



Evaluation und Metaanalyse

MSc Klinische Psychologie und Psychotherapie

SoSe 2024

Prof. Dr. Dirk Ostwald

(3) Standardisierte Effektstärkeschätzung

Motivation

Standardisierte Effektstärkeschätzung in Einzelstudien

Anwendungsbeispiel

Selbstkontrollfragen

Motivation

Standardisierte Effektstärkeschätzung in Einzelstudien

Anwendungsbeispiel

Selbstkontrollfragen

Was Eysenck right after all? A reassessment of the effects of psychotherapy for adult depression

“We used a database of randomised trials of psychotherapies for adult depression that has been described in a methods paper (Cuijpers et al. 2008a). The general methods we have used in this paper have been described in a manual that is freely available (Cuijpers, 2016b). In brief, the database was developed through a comprehensive literature search (from 1966 until 1 January 2017), and is updated every year. We searched major bibliographical databases (PsycINFO, PubMed, Embase, Cochrane Central Register of Controlled Trials). The full search string for PubMed is given in Appendix A.

In this database, we included all randomised trials in which at least one arm was a psychological treatment for adults (>18 years) with a depressive disorder according to a diagnostic interview or an elevated level of depressive symptomatology (as indicated by a score above a cut-off score on a validated self-report depression scale). In the current study, we only used trials that compared a psychotherapy for adult depression with a control group (waiting list, care-as-usual, placebo or other; this last category included control conditions that could not be categorised into one of the three categories, such as participation in online discussion forums, in workshops on other subjects, routine care in general medical care or an information booklet).

We calculated effect sizes (Hedges' g) for each comparison between a psychotherapy and a comparison group. We only included outcome measures that assess depressive symptoms. If there is more than one measure of depression, these are pooled within the study, before the overall effects are pooled across studies.”

Cuijpers et al. (2019)

Was Eysenck right after all? A reassessment of the effects of psychotherapy for adult depression

Table 1. Effects of psychotherapies compared with control groups ($k = 369$): Hedges' g^a

		k	g	95% CI
All studies		369	0.70	0.64–0.75
Extreme outliers excluded ($g > 2.0$)		352	0.61	0.57–0.66
One effect size per study (only highest)		289	0.70	0.65–0.76
One effect size per study (only lowest)		289	0.64	0.58–0.69
Only HAMD		103	0.86	0.75–0.97
Only BDI		128	0.87	0.77–0.98
Only BDI-II		80	0.68	0.57–0.80
	Subgroup analyses			
Type of therapy	CBT	192	0.70	0.63–0.77
	Behavioural activation	20	0.94	0.66–1.22
	Interpersonal psychotherapy	25	0.60	0.40–0.80
	Problem-solving therapy	27	0.77	0.54–1.01
	Third wave therapies	18	0.77	0.58–0.97
	Supportive counselling	19	0.58	0.42–0.75
	Psychodynamic therapy	11	0.40	0.17–0.63
Control group	Other	57	0.68	0.54–0.82
	Waiting list	159	0.89	0.80–0.98
Country	Care-as-usual	144	0.61	0.53–0.68
	Other	66	0.51	0.40–0.62
	Western	325	0.63	0.58–0.68
Risk of bias score	Non-Western	44	1.13	0.94–1.33
	0 (high risk)	14	1.11	0.87–1.36
	1	122	0.92	0.79–1.05
	2	63	0.73	0.60–0.86
	3	62	0.71	0.58–0.85
	4 (low risk)	108	0.46	0.41–0.52

BDI, Beck Depression Inventory; CBT, cognitive behavioural therapy; CI, confidence interval; N_{comps} , number of comparisons; NNT, numbers-needed-to-treat.

^aAccording to the random-effects model.

Cuijpers et al. (2019)

Motivation

Is psychotherapy effective? A re-analysis of treatments for depression

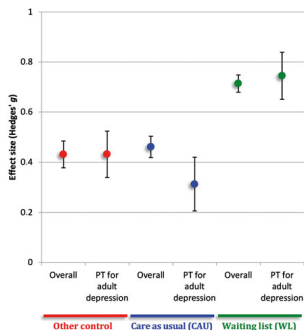
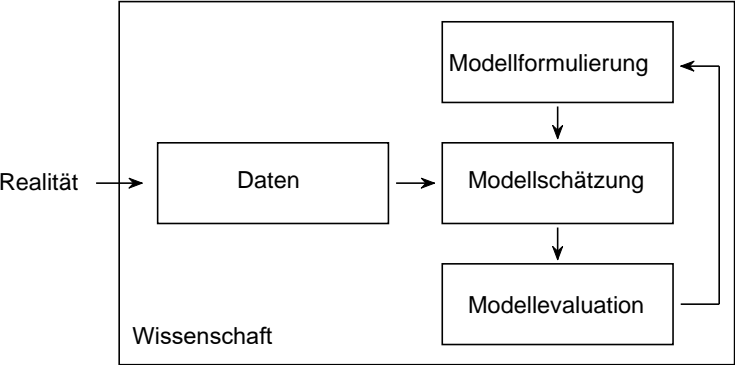


Fig. 1. Effect sizes for psychological interventions for depression. Error bars represent standard errors. PT = psychotherapy. Overall is based on all effect sizes without outliers and corrected for publication bias ($k = 146$ contrasts with WL, $k = 142$ contrasts with CAU, $k = 65$ contrasts with 'other' controls). PT for adult depression only includes (individual or group) psychotherapy for adults with a diagnosis of depression ($k = 30$ contrasts with WL, $k = 29$ contrasts with CAU, $k = 12$ contrasts with 'other control').

