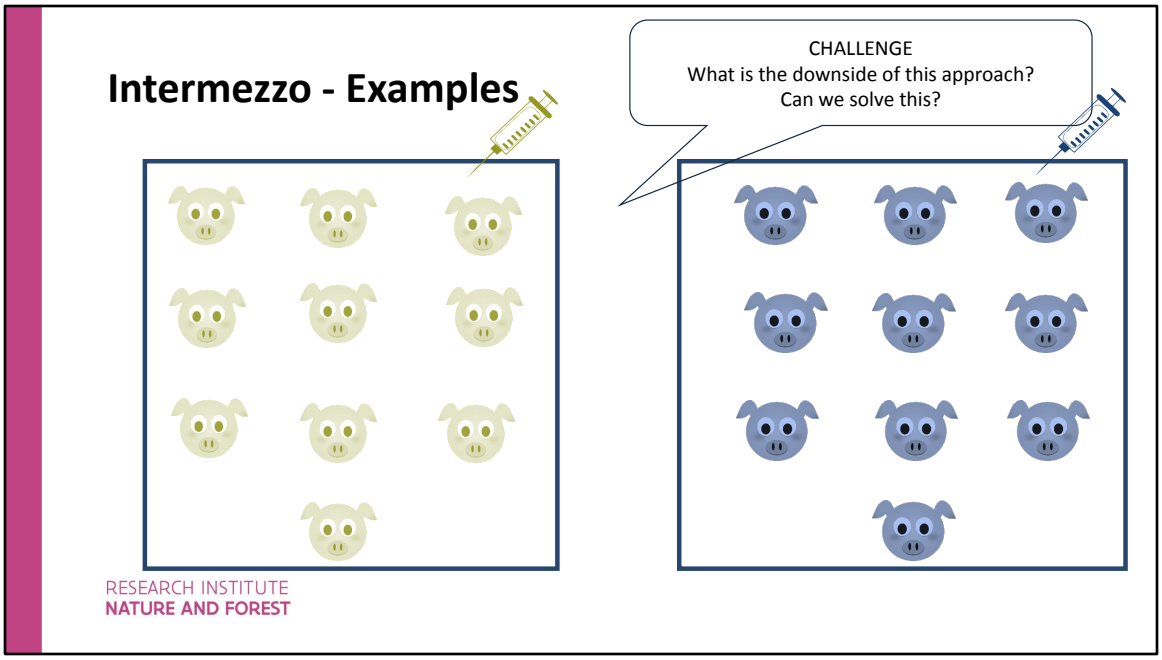


obervational unit = experimental unit = the pig

Feed intake ~ Treatment

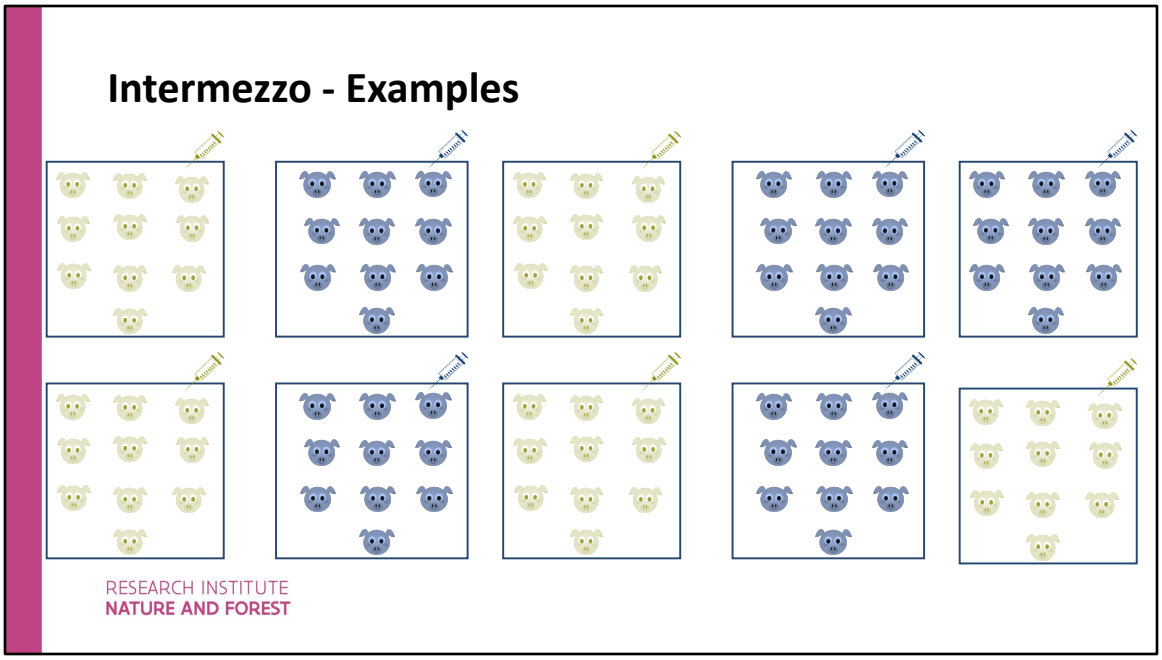
RESEARCH INSTITUTE
NATURE AND FOREST

Suppose we want to know the effect of a treatment on the feed intake of the pigs. Pigs cannot be put in a pen by themselves since they are social creatures. They are always put into pens with multiple pigs in one pen. Within each pen, there is one trough from which they all eat.
As such, we will only be able to measure feed intake at the level of the pen, not the individual pig.

