Statistical Testing and Sample Size Calculation Basics Part III

06.03.2024

Christina Abele
Center for Thrombosis and Hemostasis
University Medical Center, Mainz
christina.abele@uni-mainz.de

Agenda

- Multiple testing
- How to choose the right test
- Recap: Quantities influencing the test result
- How to report test results

Agenda

- **Multiple testing**
- How to choose the right test
- Recap: Quantities influencing the test result
- How to report test results

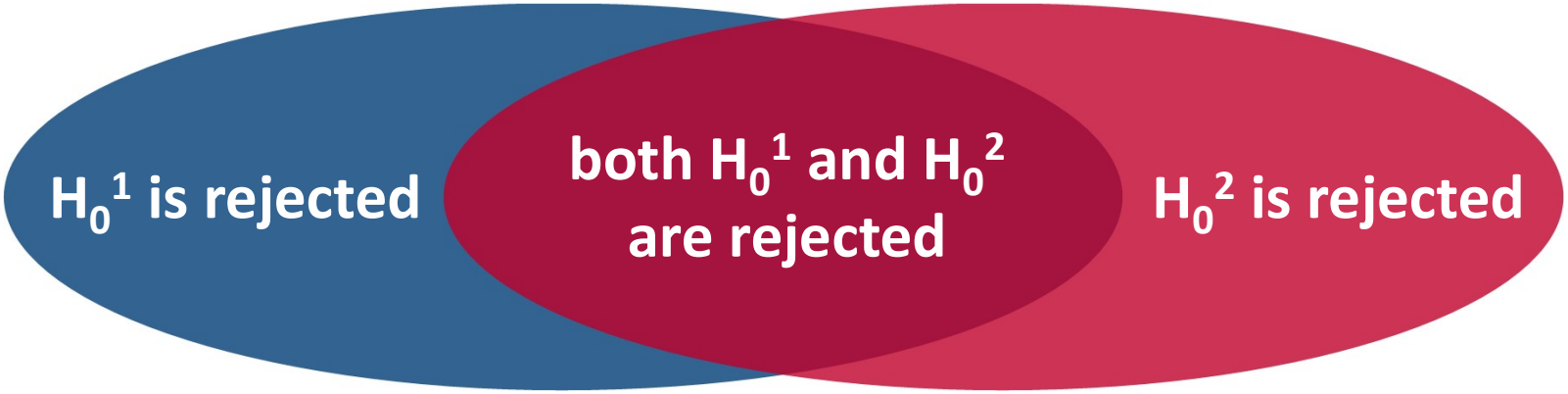
Multiple testing

What if you want to test several hypotheses at a time?

- Each time you perform a test, you run the risk of committing an error. So the total error risk increases with every additional test.
- Family of k null hypotheses H_0^1, ..., H_0^k:
  What kind of error do you want to control for?
- You usually want to control for the probability of rejecting **any** true null hypothesis (**FWER**: family-wise error rate) ...
  - ... if **all** null hypotheses are true (FWER in the **weak** sense) **or**
  - ... **no matter how many** null hypotheses are true (FWER in the **strong** sense)
  - Control of FWER in the strong sense implies control in the weak sense, but not vice versa!
- Exploratory testing: no issue

Multiple testing

Bonferroni correction



- Bonferroni inequality: FWER ≤ P (H_0^1 is rejected) + ... + P (H_0^k is rejected)
- So if P (H_0^i is rejected) < α/k for all i, FWER < α
  Testing all null hypotheses at level α/k leads to FWER control at significance level α in the stong sense.
- But: much too conservative!

Multiple testing

Holm-Bonferroni step-down procedure

- Improved Bonferroni correction
- Also controls for FWER in the strong sense
- Permits rejection of all the null hypotheses rejected by Bonferroni correction, and potentially more
- Procedure:
  - Sort hypotheses by p-values: $p^{(1)} \leq ... \leq p^{(k)}$ with corresponding $H_0^{(1)}, ..., H_0^{(k)}$
  - $p^{(1)} < \alpha/k$: reject $H_0^{(1)}$ and go on, $p^{(1)} \geq \alpha/k$: stop
  - $p^{(2)} < \alpha/(k-1)$: reject $H_0^{(2)}$ and go on, $p^{(2)} \geq \alpha/(k-1)$: stop
  - ...
  - $p^{(k)} < \alpha$: reject $H_0^{(k)}$ and go on, $p^{(k)} \geq \alpha$: stop
  - After a stop, none of the remaining null hypotheses can be rejected.

