

# Planung des verbliebenen Semesters

Termin	Thema
10.01. 14:15-15:30	Korrelierte Daten: ICC
10.01. 16:15-17:30	Systematische Literaturrecherche
17.01.	<i>Klaus Telkmann CART</i>
24.01. 14:15-15:30	Korrelierte Daten: Gemischte Modelle
24.01. 16:15-17:30	Reporting of observational studies
31.01. 14:15-15:30	Bootstrapping
31.01. 16:15-17:30	Qualitative Studien in der Epidemiologie

Studiengang M.Sc. Epidemiologie

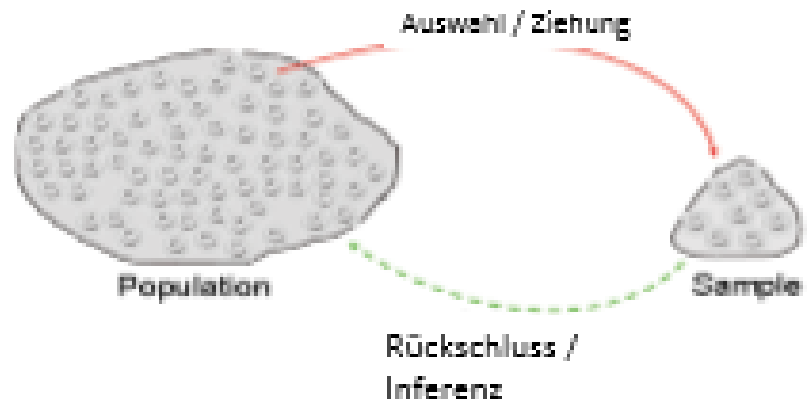
# Modul 7: Fortgeschrittene epidemiologische und statistische Methoden

WiSe 2021/2022

## **Korrelierte Daten 1**

Dozentin: Karin Bammann

- Ist die Stichprobe typisch („repräsentativ“) für die Grundgesamtheit?



# KIGGS-Studie: Wechsel des Stichprobendesigns

## Beispiel Klumpenstichprobe Prätestdesign von KIGGS

---

1. Ziehen von Schulen: (geschichtete (nach Schultyp) Zufallsauswahl)
2. Ziehen von Klassen in den ausgewählten Schulen
3. Vollerhebung der ausgewählten Klassen

Probleme, die zum Verwerfen des Designs geführt haben:

- Hoher Aufwand
- Klumpeneffekte: Schüler einer Klasse sind sich ähnlicher als Schüler einer einfachen Zufallsstichprobe.

# KIGGS: Stichprobendesign der Hauptstudie

- Zweistufiges Design:
  1. Stratifizierte Zufallsauswahl von Sample Points nach Bundesland und Gemeindegrößenklasse
  2. Zufallsstichproben in den Sample Points

→ Stratified multi stage probability sampling

Frage: Was passiert hier mit der Varianz?



# Unabhängigkeit der Beobachtungen

- Annahme der statistischen Standardverfahren:  
Beobachtungen sind i.i.d. (independent, identically distributed).
- Insbesondere gelte Unkorreliertheit:  
 **$\text{Corr}(y_j, y_k) = 0$  für jedes Beobachtungspaar**

# Unabhängigkeit der Beobachtungen in der Praxis

- Annahme i.i.d. häufig nicht gerechtfertigt, z.B.:
  - Follow-up der Probanden mit mehreren Messungen zu verschiedenen Zeitpunkten
  - Mehrere Beobachtungen pro Proband  
z.B. in der genetischen Epidemiologie
  - Geschichtete oder cluster-randomisierte Designs  
(z.B. Familien, Gemeinden, Krankenhäuser)
- Mögliche Lösungen:
  - Berechnen von Designeffekten
  - Modellierung mittels gemischter linearer Modelle

# Designeffekt (Kish 1965)

$$D_{eff} = \frac{s^2(\theta)}{s^2(\hat{\theta})_{srs}}$$

Proportion zweier (theoretischer) Varianzen



# Approximation des Designeffekts

- Lösung abhängig vom Stichprobendesign
- Für clusterrandomisierte Stichproben mit  $m$  Beobachtungen pro Cluster und ICC Intraklassenkorrelationskoeffizient gilt approximativ:

$$D_{eff} = (m - 1) * ICC + 1$$

# Intraklassenkorrelationskoeffizient (ICC)

- Entwickelt zur Beurteilung von Interrater-Agreement (statt Cohens Kappa)
- Man unterscheidet 6 verschiedene Arten (je nach Design des Experiments)
- Übertragung auf unseren Fall:
  - ICC (3,1)
  - Adjustierung für unterschiedliche Clustergrößen