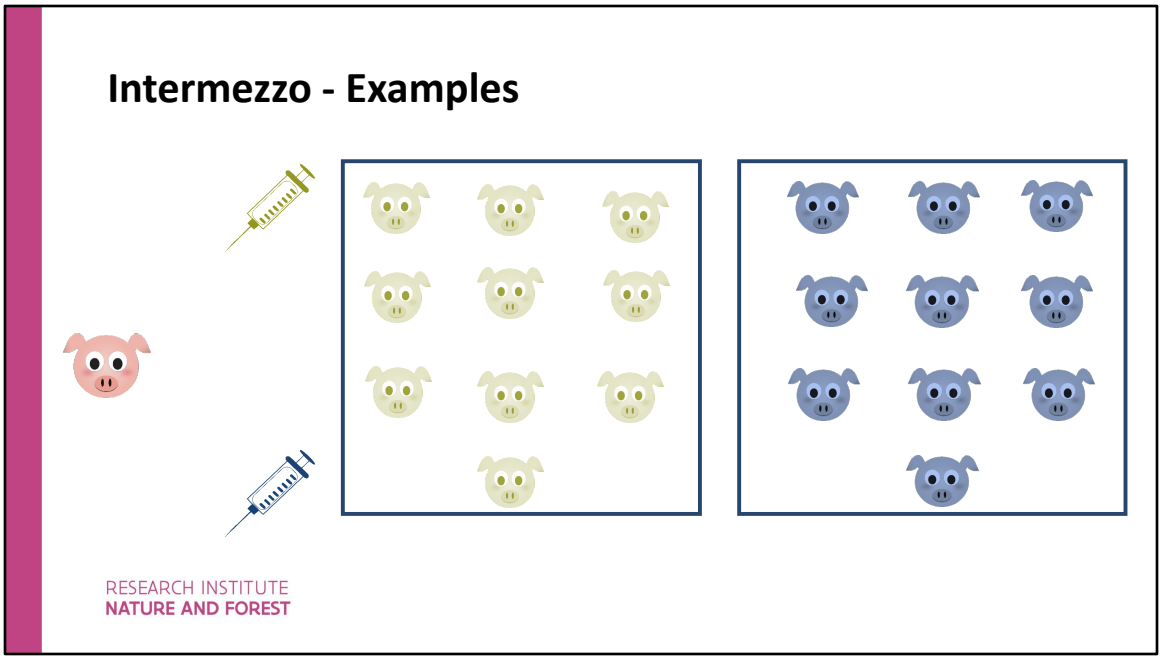


experimental unit = observational unit = pen
Downside: pen and treatment effect is completely confounded
Solutions:
- at least 3 stables for each treatment (preferably actually 10)

It doesn't matter how many pigs are in the pen, or how many fish are in the pond/ barrel. If the treatment is assigned at the higher level and your measurement is also at a higher level, you'll have a lot less observations and you will need replicates (see further on). if pig is the observational unit then all pîgs are repeated measures for stable. Including more pigs per stable will get you a more precise measurements for each stable.



at least 3 stables for each treatment (preferably actually 10)



The pigs are delivered to a farm by truck. When they step off of the truck, I randomly give them one of the treatments and randomly put the in one of the pens. However, by chance, each pen only have pigs with one type of treatment.
experimental unit = pen
observational unit = pen (if feed intake in measured) / pig (if blood is measured)

Randomization

- Foundation of correct statistical inference
- Random allocation
- Introduction of chance
- Selection bias
- More similar/comparable groups

Draw reliable conclusions

RESEARCH INSTITUTE
NATURE AND FOREST

Randomization

- Table of random numbers
- Tossing a coin
- Using computer software
- **Alternating allocation is not randomization**
- Prior to the start of the experiment

RESEARCH INSTITUTE
NATURE AND FOREST

Allocating every other subject to each group (alternating allocation) is not considered random allocation.

Randomization - experimental

- **Simple unrestricted randomization**.
  Can create imbalance in group sizes

- **Block randomization**.
  Controls better the imbalance between the different groups.

- **Clustered randomization**.
  When it is not feasible to assign the treatment to the individual.

D) Design of the study

Intervention studies	Observational studies
Researcher manipulates one or more variables	Variable of interest is not under influence by the investigator
Two treatments - New treatment vs control - New treatment vs old treatment - New treatment vs placebo	Can compare different treatments or groups
Random sampling	Random sampling
Random treatment allocation	

RESEARCH INSTITUTE
NATURE AND FOREST

In observational study random samples are obtained from two populations (subpopulations).
The observed outcomes are compared across populations. These studies have a harder to establish cause and effect relationship

Study designs

Refers to the structure of the experiment

- Explanatory factors
- Treatments
- Experimental units
- Rules and procedures of treatment allocations
- The outcome measurements

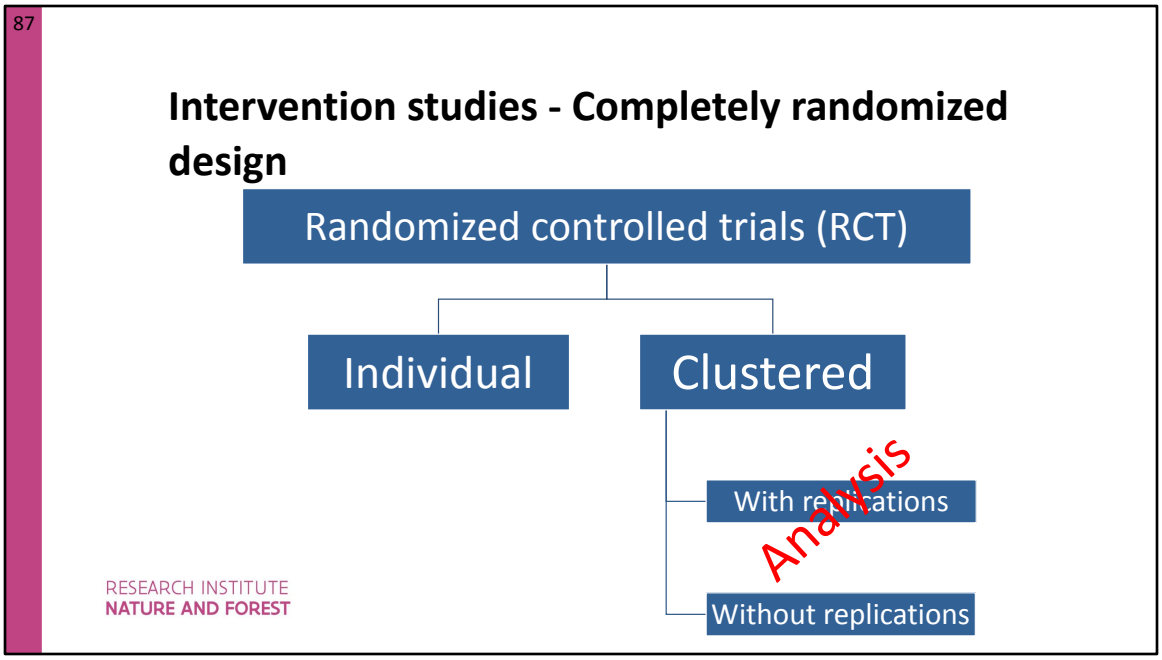
RESEARCH INSTITUTE
NATURE AND FOREST

Intervention studies – Completely randomized design

- N experimental units
- Two or more groups
- One or more control treatments
- All other variables are held constant
- Random treatment allocation

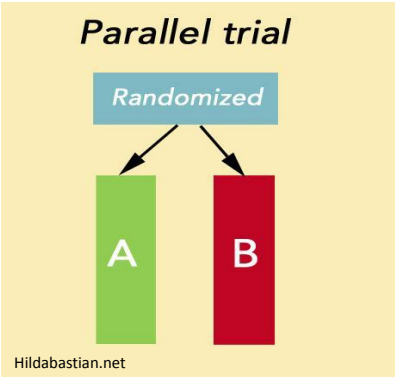
RESEARCH INSTITUTE
NATURE AND FOREST

Simplest form of designed experiment
allows for multiple treatments and different sample sizes per treatment
Not suitable for ecological observational studies with environmental heterogeneity



Clustering leads to less power than an iRCT of the same size, i.e., same number of individuals.

