

## Grundlagen sozialwissenschaftlicher Meta-Analysen

Dr. Bernd Weiß  
Prof. Dr. Michael Wagner

Forschungsinstitut für Soziologie  
Universität zu Köln

[bernd.weiss@uni-koeln.de](mailto:bernd.weiss@uni-koeln.de)  
[mwagner@wiso.uni-koeln.de](mailto:mwagner@wiso.uni-koeln.de)

13. / 14. Dezember 2012

## Teil I

# Übersicht und Vorbemerkungen



# Ziele des Workshops

1. Zentrale meta-analytische Konzepte und Ablauf einer Meta-Analyse verstanden haben.
2. Publizierte Meta-Analysen nachvollziehen und kritisch beurteilen können.
3. Vorgehen und Aufwand einer eigenen Meta-Analyse einschätzen können.



# Konzeption des Workshops

- Aufbau folgt grob dem Ablauf einer publikationsbasierten Forschungssynthese/Meta-Analyse.
- Mischung zwischen (Dozenten-)Vortrag und praktischen Lerneinheiten.
- Für die praktischen Übungen wird das Statistikpaket R benutzt (es wird eine kurze Einleitung geben).<sup>1</sup>
- Folien folgen nicht dem  $5 \times 5$ -Schema, teilweise Skriptcharakter.
- Übersicht statt Details.
- Medizinische/biostatistische Perspektive auf Meta-Analyse.

---

<sup>1</sup>Es werden aber auch Stata-Beispiele gezeigt.



# Materialien des Workshops

- Alle Materialien lassen sich von der folgenden Website herunterladen: <http://www.berndweiss.net/teaching.html>.
- Alle Materialien stehen unter einer Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Germany Lizenz.
- Die Verzeichnisstruktur ist wie folgt:

/	
assign.....	Unterlagen für Übungen
data.....	Alle Datensätze
fig.....	Abbildungen
org.....	(Öffentliche) Orga-Materialien
ref.....	Literatur (PDFs)
slide.....	Folien des Workshops
src.....	R- und Stata-Analysecode
tab .....	Tabellen, Ausgabedateien



# Zeitplan Workshop Tag 1

Zeit	Dauer	Thema
10:30-12:15	105	Begrüßung, Einführung, Literaturrecherche und -organisation <i>Mittagspause</i>
13:30-15:00	90	Datenerfassung, -organisation und Effektstärken I
15:00-15:15		<i>Pause</i>
15:15-16:30	75	Effektstärken II und Effektstärkenverteilungen I (Fixed-effect model)
16:30-16:40		<i>Pause</i>
16.40-18:15	95	Einführung in R/Stata; praktische Übungen zu den bisherigen Themen
ab 18:15		Kölsch/Wine and Cheese

# Zeitplan Workshop Tag 2

Zeit	Dauer	Thema
09:00-10:30	90	Effektstärkenverteilungen II (Random-effects model); praktische Übungen
10:30-10:45		<i>Pause</i>
10:45-12:15	90	Heterogenität, Meta-Regression, Publication bias
12:15-13:30		<i>Mittagspause</i>
13:30-15:00	90	Praktische Übungen mit R/Stata
15:00-15:15		<i>Pause</i>
15.15-16.00	45	Meta-Analyse mit Individualdaten (IPD)
16.00-16:45	45	Fragen, Diskussion & Feedback



# Unser Interesse an Meta-Analyse I

- Wagner, M. & Weiβ, B. (2003). Bilanz der deutschen Scheidungsforschung. Versuch einer Meta-Analyse. *Zeitschrift für Soziologie*, 32(1), 29–49
- Wagner, M. & Weiβ, B. (2006a). Meta-Analyse als Methode der Sozialforschung. *Methoden der Sozialforschung. Sonderheft 44 der Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 479–504
- Wagner, M. & Weiβ, B. (2006b). On the Variation of Divorce Risks in Europe: Findings from a Meta-Analysis of European Longitudinal Studies. *European Sociological Review*, 22(5), 483–500
- Weiβ, B. & Wagner, M. (2008). Potentiale und Probleme von Meta-Analysen in der Soziologie. *Sozialer Fortschritt*, 57(10-11), 250–255



# Unser Interesse an Meta-Analyse II

- Weiß, B. (2008). *Meta-Analyse als Forschungssynthese in der Soziologie. Dargestellt anhand zweier Fallbeispiele zum Schulabsentismus. Inauguraldissertation zur Erlangung des Doktorgrades der Wirtschafts- und Sozialwissenschaftlichen Fakultät der Universität zu Köln.*
- Weiß, B. & Wagner, M. (2011). The identification and prevention of publication bias in the social sciences and economics. *Journal of Economics and Statistics*, 231(5-6), 661–684
- Thompson, C. & Weiß, B. (2013). An introduction to basic meta-analysis. In B. J. Becker (Hrsg.), *Synthesizing correlations and regressions: Methodology and applications*. Unpublished manuscript.



# Unser Interesse an Meta-Analyse III

## Gegenwärtige/Zukünftige Projekte:

- How employment status affects partnership stability: A meta-analysis using longitudinal individual person data
- Untersuchung zum Publication Bias am Beispiel von drei dt. sozialwissenschaftlichen Zeitschriften (KZfSS, PVS, ZFS)
- Simulationsstudie zu den Auswirkungen von unterschiedlich spezifizierten EHA Regressionsmodellen im Rahmen einer Meta-Analyse
- Fortsetzung der Meta-Analyse zur Ehescheidung



## **2. Einführung und Motivation**

- 1. Notwendigkeit von Forschungssynthesen**
- 2. Begriffliche und konzeptionelle Grundlagen**
- 3. Vorteile von Forschungssynthesen**
- 4. Kritik an Forschungssynthesen und Meta-Analysen**



# **3. Forschungssynthese als sozialwissenschaftliche Methode**

## **5. Der sozialwissenschaftliche Forschungsprozess**

### **6. Forschungsfrage**

### **7. Datenerhebung: Literatur- und Studienrecherche**

Recherchestrategien

Bewertung und Auswahlkriterien

Dokumentation der Recherche und Verwaltung der Quellen

Stichprobenselektivität

### **8. Datenvercodung/-eingabe und Effektstärken**

### **9. Effektstärken (ES)**

Effektstärken der  $r$ -Familie

Effektstärken der  $d$ -Familie

Kategoriale Effektstärken

Regressionskoeffizienten

Effektstärkenkonvertierung



# **4. Meta-Analyse**

## **10. Befundsynthese von Aggregatdaten**

Fixed-effects model (FEM)

Random-effects model (REM)

## **11. Begriff der Heterogenität und Diagnostik**

Überblick

Die Zwischenstudienvarianz

Die Q- Statistik

Weitere Heterogenitätsmaße:  $H^2$  und  $I^2$

Zusammenfassung Heterogenitätsmaße

## **12. Heterogenitätsaufklärung**

Meta-Regression

ANOVA-ähnliche Verfahren

Weitere Bemerkungen zu Verfahren der Heterogenitätsaufklärung

## **13. Publication Bias**

## **14. Reporting Standards**

## **15. Ein Ausblick auf Meta-Analysen mit Individualdaten**

## **16. Literatur**

# 5. Anhang

## 17. Meta-Analyse-Software für R und Stata

## 18. R in 20 Minuten

## 19. Statistische Grundlagen

Der Standardfehler

Wahrscheinlichkeitsverteilungen und Q-Test



# Teil II

# Einführung



# Einführung

1. Notwendigkeit von Forschungssynthesen
2. Begriffliche und konzeptionelle Grundlagen
3. Vorteile von Forschungssynthesen
4. Kritik an Forschungssynthesen und Meta-Analysen



# Kapitelübersicht

1. Notwendigkeit von Forschungssynthesen
2. Begriffliche und konzeptionelle Grundlagen
3. Vorteile von Forschungssynthesen
4. Kritik an Forschungssynthesen und Meta-Analysen



# Gründe für die Durchführung einer Forschungssynthese

- Unsicherheit über die Wirksamkeit von Maßnahmen (Interventionen) (Politik, Schule, Medizin etc.).
- Fehlender Gesamtüberblick über einen bestimmten (wissenschaftlichen) Forschungsbereich, um sinnvoll weitere (methodische wie inhaltliche) Forschung zu initiieren.
- ...



# "Wächst der Wissenschaft das Wissen über den Kopf?"

- „But today we are experiencing a crisis of faith; many of us no longer feel sure that science, though growing explosively, is moving inexorably toward the truth. [...] increasingly chaotic output of contemporary research“ (Hunt 1997: 1).
- „Literaturflut - Informationslawine - Wissensexpllosion. Wächst der Wissenschaft das Wissen über den Kopf?“ (Marx / Gramm 2002)
- ...

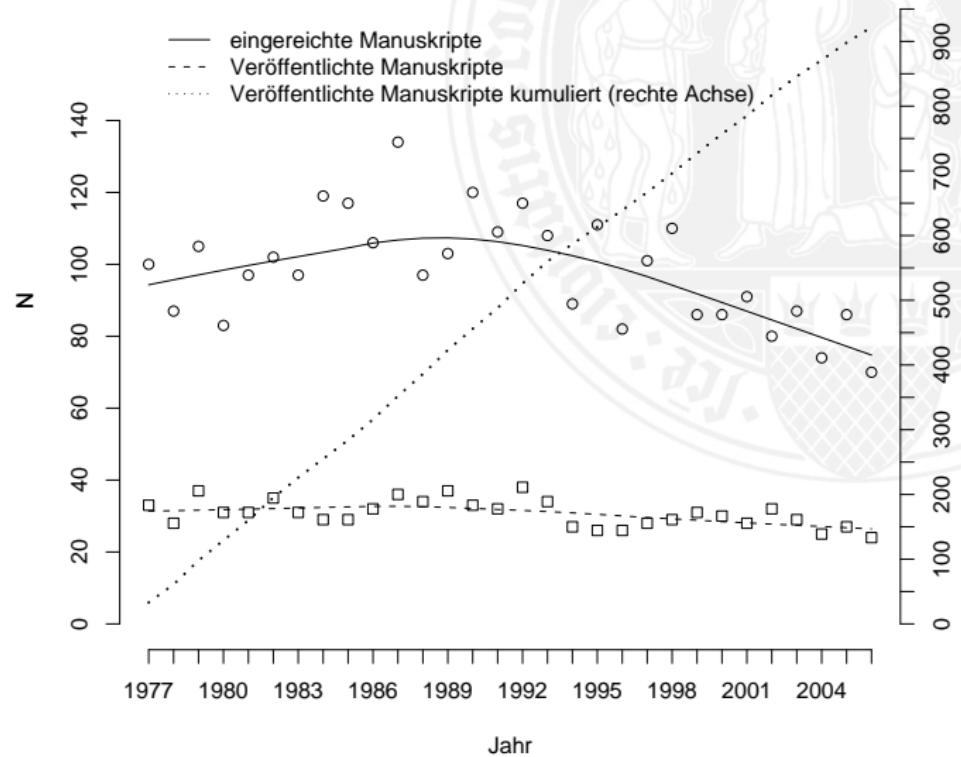


# Indikatoren für eine „Informationslawine“ in den empirischen Sozialwissenschaften

- Das Kölner Zentralarchiv für empirische Sozialforschung erfasst 150 bis 200 Studien p.a. (pers. Kommunikation mit Oliver Watteler vom 2. Mai 2008)
- 120 Publikationen zum Trennungs-/Ehescheidungsrisiko in Europa (Wagner/Weiβ 2006)
- 17 Publikationen und 20 Originaldatensätze über 24 Städte in NRW zu Wanderungsmotiven (Bleck/Wagner 2006)
- 15 Individualdatensätze mit Angaben von 80 000 Jugendlichen zum Schulschwänzen (Weiβ 2008)
- ...



# Eingereichte und veröffentlichte Manuskripte in der Kölner Zeitschrift für Soziologie und Sozialpsychologie



# Wie lassen sich große Mengen an Forschungsbefunden aufbereiten?

1. **Bibliographien** („Fischer, Sonja (Hg.), 2005: Schulmüdigkeit und Schulverweigerung. Eine annotierte Bibliografie für die Praxis. München: DJI“)
2. **Narrative Reviews** („Blumel, Susan R., 1992: Explaining Marital Success and Failure. S. 1-114 in: Bahr, Stephan J. (Hg.), Family Research. A Sixty Year Review, 1930-1990. New York: Lexington Books.“)
3. **Systematische Reviews / Forschungssynthesen**
  - Qualitative Ausrichtung
  - Quantitative Ausrichtung (Meta-Analyse)



# Grenzen von Bibliographien und narrativen (qualitativen) Reviews

- Publikationsauswahl ist häufig unsystematisch und intersubjektiv nicht nachvollziehbar (*confirmation bias*).
- Kein/Unzureichender Umgang mit *widersprüchlichen* Befunden.
- Kein/Unzureichender Umgang mit *vielen* empirischen Befunden.
- Ausmaß der Variabilität der Befunde kann nicht bestimmt werden.
- Erklärungen der Variabilität der Befunde können nicht überprüft werden.



# Kapitelübersicht

1. Notwendigkeit von Forschungssynthesen
2. Begriffliche und konzeptionelle Grundlagen
3. Vorteile von Forschungssynthesen
4. Kritik an Forschungssynthesen und Meta-Analysen

# Vielfalt an Begriffen

- Systematischer Review / Systematischer Übersichtsartikel (*Systematic review*)
- (Quantitative) Forschungssynthese ((*Quantitative Research synthesis*)
- Integrativer Review (*Integrative review*)
- Meta-Analyse, Metaanalyse (*Meta-analysis*)



# Systematischer Reviews und Meta-Analysen

## Systematic (literature) review

„A review that strives to comprehensively identify, appraise, and synthesize all the relevant studies on a given topic. Systematic reviews are often used to test just a single hypothesis, or a series of related hypothesis.“

## Meta-analysis

„A review that uses a specific statistical technique for synthesizing the results of several studies into a single quantitative estimate (i.e. a summary effect size)“ (Petticrew & H. Roberts, 2006, S. 19).



# Begriff der Meta-Analyse nach Glass (1976)

„Meta-analysis refers to the analysis of analyses. I use it to refer to the statistical analysis of a large collection on analysis results from individual studies for the purpose of integrating the findings“ (Glass, 1976, S. 3).



# Begriff der Meta-Analyse nach Drinkmann (1990)

„Mit Meta-Analyse wird eine an den Kriterien empirischer Forschung orientierte Methode zur quantitativen Integration der Ergebnisse empirischer Untersuchungen sowie zur Analyse der Variabilität dieser Ergebnisse bezeichnet“ (Drinkmann, 1990).



# Begriff der Meta-Analyse nach Borenstein et al (2009)

„Meta-analysis refers to the statistical synthesis of results from a series of studies. While the statistical procedures used in a meta-analysis can be applied to any set of data, the synthesis will be meaningful only if the studies have been collected systematically. This could be in the context of a systematic review, the process of systematically locating, appraising, and then synthesizing data from a larger number of sources.“  
(Borenstein et al. 2009: xxif.)



# Zum Verhältnis von systematischem Review, Forschungssynthese und Meta-Analyse: Ein Fazit

- Der Begriff (quantitative) Forschungssynthese (oder: quantitativer systematischer Review) beschreibt den gesamten Forschungsprozess. Forschungssynthesen haben sowohl *qualitative* als auch *quantitative* Elemente.
- Der quantitative (statistische) Teil einer Forschungssynthese heißt *Meta-Analyse*.



# Systematik von Meta-Analysen I

- (Typ I: Qualitative Zusammenfassung, klassisches narratives Review)
- Typ II: Quantitative Zusammenfassung von publizierten, aggregierten Statistiken („publikationsbasierte“ Meta-Analyse)
- Typ III: Neuauswertung auf der Basis von zusammengeführten Originaldaten („originaldatenbasierte“ Meta-Analyse)
- Typ IV: Prospektiv geplante, gepoolte Auswertungen

(Quellen: Blettner, Sauerbrei, Schlehofer, Scheuchenpflug und Friedenreich 1997, S. 149, Finney 1995, S. 91)



# Systematik von Meta-Analysen II

- Analyseebene: Aggregat- (APD) oder Individualdaten (IPD)
- Skalierung der abhängigen Variablen
- Forschungsdesign:
  - Experiment oder Beobachtung / Befragung
  - Intervention oder Zusammenhang

(Sauerbrei und Blettner 2003)



# Vier prototypische Forschungsprobleme von Forschungssynthesen

- Deskription
  - Verbreitung des Schulschwänzens (Weiβ 2008)
  - Meta-Analyse von Studien zur Motivation von Stadt-Umland-Wanderung (Bleck/Wagner 2006)
- Exploration: Meta-Analysen zum Stand der Ehescheidungsforschung (Wagner/Weiβ 2003; 2004)
- Hypothesentests: Meta-Analysen zum Ehescheidungsrisiko in Europa (Wagner/Weiβ 2006)
- Evaluation (von Interventionen): Meta-Analyse zur Wirksamkeit von Psychotherapien (Smith/Glass 1977)



# Geschichte der Meta-Analyse

- Erste Entwicklungen von Pearson (1904), Fisher (1932), Cochrane (1937).
- Zwischen 1930 und 1970 erfolgte keine Weiterentwicklung.
- Mitte der 1970er Jahre neues Forschungsinteresse: Glass (1976) prägt den Begriff „Meta-Analyse“
- In den 1980er Jahren bereits fünf Monographien: Glass et al. 1981, Hunter et al. 1982, Rosenthal 1984, Light/Pillemer 1984, Hedges/Olkin 1985.
- Ab Mitte der 1980er Jahre kontinuierliche Anwendung und Weiterentwicklung der Methode in der Psychologie, Pädagogik und Medizin.
- In einigen Fächern bereits voller Institutionalisierungsgrad.

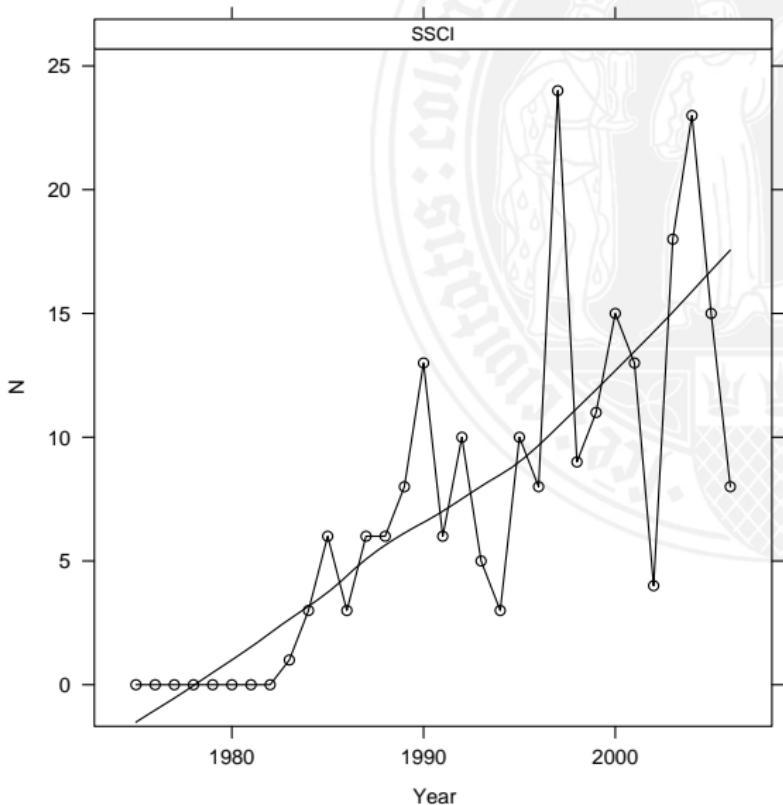


# Verschiedene Disziplinen kennen verschiedene „Meta-Analyse-Schulen“

- Psychologie
- Erziehungswissenschaften
- Medizin, Gesundheitswissenschaften, Epidemiologie
- VWL (und BWL)
- (Soziologie)



# Soziologische Forschungssynthesen (SSCI)



# Kapitelübersicht

1. Notwendigkeit von Forschungssynthesen
2. Begriffliche und konzeptionelle Grundlagen
3. Vorteile von Forschungssynthesen
4. Kritik an Forschungssynthesen und Meta-Analysen



# Vorteile von (quantitativen) Forschungssynthesen

Systematisch, strukturiert und objektiv

Systematischer, strukturierter und objektiver (intersubjektiv nachvollziehbarer) Forschungsprozess:

- Dokumentation aller Arbeitsschritte (wie Literaturrecherche, Dateneingabe etc.)
- Offenlegung aller Regeln und abgeleiteten Entscheidungen bzgl. relevanter oder irrelevanter Forschungsbefunde



# Vorteile von (quantitativen) Forschungssynthesen

## Differenzierte und aussagekräftige Ergebnisdarstellung

Forschungssynthesen erlauben eine differenziertere und aussagekräftigere Darstellung der Ergebnisse als narrative Reviews.

- Ein umfangreicher Forschungsstand kann auf wenige, klar zu interpretierende Statistiken reduziert werden.
- Quantifizierung der Ergebnisse geht aber *nicht* mit einem Ausblenden von Unterschiedlichkeit/Heterogenität einher.
- Befunde einzelner, kleiner Studien können statistisch unbedeutsam sein (niedrige stat. *power*); erst eine Forschungssynthese kann Nachweis statistischer Bedeutsamkeit liefern.



# Vorteile von (quantitativen) Forschungssynthesen

## Erklärungen für einen heterogenen Forschungsstand

- Systematische Kodierung von Studienbefunden und Studiencharakteristika.
- Dies ermöglicht die Untersuchung eines Zusammenhangs von Studienbefunden und Studiencharakteristika.
  - Unterschiede werden *entdeckt* und *erklärt*.



# Kapitelübersicht

- 
1. Notwendigkeit von Forschungssynthesen
  2. Begriffliche und konzeptionelle Grundlagen
  3. Vorteile von Forschungssynthesen
  4. Kritik an Forschungssynthesen und Meta-Analysen

# Typische Kritik an Forschungssynthesen und Meta-Analysen

- „Apples and Oranges“
- „Garbage in and Garbage out“
- Unvollständiges Datenmaterial
- Verzerrungen durch selektives Publizieren (*publication bias*)
- Bivariate und multivariate Effektstärken



# Probleme von Meta-Analysen in der Soziologie

- APD Meta-Analysen: Fehlende Werte
- Integration von Regressionskoeffizienten
- Umgang mit abhängigen Effektstärken



## Teil III

# Forschungssynthese als sozialwissenschaftliche Methode



# Forschungssynthese als sozialwissenschaftliche Methode

5. Der sozialwissenschaftliche Forschungsprozess

6. Forschungsfrage

7. Datenerhebung: Literatur- und Studienrecherche

8. Datenvercodung/-eingabe und Effektstärken

9. Effektstärken (ES)



# Kapitelübersicht

## 5. Der sozialwissenschaftliche Forschungsprozess

### 6. Forschungsfrage

## 7. Datenerhebung: Literatur- und Studienrecherche

Recherchestrategien

Bewertung und Auswahlkriterien

Dokumentation der Recherche und Verwaltung der Quellen

Stichprobenselektivität

## 8. Datenvercodung/-eingabe und Effektstärken

## 9. Effektstärken (ES)

Effektstärken der  $r$ -Familie

Effektstärken der  $d$ -Familie

Kategoriale Effektstärken

Regressionskoeffizienten

Effektstärkenkonvertierung

# Relevante Dimensionen des Forschungsdesigns

- Wahl der Untersuchungseinheiten (bspw. Individuen vs. Kollektive)
- **Untersuchungsarten**
  - Primäranalyse
  - Sekundäranalyse
  - Meta-Analyse
- Zeitdimension (Querschnitt vs. Längsschnitt)



# Untersuchungsarten

- **Primäranalyse:** Erstmalige Nutzung und Auswertung *eines* Datensatzes.
- **Sekundäranalyse:** Erneute Nutzung *eines* Datensatzes unter geänderten Bedingungen der Forschungsmethodik und / oder des theoretischen Bezugssystems.
- **Meta-Analyse:** Nutzung und Auswertung *mehrerer* Datensätze / Studien mit gemeinsamer Fragestellung (womit noch *nichts* über die Analyseeinheit gesagt wird!).



# Elemente einer quantitativen Forschungssynthese

Diekmann (1997)

1. Formulierung und Präzisierung des Forschungsproblems
2. Planung und Vorbereitung der Erhebung
3. Datenerhebung
4. Datenauswertung
5. Berichterstattung

Cooper (1982)

1. Problem Formulation
2. Data Collection
3. Data Evaluation
4. Analysis and Interpretation
5. Public Presentation

Wagner / Weiß (2003)

1. Problem / Forschungsfrage
2. Literatur- oder Datensatzrecherche (allg. Datenerhebung)
3. Dateneingabe
4. Datenanalyse (Meta-Analyse)
5. Ergebnis-präsentation



# Kapitelübersicht

5. Der sozialwissenschaftliche Forschungsprozess

6. Forschungsfrage

7. Datenerhebung: Literatur- und Studienrecherche

Recherchestrategien

Bewertung und Auswahlkriterien

Dokumentation der Recherche und Verwaltung der Quellen

Stichprobenselektivität

8. Datenvercodung/-eingabe und Effektstärken

9. Effektstärken (ES)

Effektstärken der  $r$ -Familie

Effektstärken der  $d$ -Familie

Kategoriale Effektstärken

Regressionskoeffizienten

Effektstärkenkonvertierung

# Bedeutung der Forschungsfrage

- Die Forschungsfrage *muss* die interessierenden Merkmale und Zusammenhänge präzise beschreiben, um (später) relevante und irrelevante Studien/Publikationen voneinander unterscheiden zu können.
- Aus der Forschungsfrage lassen/sollten sich Kriterien für die Wahl relevanter bzw. irrelevanter Studien ableiten.



# Kriterien für eine gelungene Forschungsfrage

- Interessierende Variablen/Merkmale wurden klar definiert („Konzeptspezifikation“).
- Relevantes Forschungsdesign spezifizieren.
- Die Forschungsfrage in einen historischen, geographischen, theoretischen, methodologischen Kontext einbetten.



## Beispiel für Fragestellung (Wagner/Weiß 2006)

„The aim of this paper is to summarize European research on various divorce risks. More precisely, we will examine how much divorce risks vary between European countries and whether such variations can be explained by country-specific macro-level factors. Are there any meaningful differences in the divorce risks between European countries?“

# Beispiel für Fragestellung (Weiβ 2008)

1. Wie viele Jugendliche schwänzen insgesamt die Schule?
2. Wie hoch ist das Schulschwänzrisiko bei Jugendlichen mit Migrationshintergrund?
3. Unterscheidet sich das Schulschwänzrisiko bei Jugendlichen mit Migrationshintergrund an Hauptschulen von den übrigen Schulformen?



# Beispiel für Fragestellung (Wilson et al. 2008)

„Objectives: To synthesize the extant empirical evidence on the effects of boot-camps and boot camp like programs on the criminal behavior (e.g., postrelease arrest, conviction, or reinstitutionalization) of convicted adult and juvenile offenders.“



# Kapitelübersicht

5. Der sozialwissenschaftliche Forschungsprozess

6. Forschungsfrage

## 7. Datenerhebung: Literatur- und Studienrecherche

Recherchestrategien

Bewertung und Auswahlkriterien

Dokumentation der Recherche und Verwaltung der Quellen

Stichprobenselektivität

8. Datenvercodung/-eingabe und Effektstärken

## 9. Effektstärken (ES)

Effektstärken der  $r$ -Familie

Effektstärken der  $d$ -Familie

Kategoriale Effektstärken

Regressionskoeffizienten

Effektstärkenkonvertierung

# Ziel der (Literatur-)Recherche

Die Literaturrecherche entspricht der Phase der Datenerhebung im Rahmen einer Primäruntersuchung.

- Erfassung eines möglichst großen Teiles der vorhandenen Literatur (Vollerhebung).
- Vermeidung von Verzerrungen:

*„The point is not to track down every paper that is somehow related to the topic. (...) The point is to avoid missing a useful paper that lies outside one's regular purview“ (White, Cooper & Hedges, 1994, S. 44).*

- Systematische und transparente Suche.



# Spezifizierung des Untersuchungsgegenstandes der Forschungssynthese

- Abgrenzung zu anderen, verwandten Themen
- Definition der Grundgesamtheit
- Wichtige Untersuchungsvariablen
- Methodische Anforderungen
- Kulturelle und sprachliche Grenzziehungen
- ...



