An Overview of Randomization

Ornella Cominetti, PhD

Proteomics Team

Nestlé Research

Outline

1. **What is randomization**
2. Why is randomization needed?
3. Types of randomization
4. Checking randomization
5. What can happen when randomization is not well-performed?
6. Conclusions

Nestlé Research

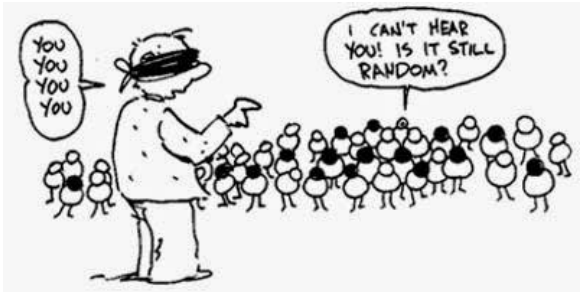
What is randomization – random allocation?

- The process of making something **random** (au hasard, aléatoire).

- In statistics, a random selection is a selection in which every item in a set has the same probability of being chosen.

- Typically, **samples** (échantillons) assigned to **batches** (lots, groupes d'expériences / d'échantillons / de sujets) or **subjects** assigned to treatments / arms.

- **Blinding**
  (à l'aveugle)



http://testofhypothesis.blogspot.com/2014/09/the-sample.html

Nestlé Research

Outline

1. What is randomization?
2. **Why is randomization needed?**
3. Types of randomization
4. Checking randomization
5. What can happen when randomization is not well-performed?
6. Conclusions

Nestlé Research

Why is randomization needed?

To minimize / control / reduce bias

- Unbalanced samples
- Sample is not representative of the general "population"
- Measurements are not consistent over time or spatially
  - Machine calibration drifts over time
  - Person learns or gets tired over time
  - Multiple persons perform the measurements
- Patients / samples are not consistent over time
  - Patients follow less strictly the protocol
  - Samples degrade the longer they are stored
  - New people are enrolled over time

Nestlé Research

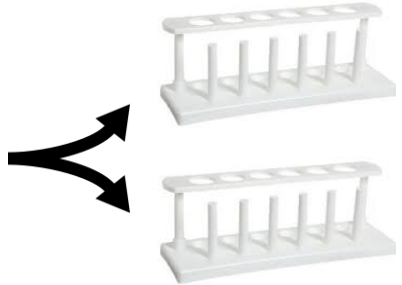
Outline

1. What is randomization?
2. Why is randomization needed?
3. **Types of randomization**
4. Checking randomization
5. What can happen when randomization is not well-performed?
6. Conclusions