# Intraklassenkorrelationskoeffizient (ICC)

Formel zitiert nach Smeeth & Ng 2002:

## *Statistical methods*

We used the methods presented by Donner and Klar that were also used in a recently published paper of ICCs from maternal and child health [3,47]. Eq. (1) given above for  $\rho$  represents a hypothetical situation in which both clusters and individuals within clusters are drawn at random from a population. An estimate of  $\rho$  ( $\rho_1$ ) can be obtained by performing a one-way analysis of variance (ANOVA) [50]. This method is valid for both binary and continuous outcomes [39,50]:

$$\rho_1 = (MS_b - MS_w) / (MS_b + (m - 1)MS_w)$$

where  $MS_b$  and  $MS_w$  are the mean squares from the ANOVA table for between and within clusters and again  $m$  is the average size of the cluster.

Because the size of each cluster varied, we substituted  $m_0$  for  $m$  [51]:

$$m_0 = (1/(k - 1))(n - (\sum m_j^2/n))$$

where  $n$  is the total sample size,  $k$  is the number of clusters and  $m_j$  is the cluster size in the “ $j$ th” cluster with  $j$  ranging from 1 to  $k$ . The standard errors for each ICC estimate were calculated using the technique derived by Donner and Koval [52]:

*L. Smeeth, E.S.-W. Ng/Controlled Clinical Trials 23 (2002) 409–421*

413

$$\text{variance}(\rho_1) = 2(1 - \rho_1)^2 [1 + (n - 1)\rho_1]^2 / n(n - 1)k$$

where  $n$  is the harmonic mean of the number of participants per cluster and again  $k$  is the number of clusters.

Because the ICC is often small and clusters are often fairly large, significance tests of the ICC have unacceptably low power. Undertaking a significance test that  $\rho$  differs from 0 and

# Vorgehen Designeffekte

- Berechnen von Intraklassenkorrelationskoeffizienten
  - Daten sind publizierbar (siehe Beispiel unten)
- Berechnen von Designeffekten
- Vornehmen von Varianzkorrekturen

## Intraclass Correlation Coefficients Typical of Cluster-Randomized Studies: Estimates From the Robert Wood Johnson Prescription for Health Projects

David M. Thompson, PhD<sup>2</sup>

Douglas H. Fernald, MA<sup>2,3</sup>

James W. Mold, MD, MPH<sup>4</sup>

<sup>1</sup>Department of Biostatistics and Epidemiology, University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma

<sup>2</sup>Department of Family Medicine, Uni-

### ABSTRACT

**PURPOSE** Researchers who conduct cluster-randomized studies must account for clustering during study planning; failure to do so can result in insufficient study power. To plan adequately, investigators need accurate estimates of clustering in the form of intraclass correlation coefficients (ICCs).

**METHODS** We used data for 5,042 patients, from 61 practices in 8 practice-based research networks, obtained from the Prescription for Health program,

# EpiSleep: Hintergrund

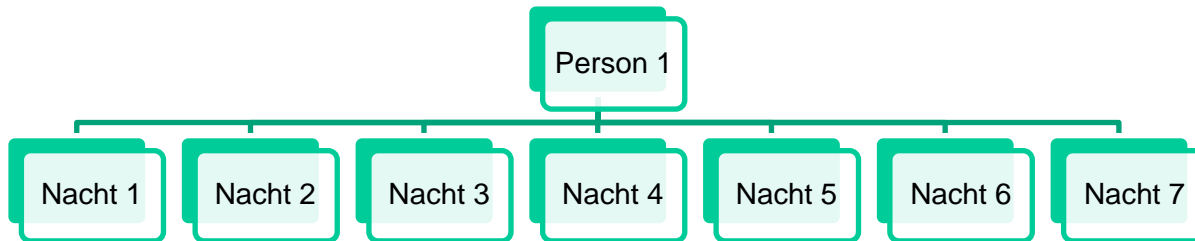
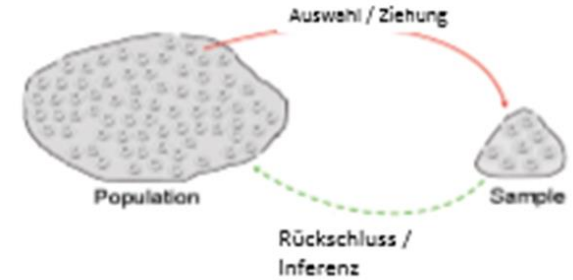
- **Seminar:** Lebensstilbedingte Erkrankungen
- **Thema:** Schlaf und Gesundheit
- **Fragestellung:** Bestimmung des Nachtschlafs (Dauer, Qualität) kann mit unterschiedlichen Erhebungsmethoden durchgeführt werden. Vergleichbarkeit der Ergebnisse?