# Abschnittsübersicht

5. Der sozialwissenschaftliche Forschungsprozess

6. Forschungsfrage

**7. Datenerhebung: Literatur- und Studienrecherche**

Recherchestrategien

Bewertung und Auswahlkriterien

Dokumentation der Recherche und Verwaltung der Quellen

Stichprobenselektivität

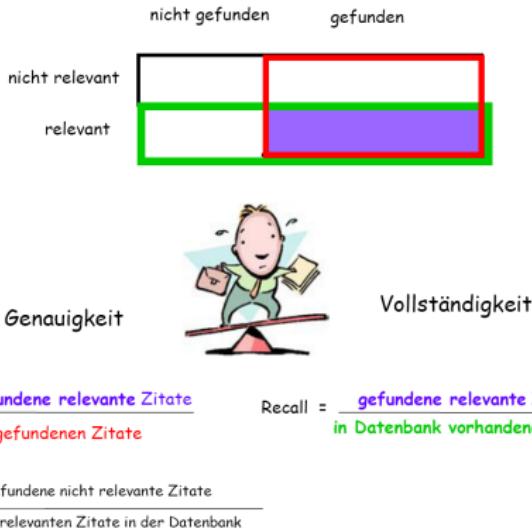
8. Datenvercodung/-eingabe und Effektstärken

9. Effektstärken (ES)



# Genauigkeit und Vollständigkeit

## Retrievalbias: Das Dilemma zwischen Genauigkeit und Vollständigkeit



(Quelle: Motschall 2004: Medline-Suche mit PubMed,  
[http://www.agmb.de/04\\_mannheim/motschall.pdf](http://www.agmb.de/04_mannheim/motschall.pdf))

# Ein unlösbares Problem

1. *Recall* (Vollständigkeit): Anteil an gefundener relevanter Literatur im Verhältnis zur Gesamtzahl (hypothetische Größe) an *relevanter* Literatur.
2. *Präzision* (Präzision, Genauigkeit): Anteil an gefundener relevanter Literatur im Verhältnis zur Anzahl *insgesamt gefundener* (relevanter und nichtrelevanter) Literatur.
3. Beide Maße stehen in einer inversen Beziehung und können nicht gleichzeitig optimiert werden:
  - Je unspezifischer die Suche, desto höher der Recall und desto höher der Anteil an irrelevanten Publikationen ⇒ hoher Arbeitsaufwand.
  - Je spezifischer die Suche, desto höher die Precision und desto höher die Wahrscheinlichkeit, relevante Untersuchungen zu übersehen ⇒ Repräsentativität der Stichprobe gefährdet.



# Suche in Literaturverweisen („footnote chasing“)

## Vorgehen

- Reviews
- thematisch verwandte Artikel, Bücher oder Zeitschriften
- Bibliographien

## Beurteilung

- + guter Einstieg in die Suche
- + meist relativ hohe Präzision
- selektive, persönliche Auswahl (eigene Untersuchungen / Literaturverweise) → Verzerrungen
- Aktualität der Literatur hängt von Referenzquelle ab.



# Suche in (Fach-)Datenbanken: Vorgehen

- z.B. Wiso 3, SOLIS/FORIS, Sociological Abstracts, SocioFile etc.
- Fachdatenbanken umfassen:
  - meist nur publizierte Artikel
  - ab einem bestimmten Jahr
  - innerhalb bestimmter geographischer Grenzen → deswegen: Charakteristika der Datenbanken beachten
- Übereinstimmung von Suchabfrage mit den Angaben in der Datenbank notwendig → richtige Benutzung Boolscher Operationen, vorsichtiger Umgang mit Begriffen



# Suche in (Fach-)Datenbanken: Beurteilung

- + Vermeidet Verzerrungen aufgrund persönlicher Auswahl.
- Relativ niedrige Präzision, sofern die Recall-Rate akzeptabel sein soll.
- Verzerrungen aufgrund der in die Datenbanken aufgenommenen Artikel.



# Suche in Zitationsdatenbanken

## Vorgehen

z.B. Social Science Citation Index

## Beurteilung

- + ermöglicht das Auffinden von Literatur aus unterschiedlichen Fachbereichen
- + umfasst neueste Publikationen
- + relativ hohe Recall-Rate
- Verzerrungen aufgrund der in die Datenbanken aufgenommenen Artikel



# Kommunikation mit den peers

## Vorgehen

- Konferenzen
- Anfragen bei Forschern
- Anfragen bei staatlichen Einrichtungen

## Beurteilung

- + ermöglicht Auffinden nicht publizierter Literatur
- + sehr hohe Präzision
- selektive Auswahl → Verzerrungen



# Systematische Suche in Bibliotheken und Zeitschriften (Browsing)

## Vorgehen

- Systematisches Durchsuchen von Zeitschriftenjahrgängen (bspw. letzte 10 Jahrgänge von JMF)

## Beurteilung

- meist geringe Präzision, zeitaufwendig
- + thematisch eng gefasste Zeitschriften oder Bibliotheksbereiche können aber die Suche sinnvoll ergänzen



# Abschnittsübersicht

5. Der sozialwissenschaftliche Forschungsprozess

6. Forschungsfrage

**7. Datenerhebung: Literatur- und Studienrecherche**

Recherchestrategien

Bewertung und Auswahlkriterien

Dokumentation der Recherche und Verwaltung der Quellen

Stichprobenselektivität

8. Datenvercodung/-eingabe und Effektstärken

9. Effektstärken (ES)

# Auswahlkriterien

Die Auswahlkriterien definieren die Grundgesamtheit und betreffen die Generalisierbarkeit der Ergebnisse der Forschungssynthese.

Relevante Dimensionen sind u.a.:

- Forschungsdesign (Längsschnitt/Querschnitt, Experiment/Umfrage/...)
- Messung zentraler Konstrukte
- Grundgesamtheit der einzelnen Studie (Stichprobenzusammensetzung, Alter, Ost/West, ...)
- Zeitlicher und geographischer Kontext der Studie (Jahr der Stichprobenerhebung, Publikationsjahr, ...)
- Qualität der Studien (inkl. ausreichend Daten für Durchführung stat. Analysen)
- Publikationstyp (un-/veröffentlichte Literatur, graue Literatur, ...)
- ...



# Beispiel für Auswahlkriterien (Wagner/Weiβ 2006)

„Publications dealing with our research question should meet the following criteria to be included in our meta-analysis: on the one hand, we were interested in publications in which marital stability is the dependent variable. On the other hand, we limited our search to publications that explicitly used European longitudinal data sets. The countries considered here are the 18 countries of the European Economic Area (EEA), the candidate countries for the European Union, and Switzerland“ (Wagner/Weiβ 2006: 487).



# Beispiel für Auswahlkriterien (Weiβ 2008)

„Letztlich orientierte sich die Auswahl geeigneter Datensätze an drei Kriterien: (1) Der Datensatz enthält Angaben zur selbst- oder fremdberichteten unentschuldigten Schulabwesenheit. Dabei ist gleichgültig, für welchen Zeitraum das Schulschwänzen erfasst oder wie präzise die Häufigkeit des Phänomens gemessen wurde. (2) Die Datensätze sind für Sekundäranalysen verfügbar. (3) Die Stichprobe wurde in Deutschland gezogen, ohne jedoch vorauszusetzen, nur Schüler mit deutscher Staatsangehörigkeit zu erfassen“ (Weiβ 2008).



# Beispiel für Auswahlkriterien (Wilson et al. 2008)

„Selection Criteria: The eligibility criteria were (a) that the study evaluated a correctional boot camp, shock incarceration, or intensive incarceration program; (b) that the study included a comparison group that received either probation or incarceration in an alternative facility; (c) that the study participants were exclusively under the supervision of the criminal or juvenile justice system; and (d) that the study reported a post-program measure of criminal behavior, such as arrest or conviction“ (Wilson et al. 2008: 2).



# Abschnittsübersicht

5. Der sozialwissenschaftliche Forschungsprozess

6. Forschungsfrage

**7. Datenerhebung: Literatur- und Studienrecherche**

Recherchestrategien

Bewertung und Auswahlkriterien

Dokumentation der Recherche und Verwaltung der Quellen

Stichprobenselektivität

8. Datenvercodung/-eingabe und Effektstärken

9. Effektstärken (ES)

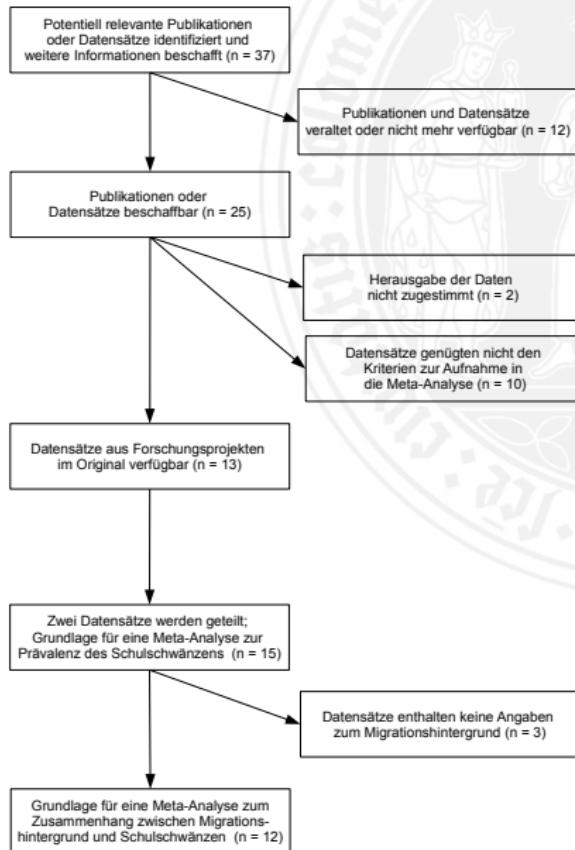
# Verwaltung der Literatur

- Tabellenkalkulation (OO-Calc, MS-Excel).
- (Kostenpflichtige) Programme zur Literaturverwaltung wie Endnote, Reference Manager etc.
- (Kostenfreie) Programme wie Mendeley, Zotero oder RevMan.

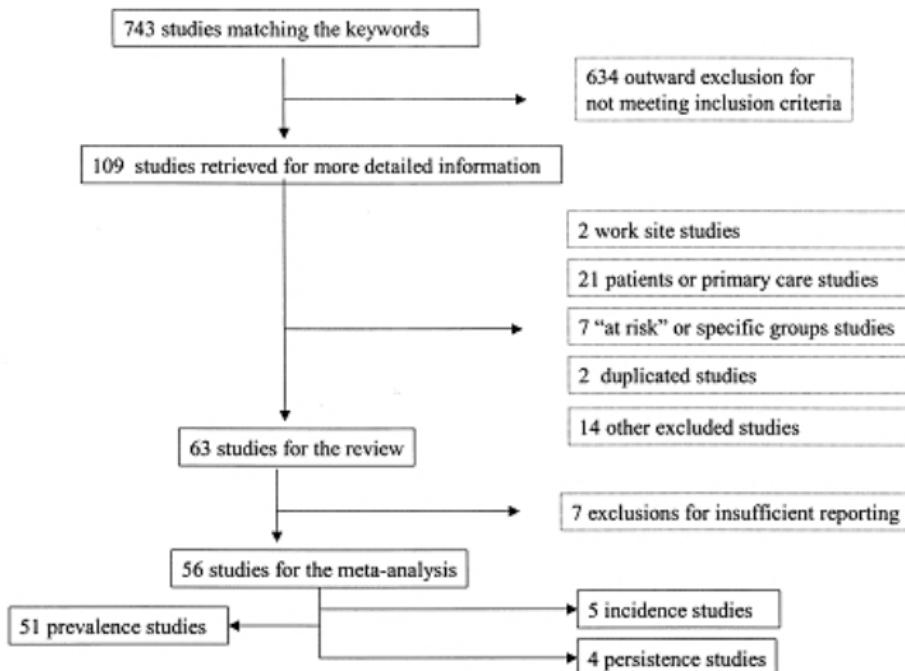
*Ein Argument für die Verwendung eines bestimmten Programms ist eine offene Schnittstelle auf die bspw. Statistiksoftware zugreifen kann (bieten etwa Mendeley oder Zotero).*



# Dokumentation der Literaturrecherche (Weiß 2008)



# Dokumentation der Literaturrecherche (Lorant et al., 2003)



# Abschnittsübersicht

5. Der sozialwissenschaftliche Forschungsprozess

6. Forschungsfrage

## 7. Datenerhebung: Literatur- und Studienrecherche

Recherchestrategien

Bewertung und Auswahlkriterien

Dokumentation der Recherche und Verwaltung der Quellen

Stichprobenselektivität

8. Datenvercodung/-eingabe und Effektstärken

9. Effektstärken (ES)



# Quellen der Verzerrungen bei der Literatursuche

- Subjektive Auswahl bei Reviews (ebenfalls eher statistisch signifikante Ergebnisse)
- Verzerrung durch Sprachbarrieren
- Verzerrung durch einseitige Literaturrecherche
- ...
- → Verzerrungen führen tendenziell zu einer Überschätzung der Effektstärken.

Weitere Informationen zum Thema „Publication Bias“ finden sich ab Folie 214.



# Kapitelübersicht

5. Der sozialwissenschaftliche Forschungsprozess

6. Forschungsfrage

7. Datenerhebung: Literatur- und Studienrecherche

Recherchestrategien

Bewertung und Auswahlkriterien

Dokumentation der Recherche und Verwaltung der Quellen

Stichprobenselektivität

8. Datenvercodung/-eingabe und Effektstärken

9. Effektstärken (ES)

Effektstärken der  $r$ -Familie

Effektstärken der  $d$ -Familie

Kategoriale Effektstärken

Regressionskoeffizienten

Effektstärkenkonvertierung

# Übersicht

- Nach der Identifikation relevanter Publikationen müssen die darin enthaltenen Informationen erfasst werden (ähnelt einer Inhaltsanalyse).
- Festlegen, welche Informationen für die Fragestellung wichtig sind.
- Fragebogen (Codierbogen) erstellen.
- Insbesondere auf die Vergleichbarkeit der empirischen Befunde achten.



# Datenerhebung und -vercodung

- Welche Informationen aus den Studien besitzen Relevanz (Synthese / Heterogenität)?
  - Publikationsbefunde (Statistiken, Effektstärken, ...)
  - Publikationsmerkmale (Publikationstyp, Publikationsjahr, ...)
  - Studien-/Stichprobenmerkmale (Erhebungsjahr, Organisation, Stichprobendesign, ...)
  - Methode und Qualität (Statistische Verfahren, bestimmte Qualitätskriterien, ...)
- Konstruktion eines ‚Fragebogens‘ (Codierbogen).
- Problem: Auswahl überflüssiger oder Auslassen bedeutsamer Variablen.



# Beispiel für Codierbogen (Auszug)

Microsoft Access - [publication1]

Daten Bearbeiten Ansicht Einfügen Format Datengäste Extras Fenster ?

Frage hier eingeben

MS Sans Serif 8 F K U

Publicationstitel: Sozistrukturelle Bestimmgründe der Ehescheidung.  
Eine empirische Überprüfung einiger Hypothesen der  
Familienforschung

Zeitschriftname: Zeitschrift für Soziologie der Erziehung und Sozialisation

pubuse: 1  
pubuse2: #Name?  
pubdate: 3.09.2000  
Codierer: 1  
pubid: 1  
pubid2: #Name?  
Theoretische Ausrichtung:  
Publikationsjahr: 1998  
Publikationssprache: 1  
Publikationstyp: 1

Theoretische Ausrichtung:  
"Der Großteil der neueren Literatur über sozistrukturelle Bestimmgrößen der Ehescheidung bezieht sich auf die Annahmen der new home economics und die daran anknüpfenden Erklärungen ethlicher Stabilität [...] Durch diese aus der Ökonomie übernommene Modellierungstechnik und die Betonung der objektiven Marktgegebenheiten als wichtige bestimmungsgröße einer Handlungsentscheidung ist es gelungen, eine Vielzahl empirischer Hypothesen abzuleiten. Die grundlegende Frage ist, inwieweit aus dem theoretisch Modell abgeleitete Hypothesen auch empirisch tragfähig sind" (S. 118)

sapu

sapuid	smpid	pubid
133	125	1
*	(AutoWert)	0

Datensatz: 1 von 1

Kommentar: Analysen beziehen sich nur auf alte Bundesländer. (123)

Formular 'abvar' öffnen Formular 'author' öffnen

Datensatz: 1 von 130

Nur auf 1 gesetzte Publikationen werden analysiert



# Beispiel für Codierbogen (Auszug)

STUDYID	<input type="text"/>	1. Study Identification Number
PUBTYPE	<input type="text"/>	2. Type of Publication
		1 Book
		2 Journal Article/Book Chapter
		3 Thesis/Dissertation
		4 Technical Report
		5 Conference Paper
		6 Other
PUBYEAR	<input type="text"/>	3. Publication Year (enter 9999 if missing or unknown)
MEANAGE	<input type="text"/>	4. Mean Age of Sample (enter 99.99 if missing or unknown)
RACE	<input type="text"/>	5. Predominate Race of Sample
		1 > 60% White
		2 > 60% Black
		3 > 60% Hispanic
		4 > 60% Other Minority
		5 Mixed (none more than 60%)
		6 Mixed, but can't estimate proportion
		9 Can't tell

**Exhibit 5.3.** Sample page of the study-level coding form from the meta-analysis of challenge programs (see Appendix E for the full coding form).



# Was ist ein (empirischer) Forschungsbefund?

- Glass (1976: 6) weist die *Effektstärke* als Mittelwertsdifferenz zwischen Experimental- und Kontrollgruppe aus, die durch Division mit der Gruppenvarianz (angenommen wird eine Gleichheit der Gruppenvarianzen) standardisiert wird.
- Nach Rosenthal (1991) beschreibt ein Forschungsbefund den Zusammenhang zwischen Variable X und Variable Y. Dieser Zusammenhang enthält zwei Elemente:
  1. ein Schätzer der Größe des Zusammenhangs (Effektstärke, engl. „effect size“).
  2. eine Maßzahl für die Reliabilität der Effektgröße (Standardfehler, Konfidenzintervalle, *p*-Wert).
- Allgemein lässt sich ein Forschungsbefund definieren als „[...] statistical representation of one empirical relationship involving the variable(s) of interest to the meta-analysis measured on a single subject sample.“ (Lipsey/Wilson 2000: 35)



# Effektstärke

- Die wesentliche Eigenschaft einer Effektstärke ist nach Hedges / Olkin (1985: 7) ihre Skaleninvarianz (ermöglicht die Vergleichbarkeit zwischen Studien).
- Wünschenswerte Eigenschaften von Effektstärken sind zudem Informationen über *Größe* und *Richtung* des Zusammenhangs zwischen Variablen.



# Von der Hypothese zur Effektstärke

- Unterschieds- oder Zusammenhangshypothesen
  - Unterschiedshypothese: „Wenn Jugendliche eine Klasse wiederholen müssen, dann steigt ihr Schwänzrisiko.“
  - Zusammenhangshypothese: „Je schlechter das Klassenklima vom einzelnen Schülern bewertet wird, desto höher auch das individuelle Schwänzrisiko.“
- Gerichtete oder ungerichtete Hypothesen
  - Unterschiedshypothesen (Wenn... dann) und Zusammenhangshypothesen sind *gerichtete* Hypothesen.
  - Ungerichtete Hypothesen behaupten lediglich einen Unterschied, ohne jedoch eine Richtung (besser – schlechter, höher – niedriger etc.) vorzugeben.



# Typen von Effektstärken / Forschungsbefunden (1)

- Kontinuierliche Merkmale:
  - Maßzahlen der  $r$ -Familie: z.B.  $r$ ,  $(\phi)$ ,  $r_{pb}$  und  $\rho$
  - Maßzahlen der  $d$ -Familie: z.B. Hedges'  $g$ , Glass'  $\Delta$  und Cohen's  $d$
- Dichotome (kategoriale) Merkmale: z.B. odds ratio, risk ratio,  $\phi$
- Teststatistiken (etwa  $t$ -,  $F$ - oder  $\chi^2$ -Werte)
- Narrative Darstellung signifikant positiver, negativer oder nichtsignifikanter Effekte
- $p$ -Werte (kond. Wahrscheinlichkeit der fälschlichen Zurückweisung der  $H_0$ , gegeben, dass in GG die  $H_0$  gilt;  $\alpha$ -Fehlerwahrscheinlichkeit)
- ...

(mehr Informationen zu ES auf den Folien 93ff verfügbar)



# Typen von Effektstärken / Forschungsbefunden (2)

Effektstärken (besser: Forschungsbefunde) lassen sich aber auch nach der Anzahl der Variablen unterscheiden:

- Univariate Effektstärken (bspw. Anteile, Mittelwerte)
- Bivariate Effektstärken (bspw.  $d$ ,  $r$ , odds ratio)
- Multivariate Effektstärken (bspw. Regressionskoeffizienten)



# Welche meta-analytischen Verfahren gibt es für die verschiedenen Formen von Forschungsbefunden?

1. Vote counting
2. Zusammenfassen von Signifikanzwerten
3. Zusammenfassen von Effektstärken  
(Effektstärkensynthese)
4. Zusammenfassen von Korrelationsmatrizen
5. ...



# Ein Vorgriff auf das Berechnen einer mittleren Effektstärke

- Auf Grundlage der Ergebnisse ( $T_j$ ) von  $k$  unabhängigen Studien wird im einfachsten Fall eine mittlere Effektstärke ( $\bar{T}$ ) als gewichtetes arithmetisches Mittel berechnet:

$$\bar{T} = \frac{\sum_{j=1}^k w_j \times T_j}{\sum_{j=1}^k w_j}$$

- Neben den ES werden auch noch **Gewichte** benötigt:
  - Fallzahl ( $w_j = n_j$ )
  - Inverse quadrierte Standardfehler (Fehlervarianz):  $w_j = \frac{1}{SE_j^2}$



# Abhängige Effektstärken

Wie entstehen abhängige Effektstärken?

- Multiple Befundstatistiken pro Person (bspw. mehrere Testergebnisse; hierarchische Regressionsmodelle).
- Verschiedene Treatment-Gruppen haben eine gemeinsame Kontrollgruppe.
- ...

Umgang mit abhängigen ES:

- Die „beste“ (oder „mittlere“) ES wählen.
- Zunächst pro Set abh. ES eine Meta-Analyse durchführen (Stufe 1) und dann eine zweite Meta-Analyse über die gemittelten ES durchführen (Stufe 2).
- Bei multiplen ES (mehrere abh. Variablen) pro Personen pro Gruppe eine Meta-Analyse durchführen.
- Multivariate Meta-Analyse (Korrelationsmatrix muss vorliegen)
- Mehrebenen-Meta-Analyse

(Quellen: Kim & Becker, 2010; Lambert & Abrams, 1996; Raudenbush, Becker & Kalaian, 1988; C. G. Thompson & Becker, 2013)

# Kapitelübersicht

5. Der sozialwissenschaftliche Forschungsprozess

6. Forschungsfrage

7. Datenerhebung: Literatur- und Studienrecherche

Recherchestrategien

Bewertung und Auswahlkriterien

Dokumentation der Recherche und Verwaltung der Quellen

Stichprobenselektivität

8. Datenvercodung/-eingabe und Effektstärken

**9. Effektstärken (ES)**

Effektstärken der *r*-Familie

Effektstärken der *d*-Familie

Kategoriale Effektstärken

Regressionskoeffizienten

Effektstärkenkonvertierung

# Effektstärkenschätzer und -parameter

Bezüglich des Umgangs mit Effektstärken im Rahmen einer Meta-Analyse ist die folgende Unterscheidung wichtig:

- (Effektstärken-) *Schätzer* werden in den Publikationen berichtet und basieren auf einer *bestimmten* Stichprobe (*sample effect sizes*).
- (Effektstärken-) *Parameter* repräsentieren den *wahren* Wert in der Grundgesamtheit (*population (or true) effect size*).

Ein Ziel von Meta-Analyse ist, mit Hilfe der Effektstärkenschätzer den Populationsparameter zu schätzen. Neben dem eigentlichen ES-Schätzer wird (fast) immer auch der *Standardfehler* als Gewicht benötigt. (Campbell Collaboration 2009: 5).



# Übersicht und weiterführende Informationen

- Hier werden
  - die Produkt-Moment-Korrelation  $r$ ,
  - Cohens  $d$ ,
  - Hedges  $g$ ,
  - das Odds Ratio sowie
  - der semi-partielle Korrelationskoeffizient (für Koeffizienten von multiplen linearen Regressionsmodellen) näher erläutert.
- Gute Übersichten liefern u.a.
  - Borenstein, Hedges, Higgins und Rothstein (2009), Lipsey und Wilson (2001)



# Abschnittsübersicht

- 5. Der sozialwissenschaftliche Forschungsprozess
- 6. Forschungsfrage
- 7. Datenerhebung: Literatur- und Studienrecherche
- 8. Datenvercodung/-eingabe und Effektstärken

## 9. Effektstärken (ES)

- Effektstärken der  $r$ -Familie
- Effektstärken der  $d$ -Familie
- Kategoriale Effektstärken
- Regressionskoeffizienten
- Effektstärkenkonvertierung

# Übersicht

- Produkt-Moment-Korrelation  $r$  von (Bravais und) Pearson (beide Merkmale haben metrisches Skalenniveau)
- Korrelationskoeffizient nach Spearman (beide Merkmale haben ordinale Skalenniveau)
- Punkt-biserale Korrelation (ein Merkmal dichotom, ein Merkmal metrisch)
- ...

(Borenstein et al. 2009: 23ff)



# Produkt-Moment-Korrelation

## Eigenschaften der Effektstärke

- Pearsons  $r$  errechnet sich aus der normierten Kovarianz und misst die Stärke des Zusammenhangs zweier Variablen.
- Der Korrelationskoeffizient liegt zwischen –1 (bei vollständig gegenläufigem Zusammenhang) und +1 (bei vollständig gleichläufigem Zusammenhang).
- Pearsons  $r$  ist ebenso wie die Kovarianz bei fehlendem Zusammenhang gleich 0.
- Der Korrelationskoeffizient kann nur für metrische Daten angewandt werden. Eine Ausnahme sind zwei dichotome Variablen, die 0/1 kodiert sind. In diesem Fall entspricht der Korrelationskoeffizient der Chi-Quadrat-basierten Maßzahl  $\phi$ .



# Der Korrelationskoeffizient r von Bravais und Pearson

## Definition der Effektstärke

Pearsons  $r$  lässt sich auch als die Kovarianz von  $X$  und  $Y$  ( $S_{XY}$ ) dividiert durch das Produkt der Standardabweichungen von  $X$  und  $Y$  ( $S_X$  und  $S_Y$ ) definieren.

$$r = \frac{S_{XY}}{S_X \cdot S_Y} \quad (1)$$

mit:

$$S_{XY} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y}) \quad (2)$$

$$S_x = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \text{ und } S_y = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2} \quad (3)$$



# Die Fisher Z-Transformation

Bei der Befundsynthese von Korrelationskoeffizienten gilt es zu beachten, dass diese (i.A.) erst transformiert werden müssen. Diese Transformation hat zwei Gründe:

1. Der Korrelationskoeffizient ist nicht normalverteilt.
2. Der Standardfehler lässt sich nur für den transformierten Korrelationskoeffizienten bestimmen, da die Z-Transformation varianzstabilisierend ist und die Varianz von  $z_r$  nicht von  $r$  abhängt.

Es gilt dann:

- Z-Transformation von  $r$ :  $z_r = 0.5 \cdot \ln\left(\frac{1+r}{1-r}\right)$ .
- Standardfehler für  $z_r$ :  $SE_{z_r} = \frac{1}{\sqrt{N-3}}$  mit  $N$ : Fallzahl.
- Rücktransformation  $z_r$  zu  $r$ :  $r = \frac{e^{2 \cdot z_r} - 1}{e^{2 \cdot z_r} + 1}$ .



# Beispiel Fisher Z-Transformation

- Gegeben sind:  $r = 0.13$  und  $n = 123$
- Fisher Z-transformierte Korrelation:  $z_r = 0.5 \cdot \ln\left(\frac{1+0.13}{1-0.13}\right) = 0.1307$
- Standardfehler:  $SE_{z_r} = \frac{1}{\sqrt{123-3}} = 0.0913$
- Rücktransformation:  $r = \frac{e^{2 \cdot 0.1307} - 1}{e^{2 \cdot 0.1307} + 1} = 0.13$



# Abschnittsübersicht

5. Der sozialwissenschaftliche Forschungsprozess

6. Forschungsfrage

7. Datenerhebung: Literatur- und Studienrecherche

8. Datenvercodung/-eingabe und Effektstärken

## 9. Effektstärken (ES)

Effektstärken der  $r$ -Familie

Effektstärken der  $d$ -Familie

Kategoriale Effektstärken

Regressionskoeffizienten

Effektstärkenkonvertierung

# Übersicht

- Die abhängige Variable ist metrisch skaliert, die unabhängige Variable ist ein (dichotomer) Faktor (Gruppenmerkmal, Pre-Post-Vergleich).
- Unstandardisierte Mittelwertdifferenz
  - Für zwei unabhängige Gruppen
  - Für abhängige Gruppen (etwa Pre-Post-Tests) (wird nicht weiter erläutert)
- Standardisierte Mittelwertdifferenzen (Cohens  $d$ , Hedges  $g$ )
  - Für zwei unabhängige Gruppen
  - Für abhängige Gruppen (etwa Pre-Post-Tests) (wird nicht weiter erläutert)
- ...