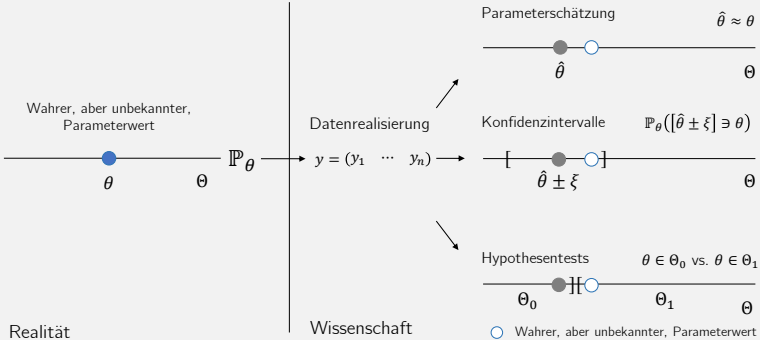
Munder et al. (2019)

Motivation



Motivation

Standardannahmen und Standardproblemstellungen der Frequentistischen Inferenz



Standardannahmen Frequentistischer Inferenz

\mathcal{M} sei ein Frequentistisches Inferenzmodell mit Stichprobenvariablen $v_1, \dots, v_n \sim p_\theta$. Es wird angenommen, dass ein konkreter Datensatz $y \in \mathbb{R}^n$ eine der möglichen Realisierungen von $y = (v_1, \dots, v_n)^T$ ist.

Aus Frequentistischer Sicht kann man eine Studie unendlich oft wiederholen und zu jedem Datensatz Schätzer oder Statistiken auswerten, z.B. das Stichprobenmittel:

$$\text{Datensatz (1)} : y^{(1)} = \left(y_1^{(1)} \quad \dots \quad y_n^{(1)} \right) \text{ mit } \bar{y}^{(1)} = \frac{1}{n} \sum_{i=1}^n y_i^{(1)}$$

$$\text{Datensatz (2)} : y^{(2)} = \left(y_1^{(2)} \quad \dots \quad y_n^{(2)} \right) \text{ mit } \bar{y}^{(2)} = \frac{1}{n} \sum_{i=1}^n y_i^{(2)}$$

$$\text{Datensatz (3)} : y^{(3)} = \left(y_1^{(3)} \quad \dots \quad y_n^{(3)} \right) \text{ mit } \bar{y}^{(3)} = \frac{1}{n} \sum_{i=1}^n y_i^{(3)}$$

$$\text{Datensatz (4)} : y^{(4)} = \left(y_1^{(4)} \quad \dots \quad y_n^{(4)} \right) \text{ mit } \bar{y}^{(4)} = \frac{1}{n} \sum_{i=1}^n y_i^{(4)}$$

$$\text{Datensatz (5)} : y^{(5)} = \dots$$

Um die Qualität statistischer Methoden zu beurteilen betrachtet die Frequentistische Statistik deshalb die Wahrscheinlichkeitsverteilungen von Schätzern und Statistiken unter Annahme von $v_1, \dots, v_n \sim p_\theta$. Was zum Beispiel ist die Verteilung der $\bar{y}^{(1)}, \bar{y}^{(2)}, \bar{y}^{(3)}, \bar{y}^{(4)}, \dots$ also die Verteilung der Zufallsvariable \bar{v} ?

Wenn eine statistische Methode im Sinne der Frequentistischen Standardannahmen "gut" ist, dann heißt das also, dass sie bei häufiger Anwendung "im Mittel gut" ist. Im Einzelfall, also im Normalfall nur eines vorliegenden Datensatzes, kann sie auch "schlecht" sein.

Standardproblemstellungen Frequentistischer Inferenz

(1) Parameterschätzung

Ziel der Parameterschätzung ist es, einen möglichst guten Tipp für wahre, aber unbekannte, Parameterwerte oder Funktionen dieser abzugeben, typischerweise mithilfe der Daten.

(2) Konfidenzintervalle

Ziel der Bestimmung von Konfidenzintervallen ist es, basierend auf der angenommenen Verteilung der Daten eine quantitative Aussage über die mit Schätzwerten assoziierte Unsicherheit zu treffen.

(3) Hypothesentests

Ziel des Hypothesentestens ist es, basierend auf der angenommenen Verteilung der Daten in einer möglichst zuverlässigen Form zu entscheiden, ob ein wahrer, aber unbekannter Parameterwert in einer von zwei sich gegenseitig ausschließenden Untermengen des Parameterraumes liegt.

Motivation

Standardisierte Effektstärkeschätzung in Einzelstudien

Anwendungsbeispiel

Selbstkontrollfragen

Standardisierte Effektstärkeschätzung in Einzelstudien

Hedges' g ist ein durch einen Multiplikationsfaktor angepasstes Cohen's d , das von Hedges (1981) und Hedges & Olkin (1985) eingeführt wurde. Grundlage für Hedges' g ist die Frequentistische Verteilung von Cohen's d in einem Normalverteilungsmodell der Daten einer einzelnen Studie. In diesem Modell ist Cohen's d ein verzerrter Schätzer der wahren, aber unbekannt, standardisierten Effektstärke, der Erwartungswert von Cohen's d entspricht also nicht der wahren, aber unbekannt, standardisierten Effektstärke, Die Differenz zwischen dem Erwartungswert und der wahren, aber unbekannt, standardisierten Effektstärke, wird *Bias* genannt.