Nestlé Research

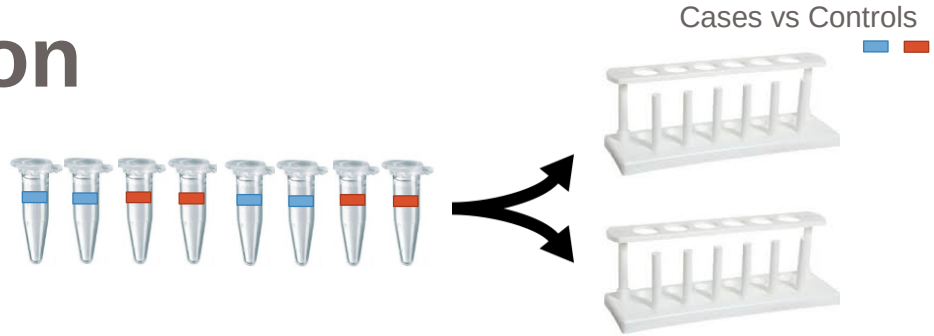
Types of randomization

- **Simple randomization**



- Block randomization
- Stratified randomization
- Adaptive randomization

Types of randomization



- Simple randomization
- **Block randomization**

Size of block: 4, 2 cases and 2 controls in each block

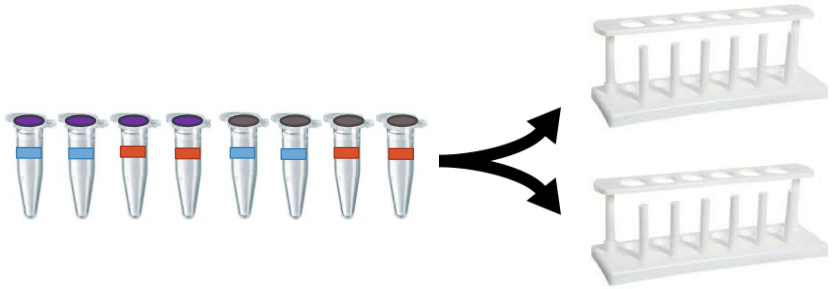
- Stratified randomization
- Adaptive randomization

Types of randomization

2 genders ♀ ♂

Cases vs Controls

- Simple randomization
- Block randomization



- **Stratified randomization**

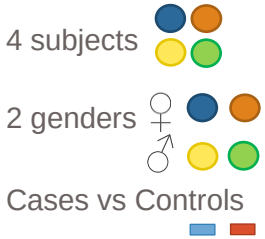
Size of block: 2, 1 case and 1 control in each block, for each gender

♀ ♂ …

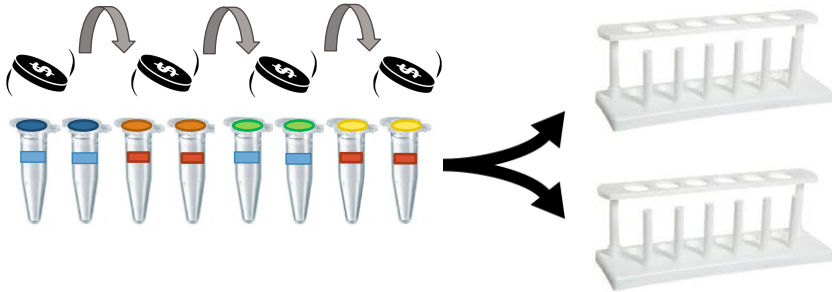
- Adaptive randomization

Nestlé Research

Types of randomization



- Simple randomization
- Block randomization
- Stratified randomization
- **Adaptive randomization**



First randomize, then perform assignment, then adapt the randomization according to results of previously analyzed samples or to additional variables of the samples in each batch

Nestlé Research

Outline

1. What is randomization?
2. Why is randomization needed?
3. Types of randomization
4. **Checking randomization**
5. What can happen when randomization is not well-performed?
6. Conclusions