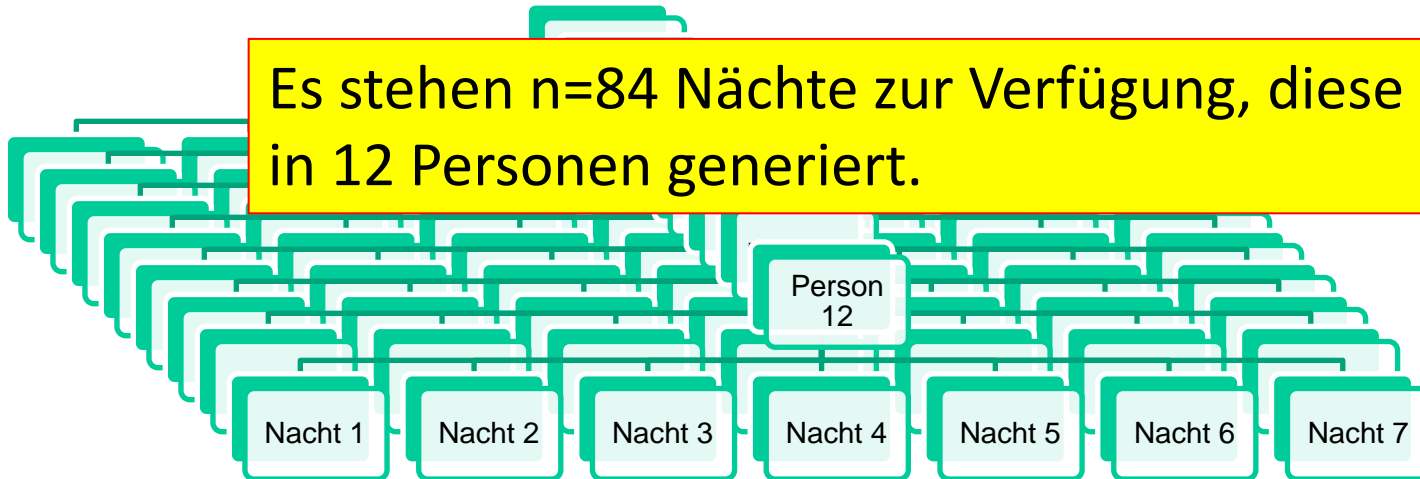
# EpiSleep: Design

- Kalibrierungsstudie
  - Akzelerometer
  - Smartphone-App
  - Kurzprotokoll
  - Langprotokoll für 1 Nacht
- Konsekutive Messung (7 Nächte)
  - Seminarteilnehmer\*innen (n=12)

# EpiSleep: Beobachtungen



Es stehen  $n=84$  Nächte zur Verfügung, diese Daten wurden in 12 Personen generiert.



# Übung

- Berechnen Sie den ICC und den Designeffekt zur Schlafdauer und interpretieren Sie die Ergebnisse.
- Hinweise
- Untersuchungseinheit EpiSleep sind Nächte, nicht Probanden
  - Beispiel KIGGS: Clustereinheit: Gemeinden; Untersuchungseinheit Probanden
  - **EpiSleep: Clustereinheit: Probanden; Untersuchungseinheit Nächte**
- Problem Cluster-Stichprobe:
  - Unterschätzung der Varianz, da sich Beobachtungen eines Clusters ähnlicher sind (sein können) als in einer normalen Zufallsstichprobe.
  - Frage: Wie prüft man das???