(Borenstein et al. 2009: 22ff)



# Unstandardisierte Mittelwertdifferenz

Die unstandardisierte Mittelwertdifferenz zweier unabhängiger Mittelwerte  $\mu_1$  und  $\mu_2$  ist definiert als:

$$\Delta = \mu_1 - \mu_2, \quad (4)$$

Die Schätzung von  $\hat{\Delta} = D$  erfolgt über zwei unabhängige Stichprobenmittelwerte  $\overline{Y_1}$  und  $\overline{Y_2}$ :

$$D = \overline{Y_1} - \overline{Y_2}. \quad (5)$$



# Varianz der unstandardisierten Mittelwertdifferenz

- Seien  $S_1$  und  $S_2$  die beiden Standardabweichungen von  $Y_1$  und  $Y_2$  und angenommen, diese unterscheiden sich *nicht* in der Population, dann ergibt sich die Varianz von  $D$  als:

$$V_D = \frac{n_1 + n_2}{n_1 n_2} S_{pooled}^2 \quad (6)$$

mit

$$S_{pooled} = \sqrt{\frac{(n_1 - 1) S_1^2 + (n_2 - 1) S_2^2}{n_1 + n_2 - 2}}. \quad (7)$$

- Für unterschiedliche Populationsstandardabweichungen ergibt sich die Varianz von  $D$  nach:

$$V_D = \frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}. \quad (8)$$

Der Standardfehler von  $D$  ergibt sich aber immer nach:

$$SE_D = \sqrt{V_D}. \quad (9)$$

# Beispiel unstandardisierte Mittelwertdifferenz

- Gegeben sind:
  - $\bar{Y}_1 = 103, S_1 = 5.5, n_1 = 50$
  - $\bar{Y}_2 = 100, S_2 = 4.5, n_2 = 50$ .
- $D$  ist:  $103 - 100 = 3$
- $SE_D$  ist (angenommen  $\sigma_1 = \sigma_2$ ):
  - $S_{pooled} = \sqrt{\frac{(50-1)5.5^2 + (50-1)4.5^2}{50+50-2}} = 5.0249$
  - $V_D = \frac{50+50}{50 \cdot 50} \cdot 5.0249 = 1.0100$
  - $SE_D = \sqrt{1.0100} = 1.0050$

(Borenstein et al., 2009, S. 22)



# Standardisierte Mittelwertdifferenz (Cohens $d$ )

Die standardisierte Mittelwertdifferenz zweier unabhängiger Mittelwerte  $\mu_1$  und  $\mu_2$  ist definiert als:

$$\delta = \frac{\mu_1 - \mu_2}{\sigma}, \text{ wobei } \sigma_1 = \sigma_2 = \sigma, \quad (10)$$

Die Schätzung von  $\hat{\delta} = d$  erfolgt über die zwei Stichprobenmittelwerte  $\overline{Y_1}$  und  $\overline{Y_2}$ :

$$d = \frac{\overline{Y_1} - \overline{Y_2}}{S_{within}} \text{ (für } S_{within} \text{ siehe Gleichung 7).} \quad (11)$$



# Varianz der standardisierten Mittelwertdifferenz

Seien  $S_1$  und  $S_2$  die beiden Standardabweichungen von  $Y_1$  und  $Y_2$  und angenommen, diese unterscheiden sich *nicht* in der Population, dann ergibt sich die Varianz von  $d$  (approximiert) als:

$$V_d = \frac{n_1 + n_2}{n_1 n_2} + \frac{d^2}{2(n_1 + n_2)} \quad (12)$$

und einem Standardfehler  $SE_d$  von

$$SE_d = \sqrt{V_d}. \quad (13)$$



# Standardisierte Mittelwertdifferenz (Hedges $g$ )

In kleinen Stichproben überschätzt  $d$  die wahre Mittelwertsdifferenz  $\delta$ . Dies lässt sich durch eine von Hedges vorgeschlagene Korrektur beheben:

$$g = J \times d, \text{ wobei } J = 1 - \frac{3}{4df - 1}. \quad (14)$$

mit  $df = n_1 + n_2 - 2$  (für unabhängige Gruppen).

Für die Varianz von  $g$  gilt:

$$V_g = J^2 \cdot V_d \quad (15)$$

und entsprechend gilt für den Standardfehler von  $g$ :

$$SE_g = \sqrt{V_g}. \quad (16)$$



# Beispiel standardisierte Mittelwertdifferenz

- Gegeben sind:
  - $\bar{Y}_1 = 103, S_1 = 5.5, n_1 = 50$
  - $\bar{Y}_2 = 100, S_2 = 4.5, n_2 = 50$ .
- $d$  ist:  $\frac{103 - 100}{5.0249} = 0.5970$
- $SE_D$  ist (angenommen  $\sigma_1 = \sigma_2$ ):
  - $S_{pooled} = \sqrt{\frac{(50-1)5.5^2 + (50-1)4.5^2}{50+50-2}} = 5.0249$
  - $V_D = \frac{50+50}{50 \cdot 50} + \frac{0.5970^2}{2(50+50)} = 0.0418$
  - $SE_D = \sqrt{0.0418} = 0.2044$
- Hedges  $g$  ist:
  - Korrekturfaktor:  $J = \left(1 - \frac{3}{4 \cdot 98 - 1}\right) = 0.9923$
  - $g = 0.9923 \cdot 0.5970 = 0.5924$
  - $V_g = 0.9923^2 \cdot 0.0418 = 0.0411$
  - $SE_g = \sqrt{0.0411} = 0.2028$

(Borenstein et al., 2009, 26ff)



# Abschnittsübersicht

- 5. Der sozialwissenschaftliche Forschungsprozess
- 6. Forschungsfrage
- 7. Datenerhebung: Literatur- und Studienrecherche
- 8. Datenvercodung/-eingabe und Effektstärken

## 9. Effektstärken (ES)

- Effektstärken der  $r$ -Familie
- Effektstärken der  $d$ -Familie
- Kategoriale Effektstärken
- Regressionskoeffizienten
- Effektstärkenkonvertierung

# Übersicht

Unter kategorialen Effektstärken werden vor allem Assoziationsmaße verstanden, die auf  $2 \times 2$ -Tabellen basieren:

- Risiko-Verhältnis (*risk ratio*)
- Chancenverhältnis (*odds ratio*)
- Risiko-Differenz (*risk difference*)
- Phi-Koeffizient
- ...



# Das Odds Ratio (1)

- Eine *Chance* (*odds*) ist definiert als die Wahrscheinlichkeit  $p(A)$ , dass ein Ereignis A eintritt, geteilt durch die Gegenwahrscheinlichkeit  $1 - p(A)$ , also:

$$Odds = \frac{p(E)}{1 - p(E)} \quad (17)$$

- Das Chancenverhältnis (Odds Ratio) vergleicht nun die Chancen für Ereignis A über zwei Gruppen (z.B. Männer/Frauen, Control/Treatment). Angenommen,
  - in der Kontrollgruppe sterben 10,
  - in der Treatment-Gruppe aber nur 5 von 100 Patienten.Daraus resultiert folgende Tabelle:

	Control (0)	Treatment (1)
Alive (0)	90	95
Death (1)	10	5



## Das Odds Ratio (2)

- Für die Kontrollgruppe gilt:

$$Odds_C = \frac{10/100}{1-(10/100)} = \frac{1/10}{9/10} = \frac{10}{90} = \frac{1}{9} \approx 0.1111$$

- Für die Treatment-Gruppe gilt:  $Odds_T = \frac{5}{95} \approx 0.0526$

- Das Odds Ratio beträgt dann:  $\frac{5/95}{1/9} = \frac{45}{95} = \frac{9}{19} \approx 0.4737$

Die Interpretation des Odds Ratio ist wie folgt:

- Das Odds ratio kann Werte von 0 bis  $+\infty$  annehmen und ist damit asymmetrisch verteilt.
- Ein Wert von 1 zeigt die statistische Unabhängigkeit der Merkmale an.
- Ein Wert kleiner 1 bzw.  $[0; 1)$  zeigt einen negativen Zusammenhang an.
- Ein Wert größer 1 bzw.  $(1; \infty)$  zeigt einen positiven Zusammenhang an.



# Das Odds Ratio (3)

Gilt allgemein die folgenden Tabellennotation

	Control (0)	Treatment (1)
Alive (0)	A	B
Death (1)	C	D

dann lässt sich das Odds Ratio ( $\alpha$ ) schneller nach folgender Formel berechnen:

$$\alpha = \frac{AD}{BC} \quad (18)$$



# Das Odds Ratio (4)

- Ähnlich wie bei der PM-Korrelation, so muss auch das Odds Ratio in einer Meta-Analyse zunächst transformiert (logarithmiert) werden (nicht normalverteilt, Standardfehler nicht bestimmt für Originalskala), also  $\ln(\text{OddsRatio}) = \ln(\alpha)$  (natürlicher Logarithmus).
- Die Rücktransformation erfolgt über die Exponentialfunktion:  $\exp(\ln(\alpha))$ .
- Die (approximierte) Varianz für  $\ln(\alpha)$  ist:

$$V_{\ln(\alpha)} = \frac{1}{A} + \frac{1}{B} + \frac{1}{C} + \frac{1}{D} \quad (19)$$

- Der (appr.) Standardfehler ist:  $SE_{\ln(\alpha)} = \sqrt{V_{\ln(\alpha)}}$



# Beispiel Odds Ratio

- Gegeben ist die folgende Tabelle:

	Control (0)	Treatment (1)
Alive (0)	90	95
Death (1)	10	5

- Das Odds Ratio ist:  $\alpha = \frac{90 \cdot 5}{10 \cdot 95} \approx 0.4737$
- Die Varianz beträgt:  $V_{ln(\alpha)} = \frac{1}{90} + \frac{1}{95} + \frac{1}{10} + \frac{1}{5} \approx 0.3216$
- Der Standardfehler:  $SE_{ln(\alpha)} = \sqrt{0.3216} = 0.5671$



# Abschnittsübersicht

5. Der sozialwissenschaftliche Forschungsprozess
6. Forschungsfrage
7. Datenerhebung: Literatur- und Studienrecherche
8. Datenvercodung/-eingabe und Effektstärken

## 9. Effektstärken (ES)

- Effektstärken der  $r$ -Familie
- Effektstärken der  $d$ -Familie
- Kategoriale Effektstärken
- Regressionskoeffizienten**
- Effektstärkenkonvertierung

# Überblick

- Die Synthese von Regressionskoeffizienten in Meta-Analysen ist schwierig (Aloe & Becker, 2011a, 333ff.):
  - Regressionskoeffizienten (RK) für den *focal predictor* von Modellen mit unterschiedlichen abhängigen Variablen (damit auch unterschiedliche Schätzverfahren) können i.d.R. nicht zusammengefasst werden.
  - Der Wert eines RK hängt u.a. von der Existenz anderer Prädiktoren sowie dem Ausmaß der Interkorrelation aller Prädiktoren ab (Kollinearität).
  - Unterschiedliche Operationalisierungen (Skalen) von abhängiger und unabhängiger (*focal predictor*) Variable.
- Bislang existieren vor allem Verfahren für das *allgemeine lineare Modell* („lineare Regression“).



# Ansätze zur Synthese linearer Regressionsmodelle

- Synthese unstandardisierter Koeffizienten mit Hilfe eines GLS-Ansatzes (benötigt aber die Varianz-Kovarianzmatrix des Regressionsmodells) (Becker & Wu, 2007)
- Standardisierte Regressionskoeffizienten (Kim, 2011)
- „Partial effect sizes“ (nur dichotome Prediktoren) (Keef & L. A. Roberts, 2004)
- t-Statistiken (Stanley & Jarrell, 1989)
- Der semi-partielle Korrelationskoeffizient ( $r_{sp}$ ; *semi-partial correlation*) (Aloe & Becker, 2011b; Aloe & Becker, 2011a)
- ...



# Der semi-partielle Korrelationskoeffizient

$$r_{sp} = \text{sgn}(t_f) \sqrt{r_Y^2 - r_{Y(f)}^2} \quad (20)$$

mit:

$\text{sgn}$ : Vorzeichenfunktion, gibt das Vorzeichen der t-Statistik zurück

$t_f$ : t-Statistik für Regressionskoeffizient des *focal predictor*

$r_Y^2$ : Erklärte Varianz Gesamtmodell

$r_{Y(f)}^2$ : Erklärte Varianz für das Modell ohne *focal predictor*

Besser geeignet ist aber die nachfolgende Formel:

$$r_{sp} = \frac{t_f \sqrt{(1 - r_Y^2)}}{\sqrt{(n - p - 1)}} \quad (21)$$

mit:

$n$ : Fallzahl des Regressionsmodells

$p$ : Gesamtanzahl unabhängiger Variablen

# Gemeinsame Meta-Analysen von $r_{sp}$ und $r$ möglich?

- $r_{sp}$  sind gewöhnlich kleiner als PM-Korrelationen ( $r$ ).
- Beide können gemeinsam in einer Meta-Analyse verwendet werden, aber im Rahmen einer Meta-Regression gilt es zu prüfen, ob sich beide Korrelationstypen systematisch unterscheiden (Aloe & Becker, 2011a, S. 346).
- Für eine Beispielanwendung siehe Aloe und Becker (2009).



# Varianz des semi-partiellen Korrelationskoeffizienten

Die Varianz für  $r_{sp}$  lässt sich wie folgt schätzen:

$$V_{r_{sp}} = \frac{r_Y^4 - 2r_Y^2 + r_{Y(f)}^2 + 1 - r_{Y(f)}^4}{n} \quad (22)$$

Unbekannt ist dabei meistens der Term  $r_{Y(f)}^2$ . Dieser ergibt sich aber aus:

$$r_{Y(f)}^2 = r_Y^2 - r_{sp}^2 \quad (23)$$



# Beispiel semi-partieller Korrelationskoeffizient

Table 38. Multiple regression to predict student test scores in science based on teacher's academic preparation and other variables

Variable Label	Parameter Estimate	Standard Error	Probability
Intercept . . . . .	42.23	0.07	0.00
<b>Teacher's academic background</b>			
Grade point average in science subject . . . . .	0.54	0.28	0.06
Courses in science education . . . . .	-0.00	0.34	1.00
Over 40 credits in earth and physical sciences. . . . .	1.10	0.45	0.02
Missing data on course grades . . . . .	-0.01	0.89	0.99
Female . . . . .	-2.07	0.24	0.00
African-American, Hispanic, or Native American. . . . .	-3.54	0.33	0.00
Missing data on race . . . . .	-2.35	0.75	0.00
Bottom SES quartile . . . . .	-2.02	0.29	0.00
R-square = 0.33			

(Quelle: Chaney, 1995, 74; die Tabelle ist nicht vollständig dargestellt)



## Beispiel semi-partieller Korrelationskoeffizient

Dem Tabellenausschnitt auf Folie 123 lassen sich folgende Informationen für den Effekt von „Grade point average in science subject“ entnehmen:

- $t$ -Statistik:  $0.54/0.28=1.93$
- $r_Y^2$  (R-square): 0.33
- N (nicht in Tabelle, siehe Fußnote)<sup>2</sup>: 26435
- p (Anzahl Prädiktoren): 26

Dann berechnet sich  $r_{sp}$  nach (trotz kleinerer Fehler in Aloe & Becker halte ich mich weitgehend an deren Darstellung):

$$r_{sp} = \frac{1.9\sqrt{(1 - 0.33)}}{\sqrt{(26435 - 26 - 1)}} = 0.0097$$

(Quelle: Aloe & Becker, 2011a, 346; kleine Typo in Beispielrechnung,  $p = 26$ , nicht 27)

---

<sup>2</sup>Aloe & Becker verwenden vermutlich die falsche Fallzahl (richtig: 24599), siehe <http://nces.ed.gov/pubs94/94378.pdf>, Fußnote 3

# Beispiel Varianz des semi-partiellen Korrelationskoeffizienten

Zur Berechnung der Varianz von  $r_{sp}$  fehlt noch  $r_{Y(f)}^2$ , d.h. die erklärte Gesamtvarianz für ein Modell ohne den *focal predictor*.

$$r_{Y(f)}^2 = r_Y^2 - r_{sp}^2 = 0.33 - 0.0097^2 = 0.3299$$

Für die Varianz von  $r_{sp}$  gilt dann:

$$V_{r_{sp}} = \frac{0.1089 - 2 \cdot 0.33 + 0.3299 + 1 - 0.1088}{26435} = 0.000025$$

Der Standardfehler beträgt:

$$SE_{r_{sp}} = 0.005$$

(Quelle: Aloe & Becker, 2011a, S. 347)



# Abschnittsübersicht

- 5. Der sozialwissenschaftliche Forschungsprozess
- 6. Forschungsfrage
- 7. Datenerhebung: Literatur- und Studienrecherche
- 8. Datenvercodung/-eingabe und Effektstärken

## 9. Effektstärken (ES)

- Effektstärken der  $r$ -Familie
- Effektstärken der  $d$ -Familie
- Kategoriale Effektstärken
- Regressionskoeffizienten
- Effektstärkenkonvertierung

# Überblick

- Unterschiedliche Studien berichten trotz gleicher/ähnlicher Fragestellung unterschiedliche Statistiken/Effektstärken.
- Viele bivariate Effektstärken lassen sich (approximativ) ineinander überführen.
- Anschließend lassen sich diese konvertierten Effektstärken gemeinsam im Rahmen einer Meta-Analyse analysieren.
- Hier wird nur eine kleine Auswahl gängiger Transformationsregeln vorgestellt. Weiterführende Darstellungen finden sich u.a. bei
  - Borenstein et al. (2009),
  - Rosenthal, Rosnow und Rubin (2000),
  - Lipsey und Wilson (2001),
  - Higgins und Green (2008),
  - ...
- Die nachfolgenden Ausführungen basieren auf Borenstein et al. (2009, 46ff.).



# Konvertierung zwischen $\ln(\text{odds ratio})$ und $d$

$$d = \ln(\text{odds ratio}) \cdot \frac{\sqrt{3}}{\pi} \quad (24)$$

$$V_d = V_{\ln(\text{odds ratio})} \cdot \frac{3}{\pi^2} \quad (25)$$

$$\ln(\text{odds ratio}) = d \frac{\pi}{\sqrt{3}} \quad (26)$$

$$V_{\ln(\text{odds ratio})} = V_d \frac{\pi^2}{3} \quad (27)$$

(Borenstein et al., 2009, 46ff)



# Konvertierung zwischen $r$ und $d$

$$d = \frac{2r}{\sqrt{1 - r^2}} \quad (28)$$

$$V_d = \frac{4V_r}{(1 - r^2)^3} \quad (29)$$

$$r = \frac{d}{\sqrt{d^2 + a}}, \quad (30)$$

wobei  $a$  ein Korrekturfaktor für  $n_1 \neq n_2$  ist:

$$a = \frac{(n_1 + n_2)^2}{n_1 n_2} \quad (31)$$

$$V_r = \frac{a^2 V_d}{(d^2 + a)^3} \quad (32)$$

(Borenstein et al., 2009, 46ff)



## Teil IV

# Statistische Verfahren der Meta-Analyse



# Statistische Verfahren der Meta-Analyse I

10. Befundsynthese von Aggregatdaten

11. Begriff der Heterogenität und Diagnostik

12. Heterogenitätsaufklärung

13. Publication Bias

14. Reporting Standards

15. Ein Ausblick auf Meta-Analysen mit Individualdaten

16. Literatur



# Meta-Analyse auf einen Blick

1. Ein Ziel einer Meta-Analyse ist das Zusammenfassen einer Effektstärkenverteilung (ESV) mit Hilfe einer (gewichteten) mittleren ES.
2. Die Wahl des Schätzverfahrens hängt u.a. vom Ausmaß der Unterschiede (Heterogenität, siehe ausführlich 156) zwischen den ES ab. Es werden (im univariaten Fall) zwei ES-Modelle unterschieden: (1) Fixed-effects model (geringe Unterschiede), (2) Random-effects model (große Unterschiede).
3. Neben der Bestimmung eines Mittelwertes geht es in MA auch/vor allem um die Darstellung/Erfassung und Aufklärung der ES-Heterogenität (Regressionsmodelle oder ANOVA-ähnliche Modelle).



# Kapitelübersicht

## 10. Befundsynthese von Aggregatdaten

Fixed-effects model (FEM)

Random-effects model (REM)

## 11. Begriff der Heterogenität und Diagnostik

Überblick

Die Zwischenstudienvarianz

Die Q- Statistik

Weitere Heterogenitätsmaße:  $H^2$  und  $P$

Zusammenfassung Heterogenitätsmaße

## 12. Heterogenitätsaufklärung

Meta-Regression

ANOVA-ähnliche Verfahren

Weitere Bemerkungen zu Verfahren der Heterogenitätsaufklärung

## 13. Publication Bias

## 14. Reporting Standards

## 15. Ein Ausblick auf Meta-Analysen mit Individualdaten

## 16. Literatur

# Befundsynthese

Die Befundsynthese von Aggregatdaten ist ein typischer Anwendungsfall einer Meta-Analyse. Dabei sind die folgenden drei Bedingungen zu beachten:

1. Die Bildung eines synthetischen Befundes ist nur sinnvoll, wenn mindestens zwei einzelne Statistiken berichtet werden,
2. diese müssen i.A. statistisch unabhängig<sup>3</sup> sein und
3. in die Mittelwertbildung gehen die einzelnen Befunde nach dem Kriterium ihrer Reliabilität (Fehlervarianz/quadrierter SE; Fallzahl) gewichtet ein.

---

<sup>3</sup>Für abhängige Effektstärken gibt es spezielle Verfahren, siehe Folie 91.



# Effektstärkenverteilungen beschreiben

- Ziel ist es, den Lageparameter („Mittelwert“) einer ES-Verteilung zu schätzen.
- Schätzung ist mit stat. Unsicherheit verbunden, u.a. sind Informationen der SEs der einzelnen Effektstärken zu berücksichtigen.
- Möglicherweise aber ist die ES-Verteilung durch (unbekannte) Faktoren wie Alter oder Geschlecht beeinflusst, d.h. überzufällige Unterschiede zwischen Effektstärkengruppen.
- Diese Heterogenität/Zwischenstudienvariation beeinflusst die Schätzung des Mittelwertes und der stat. Unsicherheit.
- Eine zentrale Frage: Ist die ES-Heterogenität so groß, dass wir sie berücksichtigen müssen?
  - Nein: Fixed-effects model
  - Ja: Random-effects model



# Abschnittsübersicht

## 10. Befundsynthese von Aggregatdaten

Fixed-effects model (FEM)

Random-effects model (REM)

## 11. Begriff der Heterogenität und Diagnostik

## 12. Heterogenitätsaufklärung

## 13. Publication Bias

## 14. Reporting Standards

## 15. Ein Ausblick auf Meta-Analysen mit Individualdaten

## 16. Literatur

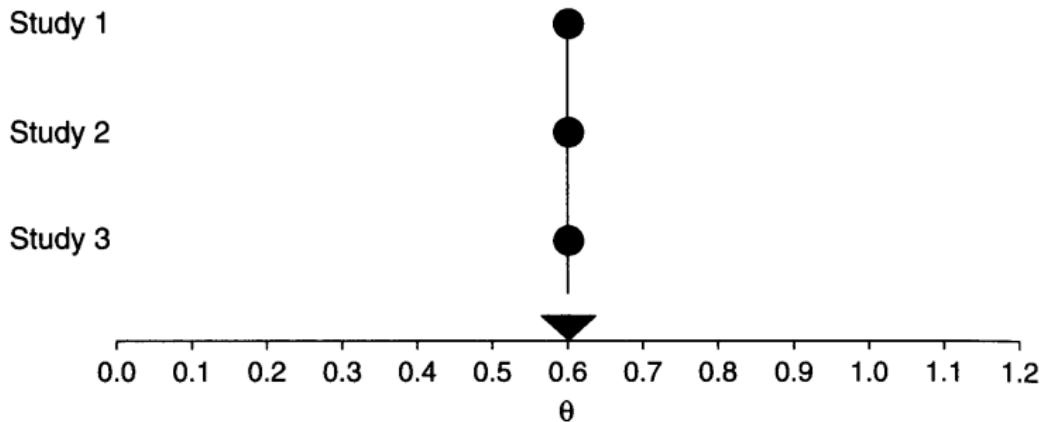


# Das *fixed-effects model* (FEM)

- Annahmen
  - Es gibt *einen* gemeinsamen (*common effect*) Populationsparameter ( $\theta$ ).
  - Alle Studien sind sich sehr ähnlich (Design, Stichprobe, Operationalisierung, ...).
  - Es gibt statische Belege für die Homogenität der Effektstärken (Test auf Heterogenität), d.h. es gibt keine (kaum) unbeobachtete Heterogenität.
  - Man kan sich sicher sein, die allermeisten Publikationen entdeckt zu haben.
- Konsequenzen
  - Der Populationsparameter kann relativ genau geschätzt werden (kleiner Standardfehler, enges Konfidenzintervall).
  - Studien mit großen Fallzahlen werden stark gewichtet.



# Verteilung der wahren Effekte im FEM



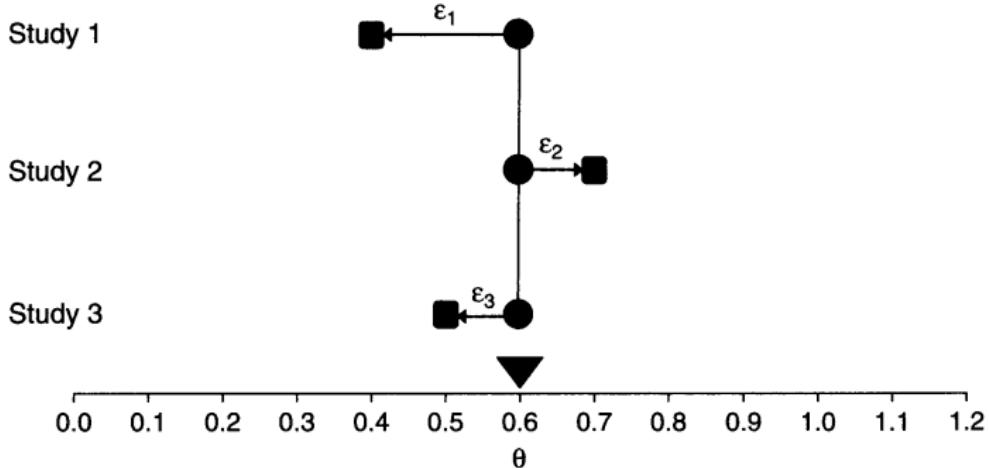
Legende:

- Wahre Effektstärke in Studie  $j$
- ▼ Wahre Effektstärke über alle Studien (combined)  
 $(\theta_1 = \theta_2 = \dots = \theta)$

(Quelle: Borenstein et al., 2009, S. 64)



# Fehler im FEM



Legende:

- Wahre Effektstärke in Studie  $j$
- ▼ Wahre Effektstärke über alle Studien (combined)  
( $\theta_1 = \theta_2 = \dots = \theta$ )
- Empirische/Beobachtete Effektstärken

(Quelle: Borenstein et al., 2009, S. 64)



# Befundsynthese von Aggregatdaten (FEM)

Gegeben seien  $k$  unabhängige Effektstärken  $T_j$  ( $j = 1, \dots, k$ ), dann berechnet sich die gewichtete mittlere Effektstärke nach:

$$\bar{T}_{FEM} = \frac{\sum_{j=1}^k w_j \times T_j}{\sum_{j=1}^k w_j}, \quad (33)$$

mit einem Gewicht  $w_j$ , das der inversen Fehlervarianz ( $1/v_j$ ) bzw. dem inversen quadrierten Standardfehler ( $1/SE_j^2$ ) der  $j$ -ten Effektstärke entspricht:

$$w_j = \frac{1}{SE_j^2} = \frac{1}{v_j}. \quad (34)$$

Die zusammengefasste Fehlervarianz  $\bar{v}_{FEM}$  selbst ergibt sich aus:

$$\bar{v}_{FEM} = \frac{1}{\sum_{j=1}^k (1/v_j)}. \quad (35)$$

Für den Standardfehler von  $\bar{T}_{FEM}$  gilt dann  $\bar{SE}_{FEM} = \sqrt{\bar{v}_{FEM}}$ .