Nestlé Research

Checking randomization

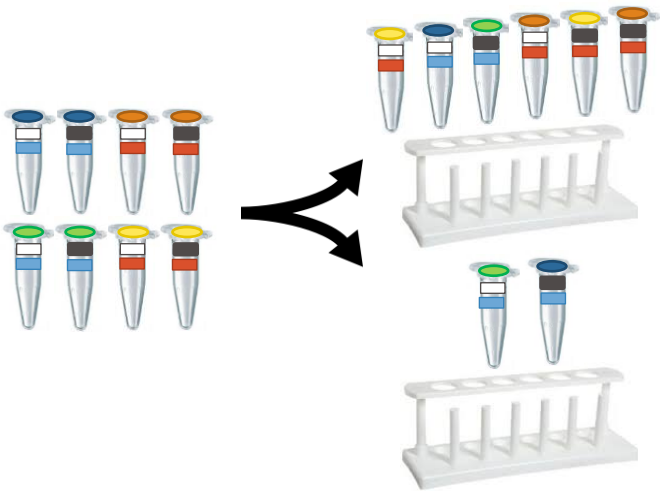
4 subjects

2 genders

2 time points

Cases vs Controls

A



B

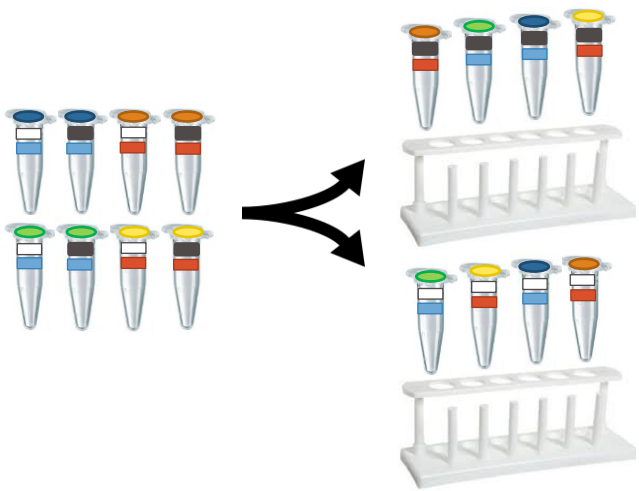


Plate	Females / Males	Time Points 1 / 2	Cases / Controls	N
1	3/3	3/3	2/4	6
2	1/1	1/1	2/0	2

Plate	Females / Males	Time Points 1 / 2	Cases / Controls	N
1	2/2	0/4	2/2	4
2	2/2	4/0	2/2	4

Nestlé Research

Checking randomization

4 subjects

2 genders

2 time points

Cases vs Controls

C

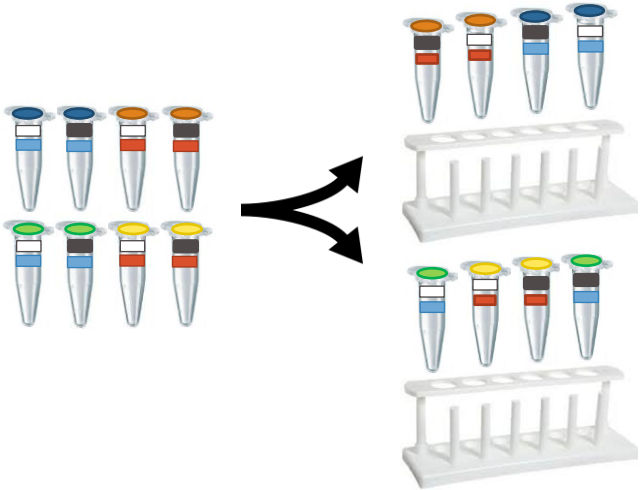


Plate	Females / Males	Time Points 1 / 2	Cases / Controls	N
1	4/0	2/2	2/2	4
2	0/4	2/2	2/2	4

D

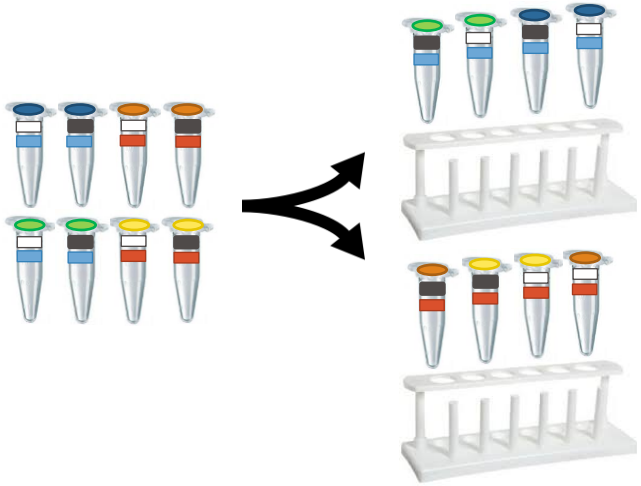


Plate	Females / Males	Time Points 1 / 2	Cases / Controls	N
1	2/2	2/2	4/0	4
2	2/2	2/2	0/4	4