Im betrachteten Normalverteilungsmodell kann der Bias von Cohen's d analytisch bestimmt und numerisch approximiert werden. Die Bekanntheit des Bias erlaubt dann seine Korrektur durch Skalierung von Cohen's d und der zur Biaskorrektur skalierte Wert von Cohen's d wird Hedges' g genannt. Dabei ist der Bias von Cohen's d für kleine Stichprobenumfänge am größten.

Letztlich ist Hedges' g einfach der Tatsache geschuldet, dass Cohen's d eng mit der T-Statistik verwandt ist, die bekanntlich bei Nicht-Zutreffen der Nullhypothese nichtzentral- t -verteilt ist und der Erwartungswert einer nichtzentralen- t -verteilten Zufallsvariable nicht mit ihrem Nichtzentralitätsparameter identisch ist. Dabei ist die Nicht-Normalität der Verteilung von Cohen's d für größere Stichprobenumfänge vernachlässigbar, was wiederum asymptotische Schätzungen der Varianz von Hedges' g ermöglicht.

Im Kontext von Metanalyse werden Hedges' g und sein Varianzschätzer meist zunächst für jede Einzelstudie bestimmt und dann über Studien, zum Beispiel mit den Fixed-Effects- und Random-Effects-Modellen aus Einheit (4) integriert.

Definition (Einzelstudienmodell)

Es seien y_{tj} für $j = 1, \dots, n_t$ und y_{cj} für $j = 1, \dots, n_c$ die Zufallsvariablen, die die primären Ergebnismaße der j ten experimentellen Einheit in der Treatment- und der Kontrollgruppe einer Einzelstudie modellieren, respektive. Dann ist das Einzelstudienmodell (ESM) gegeben durch

$$y_{tj} := \mu_t + \varepsilon_{tj} \text{ mit } \varepsilon_{tj} \sim N(0, \sigma^2) \text{ für } j = 1, \dots, n_t \quad (1)$$

$$y_{cj} := \mu_c + \varepsilon_{cj} \text{ mit } \varepsilon_{cj} \sim N(0, \sigma^2) \text{ für } j = 1, \dots, n_c$$

und die wahre, aber unbekannte, standardisierte Effektstärke (SES) ist definiert als

$$\delta := \frac{\mu_t - \mu_c}{\sigma}. \quad (2)$$

Bemerkungen

- Das Einzelstudienmodell ist offenbar identisch mit dem Zweistichproben-T-Test-Modell.
- Mit der Theorem zur linear-affinen Transformation von normalverteilten Zufallsvariablen gilt äquivalent

$$y_{tj} \sim N(\mu_t, \sigma^2) \text{ für } j = 1, \dots, n_t \text{ und } y_{cj} \sim N(\mu_c, \sigma^2) \text{ für } j = 1, \dots, n_c \quad (3)$$

- Die hier gewählte Formulierung entspricht Hedges & Olkin (1985) p. 76 (vgl. Lin & Aloe (2021))
- Hedges (1981) formuliert ein äquivalentes Modell basierend auf einer nicht-standardisierten Effektstärke.

Definition (SES-Schätzer im Einzelstudienmodell)

Gegeben sei ein Einzelstudienmodell mit wahrer, aber unbekannter, standardisierter Effektstärke $\delta \in \mathbb{R}$. Dann ist der *Standardisierte Effektstärkeschätzer (SES-Schätzer)* d von δ definiert als

$$d := \frac{\bar{y}_t - \bar{y}_c}{s}, \quad (4)$$

wobei

$$\bar{y}_t := \frac{1}{n_t} \sum_{j=1}^{n_t} y_{tj} \quad \text{und} \quad \bar{y}_c := \frac{1}{n_c} \sum_{j=1}^{n_c} y_{cj} \quad (5)$$

die Stichprobenmittel der Treatment- und der Kontrollgruppe bezeichnen und s die mithilfe der Stichprobenvarianzen

$$s_t^2 := \frac{1}{n_t - 1} \sum_{j=1}^{n_t} (y_{tj} - \bar{y}_t)^2 \quad \text{und} \quad s_c^2 := \frac{1}{n_c - 1} \sum_{j=1}^{n_c} (y_{cj} - \bar{y}_c)^2 \quad (6)$$

der Treatment- und Kontrollgruppe definierte *gepoolte Stichprobenstandardabweichung*

$$s := \sqrt{\frac{(n_t - 1)s_t^2 + (n_c - 1)s_c^2}{n_t + n_c - 2}}. \quad (7)$$

bezeichnet.

Bemerkungen

- Der SES-Schätzer ist offenbar mit Cohen's d identisch.
- Für den SES-Schätzer werden in δ lediglich μ_t , μ_c und σ durch ihre empirischen Äquivalente ersetzt.

Theorem (Eigenschaften des SES-Schätzer im ESM)

Gegeben sei ein Einzelstudienmodell, es sei d der Schätzer der wahren, aber unbekannt, standardisierten Effektstärke δ und es seien

$$n := \frac{n_t n_c}{n_t + n_c}, m := n_t + n_c - 2 \text{ und } c_m = \frac{\Gamma(m/2)}{\sqrt{m/2} \Gamma((m-1)/2)}. \quad (8)$$

Dann hat d folgende Eigenschaften:

- (1) $\sqrt{n}d \sim t(\sqrt{n}\delta, m)$,
- (2) $\mathbb{E}(d) = \frac{1}{c_m} \delta$, und
- (3) $\mathbb{V}(d) = \frac{m}{(m-2)n} (1 + n\delta^2) - \frac{1}{c_m^2} \delta^2$.