Multiple testing

Benjamini-Hochberg step-up procedure

- Controls for FWER in the strong sense **provided H_0^1, ..., H_0^k are independent or positively correlated**
- Permits rejection of all the null hypotheses rejected by the Holm-Bonferroni step-down procedure, and potentially more
- Procedure:
  - Sort hypotheses by p-values: $p^{(1)} \leq ... \leq p^{(k)}$ with corresponding $H_0^{(1)}$, ..., $H_0^{(k)}$
  - $p^{(k)} < \alpha$: reject $H_0^{(1)}$, ..., $H_0^{(k)}$ and stop, $p^{(k)} \geq \alpha$: don´t reject $H_0^{(k)}$ and go on
  - $p^{(k-1)} < \alpha/2$: reject $H_0^{(1)}$, ..., $H_0^{(k-1)}$ and stop, $p^{(k-1)} \geq \alpha/2$: don´t reject $H_0^{(k-1)}$ and go on
  - ...
  - $p^{(1)} < \alpha/k$: reject $H_0^{(1)}$, $p^{(1)} \geq \alpha/k$: don´t reject any null hypothesis

Multiple testing

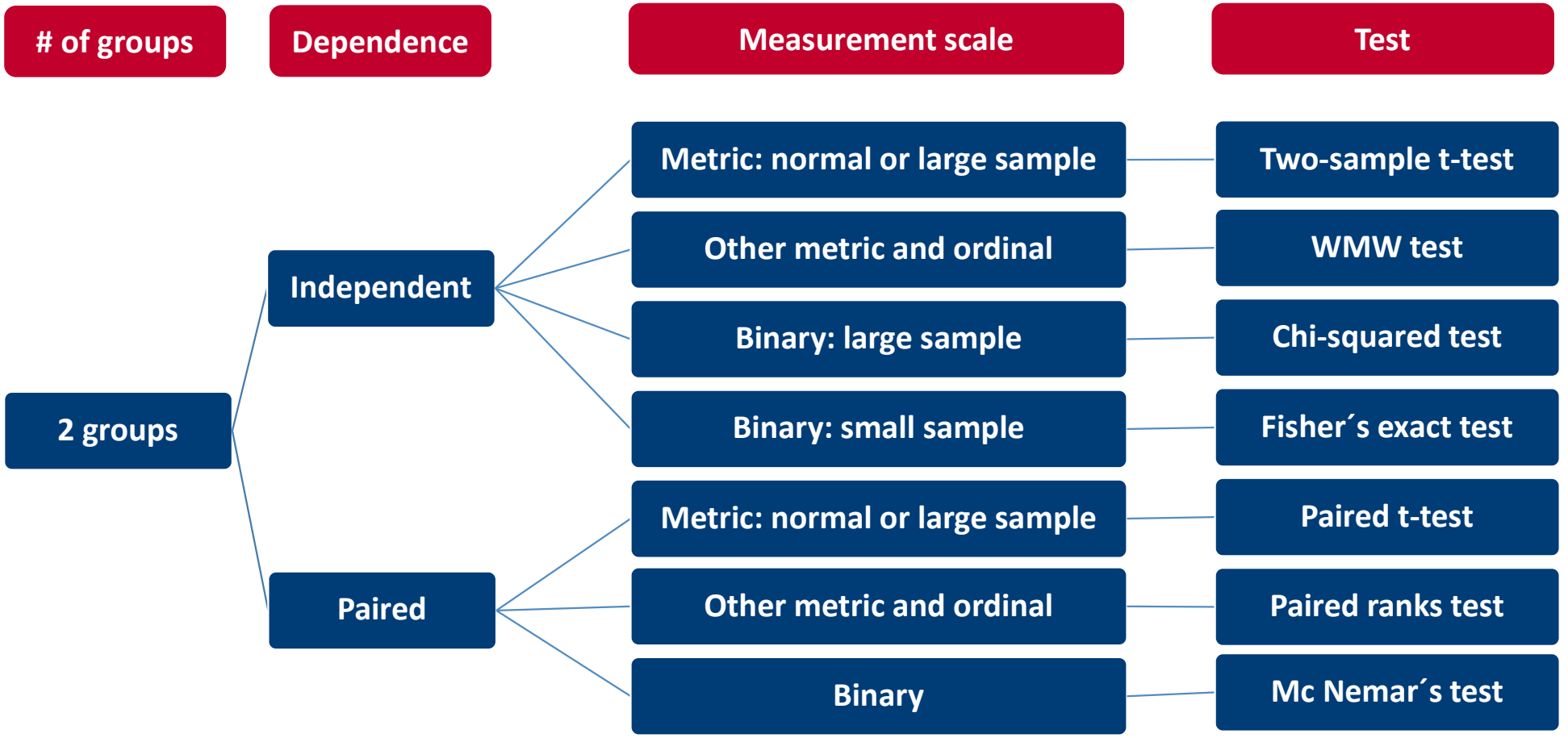
Hierarchical testing

- Controls for FWER in the strong sense
- Permits testing all hypotheses at the full level of α
- Procedure:
  - Fix a sequence of hypotheses $H_0^{(1)}$, ..., $H_0^{(k)}$ **in advance** depending on importance
  - $p^{(1)} < \alpha$: reject $H_0^{(1)}$ and go on, $p^{(1)} \geq \alpha$: stop
  - $p^{(2)} < \alpha$: reject $H_0^{(2)}$ and go on, $p^{(2)} \geq \alpha$: stop
  - ...
  - $p^{(k)} < \alpha$: reject $H_0^{(k)}$ and go on, $p^{(k)} \geq \alpha$: stop
  - After a stop, none of the remaining null hypotheses can be rejected.
- Considerable risk of rejecting hypotheses in spite of low p-values

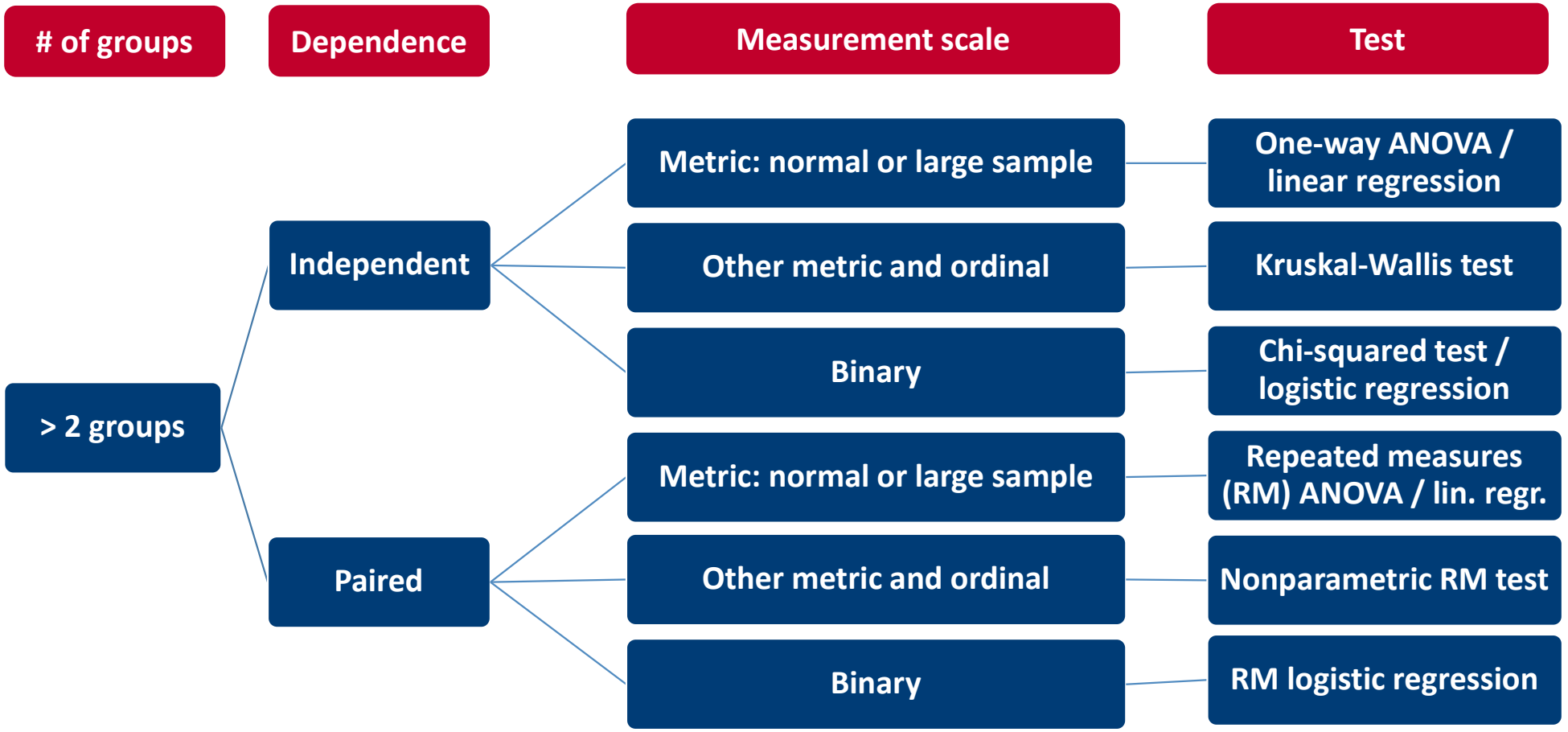
Agenda

- Multiple testing
- **How to choose the right test**
- Recap: Quantities influencing the test result
- How to report test results

