



Technische Universität München



PROTECT
Behavioral Health Promotion &
Technology Lab

Evaluation Randomisiert-Kontrollierter Studien und Experimente mit R

Multiples Testen & Alphafehler-Korrektur

Prof. Dr. David Ebert & Mathias Harrer

Graduiertenseminar TUM-FGZ

Psychology & Digital Mental Health Care, Technische Universität München

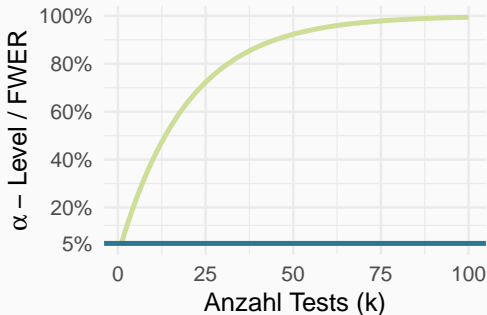
Multiples Testen

Zur Evaluation der Studienhypothesen ist in manchen Fällen **mehr als ein Test notwendig**. Zwei typische Szenarien sind:

- **Multiple primäre Endpunkte:** es wurden mehrere Outcomes gemessen (z.B. depressive Symptomschwere, Lebensqualität, Lungenfunktion, Gait Speed, ...), für die alle ein Interventionseffekt nachgewiesen werden soll.
- **Multiple Assessments des primären Endpunkts:** es gibt nur ein primäres Outcome, dieses wurde aber mehrfach gemessen, und der Interventionseffekt soll zu allen Zeitpunkten nachgewiesen werden.
- **Multiple Interventionsgruppen:** es gibt mehr als eine Interventionsgruppe, und es soll der Effekt aller untersuchten Interventionen im Vergleich zur Kontrollgruppe nachgewiesen werden.

Werden k multiple Tests mit einem nominalen Signifikanzniveau von (typischerweise) $\alpha = 0.05$ getestet, führt dies zur **Inflation der family-wise error rate** (FWER, Li et al., 2017; Vickerstaff et al., 2019):

$$P(\text{mindestens ein signifikantes Ergebnis}) = 1 - (1 - \alpha)^k$$



Die Bonferroni- und Holm-Bonferroni-Korrektur

Um eine Inflation falsch-positiver Befunde zu vermeiden und das nominale Signifikanzniveau trotz multipler Tests zu erhalten, ist daher eine α -Fehlerkorrektur notwendig.

Zwei häufig gewählte Verfahren sind die (überkonservative) **Bonferroni**- und (weniger strenge) **Holm-Bonferroni**-Korrektur (Holm, 1979). Diese lassen sich in R mit der `p.adjust`-Funktion anwenden:

```
p <- c(0.001, 0.023, 0.025)      # Sammle p-Werte multipler Tests
p.adjust(p, method = "bonferroni") # Bonferroni-Korrektur
p.adjust(p, method = "holm")      # Holm-Bonferroni-Korrektur
```

```
#> [1] 0.003 0.069 0.075
```

```
#> [1] 0.003 0.046 0.046
```

Wann muss korrigiert werden? (Li et al., 2017; Vickerstaff et al., 2019)

- **Multiple Outcomes:** eine Korrektur ist nur notwendig, wenn alle Outcome-Tests als *konfirmatorisch* betrachtet werden. Gibt es nur einen primären und mehrere sekundäre Endpunkte, ist keine Korrektur notwendig. Ergebnisse der sekundären Endpunkte sind dann aber nur explorativ, kein konfirmatorischer Nachweis!
- **Multiple Messzeitpunkte:** ist ein Messzeitpunkt als konfirmatorisch-primär präspezifiziert, ist keine Korrektur notwendig; ebenso bei Tests, bei denen über alle MZP *hinweg* modelliert wird (z.B. repeated-measures ANOVA, gemischte Modelle).
- **Multiple Interventionsgruppen:** eine Korrektur ist insbesondere dann notwendig, wenn die Interventionsgruppen “*verwandt*” sind (z.B. gleiches Medikament in unterschiedlicher Dosis).

Einseitig oder Zweiseitig?

Wir erinnern uns → Studienhypothesen bei Superiority Trials:

$$H_0 : \mu_1 \leq \mu_0 \Rightarrow \tau \leq 0$$

$$H_1 : \mu_1 > \mu_0 \Rightarrow \tau > 0$$

Dies impliziert eigentlich einen **einseitigen Test**, da wir nur testen wollen, ob μ_1 *größer* als μ_0 ist, nicht ob $\mu_1 \neq \mu_0$. “By default” ausgegebene Tests sind aber (häufig) zweiseitig (z.B. *t*-Tests)!

Unter **gewissen Voraussetzungen** kann eine einseitige Testung vorgenommen, bzw. α auf 10% erhöht werden (Dunnett & Gent, 1996; Knottnerus & Bouter, 2001):

- Das Ziel der Studie ist es nicht, **negative Effekte konfirmatorisch bestätigen zu können** (z.B. weil diese auf Basis vorheriger Forschung als unwahrscheinlich gelten).
- Wird kein positiver Effekt der Intervention festgestellt, so wird dies als die **Kontrollgruppe favorisierende Evidenz** ausgelegt.

Referenzen

- Dunnett, C. W., & Gent, M. (1996). An alternative to the use of two-sided tests in clinical trials. *Statistics in Medicine*, 15(16), 1729–1738.
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 65–70.
- Knottnerus, J. A., & Bouter, L. M. (2001). The ethics of sample size:: Two-sided testing and one-sided thinking. *Journal of Clinical Epidemiology*, 54(2), 109–110.
- Li, G., Taljaard, M., Van den Heuvel, E. R., Levine, M. A., Cook, D. J., Wells, G. A., Devereaux, P. J., & Thabane, L. (2017). An introduction to multiplicity issues in clinical trials: The what, why, when and how. *International Journal of Epidemiology*.
- Vickerstaff, V., Omar, R. Z., & Ambler, G. (2019). Methods to adjust for multiple comparisons in the analysis and sample size calculation of randomised controlled trials with multiple primary outcomes. *BMC Medical Research Methodology*, 19(1), 1–13.