Nestlé Research

Checking randomization

4 subjects

2 genders

2 time points

Cases vs Controls

E

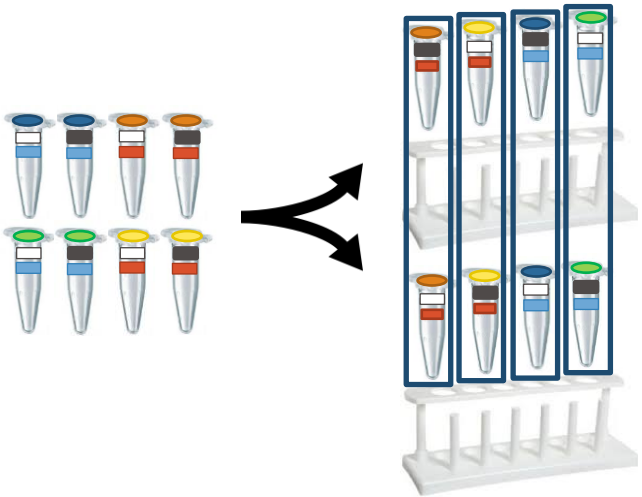


Plate	Females / Males	Time Points 1 / 2	Cases / Controls	N
1	2/2	2/2	2/2	4
2	2/2	2/2	2/2	4

F

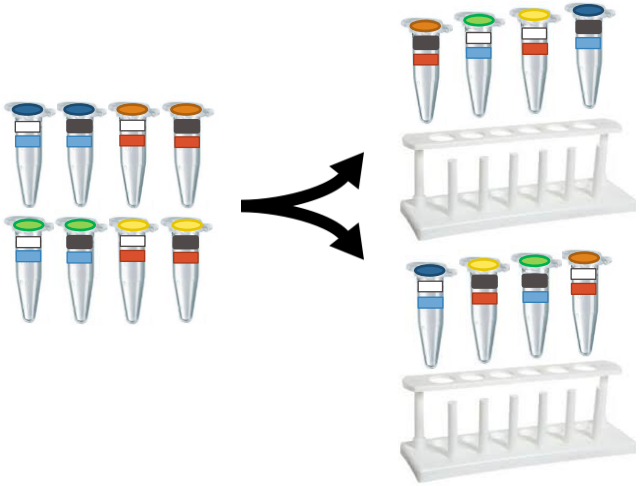


Plate	Females / Males	Time Points 1 / 2	Cases / Controls	N
1	2/2	2/2	2/2	4
2	2/2	2/2	2/2	4

Nestlé Research

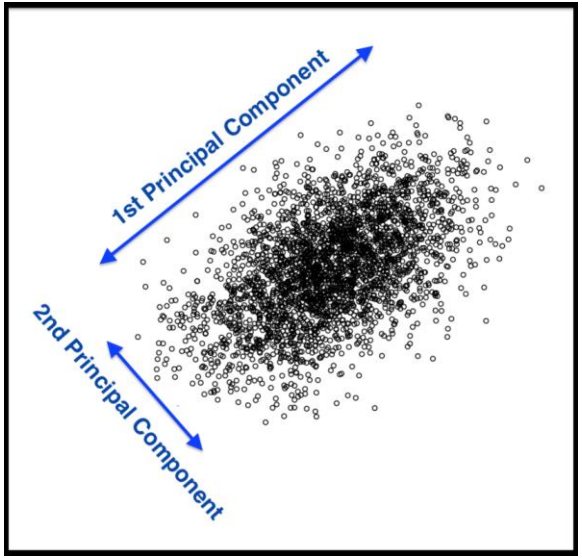
Outline

1. What is randomization?
2. Why is randomization needed?
3. Types of randomization
4. **Checking randomization in practice**
5. What can happen when randomization is not well-performed?
6. Conclusions

Nestlé Research

Checking randomization

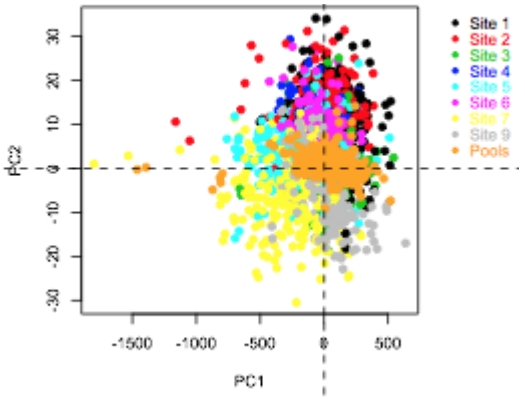
- **Principal Component Analysis (PCA) plots**
- Predicting batches from additional variables
- Checking balance between and within batches
- Checking results by order of processing / user
- Checking measurement errors / variability by batch