Remarks

- $\sqrt{n}d$ ist eine nichtzentral- t -verteilte Zufallsvariable.
- Der Nichtzentralitätsparameter von $\sqrt{n}d$ ist $\sqrt{n}\delta$, der Freiheitsgradparameter ist m .
- Für die Definition nichtzentral- t -verteilter Zufallsvariablen und ihre Kennzahlen verweisen wir auf den Appendix.
- Für $\delta \neq 0$ und $c_m \neq 1$ impliziert (2), dass $\mathbb{E}(d) \neq \delta$.
- d ist also ein verzerrter Schätzer von δ .

Standardisierte Effektstärkeschätzung in Einzelstudien

Beweis

(1) Wir verzichten auf einen Beweis und verweisen wie Hedges (1981) auf Johnson & Welch (1940).

(2) Mit dem Erwartungswert einer nichtzentral- t -verteilten Zufallsvariable ergibt sich

$$\mathbb{E}(\sqrt{nd}) = \sqrt{n}\delta \sqrt{\frac{m}{2}} \frac{\Gamma((m-1)/2)}{\Gamma(m/2)}. \quad (9)$$

Es ergibt sich also

$$\mathbb{E}(d) = \frac{\sqrt{n}}{\sqrt{n}} \mathbb{E}(d) = \frac{1}{\sqrt{n}} \mathbb{E}(\sqrt{nd}) = \delta \sqrt{\frac{m}{2}} \frac{\Gamma((m-1)/2)}{\Gamma(m/2)} = \frac{1}{c_m} \delta. \quad (10)$$

(3) Mit der Varianz von nichtzentral- t -verteilten Zufallsvariablen ergibt sich

$$\begin{aligned} \mathbb{V}(\sqrt{nd}) &= \frac{m(1 + (\sqrt{nd})^2)}{m-2} - \frac{(\sqrt{nd})^2 m}{2} \left(\frac{\Gamma((m-1)/2)}{\Gamma(m/2)} \right)^2 \\ &= \frac{m}{m-2} (1 + n\delta^2) - n\delta^2 \frac{m}{2} \left(\frac{\Gamma((m-1)/2)}{\Gamma(m/2)} \right)^2 \\ &= \frac{m}{m-2} (1 + n\delta^2) - n\delta^2 \frac{1}{c_m^2}. \end{aligned} \quad (11)$$

Es ergibt sich also

$$\mathbb{V}(d) = \frac{n}{n} \mathbb{V}(d) = \frac{1}{n} \mathbb{V}(\sqrt{nd}) = \frac{1}{n} \left(\frac{m}{m-2} (1 + n\delta^2) - n\delta^2 \frac{1}{c_m^2} \right) = \frac{m}{(m-2)n} (1 + n\delta^2) - \delta^2 \frac{1}{c_m^2}. \quad (12)$$

Standardisierte Effektstärkeschätzung in Einzelstudien

Verteilung von \sqrt{nd}

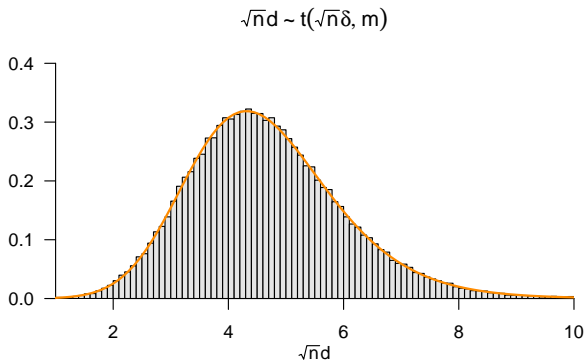
10^5 SES-Schätzerrealisierungen mit $\delta = 2, n_t = 10, n_c = 10, m = 18, n = 5$

```
sim      = 1e5                # Anzahl Realisierungen
n_t      = 10                # Treatmentgruppenumfang
n_c      = 10                # Kontrollgruppenumfang
mu_t     = 2                 # Erwartungswertparameter Treatment
mu_c     = 0                 # Erwartungswertparameter Control
sigma    = 1                 # Varianzparameter
delta    = (mu_t - mu_c)/sigma # wahre, aber unbekannte, SES
n        = (n_t*n_c/(n_t+n_c)) # Stichprobenumfangparameter
m        = n_t + n_c - 2     # Freiheitsgradparameter
c_m      = (gamma(m/2))/(sqrt(m/2)*gamma((m-1)/2)) # Biaskorrekturfaktor
ds       = rep(NaN, sim)    # SES-Schätzerarrayinitialisierung
for(i in 1:sim){           # Simulationsiterationen
  y       = matrix(rep(NaN,n_t+n_c), ncol = 2) # Datenarrayinitialisierung
  y[,1]   = rnorm(n_t, mu_t, sigma)           # Datengeneration Treatmentgruppe
  y[,2]   = rnorm(n_c, mu_c, sigma)           # Datengeneration Kontrollgruppe
  ybar    = apply(y,2,mean)                   # Gruppenmittelwerte
  s2      = apply(y,2,var)                    # Gruppenvarianzen
  s       = sqrt(((n_t-1)*s2[1] + (n_c-1)*s2[2])/m) # gepoolte Standardabweichung
  ds[i]   = (ybar[1]-ybar[2])/s              # SES-Schätzerrealisierung
```