# Verwendung des berechneten Designeffekts

- Zur Varianzkorrektur der vorliegenden Studie, z.B. bei Konfidenzintervallen:

$$\dots 1,96 * D_{eff} * ..$$

- Für Fallzahlkalkulationen zukünftiger ähnlich gelagerter Studien\*, effektive Stichprobengröße:

$$n/D_{eff}$$

PAUSE bis 16:15

# Literaturrecherchen

- Mögliche Ziele
  - Fragestellungen identifizieren
  - Forschung in den Kontext betten
- Unsystematische Recherche
  - Schneeballsystem
- Systematische Recherche; hier hilfreich neuere Methoden zu systematischen Reviews
  - Scoping reviews
  - Rapid reviews

# Materialien zu Literaturrecherchen und Systematic reviews

- Cochrane, z.B.
    - [Lunch & Learn \(Kurzwebinare\) | Cochrane Deutschland](#)  
Videos zu Pubmed, MESH, Dokumentation etc.
  - Joanna Briggs Institute (JBI):
    - Software (JBI Sumari)
    - Webinare: [JBI LIVE Webinars – YouTube](#)
    - Ask JBI: [Ask JBI... – YouTube](#)
- Anschauen: 2. (Types of SR) und 3. Folge (Scoping Review) evtl. 5. Folge (Rapid Reviews)

Studiengang M.Sc. Epidemiologie

# Modul 7: Fortgeschrittene epidemiologische und statistische Methoden

WiSe 2021/2022

## **Korrelierte Daten 2**

Dozentin: Karin Bammann

# Unabhängigkeit der Beobachtungen in der Praxis

- Annahme i.i.d. häufig nicht gerechtfertigt, z.B.:
  - Follow-up der Probanden mit mehreren Messungen zu verschiedenen Zeitpunkten
  - Mehrere Beobachtungen pro Proband  
z.B. in der genetischen Epidemiologie
  - Geschichtete oder cluster-randomisierte Designs  
(z.B. Familien, Gemeinden, Krankenhäuser)
- Mögliche Lösungen:
  - Berechnen von Designeffekten
  - Modellierung mittels gemischter linearer Modelle

# Gemischte lineare Modelle (GLMM)

- Erweiterung der GLM zu GLMM durch Einführung von Zufallseffekten:

$$\text{GLM: } f^{-1}(y) = \beta X + \varepsilon$$

Random effect

$$\text{GLMM: } f^{-1}(y) = \beta X + \gamma Z + \varepsilon$$

- Diese Erweiterung wird mittlerweile häufig benutzt um Clusterdesigns und multizentrische Studien zu analysieren.

Fixed effect

# Generalized linear mixed models (GLMM)

- Possibilities for random effects:
  - Random intercepts
  - Random slopes
  - (Combination of random intercepts AND random slopes)



# Beispiel: GLMM mit random intercept

- Hier: b Behandlung; u Klinik
- Modellgleichung:

$$\text{logit}(p_{ij} | u_i) = b_0 + b_1 x_{ij} + u_i$$

- Nun gilt: Gegeben  $u_i$  sind die Beobachtungen einer Klinik unabhängig.
- Interpretation: Identischer Behandlungseffekt in allen Kliniken, unterschiedliches Baseline zwischen den Kliniken

Random effect

Fixed effect

# Beispiel: GLMM mit random slope

- Hier: b Behandlung; u Klinik
- Modellgleichung:

$$\text{logit}(p_{ij} | u_i) = b_0 + b_1 x_{ij} + u_i x_{ij}$$

- Nun gilt: **Gegeben  $u_i$  sind die Beobachtungen einer Klinik unabhängig.**
- Interpretation: Unterschiedlicher Behandlungseffekt in allen Kliniken, **identisches Baseline-Risiko\*** zwischen den Kliniken

\* Unterschied zur Stratifikation

# Ausblick

- Wahl der Methode
  - Datenstruktur
  - Gewünschte Interpretation:
    - Unterschiedliches Baseline-Risiko
    - Unterschiedliche Treatment-Effekte

# Übung

- Berechnen Sie verschiedene gemischte Modelle GLMM für den Datensatz EPISleep.
- Vergleichen Sie diese mit den Ergebnissen von GLM.