Intervention studies – Parallel group design

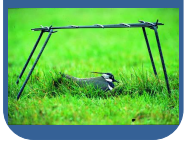


- Most commonly used
- Two or more groups
- One treatment in each group
- Groups are observed simultaneously
- One or more control treatments
- Random treatment allocation

Protection of grassland birds

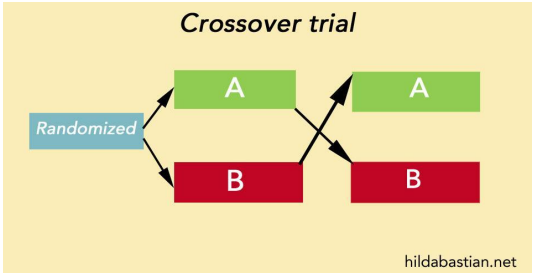
Group A: nest protection

Group B: control (no intervention)



RESEARCH INSTITUTE
NATURE AND FOREST

Intervention studies – Crossover design



- Each subject receives all treatments in a different sequence
- Number of treatments must be small
- Reduces the number of subjects given the same power
- Washout period
- Carryover effect
- Do not use when experimental unit will be changed permanently

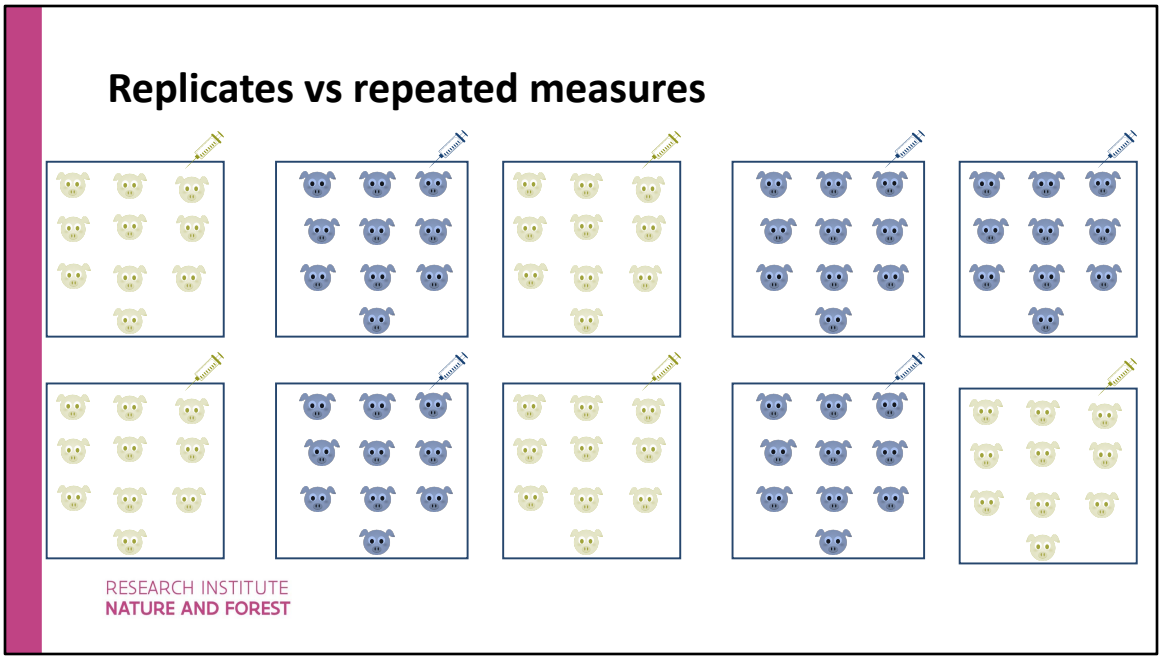
Intervention studies – Repeated measures design

- **Replicate runs or replications** are two or more experiments conducted with the same settings of the factors or independent variables, but using different experimental units.
- **Repeated measure** Response is measured multiple (in time or space) times in each subject/experimental unit

Compare trends in the response variable over time

RESEARCH INSTITUTE
NATURE AND FOREST

crossover is an example of a repeated measure design



This is an example of a clustered randomized design with replicates. The pen (cluster) has been given the treatment randomly. Pig or stable can be both observational unit.

Intervention studies – factorial design

		Treatment B	
		Yes	No
Treatment A	Yes	**AB**	**A**
	No	**B**	**No treatment**

RESEARCH INSTITUTE
NATURE AND FOREST

2*2 factorial design

Intervention studies – factorial design

		Nest protection	
		Yes	No
Mowing at a later time	Yes	**AB**	**A**
	No	**B**	**No treatment**

RESEARCH INSTITUTE
NATURE AND FOREST

Example: **Thermal plasticity in development and diapause strategy in a temperate butterfly across a latitudinal gradient**

10 females per region

		Region			
		Spain	Belgium	Estonia	Finland
Temperature	25°C				
	28°C				
	31°C				
	34°C				

2 replicate groups of 15 larvae per treatment for each female = 2groups*15larvae*10females = 300 larvae in each cell.

RESEARCH INSTITUTE
NATURE AND FOREST

4*4 full factorial
Source:
https://besjournals.onlinelibrary.wiley.com/doi/epdf/10.1111/1365-2435.14446
Foto Vildaphoto; veldparelmoervlinder
https://www.vildaphoto.net/nl/taxonomy/3805#p83771-r1

Intervention studies – factorial design

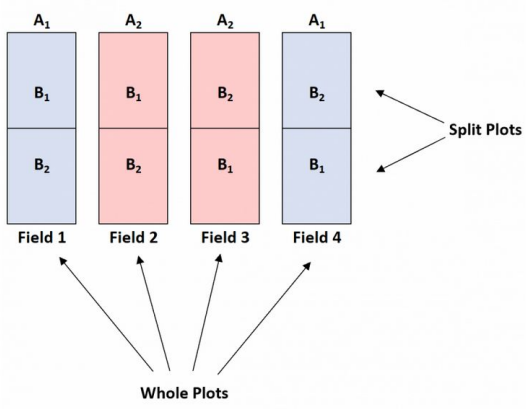
- Two or more treatments simultaneously
- All possible combinations
- Examining the interaction
- More efficient use of the experimental units
- Randomly assigned to one of the 2^k groups
- 2^k factorial
- Downside: Number of comparisons grows exponentially with the number of factors and factor levels.