Standardisierte Effektstärkeschätzung in Einzelstudien

Verteilung von \sqrt{nd}

10^5 SES-Schätzerrealisierungen mit $\delta = 2$, $n_t = 10$, $n_c = 10$, $m = 18$, $n = 5$



Theorem (Hedges' g im ESM)

Gegeben sei ein Einzelstudienmodell, d sei der verzerrte Schätzer von δ und es seien

$$n := \frac{n_t n_c}{n_t + n_c}, m := n_t + n_c - 2 \text{ und } c_m = \frac{\Gamma(m/2)}{\sqrt{m/2} \Gamma((m-1)/2)}. \quad (13)$$

wie oben definiert. Dann ist

$$g := c_m d, \quad (14)$$

genannt *Hedges' g* , ein unverzerrter Schätzer von δ und es gilt

$$\mathbb{V}(g) < \mathbb{V}(d). \quad (15)$$

Beweis

Die Unverzerrtheit von d_u ergibt sich aus

$$\mathbb{E}(g) = \mathbb{E}(c_m d) = c_m \mathbb{E}(d) = c_m \frac{1}{c_m} \delta = \delta. \quad (16)$$

Die Varianzreduktion von d_u in Bezug auf d folgt aus der Tatsache, dass $c_m < 1$ für alle m und somit

$$\mathbb{V}(g) = \mathbb{V}(c_m d) = c_m^2 \mathbb{V}(d) < \mathbb{V}(d). \quad (17)$$

□

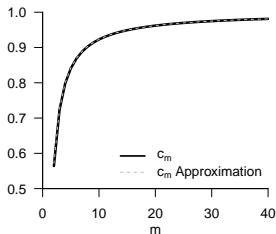
Standardisierte Effektstärkeschätzung in Einzelstudien

Bemerkungen

- Man beachte, dass der Korrekturfaktor c_m durch $n_t + n_c - 2$ vom Gesamtstichprobenumfang abhängig ist.
- Der Wert des Korrekturfaktors für einen gegebenen Wert von m kann sehr genau durch

$$c_m \approx 1 - \frac{3}{4m - 1} \quad (18)$$

angenähert werden, vgl. Hedges (1981), Borenstein (2009) und untenstehende Abbildung.

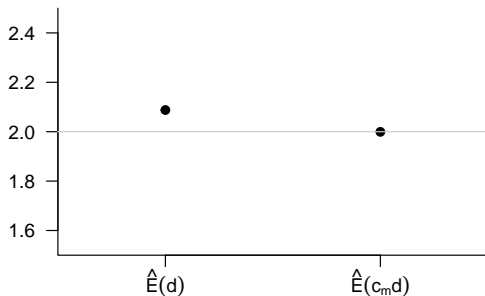


- Lin & Aloe (2021) zeigen in Appendix A1, dass $c_m \rightarrow 1$ für $m \rightarrow \infty$.
- Die Biaskorrektur ist am wichtigsten bei Gesamtgruppengrößen $n_t + n_c < 30$.
- Die Biaskorrektur ist am wichtigsten bei Treatment- und Kontrollgruppengrößen $n_t = n_c < 15$.

Standardisierte Effektstärkeschätzung in Einzelstudien

Geschätzte Erwartungswerte der verzerrten und unverzerrten SES-Schätzer im ESM

10^5 SES-Schätzerrealisierungen mit $\delta = 2$, $n_t = 10$, $n_c = 10$, $m = 18$, $n = 5$



Theorem (Asymptotische Normalverteilung von Hedges' g)

Gegeben sei ein Einzelstudienmodell und g sei der unverzerrte Schätzer von δ . Dann ist g bei einem konstanten Verhältnis von n_t und n_c für $n_t \rightarrow \infty$, $n_c \rightarrow \infty$ asymptotisch normalverteilt und es gilt speziell

$$g \stackrel{a}{\sim} N \left(\delta, \frac{1}{n_t} + \frac{1}{n_c} + \frac{\delta^2}{2(n_t + n_c)} \right). \quad (19)$$

Bemerkungen

- Wir erinnern daran, dass *Asymptotische Schätzereigenschaften* sich auf große Stichprobenumfänge beziehen.
- Das Theorem ist äquivalent dazu, dass die nichtzentrale- t -Verteilung für $\nu \rightarrow \infty$ die Normalverteilung annähert.
- Wir verzichten auf einen Beweis und verweisen wie Hedges & Olkin (1985) auf Johnson & Welch (1940).