# Abschnittsübersicht

## 10. Befundsynthese von Aggregatdaten

Fixed-effects model (FEM)

Random-effects model (REM)

## 11. Begriff der Heterogenität und Diagnostik

## 12. Heterogenitätsaufklärung

## 13. Publication Bias

## 14. Reporting Standards

## 15. Ein Ausblick auf Meta-Analysen mit Individualdaten

## 16. Literatur

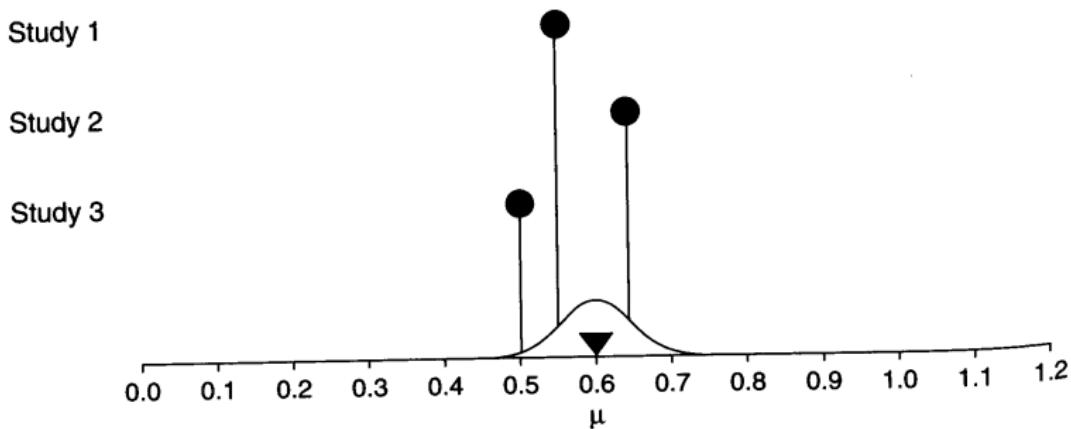


# Das *random-effects model* (REM)

- Annahmen
  - Jede Studie  $j$  besitzt ihren eigenen Populationsparameter ( $\theta_j$ ), geschätzt wird ein „mittlerer“ Effekt (*average effect*).
  - Man kann nicht annehmen, dass alle Studien ähnlich sind (Design, Stichprobe, Operationalisierung, ...).
  - Es lassen sich u.U. aus der Theorie Einflussgrößen ableiten, die für Unterschiede in den Studienbefunden verantwortlich gemacht werden können.
  - Es gibt statische Belege für die Heterogenität der Effektstärken (Test auf Heterogenität).
  - Man kan sich nicht sicher sein, die allermeisten Publikationen entdeckt zu haben.
- Konsequenzen
  - Der gemeinsame Populationsparameter kann nur ungenau geschätzt werden (großer Standardfehler, breites Konfidenzintervall).
  - Die Studiengewichte werden ähnlicher, d.h. große Studien erhalten weniger Einfluss, kleine Studien gewinnen Einfluss.



# Verteilung der wahren Effektstärken im REM



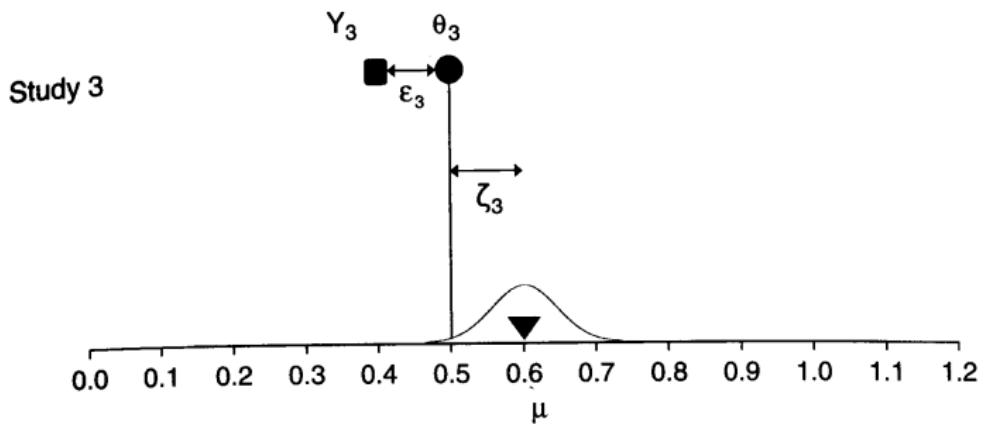
Legende:

- Wahre Effektstärke in Studie  $j$
- ▼ Mittlere („average“) Effektstärke über alle Studien (combined)

(Quelle: Borenstein et al., 2009, S. 70)



# Fehler im REM



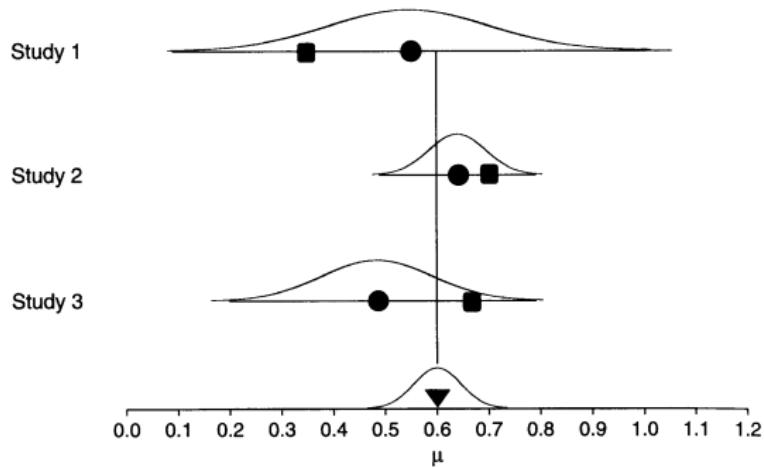
Legende:

- Wahre Effektstärke in Studie  $j$
- Empirische/Beobachtete Effektstärken
- ▼ Wahre Effektstärke über alle Studien (combined)
  - Übersetzung der unterschiedlichen Nomenklatur (siehe Folie 148):  $\zeta_3 = u_3$

(Quelle: Borenstein et al., 2009, S. 71)



# Verteilung der Stichprobenfehler im REM: Binnen- und Zwischenstudienvarianz



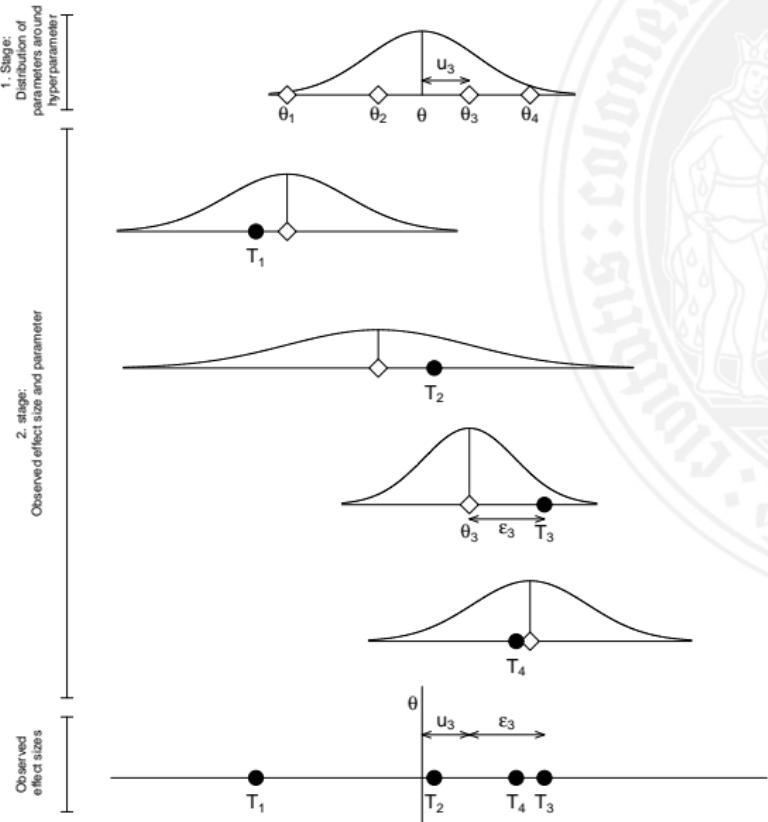
Legende:

- Wahre Effektstärke in Studie  $j$
- Empirische/Beobachtete Effektstärken (Hinweis: Studie 1 kleines N, Studie 2 großes N)
- ▼ Mittlere („average“) Effektstärke über alle Studien (combined)

(Quelle: Borenstein et al., 2009, S. 71)



# Das REM als ein zweistufiger Prozess



(Quelle: C. Thompson & Weiß, 2013; Viechtbauer, 2007)

# Vergleich von FEM und REM

- Gilt das FEM, dann liegt allen Studien ein gemeinsames  $\theta$  zugrunde. Studien mit kleinen Fallzahlen werden durch das Heruntergewichten weitgehend ignoriert, da in umfangreichen Studien bessere Informationen über  $\theta$  vorliegen.
- Dagegen wird im REM der Mittelwert einer Effektstärkenverteilung geschätzt („mean of a distribution of effects“). Kleine Studien sind daher wichtiger als im FEM und erhalten mehr Gewicht.



# Fehlerterme im FEM und REM

## Fixed-effects model

$$T_j = \theta + \epsilon_j$$

Alle Befundstatistiken  $T_j$  entstammen *einer* Population ( $\theta$ ).

## Random-effects model

$$\begin{aligned} T_j &= \theta_j + \epsilon_j \\ \theta_j &= \theta + u_j \\ \Rightarrow T_j &= \theta + \epsilon_j + u_j \end{aligned}$$

- Pro Befundstatistik  $T_j$  ein je eigener Populationsparameter  $\theta_j$ ; alle  $\theta_j$  wiederum streuen um den Parameter  $\theta$  einer „Superpopulation“.
- Berücksichtigung der Zwischenstudienvarianz  $Var(u_j) = \tau^2$ , so dass

$$w_j^* = \frac{1}{v_j^*} = \frac{1}{v_j + \tau^2}.$$



# Schätzen der Zwischenstudienvarianz $\tau^2$

Der sog. Momentenschätzer für  $\tau^2$  lautet:

$$\hat{\tau}^2 = \frac{Q - (k - 1)}{c} = \frac{Q - df}{c} \quad (36)$$

mit:

$k$ : Anzahl der Effektstärken

$df$ : Anzahl der Freiheitsgrade ( $df = k - 1$ )

$Q$ : Maß der Zwischenstudienvariation,  $Q \sim \chi_{df}^2$

(Heterogenitätstest, siehe ausführlich Folie 165)

$c$ : Skalierungsfaktor, sorgt dafür, dass  $\tau^2$  wieder die Metrik der Ausgangs-ES erhält.

$$Q = \sum_{j=1}^k \left( \frac{(T_j - \bar{T}_{FEM})^2}{v_i} \right) \quad (37)$$

Für eine ausführlichere Darstellung der Zwischenstudienvarianz siehe Folien 161ff.

# Befundsynthese von Aggregatdaten (REM)

Gegeben seien  $k$  unabhängige Effektstärken  $T_j$  ( $j = 1, \dots, k$ ), dann berechnet sich die gewichtete mittlere Effektstärke nach dem *random-effects model* nach:

$$\bar{T}_{REM} = \frac{\sum_{j=1}^k w_j^* \times T_j}{\sum_{j=1}^k w_j^*}, \quad (38)$$

mit einem Gewicht  $w_j^*$ , für das gilt:  $w_j^* = \frac{1}{v_j^*} = \frac{1}{v_j + \hat{\tau}^2}$ .

Die Gesamtfehlervarianz (bzw. der Standardfehler) für den REM-Schätzer ergibt sich nach:

$$\bar{V}_{REM} = \frac{1}{\sum_{j=1}^k w_j^*} \text{ bzw. } \bar{SE}_{REM} = \sqrt{\frac{1}{\sum_{j=1}^k w_j^*}}$$

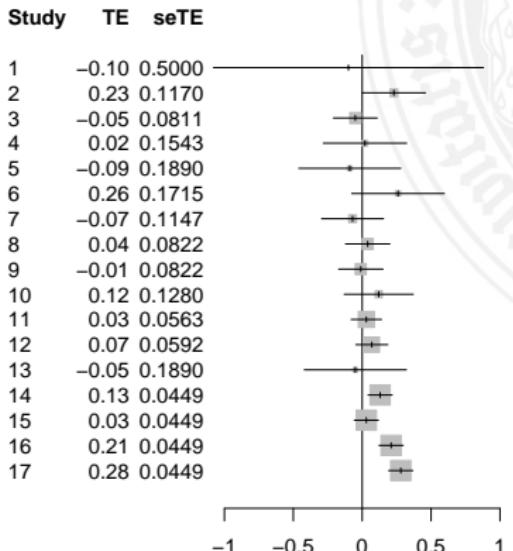


## Beispiel „Teacher verbal ability & school outcomes“

	r	z	se.z
1	-0.10	-0.10034	0.50000
5	-0.09	-0.09024	0.18898
7	-0.07	-0.07011	0.11471
3	-0.05	-0.05004	0.08111
13	-0.05	-0.05004	0.18898
9	-0.01	-0.01000	0.08220
4	0.02	0.02000	0.15430
11	0.03	0.03001	0.05634
15	0.03	0.03001	0.04486
8	0.04	0.04002	0.08220
12	0.07	0.07011	0.05923
10	0.12	0.12058	0.12804
14	0.13	0.13074	0.04486
16	0.21	0.21317	0.04486
2	0.23	0.23419	0.11704
6	0.26	0.26611	0.17150
17	0.28	0.28768	0.04486

(Aloe & Becker, 2009, Quelle: )

# Beispiel „Teacher verbal ability & school outcomes: Forestplot



# Beispiel „Teacher verbal ability & school outcomes“: FEM und REM

		95%-CI	%W(fixed)	%W(random)
1	-0.1003	[-1.0803; 0.8796]	0.12	0.38
2	0.2342	[ 0.0048; 0.4636]	2.16	4.52
3	-0.0500	[-0.2090; 0.1089]	4.49	6.72
4	0.0200	[-0.2824; 0.3224]	1.24	3.09
5	-0.0902	[-0.4606; 0.2802]	0.83	2.25
6	0.2661	[-0.0700; 0.6022]	1.01	2.63
7	-0.0701	[-0.2949; 0.1547]	2.25	4.64
8	0.0400	[-0.1211; 0.2011]	4.38	6.63
9	-0.0100	[-0.1711; 0.1511]	4.38	6.63
10	0.1206	[-0.1304; 0.3715]	1.80	4.02
11	0.0300	[-0.0804; 0.1404]	9.31	8.74
12	0.0701	[-0.0460; 0.1862]	8.43	8.49
13	-0.0500	[-0.4204; 0.3204]	0.83	2.25
14	0.1307	[ 0.0428; 0.2187]	14.70	9.75
15	0.0300	[-0.0579; 0.1179]	14.70	9.75
16	0.2132	[ 0.1253; 0.3011]	14.70	9.75
17	0.2877	[ 0.1998; 0.3756]	14.70	9.75

Number of studies combined: k=17

	95%-CI	z	p.value
Fixed effect model	0.1123 [0.0786; 0.1460]	6.530	< 0.0001
Random effects model	0.0880 [0.0265; 0.1495]	2.803	0.0051

Quantifying heterogeneity:

$\tau^2 = 0.0081$ ;  $H = 1.59$  [1.22; 2.07];  $I^2 = 60.4\%$  [32.6%; 76.7%]

Test of heterogeneity:

Q	d.f.	p.value
40.36	16	0.0007

Details on meta-analytical method:

- Inverse variance method
- DerSimonian-Laird estimator for  $\tau^2$

# Kapitelübersicht

## 10. Befundsynthese von Aggregatdaten

Fixed-effects model (FEM)

Random-effects model (REM)

## 11. Begriff der Heterogenität und Diagnostik

Überblick

Die Zwischenstudienvarianz

Die Q- Statistik

Weitere Heterogenitätsmaße:  $H^2$  und  $I^2$

Zusammenfassung Heterogenitätsmaße

## 12. Heterogenitätsaufklärung

Meta-Regression

ANOVA-ähnliche Verfahren

Weitere Bemerkungen zu Verfahren der Heterogenitätsaufklärung

## 13. Publication Bias

## 14. Reporting Standards

## 15. Ein Ausblick auf Meta-Analysen mit Individualdaten

## 16. Literatur

# Abschnittsübersicht

10. Befundsynthese von Aggregatdaten

## 11. Begriff der Heterogenität und Diagnostik

Überblick

Die Zwischenstudienvarianz

Die Q- Statistik

Weitere Heterogenitätsmaße:  $H^2$  und  $\rho^2$

Zusammenfassung Heterogenitätsmaße

12. Heterogenitätaufklärung

13. Publication Bias

14. Reporting Standards

15. Ein Ausblick auf Meta-Analysen mit Individualdaten

16. Literatur

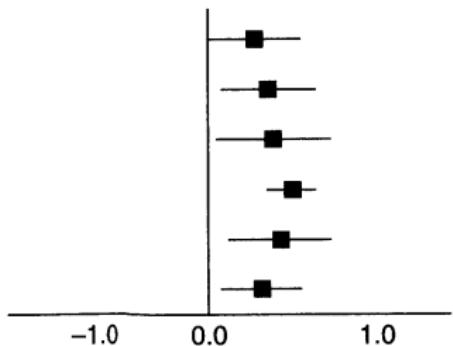
# Ist Heterogenität ein Problem?

- Die ES-Synthese ist eine (relativ) einfache Operation.
- Ist diese mittlere ES eine angemessene Zusammenfassung der zugrunde liegenden Verteilung?
- Das hängt u.a. von der „Gesamtvariation“ der ES ab:
  - Die Gesamtvariation lässt sich in zwei Teile zerlegen: (1) Stichprobenfehler („random error“; „within-study error“), (2) „wahre“ Streuung ("between study error“, „heterogeneity“)
  - „[...] we use *heterogeneity* to mean heterogeneity in true effects only“ (Borenstein et al 2009: 106).
- Ziel der Heterogenitätsdiagnostik ist u.a. das Ausmaß der „wahren“ Streuung zu bestimmen.

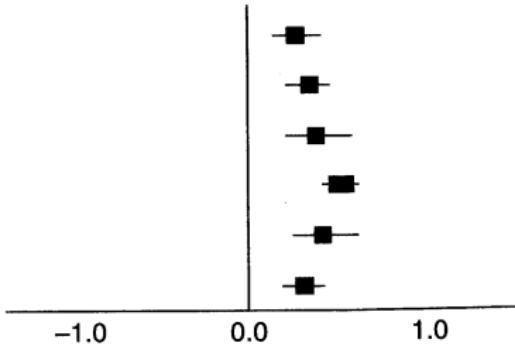


# Schätzen Sie mal ...

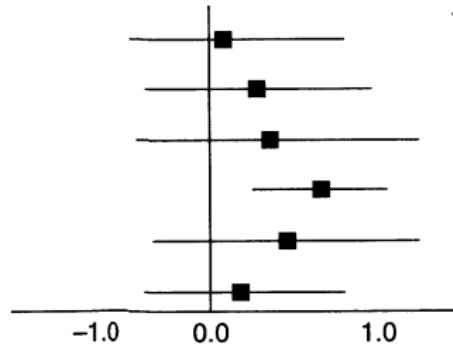
A



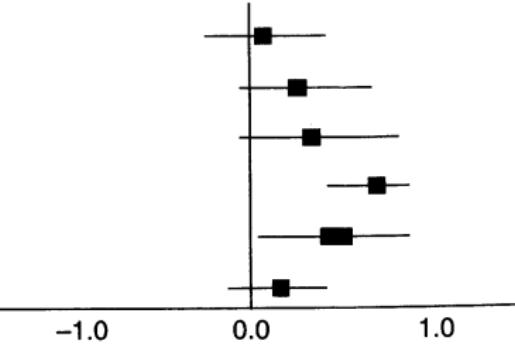
B



C



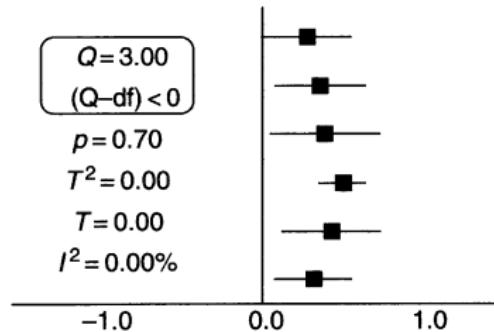
D



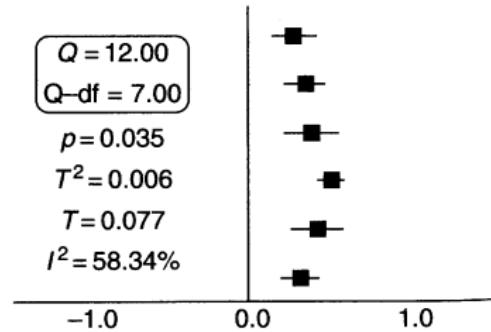
(Quelle: Borenstein et al. 2009: 108)

# Schätzen Sie mal ... die Auflösung

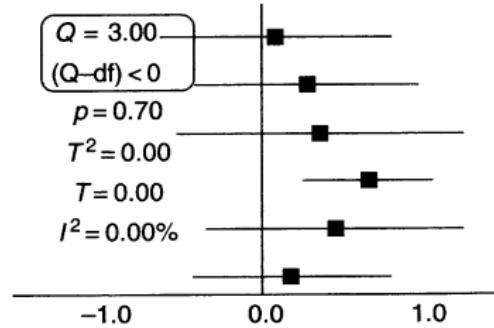
A



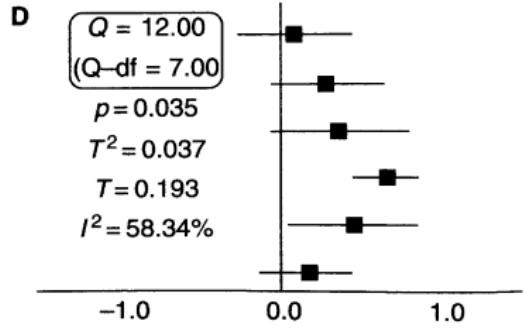
B



C



D



(Quelle: Borenstein et al. 2009: 110)



# Anliegen und Maße der Heterogenitätsdiagnostik

- „Is there evidence of heterogeneity in true effect sizes? (Q-Test)
- What is the variance of the true effects? ( $\tau^2$ ,  $\hat{\tau}^2 = T^2$ )
- What proportion of the observed dispersion is real?“ ( $I^2$ )

(Quelle: Borenstein et al., 2009, S. 105)



# Abschnittsübersicht

10. Befundsynthese von Aggregatdaten

## 11. Begriff der Heterogenität und Diagnostik

Überblick

Die Zwischenstudienvarianz

Die Q- Statistik

Weitere Heterogenitätsmaße:  $H^2$  und  $\rho^2$

Zusammenfassung Heterogenitätsmaße

## 12. Heterogenitätaufklärung

## 13. Publication Bias

## 14. Reporting Standards

## 15. Ein Ausblick auf Meta-Analysen mit Individualdaten

## 16. Literatur

# Die Zwischenstudienvarianz

„The parameter<sup>4</sup> tau-squared ( $\tau^2$ ) is defined as the variance of the true effect sizes. In other words, if we had an infinitely large sample of studies, each, itself, infinitely large (so that the estimate in each study was the true effect) and computed the variance of these effects, this variance would be  $\tau^2$ “  
(Borenstein et al. 2009: 114).

---

<sup>4</sup>In der Literatur finden sich uneinheitliche Notationen, um die Zwischenstudienvarianz zu beschreiben. Teilweise wird keine unterschiedliche Notation verwendet, um zwischen Parameter und Schätzwert zu unterscheiden. Teilweise finden sich aber auch folgende Varianten:  $\tau^2$  und  $\hat{\tau}^2$  oder  $\tau^2$  und  $T^2$ .



# Wie lässt sich die „wahre“ Zwischenstudienvarianz isolieren?

„The mechanism that we use to extract the true between-studies variation from the observed variation is as follows:

1. We compute the total amount of study-to-study variation actually observed ( $Q$ ).
2. We estimate how much the observed effects would be expected to vary from each other if the true effect was actually the same in all studies ( $k - 1 = df$ ).
3. The excess variation (if any) is assumed to reflect real differences in effect size (that is, the heterogeneity)“ (Borenstein et al. 2009: 108).



# Schätzen der Zwischenstudienvarianz $\tau^2$

Der sog. Momentenschätzer<sup>5</sup> für  $\tau^2$  lautet:

$$\hat{\tau}^2 = \frac{Q - (k - 1)}{c} = \frac{Q - df}{c}$$

mit:

$k$ : Anzahl der Effektstärken

$df$ : Anzahl der Freiheitsgrade ( $df = k - 1$ )

$Q$ : Maß der Zwischenstudienvariation,  $Q \sim \chi_{df}^2$  (Heterogenitätstest, siehe ausführlich Folie 165)

$c$ : Skalierungsfaktor, sorgt dafür, dass  $\tau^2$  wieder die Metrik der Ausgangs-ES erhält.  $c$  berechnet sich nach (zur Erinnerung:  $w_i = 1 / v_i$ ):

$$c = \sum_{j=1}^k w_i - \left( \frac{\sum_{j=1}^k w_i^2}{\sum_{j=1}^k w_i} \right) \quad (39)$$

---

<sup>5</sup>Für eine Übersicht und einen Vergleich verschiedener Schätzer von  $\tau^2$  siehe Viechtbauer (2005).



# Abschnittsübersicht

10. Befundsynthese von Aggregatdaten

## 11. Begriff der Heterogenität und Diagnostik

Überblick

Die Zwischenstudienvarianz

Die Q- Statistik

Weitere Heterogenitätsmaße:  $H^2$  und  $\rho^2$

Zusammenfassung Heterogenitätsmaße

12. Heterogenitätaufklärung

13. Publication Bias

14. Reporting Standards

15. Ein Ausblick auf Meta-Analysen mit Individualdaten

16. Literatur

# Die Heterogenitätsstatistik $Q$

- Bestimmen der Gesamtabweichung („total amount of study-to-study variation“, „observed weighted sum of squares“ (WSS)):

$$Q = \sum_{j=1}^k \left( \frac{(T_j - \bar{T}_{FEM})^2}{SE_j^2} \right) \quad (40)$$

- Erwartete Variation, wenn es keine Zwischenstudienvarianz gäbe (= FEM) (expected WSS):

$$df = k - 1 \quad (41)$$

- Ausmaß der Zwischenstudienvariation („excess variation“):

$$Q - df \quad (42)$$



# Das Prinzip des Q-Tests

- Hypothesen:
  - $H_0$ : Die Effektstärken weisen einen gemeinsamen Populationsparameter auf.
  - $H_1$ : Die Effektstärken weisen mehr als einen gemeinsamen Populationsparameter auf.
- Wenn die  $H_0$  *nicht* zurückwiesen werden kann, dann wird eine homogene Effektstärkenverteilung unterstellt.
- $Q$  ist mit  $df = k - 1$  Freiheitsgraden  $\chi^2$ -verteilt. Für jeden Freiheitsgrad lässt sich ein theoretischer  $\chi^2$ -Wert angeben, der sich mit dem empirischen  $\chi^2$ -Wert ( $= Q$ ) vergleichen lässt. Ist der empirische  $\chi^2$ -Wert größer als der theoretische (= statistisch signifikant), dann kann man *keine* homogene Effektstärkenverteilung annehmen.