RESEARCH INSTITUTE
NATURE AND FOREST

In fractional factorial, there are ways to reduce the needed sample size. The design is optimized such that you may get a less precise estimate of eg interaction effect which you know to be negligible or are not interested in.

Intervention studies – Split-Plot design

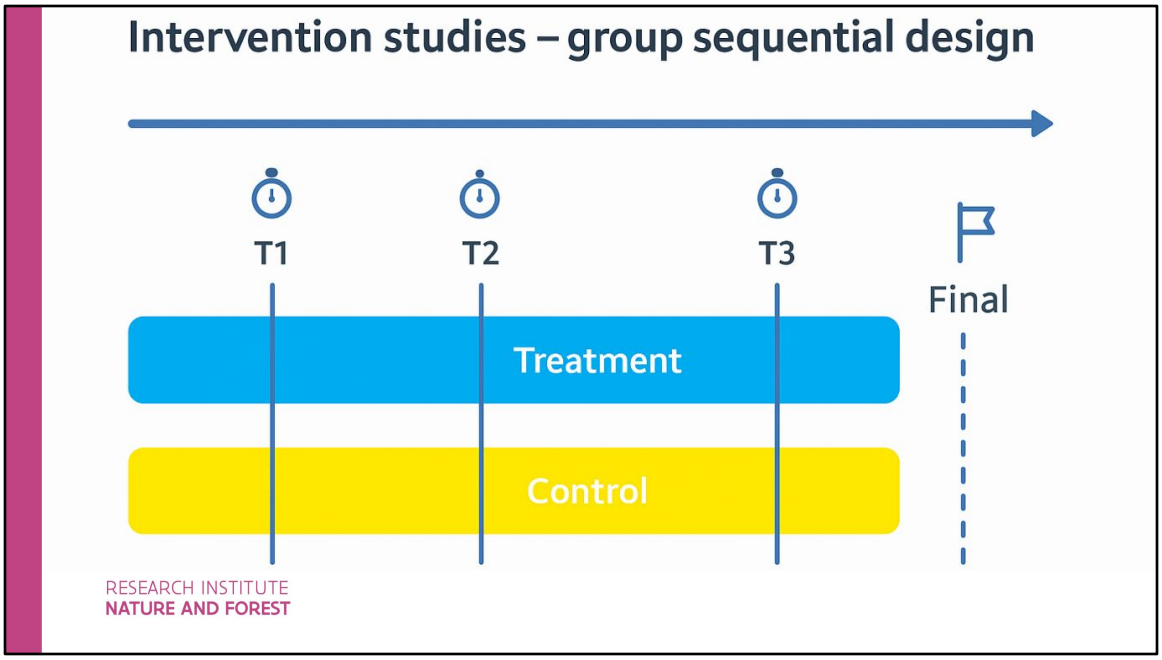
- Hard to change (whole plot)
- Easy to change (subplot)
- Higher precision in B and AB
- Lower precision in A



https://www.statology.org/split-plot-design/

RESEARCH INSTITUTE
NATURE AND FOREST

Obtain higher precision for estimating B and the interaction AB (aantal df gaat verschillend zijn tussen A, B en AxB)
at the cost of lower precision for estimating A. As an example of this A might represent varieties of wheat, and B fertilisers: if the focus is on the fertilisers, two or more very different varieties may be included primarily to examine the A B interaction thereby, hopefully, obtaining some basis for extending the conclusions about B to other varieties.



Alternative name is “interim analyse”
Each time you look, you need to “pay” in an increase of the Type 1 error.

When moving from a fixed design to a sequential design researchers need to make several decisions. The first decision is how the error rates should be controlled across multiple looks at the data. The second decision is whether the data collection is stopped when there is no support for the alternative hypothesis, and if so, which rule is used to make such a decision. After making both decisions it is possible to perform an a-priori power analysis to determine the sample size required for the sequential design.

One might choose to perform a conditional power analysis, and increase the sample size one is willing to collect, if smaller effects than a mean difference are still deemed interesting. Given that data has already been collected, it seems intuitive to perform the conditional power analysis not based on the effect size that was originally expected, but by updating the

prior belief about the true effect size given the observed dat

Example:

- Investigating fish safety of axial pumps
  - Different species of fish are send through pumps that operate at different speeds to analyze fish survival.
  - Compare results with claims made by pump manufacturer.

- Suppose sample size calculation suggests testing 4000 fish and manufacturer claims 75% survival. Should we continue the study if:
  - 980/1000 first fish survive?
  - 500/1000 first fish survive?
  - 100/1000 first fish survive?

RESEARCH INSTITUTE
NATURE AND FOREST

980/1000 -> stop for efficacy
100/1000 -> stop for futility
500/1000 -> continue

Aanbevelingen voor palingmigratie langs twee gemalen in de Bethoostersche Broecken
https://purews.inbo.be/ws/portalfiles/portal/115691654/Verhelst_etal_2024_AanbevelingenPalingmigratie.pdf
Visveiligheid pompgemaal Groot Schijn (EVINBO)
https://pureportal.inbo.be/nl/projects/fish-injury-mortality-and-passage-at-te-pumping-station-groot-sch

https://pmc.ncbi.nlm.nih.gov/articles/PMC11618692/

Intervention studies – group sequential design

- **Efficiently test a hypothesis**

- Will need **larger sample sizes** at the last look
  - Possibility to stop

- Stop at predefined timepoints and analyse the data (interim analysis)
  - Stop for futility
  - Stop for efficacy

- Control the Type-I and type-II error
  - **Multiple testing**
  - **Looks are not independent**
  - Pococks method
  - O'Brian and Fleming method

Contact team BMK for support!

RESEARCH INSTITUTE
NATURE AND FOREST

using the right alpha at each look and calculating sample size becomes much more complicated. Contact team BMK for support!