Standardisierte Effektstärkeschätzung in Einzelstudien

Asymptotische Normalverteilung von Hedges' g

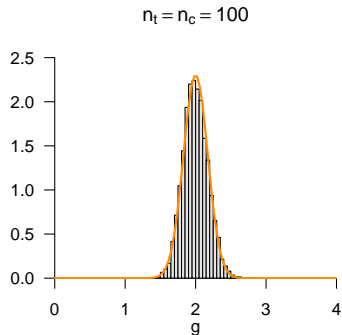
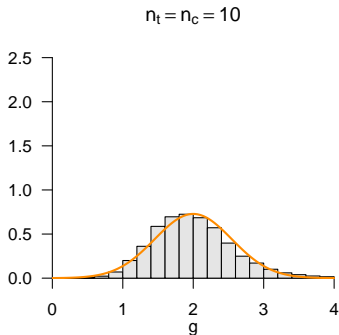
10^5 SES-Schätzerrealisierungen mit $\delta = 2$ für kleinen und großen Stichprobenumfang

```
sim           = 1e4           # Anzahl Realisierungen
mu_t         = 2             # Erwartungswertparameter Treatment
mu_c         = 0             # Erwartungswertparameter Control
sigma        = 1             # Varianzparameter
delta        = (mu_t - mu_c)/sigma # wahre, aber unbekannte, SES
ng           = c(10,100)     # Stichprobenumfänge
gs           = matrix(rep(NaN, 2*sim), ncol = 2) # Hedges' g Arrayinitialisierung
vgs          = rep(NaN,2)    # asymptotische Varianzarray
for (j in 1:2){             # Stichprobenumfangsiterationen
  n_t         = ng[j]        # Treatmentgruppenumfang
  n_c         = ng[j]        # Kontrollgruppenumfang
  n           = (n_t*n_c/(n_t+n_c)) # Stichprobenumfangparameter
  m           = n_t + n_c - 2 # Freiheitsgradparameter
  c_m         = (gamma(m/2))/(sqrt(m/2)*gamma((m-1)/2)) # Biaskorrekturfaktor
  vgs[j]      = (1/n_t)+(1/n_c)+(delta^2)/(2*(n_t+n_c)) # Asymptotische Varianz von Hedges' g
  for(i in 1:sim){         # Simulationsiterationen
    y          = matrix(rep(NaN,n_t+n_c), ncol = 2) # Datenarrayinitialisierung
    y[,1]     = rnorm(n_t, mu_t, sigma) # Datengeneration Treatmentgruppe
    y[,2]     = rnorm(n_c, mu_c, sigma) # Datengeneration Kontrollgruppe
    ybar      = apply(y,2,mean) # Gruppenmittelwerte
    s2        = apply(y,2,var) # Gruppenvarianzen
    s         = sqrt(((n_t-1)*s2[1] + (n_c-1)*s2[2])/m) # gepoolte Standardabweichung
    gs[i,j]   = c_m*(ybar[1]-ybar[2])/s} # Hedges' g
```


Standardisierte Effektstärkeschätzung in Einzelstudien

Asymptotische Normalverteilung von Hedges' g

10^5 SES-Schätzerrealisierungen mit $\delta = 2$ für kleinen und großen Stichprobenumfang



Definition (Asymptotische Varianz und Varianzschätzer von Hedges g)

Gegeben sei ein Einzelstudienmodell und für den

$$g \stackrel{a}{\sim} N \left(\delta, \frac{1}{n_t} + \frac{1}{n_c} + \frac{\delta^2}{2(n_t + n_c)} \right). \quad (20)$$

Dann wird der Varianzparameter der asymptotischen Verteilung von g die *asymptotische Varianz* von g genannt und mit

$$V_a(g) = \frac{1}{n_t} + \frac{1}{n_c} + \frac{\delta^2}{2(n_t + n_c)}. \quad (21)$$

bezeichnet. Ein häufig genutzter Schätzer für die asymptotische Varianz von Hedges' g ist

$$\hat{V}_a(g) = \frac{1}{n_t} + \frac{1}{n_c} + \frac{g^2}{2(n_t + n_c)}. \quad (22)$$

- $\hat{V}_a(g)$ wird häufig für die metanalytische Modellbildung genutzt, vgl. Einheit (4) und Lin & Aloe (2021).
- Bei konstantem Wert von Hedges g nimmt $\hat{V}_a(g)$ offenbar bei steigenden Werten von n_t und n_c ab.

Theorem (Asymptotisches Konfidenzintervall für die SES)

Gegeben sei ein Einzelstudienmodell, g sei der unverzerrte Schätzer der standardisierten Effektstärke δ und für ein Konfidenzlevel $\gamma \in]0, 1[$ sei

$$z_\gamma := \Phi^{-1} \left(\frac{1 + \gamma}{2} \right) \quad (23)$$

wobei Φ^{-1} die inverse kumulative Verteilungsfunktion der Standardnormalverteilung bezeichne. Dann ist

$$\kappa := \left[g - \sqrt{\mathbb{V}_a(g)} z_\gamma, g + \sqrt{\mathbb{V}_a(g)} z_\gamma \right] \quad (24)$$

ein asymptotisches γ -Konfidenzintervall für die standardisierte Effektstärke δ .

Bemerkungen

- Das betrachtete Konfidenzintervall wurde von Hedges & Olkin (1985) (S.86) vorgeschlagen.
- Wir nutzen hier γ für das Konfidenzlevel, da δ schon belegt ist.
- κ entspricht einem Konfidenzintervall für den Erwartungswert der Normalverteilung bei bekannter Varianz.
- In der Anwendung wird man $\mathbb{V}_a(g)$ durch seinen Schätzer $\hat{\mathbb{V}}_a(g)$ ersetzen.
- Asymptotische Konfidenzintervalle werden auch *Konfidenzintervalle von Wald-Typ* genannt.