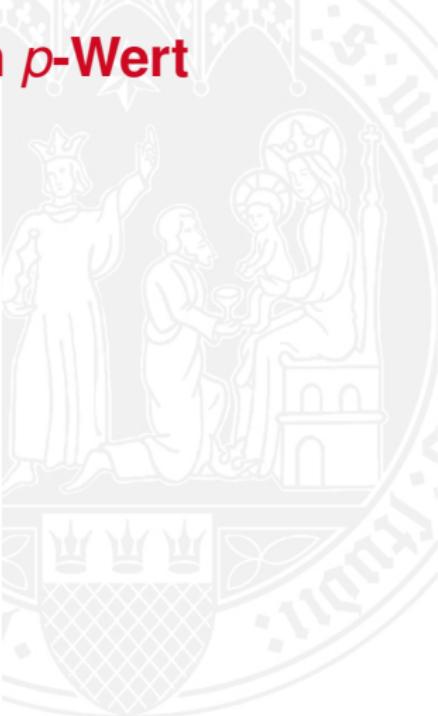
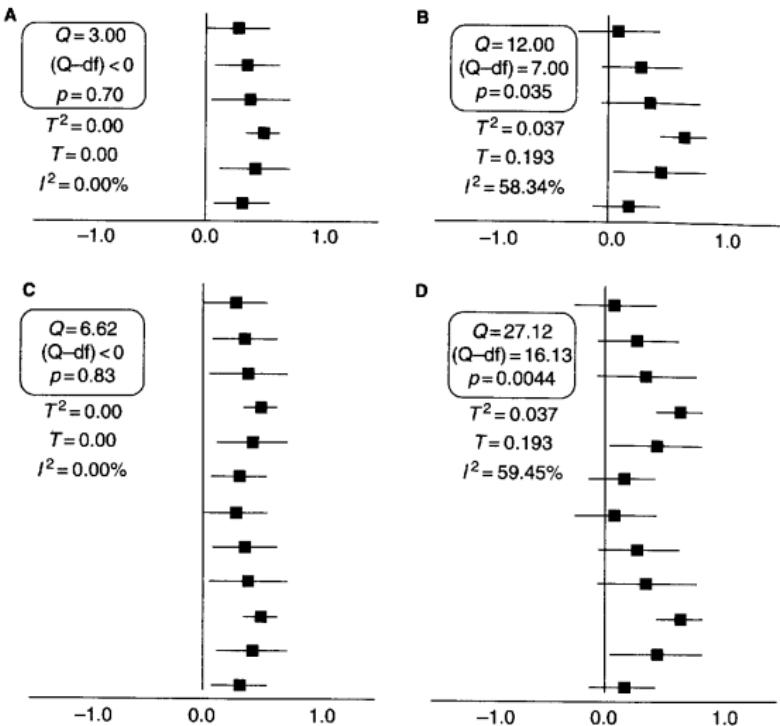


# Hinweise zur Interpretation von Q

- Der Q-Test weist eine niedrige statistische *power* auf, d.h. eine relativ hohe Wahrscheinlichkeit einen Fehler 2. Art ( $\beta$ -Fehler) zu begehen. *Dies gilt vor allem für Meta-Analysen mit kleinen Fallzahlen ( $k < 20$ )*. Mit anderen Worten: Ein nicht-signifikanter Test belegt nicht unbedingt eine homogene Effektstärkenverteilung. Deshalb wird häufig auch ein Signifikanzniveau von 10% angesetzt.
- „[...] While a significant  $p$ -value provides evidence that the true effects vary, the converse is not true. A nonsignificant  $p$ -value should not be taken as evidence that the effect sizes are consistent, since the lack of significance may be due to low power. With a small number of studies and/or large within-study variance (small studies), even substantial between-studies dispersion might yield a nonsignificant  $p$ -value“ (Borenstein et al. 2009: 113).

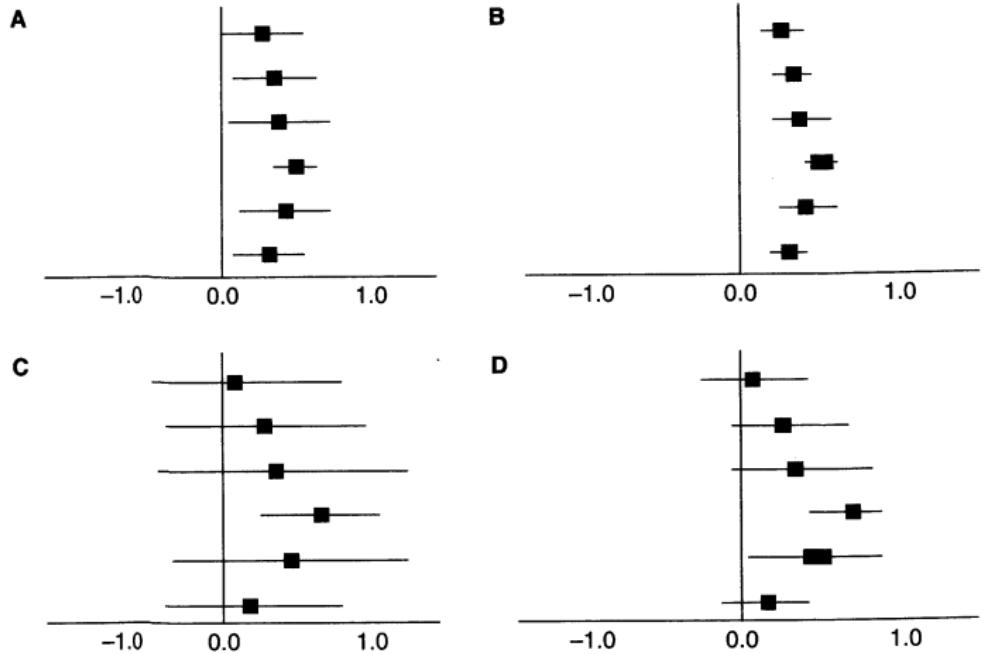


# Einfluss der Fallzahl auf Q und den p-Wert



(Quelle: Borenstein et al. 2009: 113)

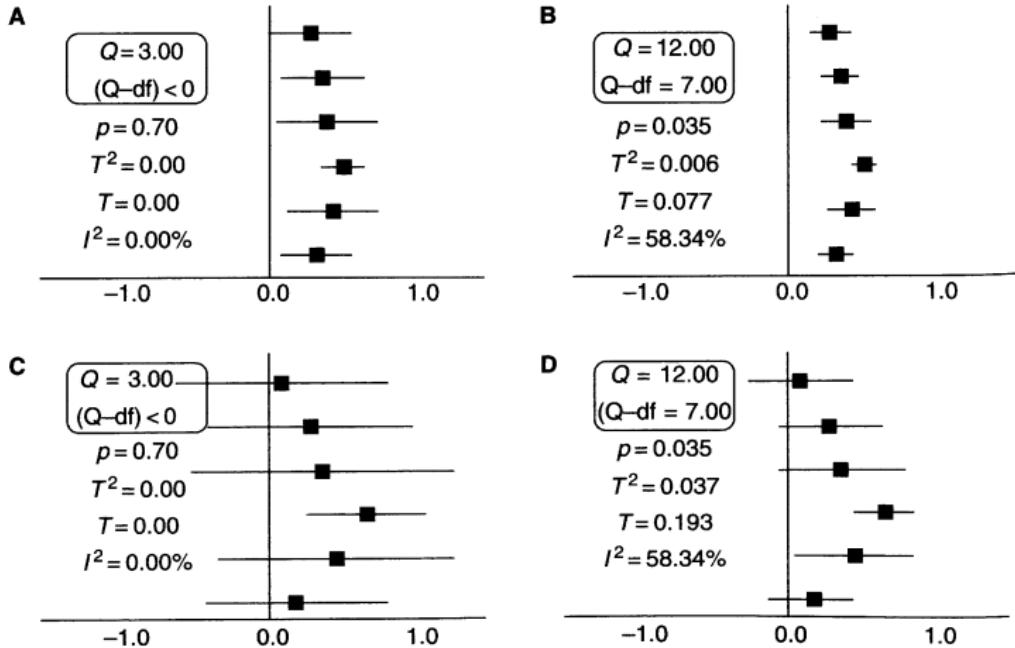
# Beispiel für das Verhältnis von Binnen- (within) und Zwischenstudienvarianz (between)



(Quelle: Borenstein et al. 2009: 108)



# Beispiel für das Verhältnis von Binnen- (within) und Zwischenstudienvarianz (between)



(Quelle: Borenstein et al. 2009: 110)



# Exkurs: Dimensionale Analyse von $Q$ und $\tau^2$

**Ausgangsproblem:** Die Zwischenstudienvarianz ergibt sich nach  $\tau^2 = \frac{Q-df}{C}$ .  $C$  wird u.a. benötigt, um die

Originalmetrik/-dimensionalität wieder herzustellen („putting the measure back into its original metric“, Borenstein et al. S. 114).

**Vorbemerkungen:** Ich werde dimensionale Gleichheit mit dem Zeichen  $\leftrightarrow$  kennzeichnen und die Dimension in eckigen Klammern darstellen, also  $[r]$  für einen Korrelationskoeffizienten. Ein leeres eckiges Klammerpaar  $[ ]$  bezeichnet Dimensionslosigkeit.

- Sei  $\tau^2 = \frac{Q-df}{C} \leftrightarrow [r^2]$  und  $w = \frac{1}{SE^2} \leftrightarrow [\frac{1}{r^2}]$ .
- Für  $Q$  und  $df$  gilt:  $Q = \sum \frac{(T_j - \bar{T}_{FEM})^2}{SE_j^2} \leftrightarrow \sum [\frac{r^2}{r^2}] = [ ]$  und  $df = [ ]$  und damit  $Q - df \leftrightarrow [ ] - [ ] = [ ]$ .
- Für  $C$  gilt:  
$$C = \sum w - \frac{\sum w^2}{\sum w} \leftrightarrow [\frac{1}{r^2}] - [\frac{\frac{1}{r^2}}{\frac{1}{r^2}}] = [\frac{1}{r^2}] - ([\frac{1}{r^2}] \cdot [\frac{r^2}{1}]) = [\frac{1}{r^2}].$$
- Daraus folgt für  $\tau^2 = \frac{Q-df}{C} \leftrightarrow [\frac{1}{r^2}] = [ ] \cdot [r^2] = [r^2]$ .

# Abschnittsübersicht

10. Befundsynthese von Aggregatdaten

## 11. Begriff der Heterogenität und Diagnostik

Überblick

Die Zwischenstudienvarianz

Die Q- Statistik

Weitere Heterogenitätsmaße:  $H^2$  und  $I^2$

Zusammenfassung Heterogenitätsmaße

12. Heterogenitätaufklärung

13. Publication Bias

14. Reporting Standards

15. Ein Ausblick auf Meta-Analysen mit Individualdaten

16. Literatur

# Kritik an den bisherigen Heterogenitätstests/-statistiken

Probleme mit den bisherigen Heterogenitätsstatistiken:

- $\tau^2$  ist nicht zwischen Meta-Analysen vergleichbar; hängt von ES ab.
- $Q$  hat geringe statistische Testpower (hoher  $\beta$ -Fehler).

Eine verbesserte Statistik sollte folgende Kriterien erfüllen:

- Abhängigkeit vom Ausmaß der Heterogenität („Dependence on the extent of heterogeneity“).
- Skaleninvarianz („Scale invariance“).
- Unabhängigkeit von der Fallzahl  $k$  („Size invariance“).

(Quelle: Higgins & S. G. Thompson, 2002)

$$H = \sqrt{\frac{Q}{k - 1}}. \quad (43)$$

$H$  wird interpretiert als eine an der Zahl der Freiheitsgrade standardisierte  $Q$ -Statistik.



# Konzeptionelle und statistische Definition von $I^2$

- $I^2$  ist *konzeptionell* das prozentuale Verhältnis der Zwischengruppenvarianz zur Gesamtvarianz (ähnlich wie der ICC für HLM):

$$I^2 = \frac{\text{Zwischenstudienvarianz}}{\text{Gesamtvarianz}} \times 100 = \frac{\tau^2}{\tau^2 + \sigma^2} \times 100$$

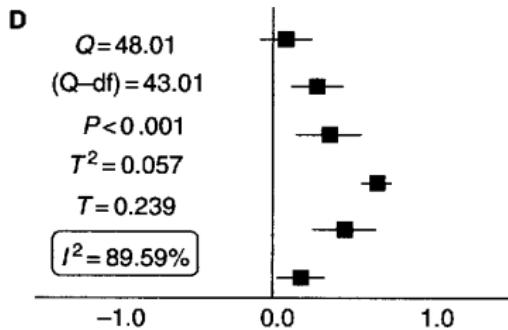
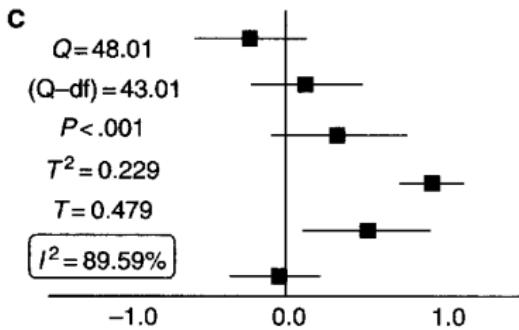
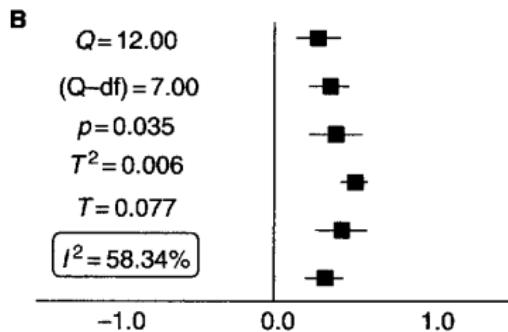
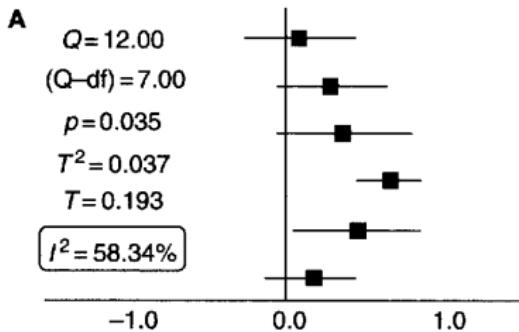
- Die Maßzahl  $I^2$  selbst lässt sich aus  $H^2$  ableiten und schätzen:

$$I^2 = \frac{H^2 - 1}{H^2} \times 100 = \frac{Q - (k - 1)}{Q} \times 100 = \frac{Q - df}{Q} \times 100.$$

- Werte von  $I^2$  um 25% ( $I^2 = 0,25$ ) entsprechen einer niedrigen Heterogenität. Dagegen sind Werte um 50% ( $I^2 = 0,50$ ) Anzeichen für eine milde und Werte um 75% Anzeichen für eine deutliche Heterogenität.



# Der Zusammenhang zwischen $Q$ und $I^2$



(Quelle: Borenstein et al. 2009: 118)



# Unsicherheits-/Konfidenzintervall (KI) für $I^2$

- Das KI für  $I^2$  kann nicht direkt berechnet werden, sondern wird über den Standardfehler (SE) von  $Q$  bzw.  $\ln(Q)$  abgeleitet.
- Bei der Berechnung des SE von  $\ln(Q)$  müssen zwei Fälle unterschieden werden: (a)  $Q > (df + 1)$  und (b)  $Q \leq (df + 1)$ .
- Anschließend lässt sich bspw. das 95%-KI für  $\ln(Q)$  konzeptionell nach ( $Statistik \pm 1,96 \times SE$ ) berechnen.
- Wenn das KI für  $I^2$  die 0 einschließt, dann wird von einer *homogenen* Effektstärkenverteilung (FEM) ausgegangen.

(Quelle: Higgins/Thompson 2002: 1554f; Borenstein et al. 2009: 124f)



# Vergleich zwischen $I^2$ und $Q$

- „With respect to the control of Type I error rate, the performance of the  $Q$  test and the  $I^2$  confidence interval was very similar“ (203).
- „With respect to statistical power, there were no notable differences between the  $Q$  test and the  $I^2$  confidence interval“ (203) .
- „In summary, our findings show that the  $I^2$  confidence interval performs in a similar way to the  $Q$  test from an inferential point of view“ (204).

(Quelle: Huedo-Medina, Botella, Sánchez-Meca & Marín-Martínez, 2006)



# Abschnittsübersicht

10. Befundsynthese von Aggregatdaten

## 11. Begriff der Heterogenität und Diagnostik

Überblick

Die Zwischenstudienvarianz

Die Q- Statistik

Weitere Heterogenitätsmaße:  $H^2$  und  $\beta^2$

Zusammenfassung Heterogenitätsmaße

12. Heterogenitätaufklärung

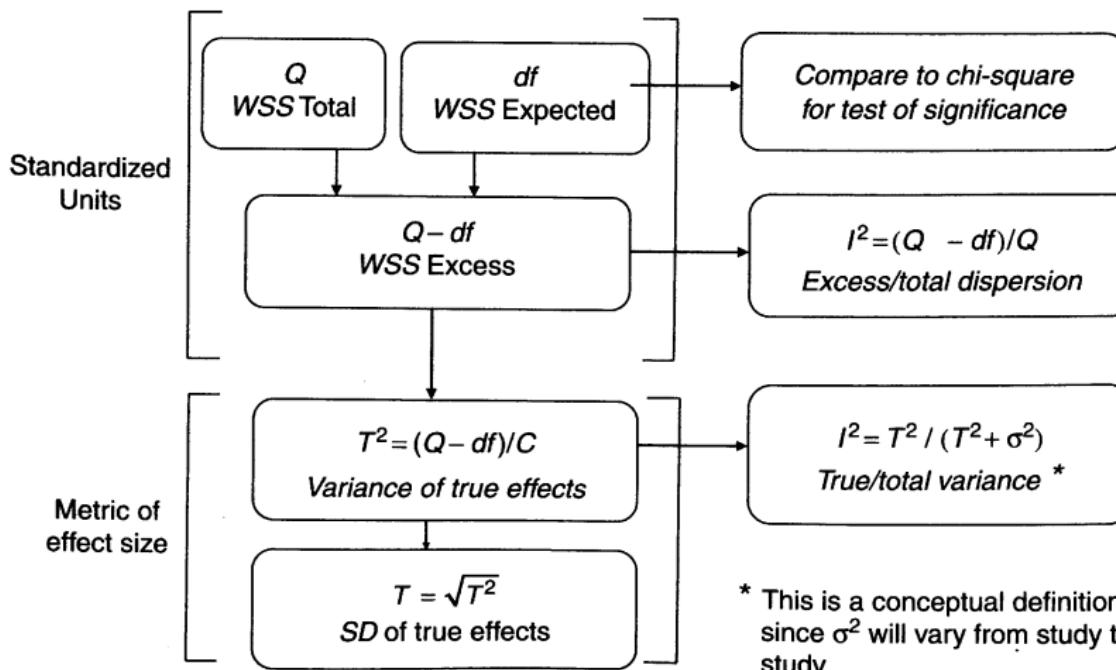
13. Publication Bias

14. Reporting Standards

15. Ein Ausblick auf Meta-Analysen mit Individualdaten

16. Literatur

# Verschiedene Heterogenitätsmaße im Überblick



(Quelle: Borenstein et al. 2009: 111)



# Zusammenfassende Bemerkungen zur Heterogenität (1)

- Im Rahmen einer (aggregatdatenbasierten) Meta-Analyse lassen sich zwei Streuungskomponenten unterscheiden:
  1. Binnenstudien- (zufallsbedingter Stichprobenfehler; *within*;  $\sigma_j^2$ ) und
  2. Zwischenstudienvariation (systematische Heterogenität; *between*;  $\tau^2$ ).
- Die für die ES ( $T_j$ ) unmittelbar beobachtbare Varianz  $Var(T_j)$  ist *nicht* in der Lage, die Gesamtvariation adäquat zu beschreiben (u.a. findet der *within*-Teil keine Berücksichtigung).



# Zusammenfassende Bemerkungen zur Heterogenität (2)

Ein sinnvolles Maß zur Beschreibung der Gesamtvariation ist die gewichtete Quadratsumme (*weighted sum of squares*)  $Q$ . Diese muss nun in die beiden Streuungskomponenten zerlegt werden:

1. Binnenstudienvariation ( $k - 1$ ): Diese Variation der  $T_j$  ist allein dem Stichprobenfehler geschuldet. Es gibt *keine* Möglichkeit, die Binnenstudienvariation in Form einer gemeinsamen (oder gesamten) Binnenstudienvarianz auszudrücken. (Allerdings lässt sich für das FEM bzw. das REM ein zusammengefasster Standardfehler ( $\bar{SE}$ ) bzw. eine Fehlervarianz angeben  $\bar{v}$ )
2. Zwischenstudienvariation ( $Q - (k - 1)$ ): Es gibt neben dem Stichprobenfehler einen (bislang<sup>6</sup>) unbekannten Einfluss, der zur Variation der  $T_j$  beiträgt. Mit Hilfe der Zwischenstudienvariation lässt sich (ein möglicher) Schätzer der Zwischenstudienvarianz ermitteln ( $\hat{\tau}^2$ ).

---

<sup>6</sup>Siehe hierzu das Thema Meta-Regression.



# Einflussgrößen der Heterogenität (1)

Drei Faktoren haben einen Einfluss auf die Gesamtvariation (sowie ihre beiden Komponenten):

1. der absolute Abstand  $|T_j - \bar{T}_{FEM}|$  der  $T_j$  zum  $\bar{T}_{FEM}$ ,
2. der Standardfehler  $SE_j$  der einzelnen  $T_j$  sowie
3. die Anzahl ( $k$ ) der  $T_j$ .



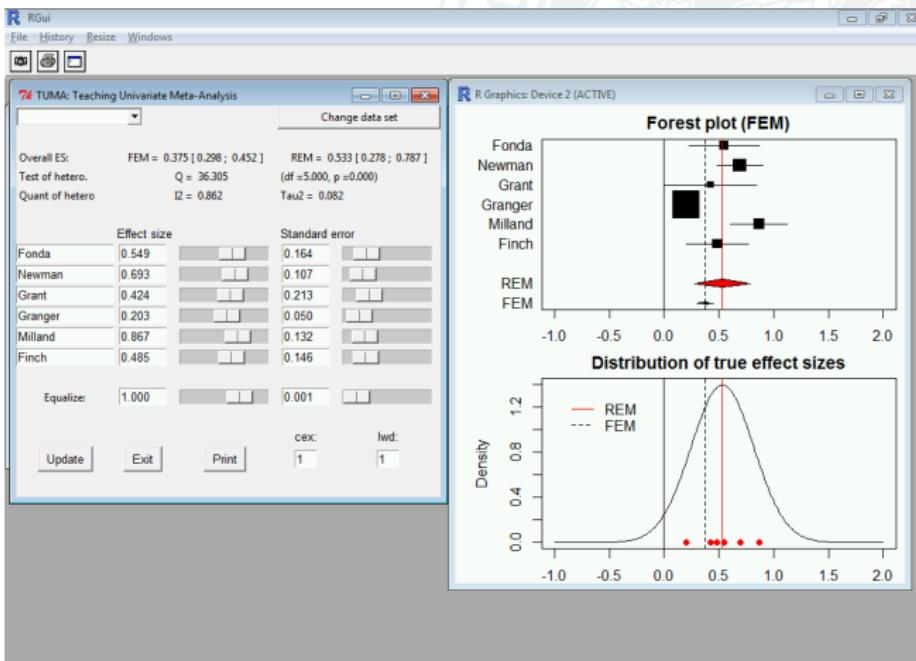
## Einflussgrößen der Heterogenität (2)

Diese drei Einflussgrößen stehen teilweise in einem entgegengesetzten Verhältnis zu  $Q$ ,  $\tau^2$  und  $I^2$ . Werden die anderen beiden Einflussgrößen konstant gehalten, dann führt ein Vergrößern von

- $|T_j - \bar{T}_{FEM}|$  zu einem höheren  $Q$  und, sofern  $Q - df > 0$  gilt, auch zu einem höheren  $\tau^2$  und  $I^2$ .
- $k$  zu einem höheren  $Q$  (da sich auch  $Q - df$  ändert, folgt nicht zwangsläufig eine Erhöhung von  $\tau^2$  oder  $I^2$ ).
- $SE_j$  zu einem kleineren  $Q$  und, sofern noch  $Q - df > 0$  gilt, auch zu einem kleineren  $\tau^2$  und damit auch  $I^2$ .



# Interaktives Erforschen der verschiedenen Heterogenitätsmaße



(Quelle: <https://github.com/berndweiss/tuma>)

# Kapitelübersicht

## 10. Befundsynthese von Aggregatdaten

Fixed-effects model (FEM)

Random-effects model (REM)

## 11. Begriff der Heterogenität und Diagnostik

Überblick

Die Zwischenstudienvarianz

Die Q- Statistik

Weitere Heterogenitätsmaße:  $H^2$  und  $I^2$

Zusammenfassung Heterogenitätsmaße

## 12. Heterogenitätsaufklärung

Meta-Regression

ANOVA-ähnliche Verfahren

Weitere Bemerkungen zu Verfahren der Heterogenitätsaufklärung

## 13. Publication Bias

## 14. Reporting Standards

## 15. Ein Ausblick auf Meta-Analysen mit Individualdaten

## 16. Literatur

# Methoden zur Heterogenitätsaufklärung

Methoden:

- Subgruppenanalysen (vergleichbar ANOVA; fixed-effects oder mixed-effects model)
- Meta-Regression (fixed-effects oder mixed-effects model)

Erklärungsfaktoren (unabh. Variablen, Prädiktoren):

- (Aggregierte) Individualmerkmale (Achtung: Gefahr eines ökologischen Fehlschlusses, siehe Folie 211)
- Studienmerkmale

Abhängige Variable: Effektstärken (Zusammenhangsmaße; damit wird ein Interaktionseffekt modelliert)



# Abschnittsübersicht

10. Befundsynthese von Aggregatdaten

11. Begriff der Heterogenität und Diagnostik

## 12. Heterogenitätsaufklärung

Meta-Regression

ANOVA-ähnliche Verfahren

Weitere Bemerkungen zu Verfahren der Heterogenitätsaufklärung

13. Publication Bias

14. Reporting Standards

15. Ein Ausblick auf Meta-Analysen mit Individualdaten

16. Literatur



# Was ist eine Meta-Regression?

„We use the term meta-regression to indicate the use of trial-level covariates, as distinct from regression analyses that are possible when individual patient data on outcomes and covariates are available“ (S. G. Thompson & Higgins, 2002, S. 1560).



# Das FEM der Meta-Regression

- Ohne Kovariaten (= Nullmodell) bestimmt sich der Schätzer des FEM nach:  $T_j = \theta + \epsilon_j$  und es wird angenommen, dass  $T_j \sim N(\theta, \sigma_j^2)$  für  $j = 1, 2, \dots, k$  Studien.
- Für  $X_1, \dots, X_p$  Prädiktorvariablen (Kovariaten, unabh. Variablen) bestimmt sich das FE Regressionsmodell nach:

$$\theta_j = \theta = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p \quad (44)$$

- Gleichung 44 lässt sich auflösen nach:

$$T_j = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p + \epsilon_j \quad (45)$$

(Quelle: Sutton, Abrams, Jonas, Sheldon & Song, 2000, S. 94)



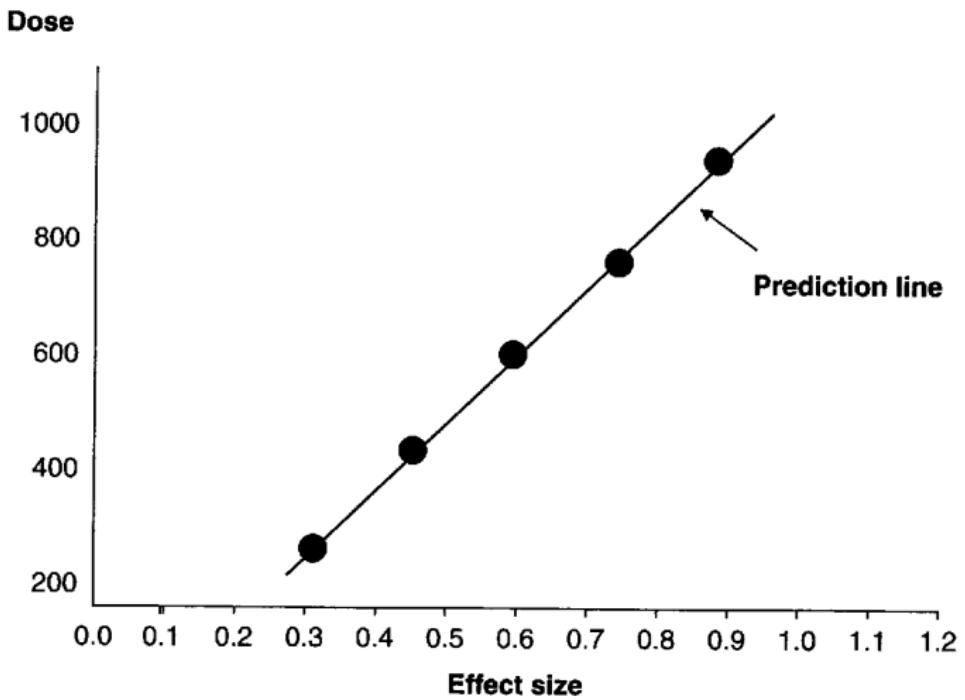
# Das FEM der Meta-Regression: Vorgehen

- Das (lineare) FEM der Meta-Regression wird nicht mit Hilfe eines OLS- (*ordinary least squares*), sondern eines WLS-Schätzers (*weighted least squares*) ermittelt.
- Als Gewichte wird die inverse Fehlervarianz verwendet, also  $w_j = \frac{1}{SE_j^2}$ .
- Eine WLS-Regression lässt sich mit den meisten Statistikpaketen durchführen, wichtig ist jedoch die anschließende Korrektur der Standardfehler (SE) der Regressionskoeffizienten ( $\beta_j$ ) nach

$$SE_j^{\text{korrigiert}} = \frac{SE_j^{\text{original}}}{\sqrt{MS_{\text{error}}}}. \quad (46)$$

- Die Gesamtvariation ( $Q$ ) lässt sich in zwei Komponenten zerlegen, in  $Q_{\text{resid}}$  ( $df = k - p - 1$ ) (oder  $Q_{\text{error}}$ , unerklärte Variation) und  $Q_{\text{model}}$  ( $df = p$ ) (erklärte Variation).