Intervention studies – group sequential design

- Looking without increasing the type-I error
  - Protocol adherence
  - Data quality check
  - Checking model assumptions
  - Monitor loss to follow up

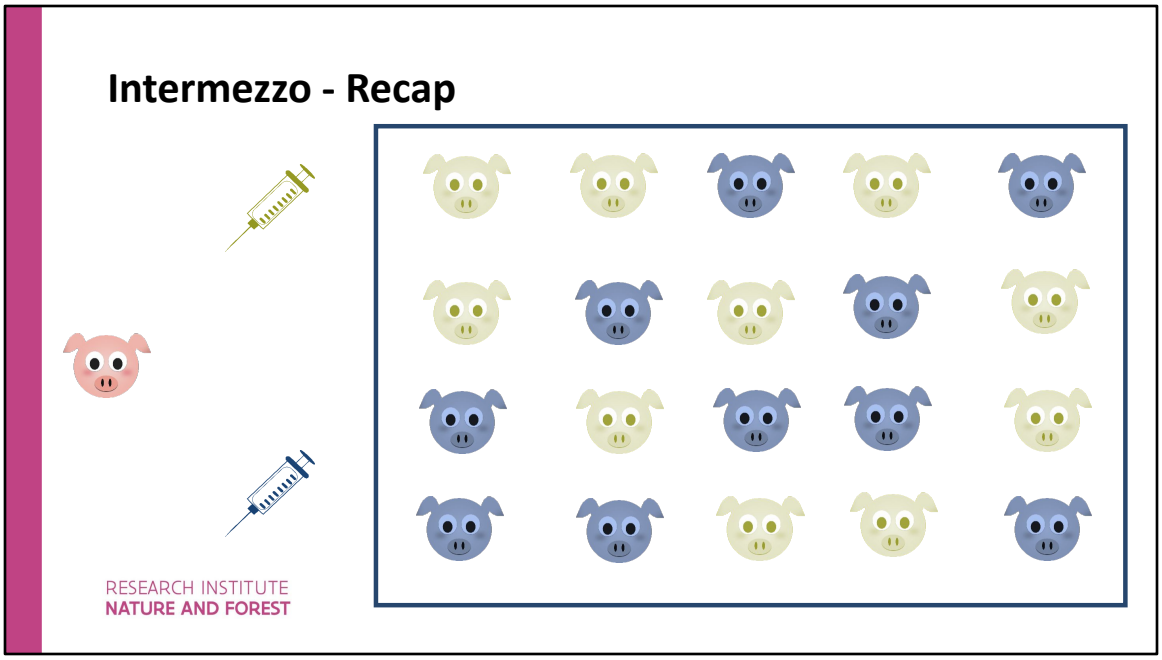
Oversight of the quality of the trial

RESEARCH INSTITUTE
NATURE AND FOREST

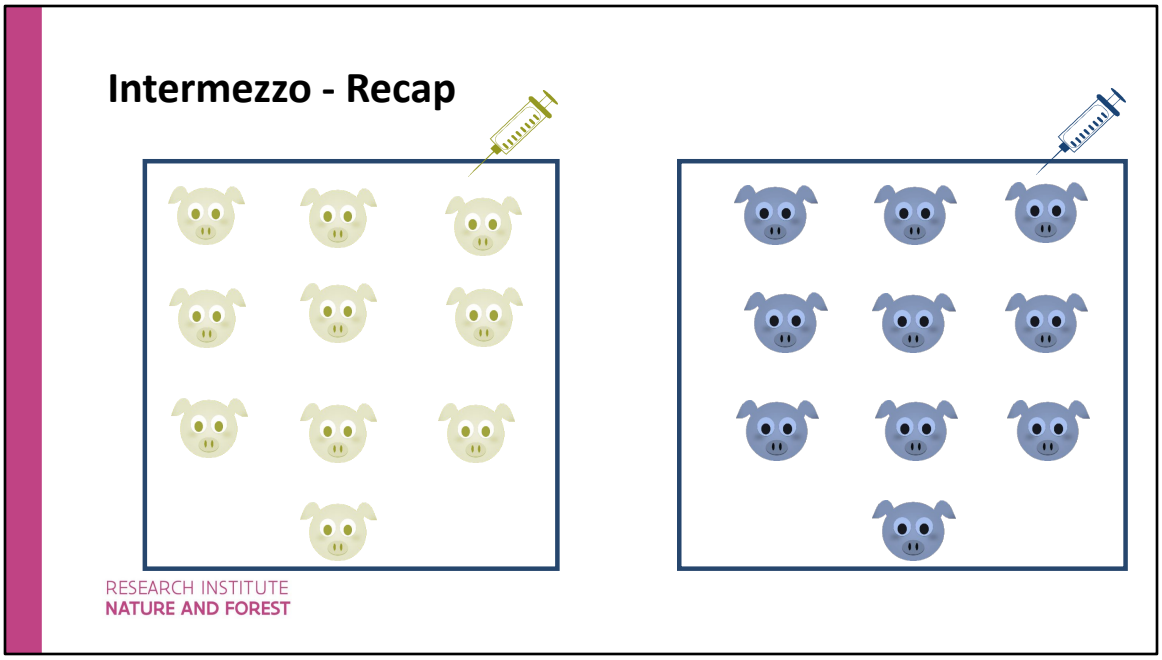
So can you never look at your data without “have to pay” in Type 1 error? Yes, you can look at the data to overlook the quality of the trial. But no, you cannot test you primary or secondary hypotheses without paying in Type1 error.

Intervention studies – Randomized block design

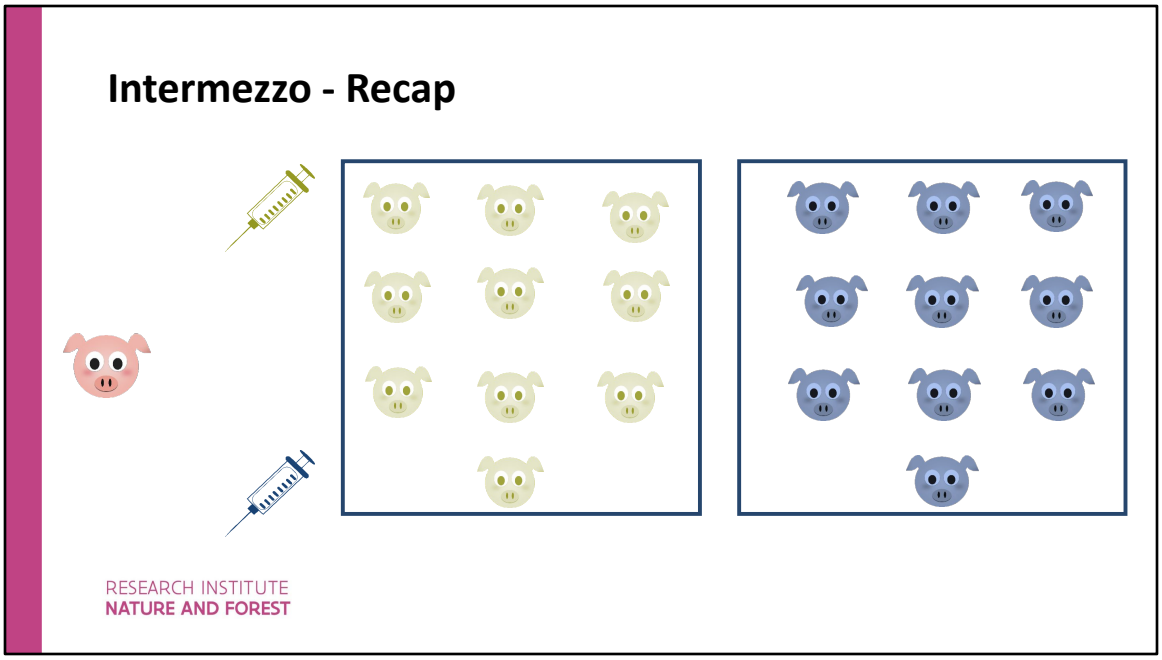
- Block = a group of homogeous experimental units
- All treatments have to be present in each block
- Between and within block variability
- Increases the power
- Number of blocks and number or experimental units within each block