# SAS Prozeduren

The screenshot displays a web browser window at the URL support.sas.com/documentation/cdl/en/statug/68162/HTML/default/viewer.htm#statug\_intromix\_sect001.htm. On the left, a vertical navigation pane lists various SAS topics, including 'Nonlinear Mixed Models: The NLMIXED Procedure' and 'Introduction to Bayesian Analysis Procedures'. The main content area begins with a paragraph about the historical development of mixed models, followed by a section titled 'The following procedures in SAS/STAT software can perform mixed and random effects analysis to various degrees:'. Below this, several procedures are listed with brief descriptions: GLM (General Linear Model), GLIMMIX (Generalized Linear Mixed Models), HPMIXED (Hierarchical Partially Matched Information eXtraction), LATTICE (Lattice Design Analysis), MIXED (Mixed-Effects Models Using Eigen and Variance Components), NESTED (Nested Random Effects Models), NLMIXED (Nonlinear Mixed Models), and VARCOMP (Variance Components). Each procedure's description highlights its primary application and key features. At the bottom of the main content area, a summary sentence states: 'The focus in the remainder of this chapter is on procedures designed for random effects and mixed model analysis: the GLIMMIX, HPMIXED, MIXED, NESTED, NLMIXED, and VARCOMP procedures.' Navigation links for 'Previous Page', 'Next Page', and 'Top of Page' are provided below the text. The browser's address bar and standard toolbar are visible at the top. A blue footer banner contains three columns of links: 'Connect', 'Get Help', and 'Account Management'. The Windows taskbar at the very bottom shows the system clock as 14:16 on 07.12.2015.

# SAS Umsetzung: Random intercept mit linearem Link

```
proc glimmix data=sleep;  
model TotalSleeptime=waso / solution;  
random intercept / subject=SubjectName  
solution;  
run;
```

## .. mit logistischem Link

```
proc glimmix data=sleep;  
model 'binäre Variablen' = .. / link=logit;  
random intercept;  
run;
```

# SAS Umsetzung: Random slope mit linearem Link

```
proc glimmix data=sleep;  
model TotalSleeptime=waso / solution;  
random waso / subject=SubjectName solution;  
run;
```



PAUSE bis 16:15

# Reporting of observational studies

- Abstracts
  - STROBE
  - Article series: Abstracts & Titles
  - Book chapter: Language do's and don'ts
- IMRaD / Weinglas
- Storylines and take-home messages

# Structure of scientific papers: IMRAD

- Introduction
- Methods
- Results
- And Discussion



# Introduction

- Teaser (one sentence)
- State of the art → Research gap → Research question of the paper
  - One page often sufficient
  - 10 to 20 references (give brief but complete overview)
  - Acceptable literature: All you can find in PubMed
  - Limited acceptability (max. 2 per paper, mostly in methods section): books
  - Do not use / only for very experienced & established authors:  
Non-scientific literature

# Methods

- Sample description
  - Population
  - Sampling procedure, Time, Response
- Measurements
  - Questionnaires
  - Examinations / Equipment
- Statistics
  - Procedures
  - Software

# Methods

- Enables reader to replicate your research
- Include any quality indicators and quality measures taken that would increase confidence in the quality of your work
- Often longest part of the manuscript

# Results

- Shortest part of a paper
- Tables and text
  - Do not double the information
  - Tables should be understandable without text and vice versa
- Tables and figures
  - DO NOT double the information
- KISS – Keep it short and simple
- Strictly neutral language

# Tables & Figures

- Do not double content
- Max. number often given by journal „Instructions for authors“
- Give titles (& subtitles) (ABOVE tables and BELOW figures)
- Introduce all abbreviations (even if already given in text)
- Use SI units
- Tables/Figures should be readable without reading the text
- Tables: Do not just copy from your statistics program



# Results

- Task: Deliver the key messages of your results
- Avoid long and confusing sentences
- Remain focused
- Start by describing your study subjects
- Provide data relevant to the research questions
- Make your point with data, not arguments

# Discussion

- Results (short)
- How are your results fitting into the literature?
- Strengths and limitations
- Implications of the research

# Story line & Take home message: Example

The body weight of my child is ..	Child has normal weight	Child with obesity	Total
<i>About right</i>	60 (75.0%)	16 (80.0%)	76 (76.0%)
<i>Too high</i>	20 (25.0%)	4 (20.0%)	24 (24.0%)
Total	80 (100%)	20 (100%)	100 (100%)

# Published papers

- Add to your CV (especially: international peer-reviewed, high impact journal).
  - ORCID, SCOPUS etc.
- Enhance visibility of your research and your research profile.

# Publishing a paper

- Drafting the manuscript
- Sending it to co-authors for comments
- Sending it to the editor of a journal (include cover letter)
- Editor pre-checks the manuscript
- If OK, editor sends manuscript to reviewers
  - Reviewers: Experienced researchers (usually post-docs) from scientific community
- Reviewers read the paper and write a review report:
  - Accept, accept with minor/major changes, reject the paper
- Editor sends review reports to the authors.
- Authors revise the manuscript and prepare a rebuttal letter (point-by-point reply).
- Editor pre-checks the revision ..

# Instructions for authors

- Follow them!