# Das FEM der Meta-Regression



(Quelle: Borenstein et al. 2009: 194)



# „Mixed-effects“ Regression (das „REM“ der Meta-Regression)

- Ohne Kovariaten (= Nullmodell) bestimmt sich der Schätzer des REM nach:  $T_j = \theta_j + \epsilon_j = \theta + \epsilon_j + u_j$  und es wird angenommen, dass  $\theta_j \sim N(\theta, \tau^2)$  für  $j = 1, 2, \dots, k$ .
- Für  $X_1, \dots, X_p$  Prädiktorvariablen (Kovariaten, unabh. Variablen) ergibt sich das *mixed-effects model* nach:

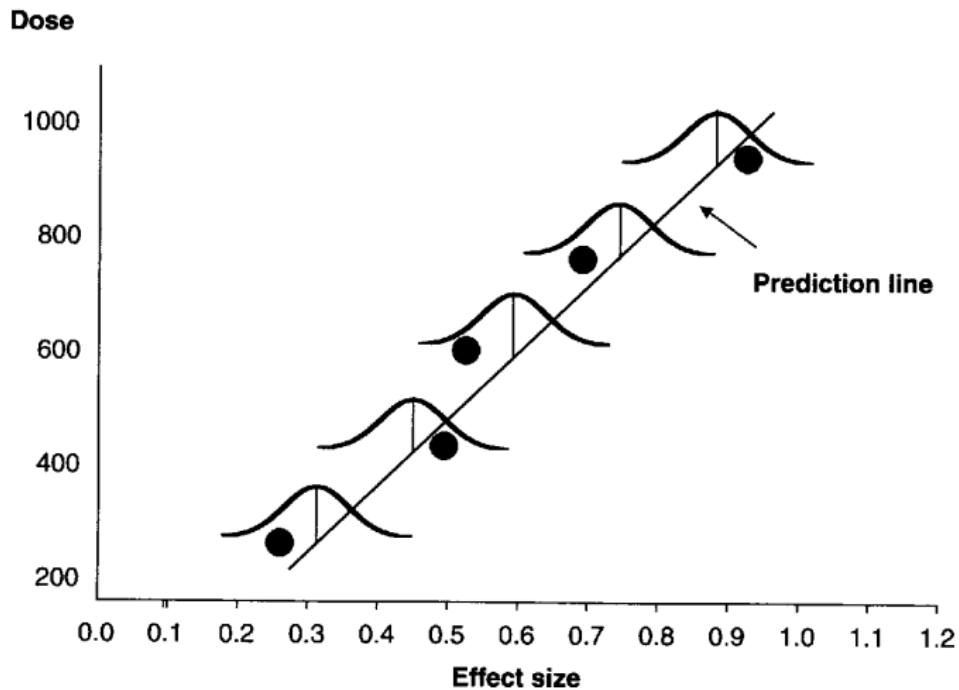
$$T_j = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p + \epsilon_j + u_j. \quad (47)$$

- Im Fall von erkläzungskräftigen Prädiktoren reduziert sich  $\tau^2$  und es lässt sich der Anteil an erklärter Zwischenstudienvarianz berechnen:

$$R^2 = 1 - \frac{\tau^2_{\text{unerklaert}}}{\tau^2_{\text{Gesamt / Nullmodell}}}.$$

(Quelle: Sutton, Abrams et al., 2000, S. 98)

# Das „REM“ der Meta-Regression



(Quelle: Borenstein et al. 2009: 194)



# Beispiel: Wirksamkeit des BCG-Impfstoffs zur Bekämpfung der Tuberkulose

## Fragestellung

- Es gibt zahlreiche (13) Studien, welche die Wirksamkeit des BCG-Impfstoff zur Verhinderung der Tuberkulose untersuchen.
- Kontrollgruppe sind ungeimpfte Personen, als Effektstärke wird das *relative risk* (RR) verwendet. Beispiel:
  - Geimpft: 4 erkrankt, 119 nicht erkrankt
  - Nicht geimpft: 11 erkrankt, 128 nicht erkrankt
  - RR:  $(4/(4+119))/(11/(11+128)) = 0.4109$
- Es wird vermutet, dass der Impfstoff in kälteren Klimazonen wirksamer ist und daher das RR mit der Entfernung zum Äquator (hier: geographische Breite) zunimmt.



# Beispiel: Wirksamkeit des BCG-Impfstoffs zur Bekämpfung der Tuberkulose

## Datensatz

	tpos	tneg	cpos	cneg	ablat	yi	vi
1	4	119	11	128	44	-0.88931	0.325585
2	6	300	29	274	55	-1.58539	0.194581
3	3	228	11	209	42	-1.34807	0.415368
4	62	13536	248	12619	52	-1.44155	0.020010
5	33	5036	47	5761	13	-0.21755	0.051210
6	180	1361	372	1079	44	-0.78612	0.006906
7	8	2537	10	619	19	-1.62090	0.223017
8	505	87886	499	87892	13	0.01195	0.003962
9	29	7470	45	7232	27	-0.46942	0.056434
10	17	1699	65	1600	42	-1.37134	0.073025
11	186	50448	141	27197	18	-0.33936	0.012412
12	5	2493	3	2338	33	0.44591	0.532506
13	27	16886	29	17825	33	-0.01731	0.071405

tpos = treatment group, erkrankt; tneg = treatment group, nicht erkrankt; cpos = control group, erkrankt; cneg = control group, nicht erkrankt; ablat = Abs. Breitengrad

# Beispiel: Wirksamkeit des BCG-Impfstoffs zur Bekämpfung der Tuberkulose

## Nullmodell

```
Random-Effects Model (k = 13; tau^2 estimator: DL)

tau^2 (estimate of total amount of heterogeneity): 0.3088
tau (sqrt of the estimate of total heterogeneity): 0.5557
I^2 (% of total variability due to heterogeneity): 92.12%
H^2 (total variability / sampling variability): 12.69

Test for Heterogeneity:
Q(df = 12) = 152.2330, p-val < .0001

Model Results:

estimate      se      zval     pval    ci.lb    ci.ub
-0.7141  0.1787  -3.9952  <.0001  -1.0644  -0.3638      ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



# Beispiel: Wirksamkeit des BCG-Impfstoffs zur Bekämpfung der Tuberkulose

## Fixed-effects model

```
Fixed-Effects with Moderators Model (k = 13)
```

```
Test for Residual Heterogeneity:  
QE(df = 11) = 30.7331, p-val = 0.0012
```

```
Test of Moderators (coefficient(s) 2):  
QM(df = 1) = 121.4999, p-val < .0001
```

```
Model Results:
```

	estimate	se	zval	pval	ci.lb	ci.ub	
intrcpt	0.3436	0.0810	4.2390	<.0001	0.1847	0.5024	***
ablat	-0.0292	0.0027	-11.0227	<.0001	-0.0344	-0.0240	***

```
---
```

```
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



# Beispiel: Wirksamkeit des BCG-Impfstoffs zur Bekämpfung der Tuberkulose

Mixed-effects model („method of moments“)

```
Mixed-Effects Model (k = 13; tau^2 estimator: DL)
```

```
tau^2 (estimate of residual amount of heterogeneity): 0.0633  
tau (sqrt of the estimate of residual heterogeneity): 0.2516
```

```
Test for Residual Heterogeneity:  
QE(df = 11) = 30.7331, p-val = 0.0012
```

```
Test of Moderators (coefficient(s) 2):  
QM(df = 1) = 18.8452, p-val < .0001
```

Model Results:

	estimate	se	zval	pval	ci.lb	ci.ub	
intrcpt	0.2595	0.2323	1.1172	0.2639	-0.1958	0.7149	
ablat	-0.0292	0.0067	-4.3411	<.0001	-0.0424	-0.0160	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1