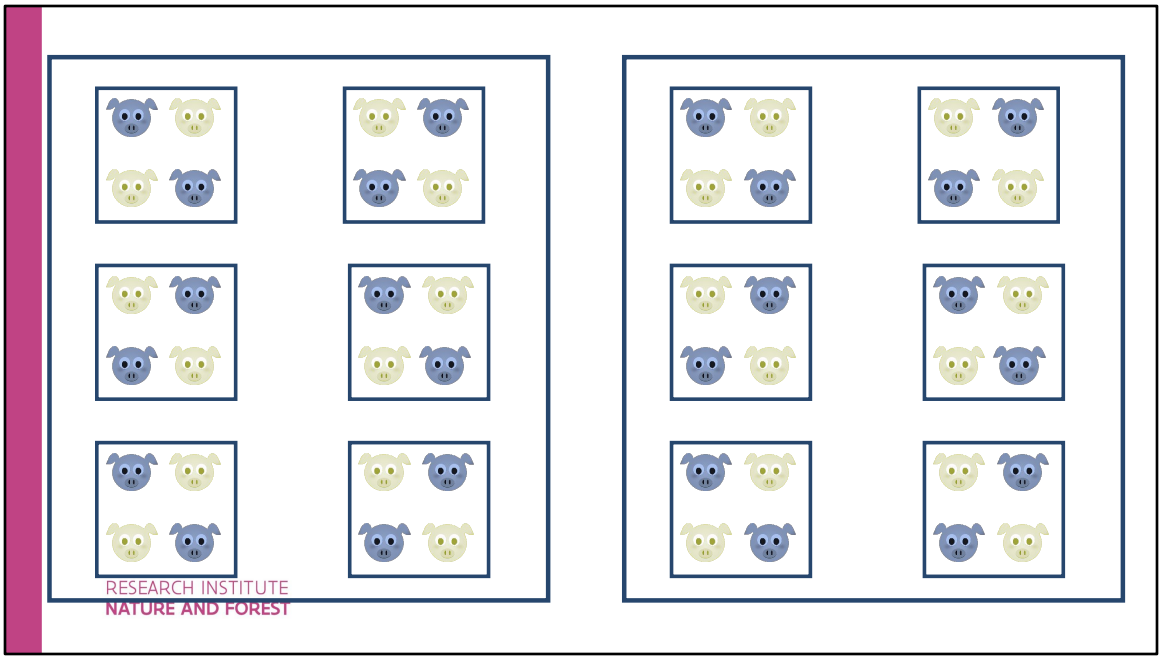
This is an example of an individually randomized trial. The experimental unit is pig.



This is an example of clustered randomized trial with no replications. The stable is the experimental unit. Depending on the outcome variable either stable or pig can be the observational unit.
Downside: stable and treatment effect is completely confounded
Solutions:
- at least 3 stables for each treatment (preferably actually 10)



This is an example of clustered randomized trial with no replications. The stable is the experimental unit. Depending on the outcome variable either stable or pig can be the observational unit. Although you have administered the treatment at pig level, due to the design the experimental unit is stable and not pig.



This is an randomized block design. Each treatment is present within each pen (block). The pig is the experimental unit. The stables are replicates
obs = pig (assuming we're going to measure something in hte blood)
exp = pig
penn = block
stables = replicates

Intervention studies – Latin square

- Blocking in 2 dimensions
- Reduce variability
- Square arrangement
- Randomisation
- Graeco latin square

$$\begin{matrix} T_4 & T_2 & T_3 & T_1 \\ T_2 & T_4 & T_1 & T_3 \\ T_1 & T_3 & T_4 & T_2 \\ T_3 & T_1 & T_2 & T_4. \end{matrix}$$

RESEARCH INSTITUTE
NATURE AND FOREST

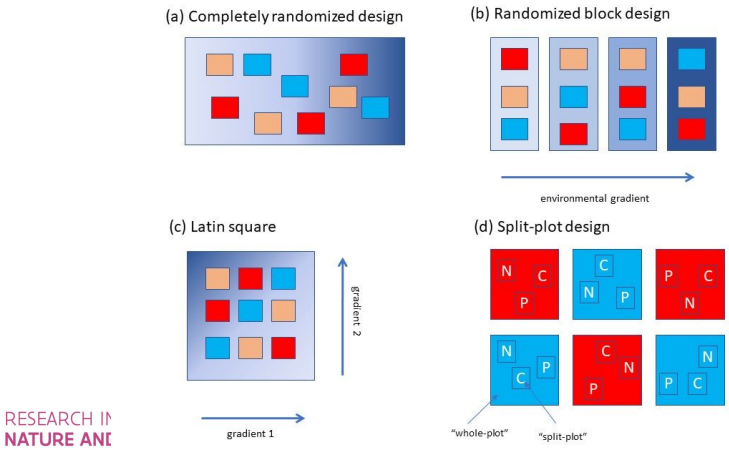
INBO example:
Greenhouse/climate chambers:

- Heat is blown in along one side; drier and warmer along one side of the greenhouse.
- Sun is on one side -> second dimension of gradient in the greenhouse.

Each treatment in one row and one column.

Randomization by choosing one design at random from all possible latin square designs.