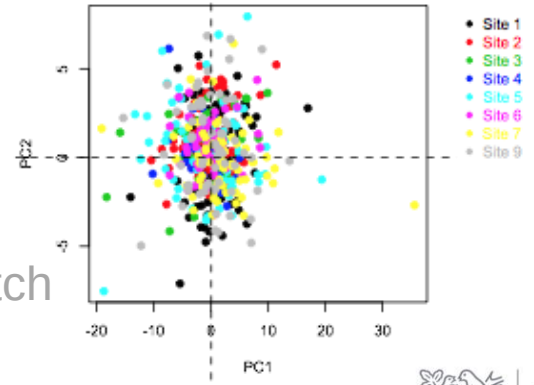
Checking randomization

- **Principal Component Analysis (PCA) plots**
- Predicting batches from additional variables
- Checking balance between and within batches
- Checking results by order of processing / user
- Checking measurement errors / variability by batch

Batch effect



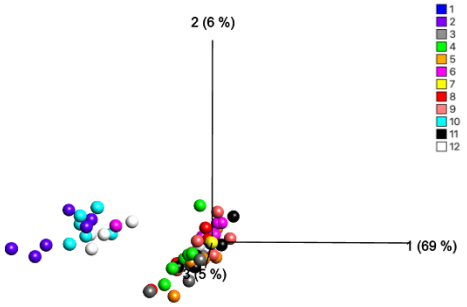
No batch effect

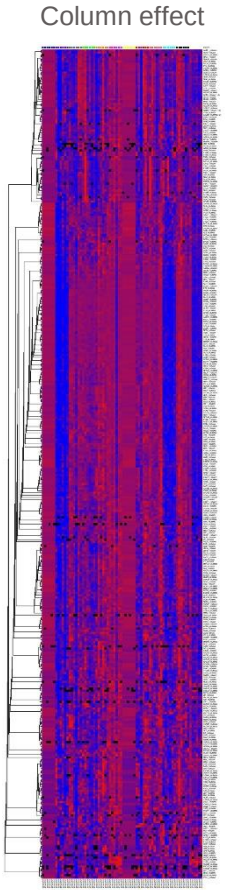


Nestlé Research

Checking randomization

- **Principal Component Analysis (PCA) plots**
- Predicting batches from additional variables
- Checking balance between and within batches
- Checking results by order of processing / user
- Checking measurement errors / variability by batch





Checking randomization

- Principal Component Analysis (PCA) plots
- **Predicting batches from variables**
- Checking balance between and within batches
- Checking results by order of processing / user
- Checking measurement errors / variability by batch

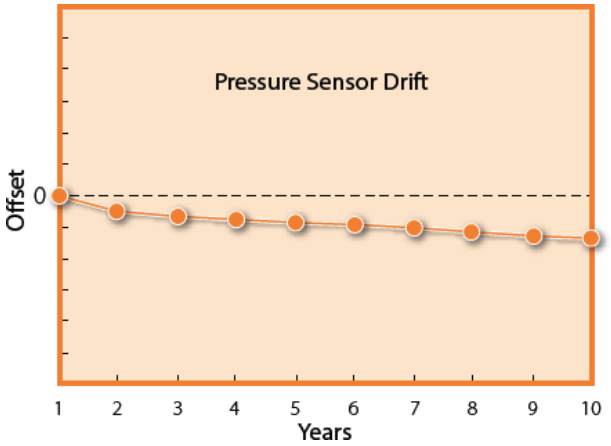
Nestlé Research

Checking randomization

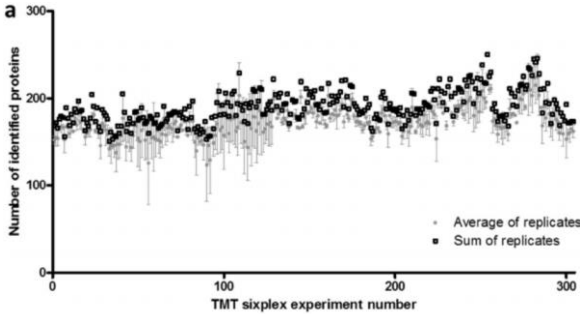
- Principal Component Analysis (PCA) plots
- Predicting batches from additional variables
- **Checking balance between and within batches**
- Checking results by order of processing / user
- Checking measurement errors / variability by batch