How to choose the right test



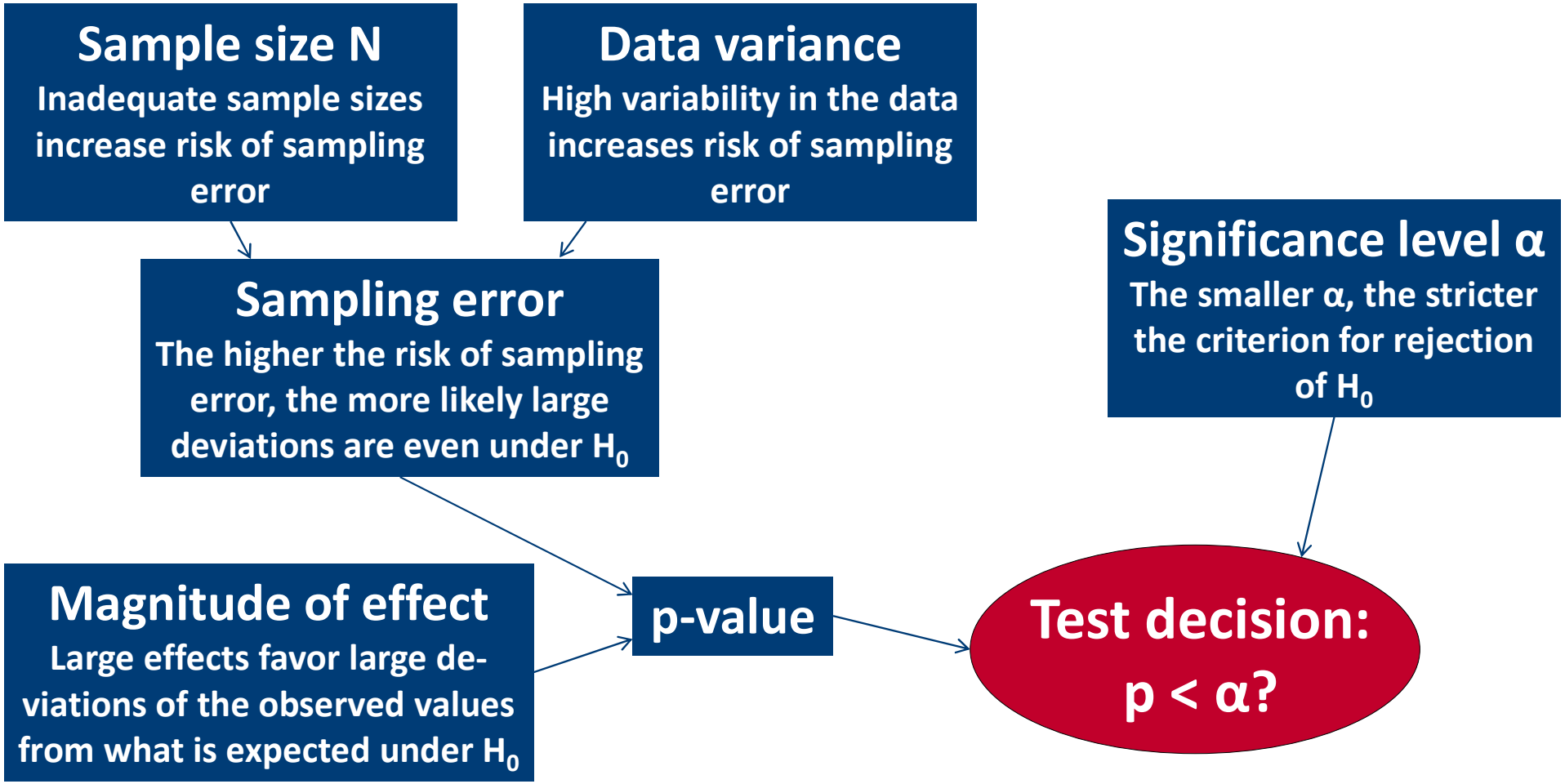
How to choose the right test



Agenda

- Multiple testing
- How to choose the right test
- **Recap: Quantities influencing the test result**
- How to report test results

Recap: Quantities influencing the test result



Agenda

- Multiple testing
- How to choose the right test
- Recap: Quantities influencing the test result
- **How to report test results**

How to report test results

CONsolidated **S**tandards **O**f **R**eporting **T**rials checklist:

- Statistical methods used to compare groups for primary and secondary outcomes
- Methods for additional analyses, such as subgroup analyses and adjusted analyses