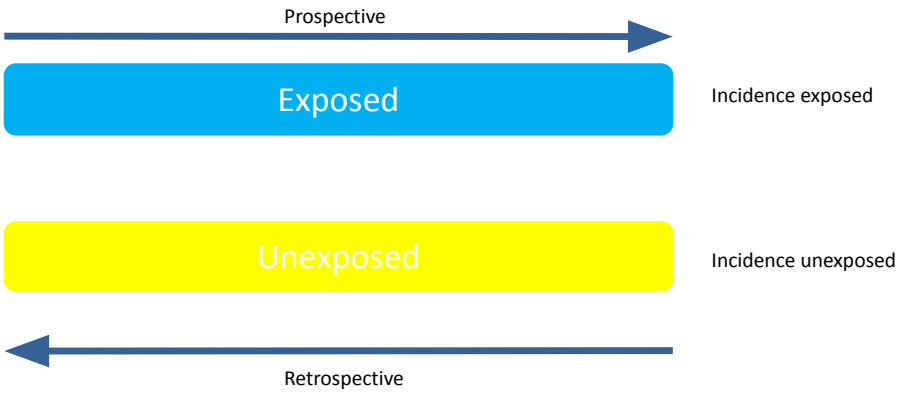
Recap:



RESEARCH IN
NATURE AND

Source: https://www.davidzeleny.net/anadat-r/doku.php/en:sampling_design

Observational studies – Cohort



RESEARCH INSTITUTE
NATURE AND FOREST

Observational studies – Cross-sectional study

- Prevalance
- Observation at a single point in time
- Generate hypotheses
- Examine evolving trend
- No loss to follow up
- Inexpensive and quick
- Selective survival bias

Observational studies – Case control

- Always retrospective
- Two groups: cases and controls
- Faster and cheaper
- Prone to bias
- odds

RESEARCH INSTITUTE
NATURE AND FOREST

Recall bias

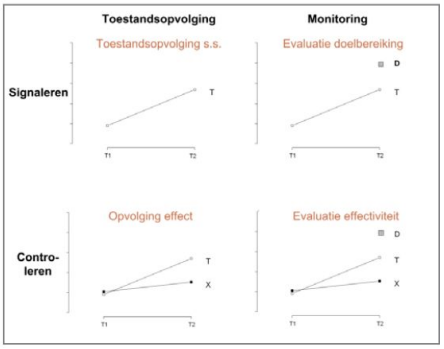
Monitoring networks (Meetnetten)

*A monitoring network is a tool for tracking the relevant characteristics of a **target population** by measuring carefully selected **variables** at **regular intervals**, using a **standardized method**, in a system of **measuring locations**.*

•AIM: Surveillance vs monitoring

•FUNCTION: signaling vs controlling

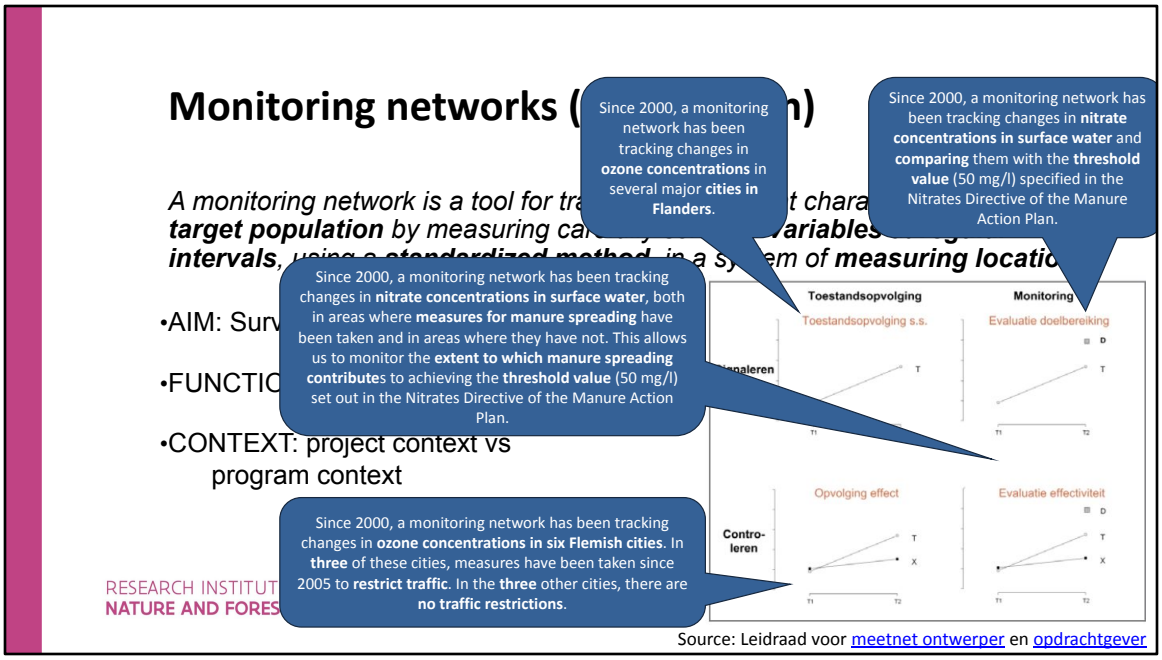
•CONTEXT: project context vs program context



RESEARCH INSTITUTE
NATURE AND FOREST

Source: Leidraad voor meetnetontwerper en opdrachtgever

By conducting measurements consistently over a sufficiently long period of time, we can observe possible changes in the status of the target population.
- PURPOSE of the monitoring network:
    - **status monitoring** = describing the status and evolution of the target population ><
    - **monitoring**: the evolution of the target population is compared with a priori established standards, reference values, or objectives. The design must be tailored to the possibility of detecting deviations from the set standards with a reasonable probability.
- FUNCTION:
    - monitoring networks with a **signaling** function detect any (negative) developments in a timely manner ><
    - a **controlling** function monitors the effect of certain policy or management measures, social activities, or developments. However, if a causal effect of policy measures is to be demonstrated, control is also necessary!



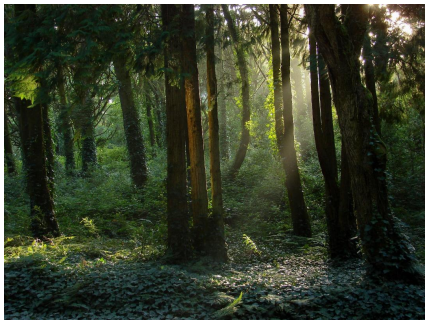
Source: Leidraad voor meetnet ontwerper en opdrachtgever

Here we show some examples of the different types of monitoring networks, in terms of their aim and function.
Three characteristics of monitoring networks (aim, function, and context) are essential for defining
the information requirements and have important consequences for the design of the monitoring network.
We define the objective as surveillance or monitoring.
We distinguish between monitoring networks with a signaling
and a controlling function. The monitoring network information can be used within a program or
project context.