Checking randomization

- Principal Component Analysis (PCA) plots
- Predicting batches from additional variables
- Checking balance between and within batches
- **Checking results by order of processing / user**
- Checking measurement errors / variability by batch



https://www.solinst.com/products/dataloggers-and-telemetry/3001-leve logger-series/technical-bulletins/understanding-pressure-sensor-drift.php

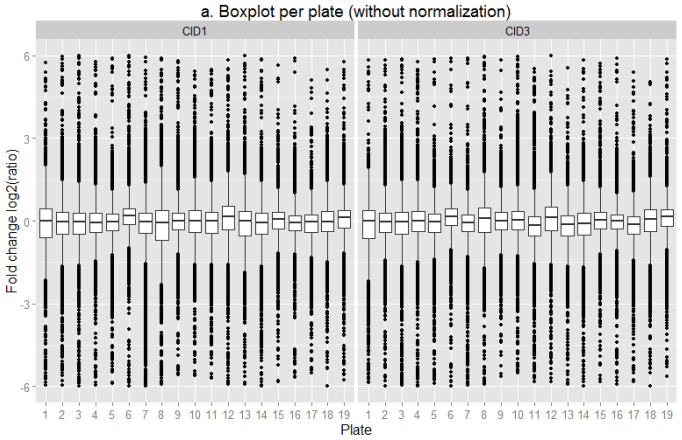


Cominetti et al. (2016) Proteomic Biomarker Discovery in 1000 Human Plasma Samples with Mass Spectrometry, *J. Proteome Res.* 2016, 15.

Nestlé Research

Checking randomization

- Principal Component Analysis (PCA) plots
- Predicting batches from additional variables
- Checking balance between and within batches
- Checking results by order of processing / user
- **Checking measurement errors / variability by batch**



Cominetti et al. (2016) Proteomic Biomarker Discovery in 1000 Human Plasma Samples with Mass Spectrometry, *J. Proteome Res.* 2016, 15.

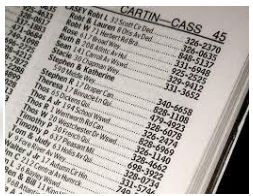
Nestlé Research

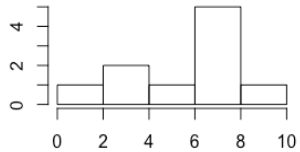
How do we use computers to randomize?

Computers use pseudo random number generators: sequence of numbers that are approximately random but are completely determined by an initial value: seed

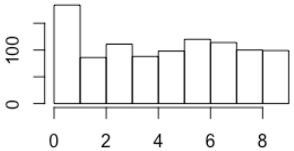
This allows reproducibility.

Low-tech random number generator:

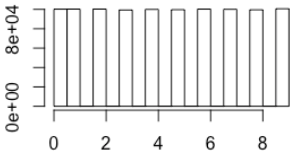




N = 10



N = 1'000



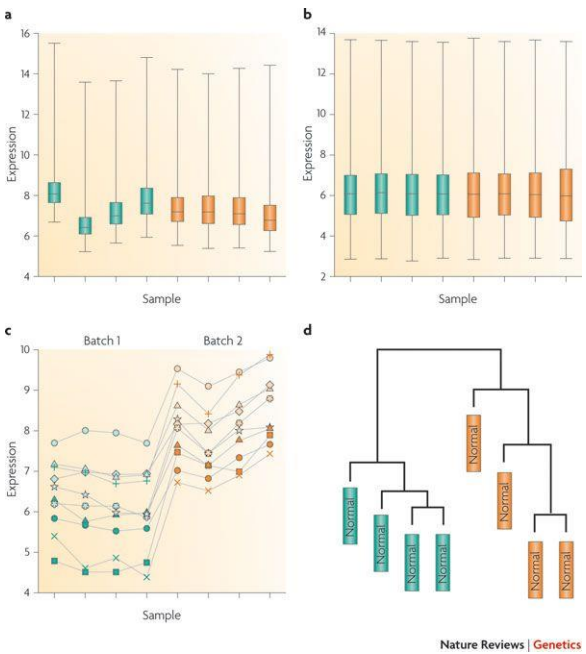
N = 1'000'000

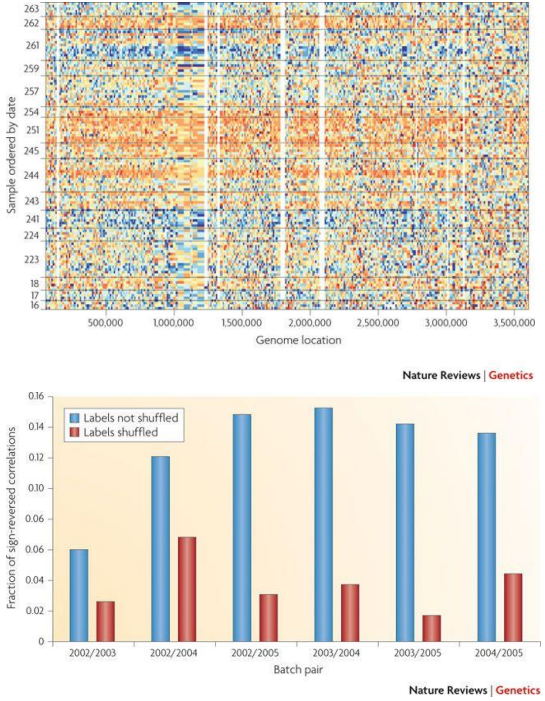
Outline

1. What is randomization?
2. Why is randomization needed?
3. Types of randomization
4. Checking randomization
5. **What can happen when randomization is not well-performed?**
6. Conclusions

Nestlé Research

What can happen when randomization is not well-performed?





Leek et al. (2010) Tackling the widespread and critical impact of batch effects in high-throughput data, *Nature Reviews Genetics* volume11, pages733–739.

Nestlé Research

Outline

1. What is randomization?
2. Why is randomization needed?
3. Types of randomization
4. Checking randomization
5. What can happen when randomization is not well-performed?
6. **Conclusions**

Nestlé Research

Conclusions

- Randomization is very important, and it should be performed **before** running the analyses!

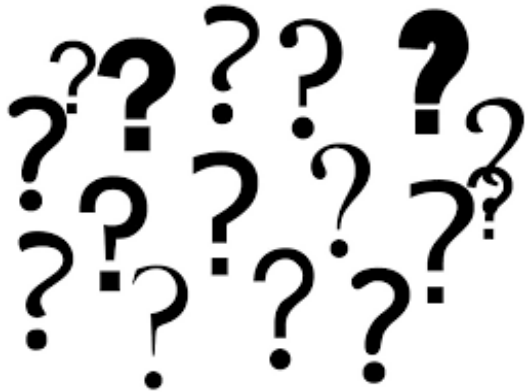
> “To consult the statistician after an experiment is finished is often merely to ask him to conduct a post mortem examination. He can perhaps say what the experiment died of.”
>
> Ronald Fisher

Conclusions

- Randomization is very important, and it should be performed **before** running the analyses!
- Randomization should be performed considering as many relevant variables as one can think of (date, site, gender, age, case/control, BMI, etc).
- After randomization has been performed, it should be checked that indeed batches are well randomized.
- Wish for the best! Even after randomizing there could still be still-unknown lurking (cachée / imprévue) variables that could invalidate the study and its conclusions.

Nestlé Research

Questions?



Nestlé Research