Standardisierte Effektstärkeschätzung in Einzelstudien

Beweis

Wir müssen zeigen, dass

$$\mathbb{P}(\kappa \ni \delta) = \gamma. \quad (25)$$

Dazu halten wir zunächst fest, dass

$$g \stackrel{a}{\sim} N(\delta, \mathbb{V}_a(g)). \quad (26)$$

Mit dem Theorem zur Z -Transformation gilt dann

$$Z_g := \frac{g - \delta}{\sqrt{\mathbb{V}_a(g)}} \sim N(0, 1). \quad (27)$$

Weiterhin gilt per Definition von z_γ , dass

$$\mathbb{P}(-z_\gamma \leq Z_g \leq z_\gamma) = \gamma. \quad (28)$$

Beweis (fortgeführt)

Aus der Definition eines γ -Konfidenzintervalls folgt dann

$$\begin{aligned}\gamma &= \mathbb{P}(-z_\gamma \leq Z_g \leq z_\gamma) \\ &= \mathbb{P}\left(-z_\gamma \leq \frac{g - \delta}{\sqrt{V_a(g)}} \leq z_\gamma\right) \\ &= \mathbb{P}\left(-z_\gamma \sqrt{V_a(g)} \leq g - \delta \leq z_\gamma \sqrt{V_a(g)}\right) \\ &= \mathbb{P}\left(-g - z_\gamma \sqrt{V_a(g)} \leq -\delta \leq -g + z_\gamma \sqrt{V_a(g)}\right) \\ &= \mathbb{P}\left(g + z_\gamma \sqrt{V_a(g)} \geq \delta \geq g - z_\gamma \sqrt{V_a(g)}\right) \\ &= \mathbb{P}\left(g - z_\gamma \sqrt{V_a(g)} \leq \delta \leq g + z_\gamma \sqrt{V_a(g)}\right) \\ &= \mathbb{P}\left(\left[g - \sqrt{V_a(g)}z_\gamma, g + \sqrt{V_a(g)}z_\gamma\right] \ni \delta\right) \\ &= \mathbb{P}(\kappa \ni \delta)\end{aligned}\tag{29}$$

und damit ist alles gezeigt.

Standardisierte Effektstärkeschätzung in Einzelstudien

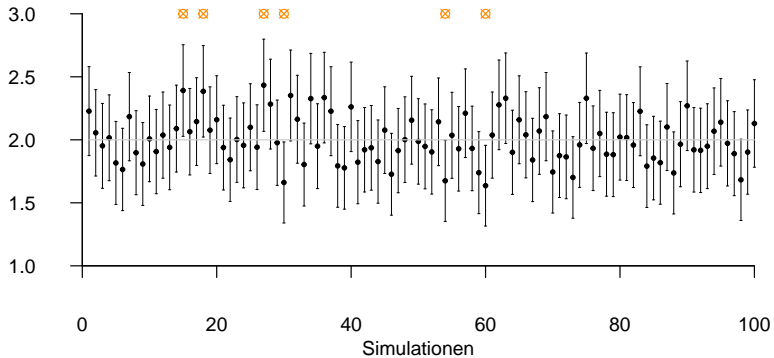
Simulation eines asymptotischen 95%-Konfidenzintervalls für die SES

```
set.seed(0)
sim          = 1e2
mu_t        = 2
mu_c        = 0
sigma       = 1
delta       = (mu_t - mu_c)/sigma
gamma       = 0.95
z_gamma     = qnorm((1+gamma)/2)
n_t         = 100
n_c         = 100
n           = (n_t*n_c)/(n_t+n_c)
m           = n_t + n_c - 2
c_m         = (gamma(m/2))/(sqrt(m/2)*gamma((m-1)/2))
gs          = rep(NaN, sim)
vs          = rep(NaN, sim)
kappa       = array(rep(NaN, sim*2), dim = c(2,sim))
for(i in 1:sim){
  y          = matrix(rep(NaN,n_t+n_c), ncol = 2)
  y[,1]     = rnorm(n_t, mu_t, sigma)
  y[,2]     = rnorm(n_c, mu_c, sigma)
  ybar      = apply(y,2,mean)
  s2        = apply(y,2,var)
  s         = sqrt(((n_t-1)*s2[1] + (n_c-1)*s2[2])/m)
  gs[i]     = c_m*(ybar[1]-ybar[2])/s
  vs[i]     = ((n_t+n_c)/(n_t*n_c))*(gs[i]^2)/(2*(n_t+n_c))
  kappa[1,i] = gs[i] - sqrt(vs[i])*z_gamma
  kappa[2,i] = gs[i] + sqrt(vs[i])*z_gamma
}
```

Zufallsgeneratorstate
Anzahl Realisierungen
Erwartungswertparameter Treatment
Erwartungswertparameter Control
Varianzparameter
wahre, aber unbekannte, SES
Konfidenzlevel
Konfidenzlevelkonstante
Treatmentgruppenumfang
Kontrollgruppenumfang
Stichprobenumfangparameter
Freiheitsgradparameter
Biaskorrekturfaktor
Hedges' g Arrayinitialisierung
asymptotische Varianzarray
Konfidenzintervallarray
Simulationsiterationen
Datenarrayinitialisierung
Datengeneration Treatmentgruppe
Datengeneration Kontrollgruppe
Gruppenmittelwerte
Gruppenvarianzen
gepoolte Standardabweichung
Hedges' g
Varianzschätzer
untere Konfidenzintervallgrenze
obere Konfidenzintervallgrenze

Standardisierte Effektstärkeschätzung in Einzelstudien

Simulation eines asymptotischen 95%-Konfidenzintervalls für die SES



Motivation

Standardisierte Effektstärkeschätzung in Einzelstudien

Anwendungsbeispiel

Selbstkontrollfragen



A systematic review and meta-analysis of transdiagnostic cognitive behavioural therapies for emotional disorders