Monitoring networks (Meetnetten)

*A monitoring network is a tool for tracking the relevant characteristics of a **target population** by measuring carefully selected **variables** at **regular intervals**, using a **standardized method**, in a system of **measuring locations**.*

- AIM: Surveillance vs monitoring
- FUNCTION: signaling vs controlling
- CONTEXT: project context vs program context



Forest vitality

RESEARCH INSTITUTE
NATURE AND FOREST

Source: Leidraad voor meetnetontwerper en opdrachtgever

CONTEXT: In a project context, the aim is usually to obtain information about a few specific elements, often because of concrete bottlenecks or measures. In most cases, the measurements also stop once we have an answer to our questions. In a program (strategic) context, the client wants to have figures describing the overall situation and the evolution of a collection of elements to support strategic decisions.
EXAMPLE: An example of a program context is the Level-I monitoring network for forest vitality, which has been monitoring the health of forests on a European scale since 1985. In the Flemish Region, there are around 80 test areas with a total of almost 1,500 sample trees. In terms of fieldwork, a crown assessment of different tree species is carried out from the ground using binoculars in order to identify a general trend in tree health across Flanders. However, this monitoring network does not allow conclusions to be drawn about the effect of, for example, management measures in specific forests. Although the fieldwork will be virtually identical, the target population is different and therefore the sampling frame and thus also the sampling will be different. For this purpose, there are other monitoring networks in the project context with trees that are representative of the forests where these management measures are applied (and possibly also for statistical control).

EXAMPLE from the guideline: The ozone concentration in the air is measured at a series of randomly selected measuring points spread across Flanders. Such a monitoring network in a program context and with a signaling function allows time-related changes in ozone concentration, based on daily maximums, to be detected across Flanders. Depending on the geographical density of the measuring points, it also allows regions to be located where the ozone concentration exceeds certain threshold values. However, the monitoring network is not suitable for providing accurate and reliable information for specific local bottlenecks.

The ozone concentration in the air is measured at a series of measuring points selected on the basis of known problem areas. Such a monitoring network in a project context and with a signaling function provides information about changes in ozone concentration, based on daily maximum values, at these locations and allows the frequency of exceedances of certain standards to be detected. However, the monitoring network does not provide unbiased information about the overall situation in Flanders, because measurements were deliberately not taken at locations with low concentrations.
Photo by freepik

E) Discussion

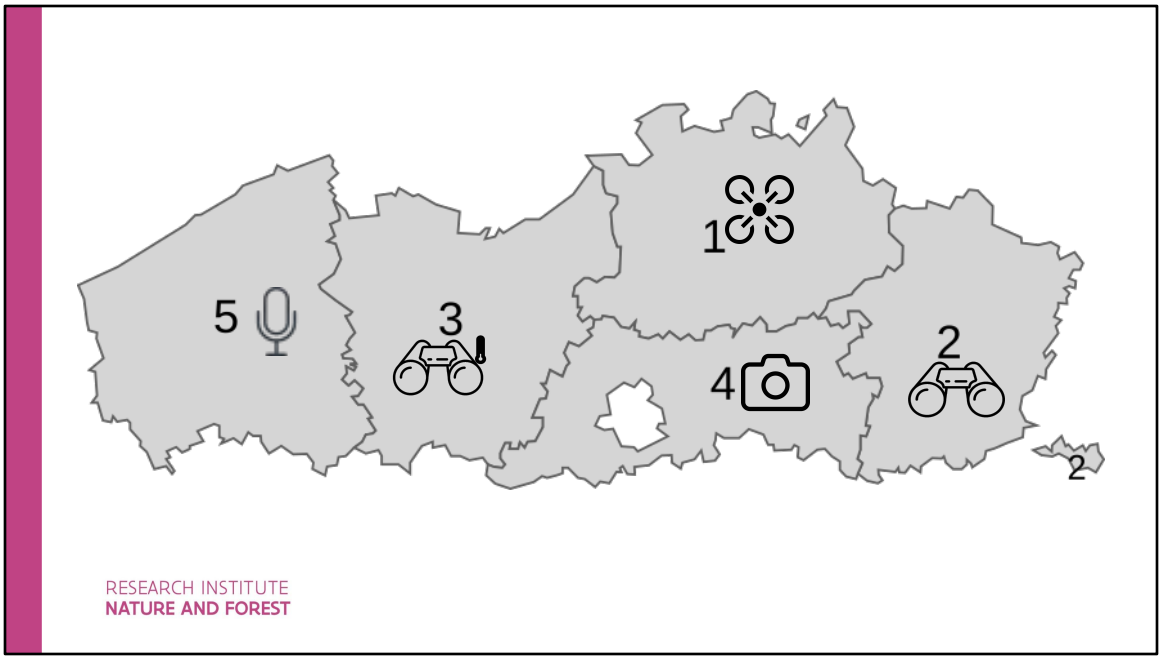
This is a fictional example!
Suppose we want to know which technique is best in detecting partridges

Hypothese

Is er een verschil tussen 5 verschillende detectie technieken ?

RESEARCH INSTITUTE
NATURE AND FOREST

There are 5 possible techniques and we want to know whether some techniques will allow us to see/detect more partridges than others.



We randomly assign each of the techniques to one of the provinces of Flanders.

provincie	techniek	echte_dichtheid	waargenomen
Provincie1	Visueel	5	2
Provincie1	Visueel	8	6
Provincie1	Visueel	5	4
Provincie1	Visueel	9	7
Provincie1	Visueel	10	9
Provincie1	Visueel	2	2
Provincie1	Visueel	6	4
Provincie1	Visueel	9	7
Provincie1	Visueel	6	5

RESEARCH INSTITUTE
NATURE AND FOREST

We wouldn't know the true partridge density in reality.

However, we do know that higher densities are typically seen in the West and lower in the and of Flanders. We also didn't consider habitat type in the locations where we use the techniques.

We CAN technically analyse the data like this but we won't be able to separate the effect of the technique from the effect of the province (they are confounded).

What if you analyse it like that ?

- Inflation of type-I error
- Confounding
- Analysis should reflect the design

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

RESEARCH INSTITUTE
NATURE AND FOREST

There is a close connection between design and analysis in that an objective of design is to make both analysis and interpretation as simple and clear as possible. Equally, while some defects in design may be corrected by more elaborate analysis, there is nearly always some loss of security in the interpretation, i.e. in the underlying subject-matter meaning of the outcomes.

Ask a statistician

RESEARCH INSTITUTE
NATURE AND FOREST

What is the job of the statistician

- Listen to the client
- Choose a model for the data
- Fit the model to the data
- Checks if the model fits
- Gives a report that is hard to understand

RESEARCH INSTITUTE
NATURE AND FOREST

What some people think a statistician does

What is the ACTUAL job of the statistician

- Identify the research question and refine it with the client
- Set up the estimand
- Take a look at the data and the design and set up some assumptions
- Do the analysis which may involve fitting of a model
- Give an interpretation to the results
- Give an answer to the refined research question
- Feedback with client

RESEARCH INSTITUTE
NATURE AND FOREST

What we aim to do for you.

RESEARCH INSTITUTE
NATURE AND FOREST

Q&A

123

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