# Beispiel: Wirksamkeit des BCG-Impfstoffs zur Bekämpfung der Tuberkulose

## Mixed-effects model (REML)

```
Mixed-Effects Model (k = 13; tau^2 estimator: REML)

tau^2 (estimate of residual amount of heterogeneity): 0.0764 (SE = 0.0591)
tau (sqrt of the estimate of residual heterogeneity): 0.2763

Test for Residual Heterogeneity:
QE(df = 11) = 30.7331, p-val = 0.0012

Test of Moderators (coefficient(s) 2):
QM(df = 1) = 16.3571, p-val < .0001

Model Results:

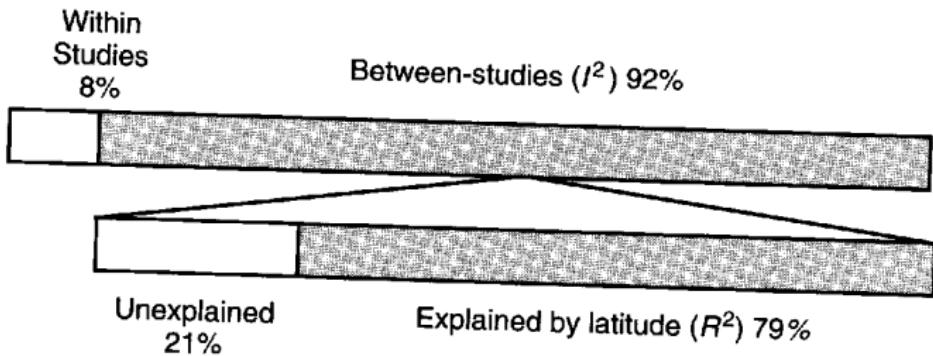
            estimate      se     zval    pval    ci.lb    ci.ub
intrcpt     0.2515  0.2491   1.0095  0.3127  -0.2368  0.7397
ablat       -0.0291  0.0072  -4.0444 <.0001  -0.0432  -0.0150  ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



# Das „REM“ der Meta-Regression: Erklärte Varianz

$$R^2 = 1 - \frac{\hat{T}_{unexplained}}{\hat{T}_{total}} = 1 - \frac{0.0633}{0.3088} = 0.795 \quad (48)$$

(Basiert auf den Ergebnissen des MM-Schätzer.)



(Quelle: Borenstein et al. 2009: 202)



# Abschnittsübersicht

10. Befundsynthese von Aggregatdaten

11. Begriff der Heterogenität und Diagnostik

## 12. Heterogenitätsaufklärung

Meta-Regression

ANOVA-ähnliche Verfahren

Weitere Bemerkungen zu Verfahren der Heterogenitätsaufklärung

13. Publication Bias

14. Reporting Standards

15. Ein Ausblick auf Meta-Analysen mit Individualdaten

16. Literatur



# ANOVA-ähnliche Modelle

## Grundprinzip

- Die  $T_{ji}$  lassen sich ( $p$ ) Gruppen zuweisen, mit  $j = 1, \dots, p$  Gruppen,  $i = 1, \dots, k_j$  Studien und  $k = \sum k_j$ .
- Sei weiterhin  $T_{j\bullet}$  der geschätzte mittlere Effekt in Gruppe  $j$  und  $\theta_{j\bullet}$  der Parameter für Gruppe  $j$ , dann lautet die erste Frage in einer ANOVA: Unterscheiden sich die  $\theta_{j\bullet}$  voneinander? Die  $H_0$  lautet:  $H_0 : \theta_{1\bullet} = \theta_{2\bullet} = \dots = \theta_{p\bullet}$ .
- Die zweite Frage lautet: Unterscheiden sich die  $\theta_{ji}$  in der  $j$ -ten Gruppe? Die entsprechende  $H_{0j}$  für Gruppe  $j$  lautet:  $\theta_{j1} = \theta_{j2} = \dots = \theta_{jk_j} = \theta_{j\bullet}$ .
- Es lassen sich verschiedene Q-Tests durchführen, um die „between“- und die „within“-Variation zu testen. Ist der kat. Prädiktor bedeutsam, dann erwarten wir die Zurückweisung von  $H_0 : \theta_{1\bullet} = \theta_{2\bullet} = \dots = \theta_{p\bullet}$  („between-studies variation“), nicht aber von  $H_{0j} : \theta_{j1} = \theta_{j2} = \dots = \theta_{jk_j} = \theta_{j\bullet}$  („within variation“).

(Quelle: Notation vertauscht,  $i$  = Studiensubskript Becker, 2010)

# ANOVA-ähnliche Verfahren

## Variationszerlegung

- Die „Variationszerlegung“ in  $Q_{within}$ ,  $Q_{between}$  und  $Q_{total}$  folgt der aus der Varianzanalyse bekannten Zerlegung in „sum of squares within/resid“ (SSW), „sum of squares between/model“ (SSB) and „sum of squares total“ (SST = SSW+SSB). Allerdings werden hier gewichtete Quadratsummen verwendet.
- $Q_{total}$ : Gesamtvariation; entspricht dem bekannten  $Q$  für alle  $k$  Studien (Nullmodell; ohne Prädiktoren) ( $df = k - 1$ ) und es gilt:  $Q_{total} = Q_{between} + Q_{within}$ .
- $Q_{within/resid}$ : Nicht durch das Modell erklärte Variation ( $df = k - p$ ;  $p$  = Anzahl der zu schätzenden Parameter)
- $Q_{between/model}$ : Durch das Modell erklärte Variation ( $df = p - 1$ )



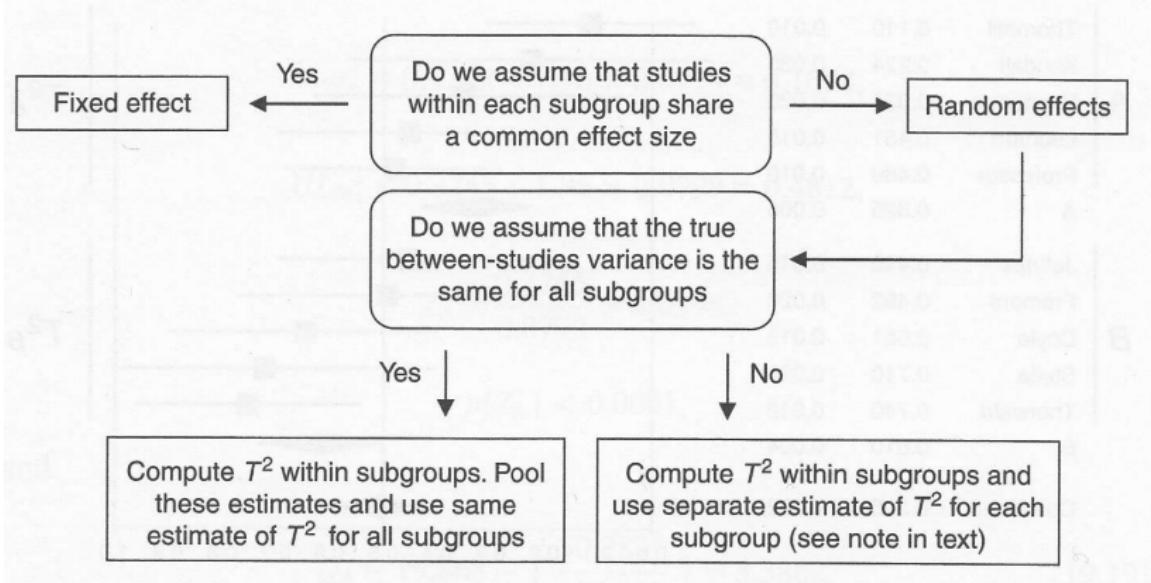
# ANOVA-ähnliche Verfahren

## Modellwahl: FEM oder REM

- Wird das FEM angenommen, dann gilt, dass allen Studien in Subgruppe A bzw. B. jeweils eine eigene *gemeinsame* ES zugrundeliegt.
- Gibt es auch nach Einbeziehung des kategorialen Prädiktors eine stat. signifikante „within“-Variation, dann ist das REM angemessen. Wie aber wird die Zwischenstudienvarianz geschätzt?
  - Pro Subgruppe ein eigener Schätzer von  $\hat{\tau}^2$  (wird im Bsp. auf Folie 208 angenommen)
  - Ein über alle Subgruppen „gepoolter“ Schätzer von  $\hat{\tau}^2$



# Flowchart zur Wahl FEM oder REM



(Quelle: Borenstein et al., 2009, S. 163)



# Beispiel für meta-analytische Subgruppenanalyse

## Datensatz

	y	V	g	SE
1	0.110	0.010	A	0.1000
2	0.224	0.030	A	0.1732
3	0.338	0.020	A	0.1414
4	0.451	0.015	A	0.1225
5	0.480	0.010	A	0.1000
6	0.440	0.015	B	0.1225
7	0.492	0.020	B	0.1414
8	0.651	0.015	B	0.1225
9	0.710	0.025	B	0.1581
10	0.740	0.012	B	0.1095

y = Hedges' g, V=Varianz, g=Gruppe, SE=Standardfehler  
(Quelle: Borenstein et al., 2009, S. 152)



# Beispiel für meta-analytische Subgruppenanalyse

		95%-CI	%W(fixed)	%W(random)
1	0.110	[-0.0860; 0.3060]	15.23	11.61
2	0.224	[-0.1155; 0.5635]	5.08	7.74
3	0.338	[ 0.0608; 0.6152]	7.61	9.29
4	0.451	[ 0.2110; 0.6910]	10.15	10.32
5	0.480	[ 0.2840; 0.6760]	15.23	11.61
6	0.440	[ 0.2000; 0.6800]	10.15	10.32
7	0.492	[ 0.2148; 0.7692]	7.61	9.29
8	0.651	[ 0.4110; 0.8910]	10.15	10.32
9	0.710	[ 0.4001; 1.0199]	6.09	8.44
10	0.740	[ 0.5253; 0.9547]	12.69	11.06

Number of studies combined: k=10

95%-CI z p.value  
Fixed effect model 0.4581 [0.3816; 0.5346] 11.740 < 0.0001  
Random effects model 0.4638 [0.3305; 0.5972] 6.816 < 0.0001

Quantifying heterogeneity:

tau^2 = 0.0299; H = 1.71 [1.23; 2.4]; I^2 = 66% [33.4%; 82.6%]

Test of heterogeneity:

Q d.f. p.value  
26.44 9 0.0017

Results for subgroups (fixed effect model):

	k	95%-CI	Q	tau^2	I^2
g = A	5	0.3241	[0.2193; 0.4289]	8.43	0.0164 52.6%
g = B	5	0.6111	[0.4992; 0.7230]	4.54	0.0022 12%

Test for subgroup differences (fixed effect model):

	Q	d.f.	p.value
Between groups	13.46	1	0.0002
Within groups	12.97	8	0.1129

Results for subgroups (random effects model):

	k	95%-CI	Q	tau^2	I^2
g = A	5	0.3245	[0.1678; 0.4812]	8.43	0.0164 52.6%
g = B	5	0.6100	[0.4903; 0.7298]	4.54	0.0022 12%

Test for subgroup differences (random effects model):

	Q	d.f.	p.value
Between groups	8.05	1	0.0045

Details on meta-analytical method:

- Inverse variance method
- DerSimonian-Laird estimator for tau^2

# Publikationsbeispiel für meta-analytische Subgruppenanalyse

Categories	k	Weighted Mean (SE)	$Q_{\text{Within}}$
Data source, $Q_{\text{Between}}(1) = 14.59^*$			
Effects from studies not using EEO data	13	.04 (.026) <sup>a</sup>	9.20
Effects from studies using EEO data	4	.17 (.054) <sup>b</sup>	18.60*
Measure of verbal ability, $Q_{\text{Between}}(4) = 16.57^*$			
College admissions tests	3	.08 (.046) <sup>a</sup>	2.78
Researcher-developed tests	2	-.03 (.070) <sup>a</sup>	0.17
Verbal fluency test	5	.17 (.022) <sup>b</sup>	19.03*
GPA in college English	2	.11 (.118) <sup>a</sup>	2.15
Other standardized tests	5	-.05 (.038) <sup>a</sup>	1.68
Type of school outcome, $Q_{\text{Between}}(2) = 11.46^*$			
Students' academic achievement	5	.13 (.061) <sup>b</sup>	22.91*
Principal/researcher ratings	8	.06 (.038) <sup>a</sup>	7.25
Teacher performance	4	.03 (.038) <sup>a</sup>	0.78
School level, $Q_{\text{Between}}(3) = 7.03$			
Elementary/middle school	6	.12 (.027) <sup>b</sup>	13.41*
High school	5	.14 (.054) <sup>b</sup>	19.51*
Elementary, middle, and high school	2	.11 (.052) <sup>a</sup>	1.66
No information	4	.03 (.038) <sup>a</sup>	0.78
Publication status, $Q_{\text{Between}}(1) = 10.68^*$			
Published	12	.13 (.028) <sup>b</sup>	30.59
Unpublished	5	-.01 (.041) <sup>a</sup>	1.12

Note. The degrees of freedom for  $Q_{\text{Within}}$  are  $k - 1$ . EEO = Equality of Educational Opportunity study; k = number of effects.

<sup>a</sup>Fixed-effects model.

<sup>b</sup>Random- or mixed-effect models.

\* $p < .05$ .

(Quelle: Aloe & Becker, 2009, S. 618)

# Abschnittsübersicht

10. Befundsynthese von Aggregatdaten

11. Begriff der Heterogenität und Diagnostik

## 12. Heterogenitätsaufklärung

Meta-Regression

ANOVA-ähnliche Verfahren

Weitere Bemerkungen zu Verfahren der Heterogenitätsaufklärung

13. Publication Bias

14. Reporting Standards

15. Ein Ausblick auf Meta-Analysen mit Individualdaten

16. Literatur

# Ökologischer Fehlschluss

„Caution is required, however, because aggregation bias has the potential to cause spurious results when covariates relating to average patient characteristics, such as age are used“  
(Sutton et al. 2000: 105).



# Beispiel für ökologischen Fehlschluss

Women		Men		Combined			
		I	N	I	N		
Trt	30	15	45	Trt	7	7	14
Plc	13	26	39	Plc	3	9	12
	$RR = 2$		$RR = 2$		$RR = 2$		
Study 1							

# Kapitelübersicht

## 10. Befundsynthese von Aggregatdaten

Fixed-effects model (FEM)

Random-effects model (REM)

## 11. Begriff der Heterogenität und Diagnostik

Überblick

Die Zwischenstudienvarianz

Die Q- Statistik

Weitere Heterogenitätsmaße:  $H^2$  und  $I^2$

Zusammenfassung Heterogenitätsmaße

## 12. Heterogenitätsaufklärung

Meta-Regression

ANOVA-ähnliche Verfahren

Weitere Bemerkungen zu Verfahren der Heterogenitätsaufklärung

## 13. Publication Bias

## 14. Reporting Standards

## 15. Ein Ausblick auf Meta-Analysen mit Individualdaten

## 16. Literatur

# Publication Bias als ein *Missing-data-Problem*<sup>7</sup>

- Selbst eine perfekte Literaturrecherche kann nur tatsächlich publizierte Studien finden.
- Nichtpublizierte Studien sind dann problematisch, wenn die Gründe für die Nichtpublikation mit Eigenheiten der Studien zusammenhängen.
- Unter einem *Publication Bias* wird „the tendency toward preparation, submission and publication of research findings based on the nature and direction of the research results“ (Dickersin, Rothstein, Sutton & Borenstein, 2005, S. 13) verstanden.

---

<sup>7</sup>Große Teile dieses Abschnitts basieren auf Weiß und Wagner (2011).



# Formen des *Dissemination Bias*

Ein umfassender Begriff, der verschiedene „information suppression mechanisms“ einschließt, ist *Dissemination Bias*:

- Publication Bias (nur signifikante, hypothesen-konforme Ergebnisse werden berichtet)
- Positive Results Bias
- Outcome Bias (selektives Berichten von mehreren möglichen Befunden)
- Time-lag-bias, Pipeline Effect
- Language Bias
- Availability Bias
- Cost Bias
- Familiarity Bias
- ...



# Ursachen für Publication Bias

- File-drawer-Problematik: Forscher neigen dazu, nicht signifikante oder nicht hypothesen-konforme Ergebnisse zurückzuhalten.
- Analysen und Modelle werden „zurechtgebogen“ (*model tweaking*), um die gewünschten Befunde zu erhalten.
- Herausgeber wissenschaftlicher Zeitschriften bevorzugen Studien mit signifikanten Ergebnissen.
- Dissertationen, Konferenzbeiträge u.ä. werden z.T. nicht veröffentlicht.

(Für eine theoretische Erörterung möglicher Motive siehe auch Auspurg & Hinz, 2011)



# Publication Bias ... ein wahres Beispiel

„But reviewers in general punish honesty and reward rhetoric. In my own area (psycholinguistics), many reviewers (many of them senior researchers who've been around for decades) do not even understand basic things in hypothesis testing. For example, **an editor-in-chief of a major journal recently rejected a paper of mine on the grounds that the p values were not small enough below the upper bound of 0.05 (in his words, ,the effects were not strong enough')**. If you ask a random researcher what a confidence interval is, 95% will give you the wrong answer; many don't even know what a p value tells you (many think it's the probability of the null being false). I could go on and on.“ (Shravan Vasishth in  
[http://www.stat.columbia.edu/~cook/movabletype/archives/2007/04/some\\_thoughts\\_o\\_3.html](http://www.stat.columbia.edu/~cook/movabletype/archives/2007/04/some_thoughts_o_3.html))



# Verfahren zur Identifikation von Publication Bias

- Funnel Plot
- Trim and Fill
- ~~Fail-safe N~~ (nicht empfohlen!)
- Regressions- und Korrelationsverfahren
- Selection Models
- Caliper Test

(Quelle: Für eine ausführliche Diskussion der Vor- und Nachteile siehe Weiß & Wagner, 2011; Auspurg & Hinz, 2011).



# Identifizierung von PB: Funnel Plot I

Annahme: Es besteht ein Effekt, der klein, aber nicht Null ist.

- Untersuchungen mit kleinen Stichproben weisen eine stärkere zufallsbedingte Streuung der Effekte auf; nur die relativ starken Effekte sind statistisch signifikant.
- Untersuchungen mit großen Stichproben hingegen weisen eher geringe Effektstärken auf, die kaum um den wirklichen Wert streuen. Diese relativ kleinen Effekte sind aufgrund der großen Stichprobe (meist) signifikant.
  - Publikationsverzerrungen liegen vor, wenn kaum Studien mit kleinen Stichproben (großer Standardfehler) und geringen Effekten publiziert werden und ein „Loch“ in der Graphik zu erkennen ist.

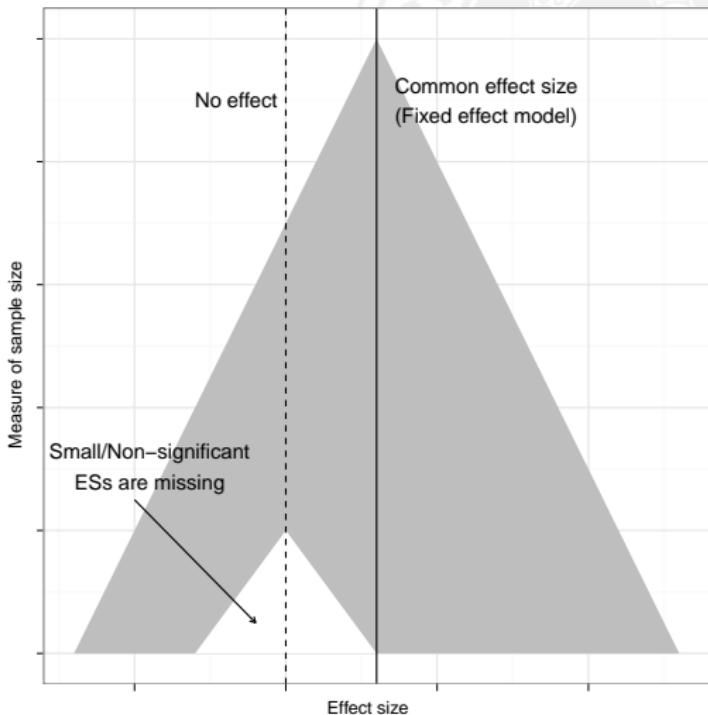


## Identifizierung von PB: Funnel Plot II

Ein Funnel Plot wird allgemein nach folgenden Prinzipien konstruiert:

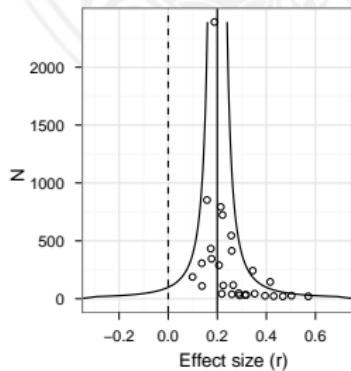
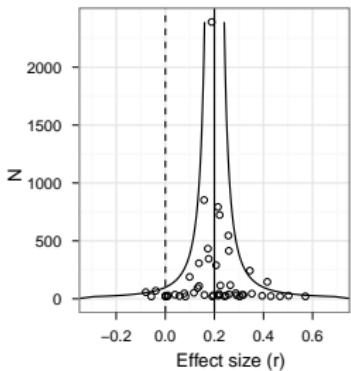
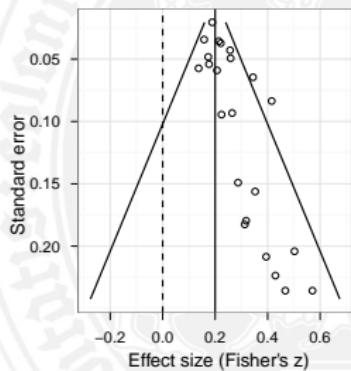
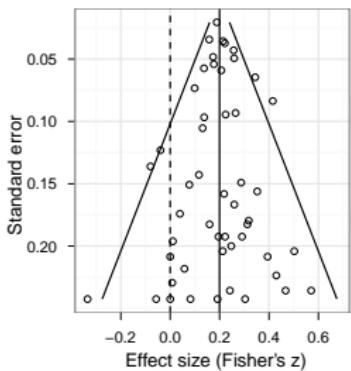
- Auf der X-Achse werden die Effektstärken abgetragen (möglichst mit symmetrischer Skala, also  $\log(\text{OR})$ ,  $\log(\text{RR})$  etc.).
- Auf der Y-Achse findet sich ein Maß für die Stichprobengröße; meistens wird hier der Standardfehler (statt der Fallzahl) der Effektstärken verwendet (nach oben hin abnehmend!).
- Eine vertikale Linie veranschaulicht den Schätzer des FEM.
- Als visuelle Hilfe kann um den Schätzer des FEM pro Einheit des Standardfehlers das *erwartete* Konfidenzintervall gezeichnet werden. Punkte (95% innerhalb) außerhalb des entstandenen Trichters weisen auf Heterogenität hin.

# Schematischer Funnel Plot



(Quelle: Weiß & Wagner, 2011, S. 666)

# Funnel Plot (simulierte Daten)



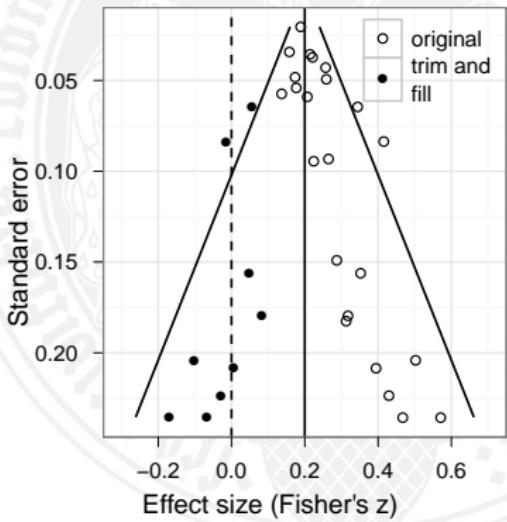
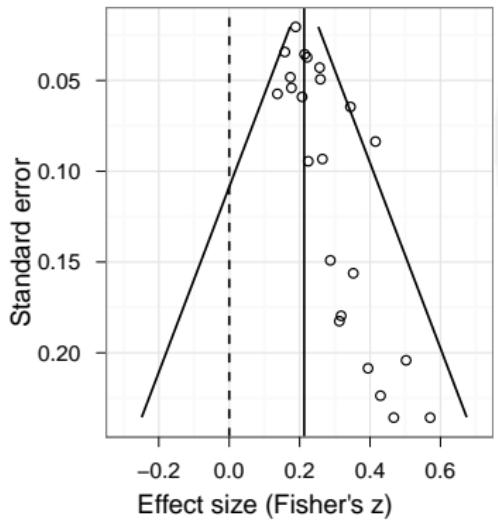
(Quelle: n.s.:  $p > 0.10$  Weiß & Wagner, 2011, S. 666)

# Trim and Fill

- Imputationsansatz: Neben den beobachteten  $k$  Studien existieren noch  $k_0$  Studien, die aufgrund des *Publication Bias* nicht ermittelt werden konnten.
- Aufgabe des Algorithmus ist, die Anzahl und die Ausprägungen dieser  $k_0$  Studien zu schätzen.
- Ein Funnelplot mit den beobachteten und den geschätzten Effektstärken vermittelt einen guten Eindruck einer symmetrischen Effektstärkenverteilung.
- Gleichwohl belegt die Existenz von  $k_0$  Studien nicht einen *Publication Bias*, es wird erst einmal nur auf eine Asymmetrie der Effektstärkenverteilung in Abhängigkeit von deren Reliabilität hingewiesen.



# Beispiel für Trim and Fill



(Quelle: Weiß & Wagner, 2011, S. 671).



# Korrelationsverfahren

Berechnung der Rangkorrelation (Kendall's Rang-Korrelation) zwischen (z-standardisierten) Effektstärken und Fehlervarianzen. Ein nicht-signifikanter Test deutet auf eine *unverzerrte* Stichprobe hin.



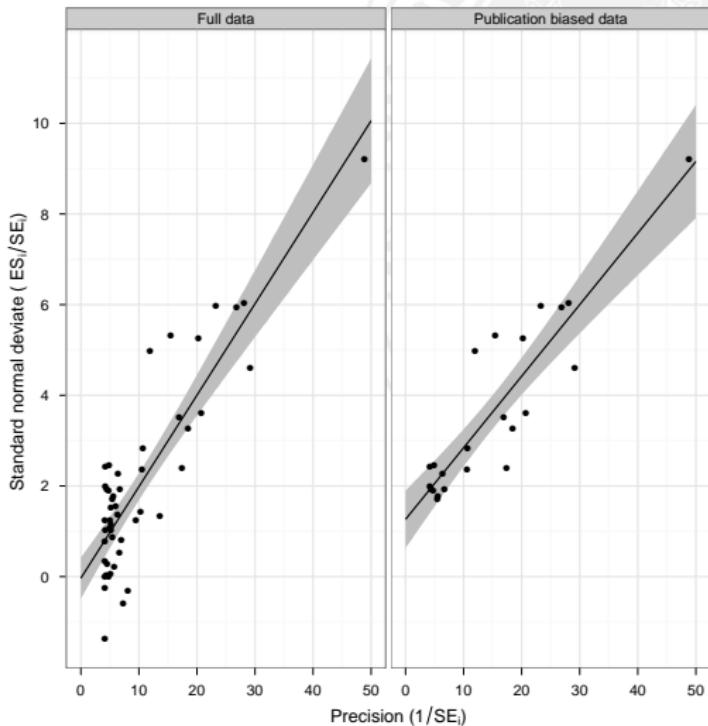
# Eggers lineares Regressionsmodell

- Es wird ein bivariate lineares Regressionsmodell geschätzt.
- Als abhängige Variable wird die am Standardfehler normierte Effektstärke ( $\frac{T_j}{SE_j}$ ) benutzt.
- Das unabhängige Merkmal ist der inverse Standardfehler ( $\frac{1}{SE_j}$ ) (*precision*).
- Liegt *keine* Verzerrung vor, sollte die Regressionsgerade durch den Ursprung laufen ( $H_0 : \beta_0 = 0$ ). Warum?:
  - $\frac{1}{SE_j}$  nähert sich für kleine Studien 0 (x-Achse).
  - $\frac{T_j}{SE_j}$  nähert sich für kleine Studien auch 0 (y-Achse).

(Quellen: Egger, Davey Smith, Schneider & Minder, 1997;  
J. Sterne, Egger, Rothstein, Sutton & Borenstein, 2005)



# Beispiel für „Egger's linear regression method“



(Quelle: basiert auf simulierten Daten, siehe Folie 222; Weiß & Wagner, 2011, S. 669).

# Kapitelübersicht

## 10. Befundsynthese von Aggregatdaten

Fixed-effects model (FEM)

Random-effects model (REM)

## 11. Begriff der Heterogenität und Diagnostik

Überblick

Die Zwischenstudienvarianz

Die Q- Statistik

Weitere Heterogenitätsmaße:  $H^2$  und  $I^2$

Zusammenfassung Heterogenitätsmaße

## 12. Heterogenitätsaufklärung

Meta-Regression

ANOVA-ähnliche Verfahren

Weitere Bemerkungen zu Verfahren der Heterogenitätsaufklärung

## 13. Publication Bias

## 14. Reporting Standards

## 15. Ein Ausblick auf Meta-Analysen mit Individualdaten

## 16. Literatur

# Übersicht Reporting Standards I

- APA Publications and Communications Board Working Group on Journal Article Reporting Standards. (2008). Reporting standards for research in psychology: Why do we need them? What might they be? *American Psychologist*, 63(9), 839–851.  
doi:10.1037/0003-066X.63.9.839
- Stroup, D. F., Berlin, J. A., Morton, S. C., Olkin, I., Williamson, D. G., Rennie, D., ... & Thackeray, S. B. (2000). Meta-analysis of observational studies in epidemiology. A proposal for reporting. *Journal of the American Medical Association*, 283(15), 2008–2012
- Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., ... & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Medicine*, 6(7). doi:10.1371/journal.pmed.1000100



# Übersicht Reporting Standards II

- Riley, R. D., Lambert, P. C. & Abo-Zaid, G. (2010). Meta-analysis of individual participant data: rationale, conduct, and reporting. *BMJ*, 340, c221–c221. doi:10.1136/bmj.c221
- MEAR Network. (2012). Maer network reporting standards. Zugriff am 29. November 2012, unter <http://www.hendrix.edu/MAER-Network/post.aspx?id=62556&blogid=51160>

(Alle Standards finden sich als PDF-Dateien auch im Verzeichnis ref/)



# MAER-Net Reporting Standards I

Quelle: <http://www.hendrix.edu/MAER-Network/post.aspx?id=62556&blogid=51160>

MAER-Net (Meta-Analysis of Economics Research (MAER) Network) recommends that all meta-analyses in economics should comply with the following reporting protocols.

- Research Questions and Effect Size
  - A clear statement of the specific economic theories, hypotheses, or effects studied.
  - A precise definition of how effects are measured (the ‘effect size’), accompanied by any relevant formulas.
  - An explicit description about how measured effects are comparable, including any methods used to standardize or convert them to a common metric.
- Research Literature, Searching, Compilation and Coding



# MAER-Net Reporting Standards II

- A full report of how the research literature was searched. This report should include:
  - the exact databases or other sources used;
  - the precise combination of keywords employed; and
  - the date that the search was completed.
- A full disclosure of the rules for study (or effect size) inclusion/exclusion. It is also useful to provide a list of all studies included and a description of why others were excluded.
- A statement addressing who searched, read, and coded the research literature. Two or more reviewers should code the relevant research.
- A complete list of the information coded for each study or estimate. At a minimum, we recommend that reviewers code:
  - the estimated effect size;



# MAER-Net Reporting Standards III

- its standard error, when feasible, and the degrees of freedom (or sample size);
- variables that distinguish which type of econometric model, methods and techniques were employed;
- dummy (i.e., 0/1) variables for the omission of theoretically relevant variables in the research study investigated;
- empirical setting (e.g., region, market, industry);
- data types (panel, cross-sectional, time series, . . . );
- year of the data used and/or publication year;
- type of publication (journal, working paper, book chapter, etc.); and
- the primary study, publication and/or dataset from which an observation is drawn.
- MRA Modeling Issues



# MAER-Net Reporting Standards IV

- A table of descriptive statistics of the variables that are coded (means and standard deviations) and graph(s) displaying the effect sizes (e.g., funnel graphs, forest plots, bar charts).
- A fully reported multiple MRA, along with the exact strategy used to simplify it (e.g., general-to-specific, Bayesian).
- An investigation of publication, selection, and misspecification biases. When suspected, these should be controlled for in subsequent MRA models.
- Methods to accommodate heteroscedasticity and within-study dependence.
- Results from MRA model specification tests, robustness checks, or sensitivity analyses.



# MAER-Net Reporting Standards V

With one possible exception, MAER-Net has come to a clear consensus about these reporting guidelines. The requirement to have two reviewers code all the relevant research has received the most comment and discussion. As economists, we all are acutely aware of the trade-off between the improved quality that the second coder will likely add (through catching mistakes and resolving ambiguities) and the increased cost (in weeks of highly skilled professional labour). We understand that the highest standards of scientific rigor demand at least two highly-knowledgeable researchers code the relevant research base. Nonetheless, MAER-Net does not wish to prohibit Ph.D. students and researchers at resource-challenged institutions from employing this important tool to understand their areas of research. To finesse these opposing concerns, the above



# MAER-Net Reporting Standards VI

statement is sufficiently broad to encompass a second reviewer randomly checking a substantial proportion of the research literature if their coding protocol is stated explicitly and justified.



# Kapitelübersicht

## 10. Befundsynthese von Aggregatdaten

Fixed-effects model (FEM)

Random-effects model (REM)

## 11. Begriff der Heterogenität und Diagnostik

Überblick

Die Zwischenstudienvarianz

Die Q- Statistik

Weitere Heterogenitätsmaße:  $H^2$  und  $I^2$

Zusammenfassung Heterogenitätsmaße

## 12. Heterogenitätsaufklärung

Meta-Regression

ANOVA-ähnliche Verfahren

Weitere Bemerkungen zu Verfahren der Heterogenitätsaufklärung

## 13. Publication Bias

## 14. Reporting Standards

## 15. Ein Ausblick auf Meta-Analysen mit Individualdaten

## 16. Literatur

# Was ist eine IPD Meta-Analyse?

*„As with any meta-analysis, an individual participant data meta-analysis aims to summarise the evidence on a particular clinical question from multiple related studies, such as whether a treatment is effective. The statistical implementation of an individual participant data meta-analysis crucially **must preserve the clustering of patients within studies**; it is inappropriate to simply analyse individual participant data as if they all came from a single study. Clusters can be retained during analysis by using a two step or a one step approach.“*

(Quelle: Riley, Lambert & Abo-Zaid, 2010, S. 1)



# APD und IPD Meta-Analysen

- Meta-Analysen auf der Grundlage von *Aggregate Person Data* (APD)
- Meta-Analysen auf der Grundlage von *Individual Person/Participant/Patient Data*
  - Einstufige Ansätze: (Mehrebenen-)Regressionsmodelle auf der Grundlage eines gepoolten Datensatzes
  - Zweistufige Ansätze: pro Datensatz eine Effekstärke berechnen und anschließend eine APD Meta-Analyse durchführen
- Meta-Analysen, die sowohl APD und IPD verwenden.



# Vorteile von IPD Meta-Analysen

- Erhöhte Auswahl angemessener Analyseverfahren, etwa für die Analyse von nicht-experimentellen Daten (multivariate Auswertungstechniken).
- Zentrale Merkmale lassen sich vereinheitlichen, erhöht die Vergleichbarkeit der Konstrukte.
- Bei zweistufigem Vorgehen keine Probleme mit fehlenden ES-Angaben.
- Es lassen sich zusätzliche Hypothesen untersuchen, insbesondere solche, die mit individuellen Charakteristika zusammenhängen.
- U.U. größere statistische *power*, um Interaktionen zwischen Studienmerkmalen und ES zu finden.
- Schließlich besteht bei der Verwendung von Aggregatdaten immer die Gefahr eines ökologischen Fehlschlusses.

(Quelle: Weiß & Wagner, 2008; Pigott, 2012, S. 252)



# Nachteile von IPD Meta-Analysen

- Hoher Kosten- und Zeitaufwand (vor allem Medizin, teilw. Kooperationen von mehreren dutzend Partnern).
- Nicht immer sind *alle* vorhandenen Datensätze auch verfügbar.
- Große Datensätze und anspruchsvolle Analyseverfahren können Desktop-PCs überfordern.



# Eigene Erfahrungen mit einer IPD Meta-Analysen

- Forschungsprojekt: „How employment status affects partnership stability: A meta-analysis using longitudinal individual person data“
- 16 Längsschnittdatensätze aufbereiten
- Gepoolter Datensatz wird (im „person-period format“) mehrere Millionen Zeilen umfassen
- (Sehr/Zu?) Komplexe Dokumentation zur Datenaufbereitung (diskrete EHA)
- Offene Probleme:
  - Umgang mit Gewichten
  - Sehr aufwändige Datenaufbereitung (techn. Infrastruktur, Organisation der Skripte, Dokumentation, . . . )
- ...



# Literatur zu IPD Meta-Analysen I

- Stewart, L. A. & Tierney, J. F. (2002). To IPD or not to IPD? Advantages and disadvantages of systematic reviews using individual patient data. *Evaluation and The Health Professions*, 25(1), 76–97. Zugriff am unter [http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list\\_uids=11868447](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11868447)
- Simmonds, M. C., Higgins, J. P., Stewart, L. A., Tierney, J. F., Clarke, M. J. & Thompson, S. G. (2005). Meta-analysis of individual patient data from randomized trials: a review of methods used in practice. *Clinical Trials*, 2(3), 209–217. Zugriff am unter [http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list\\_uids=16279144](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16279144)
- Riley, R. D., Lambert, P. C., Staessen, J. A., Wang, J., Gueyffier, F., Thijs, L. & Boutitie, F. (2007). Meta-analysis of continuous outcomes combining individual patient data and aggregate data. *Statistics in Medicine*, 27(11), 1870–1893. doi:10.1002/sim.3165



# Literatur zu IPD Meta-Analysen II

- Sutton, A. J., Kendrick, D. & Coupland, C. A. (2008). Meta-analysis of individual- and aggregate-level data. *Statistics in Medicine*, 27(5), 651–669<sup>8</sup>
- Cooper, H. & Patall, E. A. (2009). The relative benefits of meta-analysis conducted with individual participant data versus aggregated data. *Psychological methods*, 14(2), 165–176. PMID: 19485627.  
doi:10.1037/a0015565
- Curran, P. J. & Hussong, A. M. (2009). Integrative data analysis: the simultaneous analysis of multiple data sets. *Psychological methods*, 14(2), 81–100. PMID: 19485623. doi:10.1037/a0015914
- Riley, R. D., Lambert, P. C. & Abo-Zaid, G. (2010). Meta-analysis of individual participant data: rationale, conduct, and reporting. *BMJ*, 340, c221–c221. doi:10.1136/bmj.c221
- Pigott, T. D. (2012). *Advances in meta-analysis*. New York: Springer (insbes. Kapitel 8)

---

<sup>8</sup>Mit einem Bayesianischen Modell Kombination von IPD und APD Meta-Analyse.



# Kapitelübersicht

## 10. Befundsynthese von Aggregatdaten

Fixed-effects model (FEM)

Random-effects model (REM)

## 11. Begriff der Heterogenität und Diagnostik

Überblick

Die Zwischenstudienvarianz

Die Q- Statistik

Weitere Heterogenitätsmaße:  $H^2$  und  $I^2$

Zusammenfassung Heterogenitätsmaße

## 12. Heterogenitätsaufklärung

Meta-Regression

ANOVA-ähnliche Verfahren

Weitere Bemerkungen zu Verfahren der Heterogenitätsaufklärung

## 13. Publication Bias

## 14. Reporting Standards

## 15. Ein Ausblick auf Meta-Analysen mit Individualdaten

## 16. Literatur

# Literaturempfehlungen und weitere Informationsquellen I

- Literatur
  - Borenstein, M., Hedges, L. V., Higgins, J. P. T. & Rothstein, H. R. (2009 Mai). *Introduction to meta-analysis*. John Wiley und Sons
  - Lipsey, M. W. & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, California: Sage Publications
  - Cooper, H. M. (2010). *Research synthesis and meta-analysis. A step-by-step approach*. Thousand Oaks, California: Sage Publications
  - Petticrew, M. & Roberts, H. (2006). *Systematic reviews in the social sciences. A practical guide*. Malden: Blackwell



# Literaturempfehlungen und weitere Informationsquellen II

- Cooper, H. M., Hedges, L. V. & Valentine, J. C. (2009). *The handbook of research synthesis and meta-analysis*. New York: Russel Sage Foundation
- Zeitschriften
  - Research Synthesis Methods <<http://onlinelibrary.wiley.com/journal/10.1002/%28ISSN%291759-2887>>
  - Statistics in Medicine <<http://onlinelibrary.wiley.com/journal/10.1002/%28ISSN%291097-0258>>
  - Systematic Reviews  
<<http://www.systematicreviewsjournal.com/>>
- Organisationen
  - The Cochrane Collaboration <<http://www.cochrane.org/>>
  - The Campbell Collaboration  
<<http://www.campbellcollaboration.org/>>



# Literatur I

-  Aloe, A. M. & Becker, B. J. (2011a). Advances in combining regression results in meta-analysis. In M. Williams & W. P. Vogt (Hrsg.), *The sage handbook of innovation in social research methods* (S. 331–352).
-  Aloe, A. M. & Becker, B. J. (2009 November). Teacher verbal ability and school outcomes. *Educational Researcher*, 38(8), 612–624. doi:10.3102/0013189X09353939
-  Aloe, A. M. & Becker, B. J. (2011b Juni). An effect size for regression predictors in meta-analysis. *Journal of Educational and Behavioral Statistics*. doi:10.3102/1076998610396901
-  APA Publications and Communications Board Working Group on Journal Article Reporting Standards. (2008). Reporting standards for research in psychology: Why do we need them? What might they be? *American Psychologist*, 63(9), 839–851. doi:10.1037/0003-066X.63.9.839
-  Auspurg, K. & Hinz, T. (2011). What uels publication bias? Theoretical and empirical analyses of risk factors using the caliper-test. *Journal of Economics and Statistics (Jahrbuecher fuer Nationaloekonomie und Statistik)*, 231(5-6), 630–660. Zugriff am 18. Februar 2012, unter <http://ideas.repec.org/a/jns/jbstat/v231y2011i5-6p661-684.html>
-  Becker, B. J. (2010). EDF 6937 seminar on meta-analysis: modeling study results. Florida State University.
-  Becker, B. J. & Wu, M.-J. (2007). The synthesis of regression slopes in meta-analysis. *Statistical Science*, 22(3), 414–429.
-  Blettner, M., Sauerbrei, W., Schlehofer, B., Scheuchenpflug, T. & Friedenreich, C. (1997). Vergleich von traditionellen reviews, metaanalysen und gepoolten analysen zur bewertung von risikofaktoren. *Informatik, Biometrie und Epidemiologie in Medizin und Biologie*, 28(3), 148–166.

# Literatur II

-  Borenstein, M., Hedges, L. V., Higgins, J. P. T. & Rothstein, H. R. (2009 Mai). *Introduction to meta-analysis*. John Wiley und Sons.
-  Chaney, B. (1995). *Student outcomes and the professional preparation of eighth-grade teachers in science and mathematics. NSF/NELS:88 teacher transcript analysis*. Zugriff am 26. November 2012, unter <http://www.eric.ed.gov/ERICWebPortal/detail?accno=ED389530>
-  Cooper, H. M. (2010). *Research synthesis and meta-analysis. A step-by-step approach*. Thousand Oaks, California: Sage Publications.
-  Cooper, H. M., Hedges, L. V. & Valentine, J. C. (2009). *The handbook of research synthesis and meta-analysis*. New York: Russel Sage Foundation.
-  Cooper, H. & Patall, E. A. (2009). The relative benefits of meta-analysis conducted with individual participant data versus aggregated data. *Psychological methods*, 14(2), 165–176. PMID: 19485627.  
doi:10.1037/a0015565
-  Curran, P. J. & Hussong, A. M. (2009). Integrative data analysis: the simultaneous analysis of multiple data sets. *Psychological methods*, 14(2), 81–100. PMID: 19485623. doi:10.1037/a0015914
-  Dickersin, K., Rothstein, H., Sutton, A. J. & Borenstein, M. (2005). Publication bias: recognizing the problem, understanding its origins and scope, and preventing harm. In *Publication bias in meta-analysis : prevention, assessment and adjustments*. Chichester, England ; Hoboken, NJ: Wiley. Zugriff am unter <http://www.loc.gov/catdir/toc/ecip0516/2005021394.html>
-  Drinkmann, A. (1990). *Methodenkritische untersuchungen zur metaanalyse*. Weinheim: Deutscher Studienverlag.

# Literatur III

-  Egger, M., Davey Smith, G., Schneider, M. & Minder, C. (1997 September). Bias in meta-analysis detected by a simple, graphical test. *BMJ : British Medical Journal*, 315(7109), 629–634. PMID: 9310563 PMCID: 2127453.
-  Finney, D. J. (1995). A statistician looks at meta-analysis. *Journal of Clinical Epidemiology*, 48(1), 87–103.
-  Glass, G. V. (1976). Primary, secondary and meta-analysis of research. *Educational Researcher*, 5, 3–8.
-  Higgins, J. P. T. & Green, S. (2008). *Cochrane handbook for systematic reviews of interventions*. Chichester: John Wiley & Sons.
-  Higgins, J. P. T. & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, 21, 1539–1558.
-  Huedo-Medina, T. B., Botella, J., Sánchez-Meca, J. & Marín-Martínez, F. (2006). Assessing heterogeneity in meta-analysis: q statistic or i<sup>2</sup> index? *Psychological Methods*, 11(2), 193–206.
-  Keef, S. P. & Roberts, L. A. (2004 Mai). The meta-analysis of partial effect sizes. *The British Journal of Mathematical and Statistical Psychology*, 57(Pt 1), 97–129. PMID: 15171803. doi:10.1348/000711004849303
-  Kim, R.-S. (2011). *Standardized regression coefficients as indices of effect sizes in meta-analysis*. (Diss., Florida State University, College of Education).
-  Kim, R.-S. & Becker, B. J. (2010). The degree of dependence between multiple-treatment effect sizes. *Multivariate Behavioral Research*, 45(2), 213. doi:10.1080/00273171003680104

# Literatur IV

-  Lambert, P. C. & Abrams, K. R. (1996). Meta-analysis using multilevel models. *Multilevel Modelling Newsletter*, 7(2), 16–18.
-  Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., ... Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Medicine*, 6(7). doi:10.1371/journal.pmed.1000100
-  Lipsey, M. W. & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, California: Sage Publications.
-  Lorant, V., Deliège, D., Eaton, W., Robert, A., Philippot, P. & Ansseau, M. (2003). Socioeconomic inequalities in depression: a meta-analysis. *American Journal of Epidemiology*, 157(2), 98–112.
-  MEAR Network. (2012). Maer network reporting standards. Zugriff am 29. November 2012, unter <http://www.hendrix.edu/MAER-Network/post.aspx?id=62556&blogid=51160>
-  Petticrew, M. & Roberts, H. (2006). *Systematic reviews in the social sciences. A practical guide*. Malden: Blackwell.
-  Pigott, T. D. (2012). *Advances in meta-analysis*. New York: Springer.
-  Raudenbush, S. W., Becker, B. J. & Kalaian, H. (1988). Modeling multivariate effect sizes. *Psychological Bulletin*, 103(1), 111–120. doi:10.1037/0033-2909.103.1.111
-  Riley, R. D., Lambert, P. C. & Abo-Zaid, G. (2010). Meta-analysis of individual participant data: rationale, conduct, and reporting. *BMJ*, 340, c221–c221. doi:10.1136/bmj.c221

# Literatur V

-  Riley, R. D., Lambert, P. C., Staessen, J. A., Wang, J., Gueyffier, F., Thijs, L. & Boutitie, F. (2007). Meta-analysis of continuous outcomes combining individual patient data and aggregate data. *Statistics in Medicine*, 27(11), 1870–1893. doi:10.1002/sim.3165
-  Rosenthal, R., Rosnow, R. L. & Rubin, D. B. (2000). *Contrasts and effect sizes in behavioral research : a correlational approach*. Cambridge u.a.: Cambridge Univ. Press.
-  Simmonds, M. C., Higgins, J. P., Stewart, L. A., Tierney, J. F., Clarke, M. J. & Thompson, S. G. (2005). Meta-analysis of individual patient data from randomized trials: a review of methods used in practice. *Clinical Trials*, 2(3), 209–217. Zugriff am unter [http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list\\_uids=16279144](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16279144)
-  Stanley, T. D. & Jarrell, S. B. (1989 Juni). Meta-regression analysis: a quantitative method of literature surveys. *Journal of Economic Surveys*, 3(2), 161–170. doi:10.1111/j.1467-6419.1989.tb00064.x
-  Sterne, J. A. C. (Hrsg.). (2009). *Meta-analysis in stata: an updated collection from the stata journal*. College Station, Tex.: Stata Press.
-  Sterne, J., Egger, M., Rothstein, H., Sutton, A. J. & Borenstein, M. (2005). Regression methods to detect publication and other bias in meta-analysis. In *Publication bias in meta-analysis: prevention, assessment and adjustments* (S. 99–110). Chichester, England; Hoboken, NJ: Wiley. Zugriff am unter <http://www.loc.gov/catdir/toc/ecip0516/2005021394.html>
-  Stewart, L. A. & Tierney, J. F. (2002). To IPD or not to IPD? Advantages and disadvantages of systematic reviews using individual patient data. *Evaluation and The Health Professions*, 25(1), 76–97. Zugriff am unter [http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list\\_uids=11868447](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11868447)

# Literatur VI



Stroup, D. F., Berlin, J. A., Morton, S. C., Olkin, I., Williamson, D. G., Rennie, D., ... Thacker, S. B. (2000). Meta-analysis of observational studies in epidemiology. A proposal for reporting. *Journal of the American Medical Association*, 283(15), 2008–2012.



Sutton, A. J., Abrams, K. R., Jonas, M., Sheldon, T. A. & Song, F. (2000). *Methods for meta-analysis in medical research*. Wiley series in probability and mathematical statistics. Chichester; New York: J. Wiley. Zugriff am unter <http://www.loc.gov/catdir/toc/onix07/00042889.html>



Sutton, A. J., Kendrick, D. & Coupland, C. A. (2008). Meta-analysis of individual- and aggregate-level data. *Statistics in Medicine*, 27(5), 651–669.



Thompson, C. & Weiβ, B. (2013). An introduction to basic meta-analysis. In B. J. Becker (Hrsg.), *Synthesizing correlations and regressions: Methodology and applications*. Unpublished manuscript.



Thompson, C. G. & Becker, B. J. (2013). *The impact of multiple endpoint dependency on the statistic in meta-analysis*.



Thompson, S. G. & Higgins, J. P. T. (2002). How should meta-regression analyses be undertaken and interpreted? *Statistics in Medicine*, 21, 1559–1573.



Viechtbauer, W. (2005). Bias and efficiency of meta-analytic variance estimators in the random-effects model. *Journal of Educational and Behavioral Statistics*, 30(3), 261–293. doi:10.3102/10769986030003261



Viechtbauer, W. (2007). Accounting for heterogeneity via random-effects models and moderator analyses in meta-analysis. *Zeitschrift für Psychologie/Journal of Psychology*, 215(2), 104–121. doi:10.1027/0044-3409.215.2.104

# Literatur VII

-  Wagner, M. & Weiß, B. (2003). Bilanz der deutschen Scheidungsforschung. Versuch einer Meta-Analyse. *Zeitschrift für Soziologie*, 32(1), 29–49.
-  Wagner, M. & Weiß, B. (2006a). Meta-Analyse als Methode der Sozialforschung. *Methoden der Sozialforschung. Sonderheft 44 der Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 479–504.
-  Wagner, M. & Weiß, B. (2006b). On the Variation of Divorce Risks in Europe: Findings from a Meta-Analysis of European Longitudinal Studies. *European Sociological Review*, 22(5), 483–500.
-  Weiß, B. (2008). *Meta-Analyse als Forschungssynthese in der Soziologie. Dargestellt anhand zweier Fallbeispiele zum Schulabsentismus. Inauguraldissertation zur Erlangung des Doktorgrades der Wirtschafts- und Sozialwissenschaftlichen Fakultät der Universität zu Köln.*
-  Weiß, B. & Wagner, M. (2008). Potentiale und Probleme von Meta-Analysen in der Soziologie. *Sozialer Fortschritt*, 57(10-11), 250–255.
-  Weiß, B. & Wagner, M. (2011). The identification and prevention of publication bias in the social sciences and economics. *Journal of Economics and Statistics*, 231(5-6), 661–684.
-  White, H. D., Cooper, H. M. & Hedges, L. V. (1994). Scientific communication and literature retrieval. In *The handbook of research synthesis* (S. 41–55). New York: Russel Sage Foundation.

# Teil V

## Anhang



# Anhang

17. Meta-Analyse-Software für R und Stata

18. R in 20 Minuten

19. Statistische Grundlagen



# Kapitelübersicht

17. Meta-Analyse-Software für R und Stata

18. R in 20 Minuten

19. Statistische Grundlagen

Der Standardfehler

Wahrscheinlichkeitsverteilungen und Q-Test



# Meta-Analyse-Software für R

R Pakete (sozialwissenschaftliche Auswahl):

- **meta** (uni- und bivariate Meta-Analyse; mächtiger als **rmeta**)
- **rmeta** (univariate Meta-Analyse)
- **metafor** (Meta-Regression; sehr umfangreich, siehe auch <http://www.metafor-project.org/>)
- **mvmeta** (Multivariate Meta-Analyse)
- **metacor**
- **MAc**, **MAclinical**, **MAd** (für eine Einführung siehe [http://rwiki.sciviews.org/doku.php?id=packages:cran:ma\\_meta-analysis](http://rwiki.sciviews.org/doku.php?id=packages:cran:ma_meta-analysis))
- **compute.es** (zur Berechnung und Konvertierung von Effektstärken)



# Meta-Analyse-Software Stata

- Eine Liste von Stata add-ons ist unter <http://www.stata.com/support/faqs/statistics/meta-analysis/> (What meta-analysis features are available in Stata?) verfügbar.
- J. A. C. Sterne (2009) hat „Meta-Analysis in Stata: An Updated Collection from the Stata Journal“ herausgegeben.



# Sonstige Meta-Analyse-Software

- SPSS: „meta-analysis stuff“ von David B. Wilson (<http://mason.gmu.edu/~dwilsonb/ma.html>)
- Comprehensive Meta-Analysis:  
<http://www.meta-analysis.com/>
- OpenMeta[Analyst]: [http://tuftscaes.org/open\\_meta/](http://tuftscaes.org/open_meta/) (interessant, basiert auf R und verwendet u.a. metafor)
- ...



# Kapitelübersicht

17. Meta-Analyse-Software für R und Stata

18. R in 20 Minuten

19. Statistische Grundlagen

Der Standardfehler

Wahrscheinlichkeitsverteilungen und Q-Test



# Grundlagen von R

- R ist ein Statistikprogramm *und* eine Programmiersprache <<http://www.r-project.org/>>.
- (Fast) Alles in R sind *Objekte* (etwa Datensätze oder Ergebnisse von Berechnungen).
- Sofern der Hauptspeicher ausreicht, können dort beliebig viele R Objekte *gleichzeitig* existieren (d.h. auch mehrere Datensätze).
- Das Laden von Datensätzen oder Speichern von Ergebnissen ist daher *immer* mit einer Zuweisung (*assignment*) (R-Befehl: `<-`) verbunden.
- R ist modular aufgebaut. Bestimmte (statistische) Funktionen (Meta-Analyse, Mehrebenenmodelle, Überlebensmodelle etc.) müssen erst (via `library(Paketname)`) geladen werden.
- R ist *case sensitive*, d.h. `mean(x)`  $\neq$  `Mean(x)`.
- Unter MS Windows werden in R für Dateipfade immer *slashes* (/) verwendet, *backslash* (\) ist nur als doppelter *backslash* (\\) erlaubt.
- R ist streng im Umgang mit fehlenden Werten (mit NA gekennzeichnet) und bspw. muss in vielen Funktionen ein listenweiser Ausschluss ausdrücklich angegeben werden.
- Kommentare werden mit dem Zeichen # eingeleitet.

# Zuweisungen und R Ausgabe

Zuweisungen erfolgen in R mit dem Befehl <-:

```
x <- 22  
x
```

```
R> [1] 22
```

Mit R> wird hier die R-Ausgabe bezeichnet. [1] kennzeichnet in R das i-te Elemente in der jeweiligen Zeile (hier, nur 1 Element).

```
y <- 1:15
```

```
y
```

```
R> [1] 1 2 3 4 5 6 7 8 9 10 11 12  
R> [13] 13 14 15
```



# R als Taschenrechner

```
## Mit Skalaren rechnen
x <- 123
y <- 7
x + y
```

```
R> [1] 130
```

```
## Mit Vektoren (= Variablen) rechnen
## c() erzeugt einen Vektor (c = concatenate)
x <- c(3, 6, 9)
x
```

```
R> [1] 3 6 9
```

```
y <- x/3
y
```

```
R> [1] 1 2 3
```

# Wichtige Funktionen

- Hilfeseiten aufrufen: `help(Funktionsname)` oder `?Funktionsname` (Bsp.: `help(mean)` oder `?mean`)
- Welche Objekte sind im *workspace*: `ls()`
- Arbeitsverzeichnis definieren/abfragen:  
`setwd("Pfad")` und `getwd()`
- Beschreibung eines R-Objektes: `str(Robjekt)`
- (Variablen-)Namen eines R-Objektes:  
`names(Robjekt)`
- Die ersten n Fälle anzeigen: `head(Robjekt)`



# Hilfeausgabe für ?mean

mean

package:base

R Documentation

Arithmetic Mean

Description:

Generic function for the (trimmed) arithmetic mean.

Usage:

```
mean(x, ...)

## Default S3 method:
mean(x, trim = 0, na.rm = FALSE, ...)
```

Arguments:

x: An R object. Currently there are methods for numeric/logical vectors and date, date-time and time interval objects, and for data frames all of whose columns have a method. Complex vectors are allowed for 'trim = 0', only.

trim: the fraction (0 to 0.5) of observations to be trimmed from each end of 'x' before the mean is computed. Values of trim outside that range are taken as the nearest endpoint.

na.rm: a logical value indicating whether 'NA' values should be stripped before the computation proceeds.

...: further arguments passed to or from other methods.

Value:

If 'trim' is zero (the default), the arithmetic mean of the values in 'x' is computed, as a numeric or complex vector of length one. If 'x' is not logical (coerced to numeric), numeric (including integer) or complex, 'NA\_real\_' is returned, with a warning.

If 'trim' is non-zero, a symmetrically trimmed mean is computed with a fraction of 'trim' observations deleted from each end before the mean is computed.

References:

Becker, R. A., Chambers, J. M. and Wilks, A. R. (1988) *The New S Language*. Wadsworth & Brooks/Cole.

See Also:

'weighted.mean', 'mean.POSIXct', 'colMeans' for row and column means.

Examples:

```
x <- c(0:10, 50)
xm <- mean(x)
c(xm, mean(x, trim = 0.10))
```

# Daten laden und untersuchen I

```
## Zuweisung nicht vergessen!
dat <- read.csv(file = ".../.../data/dVerbAb.csv",
                 sep = ";", dec = ",")

## head() zeigt die ersten 6 Zeilen eines
## R Objektes an
head(dat)
```

```
R>   ID year pub      r      N     Var      SE
R> 1  1 1980  0 -0.10    7 0.163350 0.40417
R> 2  2 2005  1  0.23   76 0.011960 0.10936
R> 3  4 1968  0 -0.05  155 0.006461 0.08038
R> 4  5 1968  0  0.02   45 0.022709 0.15070
R> 5  6 1969  1 -0.09   31 0.032796 0.18110
R> 6  7 1969  1  0.26   37 0.024149 0.15540
```



# Daten laden und untersuchen II

```
## Struktur anzeigen
str(dat)

R> 'data.frame': 17 obs. of 7 variables:
R> $ ID  : int  1 2 4 5 6 7 9 10 11 12 ...
R> $ year: int  1980 2005 1968 1968 1969 1969 1970 1987 1987 1993 ...
R> $ pub : int  0 1 0 0 1 1 0 1 1 1 ...
R> $ r   : num -0.1 0.23 -0.05 0.02 -0.09 0.26 -0.07 0.04 -0.01 0.12 ...
R> $ N   : int  7 76 155 45 31 37 79 151 151 64 ...
R> $ Var : num 0.16335 0.01196 0.00646 0.02271 0.0328 ...
R> $ SE  : num 0.4042 0.1094 0.0804 0.1507 0.1811 ...

## Dimensionen
dim(dat)

R> [1] 17 7

## Variablennamen ausgeben lassen
names(dat)

R> [1] "ID"      "year"    "pub"     "r"       "N"       "Var"
R> [7] "SE"
```

# Mit Datenobjekten (data.frame) arbeiten I

```
## Einzelne Elemente eines data.frame  
## ansprechen: mit $  
dat$r  
  
R> [1] -0.10  0.23 -0.05  0.02 -0.09  0.26 -0.07  
R> [8]  0.04 -0.01  0.12  0.03  0.07 -0.05  0.13  
R> [15] 0.03  0.21  0.28  
  
## Einzelne Elemente des data.frame  
## ansprechen: Indexnotation  
dat[, "r"]  
  
R> [1] -0.10  0.23 -0.05  0.02 -0.09  0.26 -0.07  
R> [8]  0.04 -0.01  0.12  0.03  0.07 -0.05  0.13  
R> [15] 0.03  0.21  0.28
```



# Mit Datenobjekten (data.frame) arbeiten II

```
## Mehrere Elemente eines data.frame  
## ansprechen: nur noch Indexnotation  
## Fälle 1:5, Variablen: r und Var  
dat[1:5, c("r", "Var")]
```

```
R>      r      Var  
R> 1 -0.10 0.163350  
R> 2  0.23 0.011960  
R> 3 -0.05 0.006461  
R> 4  0.02 0.022709  
R> 5 -0.09 0.032796
```



# Deskriptive Statistik

```
summary(dat[, c("r", "pub")])
```

```
R>          r                  pub
R> Min.   :-0.1000      Min.   :0.000
R> 1st Qu.:-0.0500      1st Qu.:0.000
R> Median : 0.0300      Median :1.000
R> Mean   : 0.0618      Mean   :0.706
R> 3rd Qu.: 0.1300      3rd Qu.:1.000
R> Max.   : 0.2800      Max.   :1.000
```

```
sd(dat$r)
```

```
R> [1] 0.1241
```

```
table(dat$pub)
```

```
R>
R> 0  1
R> 5 12
```

# Einen Teildatensatz erstellen

SPSS/Stata: if

```
subset(dat, pub==0)
```

```
R>      ID year pub      r      N       Var       SE
R> 1 1980 0 -0.10 7 0.163350 0.40417
R> 3 1968 0 -0.05 155 0.006461 0.08038
R> 4 1968 0 0.02 45 0.022709 0.15070
R> 7 1970 0 -0.07 79 0.012695 0.11267
R> 11 13 1991 0 0.03 318 0.003149 0.05612
```

```
subset(dat, pub==0 & r > 0)
```

```
R>      ID year pub      r      N       Var       SE
R> 4 1968 0 0.02 45 0.022709 0.15070
R> 11 13 1991 0 0.03 318 0.003149 0.05612
```



# Bedingtes Recodieren

```
## Kopie erstellen
```

```
tmp <- dat
```

```
tmp$pub
```

```
R> [1] 0 1 0 0 1 1 0 1 1 1 0 1 1 1 1 1 1
```

```
## pub == 0 mit 99 ersetzen
```

```
tmp$pub <- ifelse(tmp$pub == 0, 99, tmp$pub)
```

```
tmp$pub
```

```
R> [1] 99 1 99 99 1 1 99 1 1 1 99 1 1 1 1
```

```
R> [16] 1 1
```



# Umgang mit Missing

```
## Kopie erstellen
tmp <- dat
## 0 wird zu missing (NA)
tmp$pub <- ifelse(tmp$pub == 0, NA, tmp$pub)

## subset nur missings in pub
subset(tmp, is.na(pub))
```

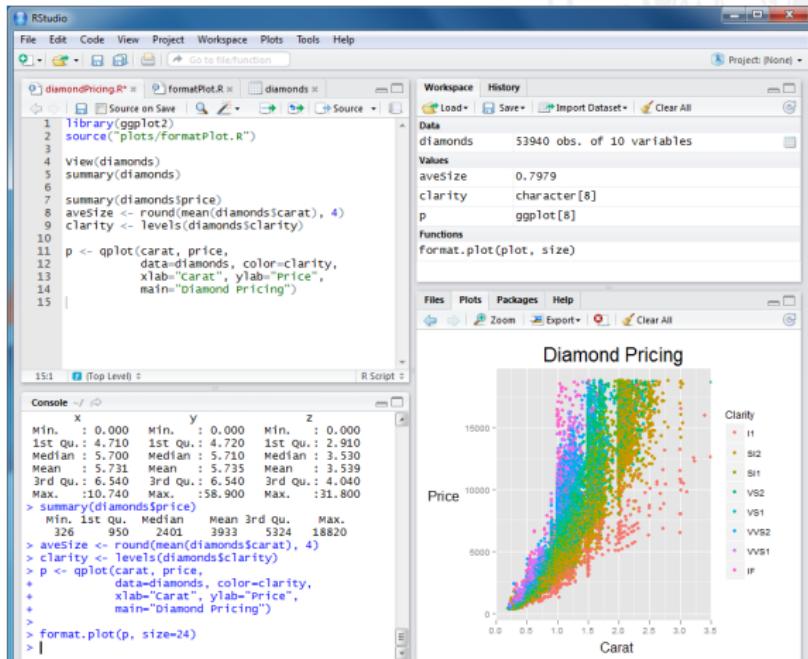
```
R>      ID year pub      r      N       Var        SE
R> 1    1 1980  NA -0.10     7 0.163350 0.40417
R> 3    4 1968  NA -0.05   155 0.006461 0.08038
R> 4    5 1968  NA  0.02    45 0.022709 0.15070
R> 7    9 1970  NA -0.07   79 0.012695 0.11267
R> 11  13 1991  NA  0.03  318 0.003149 0.05612
```

```
## is.na-Funktion
is.na(tmp$pub)
```

```
R> [1] TRUE FALSE  TRUE  TRUE FALSE FALSE FALSE TRUE
R> [8] FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE
R> [15] FALSE FALSE FALSE
```

# RStudio und R

- Unter MS-Windows gibt es zur Bedienung von R nur die spartanische R Console.
- Empfehlenswert ist RStudio, eine kostenlose integrierte Entwicklungsumgebung für R  
<http://www.rstudio.com/ide/>.



The screenshot shows the RStudio interface with the following components:

- Script Editor:** Contains R code for loading ggplot2, sourceing a file, viewing diamonds, summarizing diamonds, calculating average size, and creating a scatter plot. The plot shows Price vs. Carat with points colored by Clarity.
- Console:** Displays the output of the R code, including summary statistics for diamonds and the resulting scatter plot.
- Workspace Browser:** Shows the `diamonds` dataset (53940 obs. of 10 variables), the `avsize` variable (0.7979), the `clarity` variable (character [8]), and the `p` function (ggplot [8]).
- Plots:** A scatter plot titled "Diamond Pricing" showing Price on the Y-axis (0 to 15000) versus Carat on the X-axis (0.0 to 3.5). The points are colored according to their clarity grade, as defined in the legend: I1 (red), SI2 (orange), SI1 (yellow), VS2 (green), VS1 (light green), VVS2 (blue), VVS1 (light blue), and IF (pink).

# Kapitelübersicht

17. Meta-Analyse-Software für R und Stata

18. R in 20 Minuten

**19. Statistische Grundlagen**

Der Standardfehler

Wahrscheinlichkeitsverteilungen und Q-Test



# Abschnittsübersicht

17. Meta-Analyse-Software für R und Stata

18. R in 20 Minuten

## 19. Statistische Grundlagen

Der Standardfehler

Wahrscheinlichkeitsverteilungen und Q-Test



# Was ist Inferenzstatistik?

- Inferenzstatistik befasst sich mit dem Schluss von der Stichprobe auf die Grundgesamtheit (Population).
- Dieser Schluss ist (immer) fehlerbehaftet; neben dem Schätzen des Populationsparameters („Punktschätzung“) ist daher von zentraler Bedeutung, die Unsicherheit bei der Parameterschätzung beschreiben zu können.
- Der *Standardfehler* ist ein Streuungsmaß, das die Unsicherheit der Parameterschätzung beschreibt.
- Mit Hilfe des Standardfehlers lässt sich das Konfidenzintervall („Intervallschätzung“) berechnen sowie die Irrtumswahrscheinlichkeit bestimmen.



# Der Standardfehler für das arithmetische Mittel

- Der SE beschreibt die Güte des ermittelten Stichprobenwertes: Je größer die Fallzahl, desto kleiner der SE. Je kleiner der Standardfehler, desto besser die Schätzung.
- Der Standardfehler (SE) des Stichprobenmittelwertes lautet:

$$\frac{\text{Standardabweichung des Merkmals}}{\sqrt{\text{Fallzahl}}} = \frac{\sigma}{\sqrt{N}}$$

- Der SE gibt die Streuung der Stichproben-Mittelwerte von gleichgroßen und zufällig aus einer Grundgesamtheit gezogenen Stichproben um den wahren Populationsmittelwert  $\mu$  (sprich „mü“; allgemein:  $\theta$ , sprich: theta) an.



# Abschnittsübersicht

17. Meta-Analyse-Software für R und Stata

18. R in 20 Minuten

## 19. Statistische Grundlagen

Der Standardfehler

Wahrscheinlichkeitsverteilungen und Q-Test



# Exkurs: Die statistische Beschreibung von Wahrscheinlichkeitsverteilungen

Die folgenden Ausführungen zur statistischen Signifikanz von Q verweisen (teilweise implizit, teilweise explizit) auf die Beschreibung der Wahrscheinlichkeitsverteilung einer Zufallsvariablen. Die folgenden Begriffe sollten Sie ggf. noch einmal auffrischen:

- Diskrete Zufallsvariablen: Wahrscheinlichkeitsfunktion; stetige Zufallsvariablen: Wahrscheinlichkeitsdichtefunktion (teilw. mit „Dichte“ abgekürzt; engl.: *probability density function*)
- Mit Hilfe von Wahrscheinlichkeitsverteilungen lassen sich Ereignissen Wahrscheinlichkeiten zuordnen.
- (Kumulative) Verteilungsfunktion
- Quantilfunktion



# Idee des Q-Tests (I)

- $Q$  ist die Gesamtvariation (nicht -varianz!) (beobachtete Variation),  $df = k - 1$  die erwartete Variation im Falle des FEM (es gibt *einen gemeinsamen* Populationsparameter) und  $Q - df$  die Abweichung zwischen beobachteter und erwarteter Variation.
- Wenn  $Q \leq df$  gilt, dann wird definitionsgemäß  $Q = 0$  und es wird eine homogene Effektstärkenverteilung unterstellt.
- Wenn aber  $Q > df$  ( $Q - df > 0$ ) gilt, müssen wir dann *immer* von einer heterogenen ES-Verteilung ausgehen? Oder muss  $Q$  ( $Q - df$ ) nicht „ziemlich groß sein“, damit die Idee einer homogenen Effektstärkenverteilung zurückgewiesen werden kann? Was aber heißt „ziemlich groß“?



## Idee des Q-Tests (II)

- Der  $Q$ -Wert folgt einer sogenannten  $\chi^2$ -Verteilung (gehört zu den stetigen Wahrscheinlichkeitsverteilungen).
- Mit dem Wissen, dass  $Q$   $\chi^2$ -verteilt ist, lässt sich für einen gegebenen Freiheitsgrad ( $df$ ) bspw. die Wahrscheinlichkeit  $Pr(0 \leq Q)$  bestimmten, d.h., dass ein bestimmter  $Q$ -Wert im Intervall  $[0, Q]$  liegt. Die Wahrscheinlichkeit, dass  $Q$  außerhalb des Intervalls liegt (also größer als  $Q$  ist bzw.  $(Q; +\infty]$ ), beträgt  $1 - Pr(0 \leq Q)$ . Anders formuliert:  $1 - Pr(0 \leq Q)$  ist die Wahrscheinlichkeit, einen Wert größer als  $Q$  zu bekommen.
- „Ziemlich groß“ (siehe vorherige Folie) ist ein  $Q$ -Wert aber doch dann, wenn  $1 - Pr(0 \leq Q)$  (= p-Wert = Irrtumswahrscheinlichkeit) klein ist, wobei verschiedene „Kleinheitsgrade“ (10%, 5%, 1% = Signifikanzniveaus) unterschieden werden. Zu jedem dieser Signifikanzniveaus lässt sich ein theoretischer  $\chi^2$ -Wert berechnen und wenn der empirische  $\chi^2$ -Wert (als  $Q$ ) größer ist, dann sagen wir, dass der Wert signifikant auf dem 10%--, 5%--, 1%-Niveau ist.
- Üblicherweise wird in der Meta-Analyse-Literatur das 10%-Niveau angenommen.

# Bestimmung der Irrtumswahrscheinlichkeit von Q

Mit R lässt sich für verschiedene Wahrscheinlichkeitsverteilungen die Wahrscheinlichkeit  $Pr(X \leq x)$  ( $= F(x)$  = kumulative Verteilungsfunktion) bestimmten – und nur diese. Da im Falle des Q-Testes aber  $1 - Pr(0 \leq Q) = p$ -Wert gesucht wird, muss folglich immer  $1 - Pr(0 \leq Q)$  gerechnet werden.

```
## Q = 10 (df = 5)  
1-pchisq(q = 10, df = 5)
```

```
R> [1] 0.07524
```

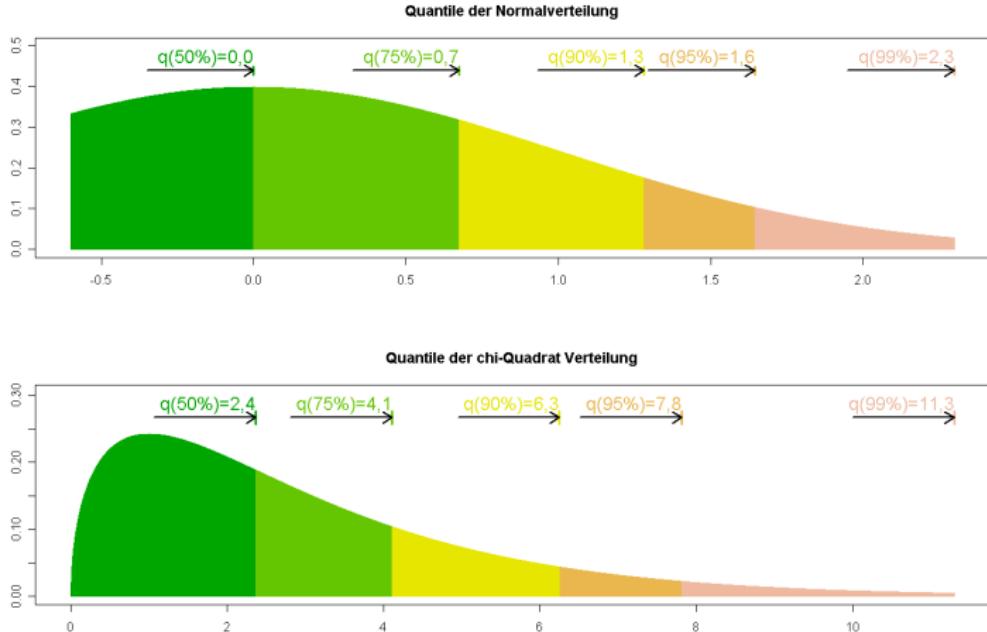
```
## Q = 36.1437 (df = 5), Borenstein Tab 14.7  
1-pchisq(q = 36.1437, df = 5)
```

```
R> [1] 8.89e-07
```

```
## Welches Q fuer alpha <= 0.1?  
qchisq(0.9, df = 5)
```

```
R> [1] 9.236
```

# Quantile der $\chi^2$ – Verteilung (df = 3)



(Quelle: <http://de.wikipedia.org/wiki/Chi-Quadrat-Verteilung>)