Received: 12 April 2023

Accepted: 17 November 2023

Published online: 16 January 2024



Carmen Schaeuffele ^{1,2}, Laura E. Meine ^{2,3,4}, Ava Schulz ^{2,5},
Maxi C. Weber¹, Angela Moser^{2,5}, Christina Paersch^{2,5,6}, Dominique Recher ^{2,5},
Johanna Boettcher¹, Babette Renneberg¹, Christoph Flückiger⁸ & Birgit Kleim^{2,8}

Transdiagnostic cognitive behavioural psychotherapy (TD-CBT) may facilitate the treatment of emotional disorders. Here we investigate short- and long-term efficacy of TD-CBT for emotional disorders in individual, group and internet-based settings in randomized controlled trials (PROSPERO CRD42019141512). Two independent reviewers screened results from PubMed, MEDLINE, PsycINFO, Google Scholar, medRxiv and OSF Preprints published between January 2000 and June 2023, selected studies for inclusion, extracted data and evaluated risk of bias (Cochrane risk-of-bias tool 2.0). Absolute efficacy from pre- to posttreatment and relative efficacy between TD-CBT and control treatments were investigated with random-effects models. Of 56 identified studies, 53 (6,705 participants) were included in the meta-analysis. TD-CBT had larger effects on depression ($g = 0.74$, 95% CI = 0.57–0.92, $P < 0.001$) and anxiety ($g = 0.77$, 95% CI = 0.56–0.97, $P < 0.001$) than did controls. Across treatment formats, TD-CBT was superior to waitlist and treatment-as-usual. TD-CBT showed comparable effects to disorder-specific CBT and was superior to other active treatments for depression but not for anxiety. Different treatment formats showed comparable effects. TD-CBT was superior to controls at 3, 6 and 12 months but not at 24 months follow-up. Studies were heterogeneous in design and methodological quality. This review and meta-analysis strengthens the evidence for TD-CBT as an efficacious treatment for emotional disorders in different settings.

Schaeuffele et al. (2024)

A systematic review and meta-analysis of TD-CBT for emotional disorders

“All analyses were conducted in R (v.4.3.1), using the metafor (v.4.2-0), meta (v.6.5-0) and dmetar packages (v.0.1.0). We calculated controlled effect sizes for the difference between the transdiagnostic treatment and the control conditions in main outcomes (depression and anxiety) at posttreatment (relative efficacy), using the bias-corrected Hedges’ g and the 95% CI. These were calculated by subtracting the mean posttreatment score of the transdiagnostic condition from the mean score of the control condition, divided by the pooled standard deviation of both conditions. Values of 0.2, 0.5 and 0.8 of Hedges’ g represent a small, moderate and large effect size, respectively.”

Data and Code availability. The data that support the findings of this study, along with data collection templates, are publicly available at the Open Science Framework and can be accessed at <https://osf.io/ta4fg/>. Custom analysis code that supports the findings of this study is publicly available at the Open Science Framework and can be accessed at <https://osf.io/ta4fg/>.

Schaeuffele et al. (2024)

Anwendungsbeispiel

Identifikation von 56 relevanten RCTs mit n = 6916 Teilnehmer:innen (PRISMA Flowchart)

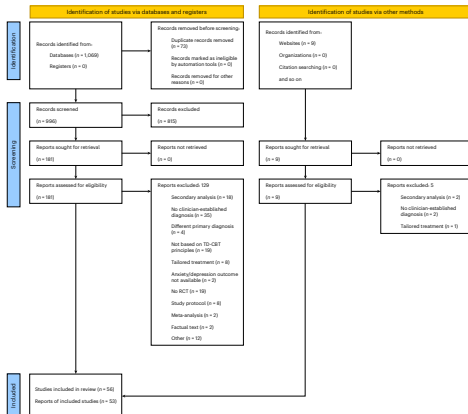


Fig. 1 | PRISMA flowchart of the literature search and screening procedure. Three studies could not be included in the meta-analysis because either no self-report of anxiety or depression was available¹⁰ or no data were available^{11,12}. For one study¹³, treatment effects at 12 months follow-up were reported in a

separate publication¹⁴ which was not included in the final number of studies as this reflects the number of RCTs identified. However, we included the follow-up values in our meta-analysis.

Schaeuffele et al. (2024)

Anwendungsbeispiel

Einschätzung des Verzerrungsrisikos anhand des [Cochrane Risk of Bias Tools](#)

	Risk of bias domains					Overall
	D1	D2	D3	D4	D5	
Ref. ¹⁰	+	-	+	-	-	-
Ref. ¹¹	+	-	-	-	-	-
Ref. ¹²	+	-	-	-	-	-
Ref. ¹³	+	-	-	-	-	-
Ref. ¹⁴	+	-	+	-	-	-
Ref. ¹⁵	+	-	+	-	-	-
Ref. ¹⁶	+	-	-	-	-	-
Ref. ¹⁷	+	-	-	-	-	-
Ref. ¹⁸	+	-	-	-	-	-
Ref. ¹⁹	+	-	-	-	-	-
Ref. ²⁰	+	-	+	-	+	-
Ref. ²¹	+	-	+	-	-	-
Ref. ²²	+	-	+	-	-	-
Ref. ²³	+	-	-	-	-	-
Ref. ²⁴	+	-	-	-	-	-
Ref. ²⁵	+	-	-	-	-	-
Ref. ²⁶	+	-	-	-	-	-
Ref. ²⁷	+	-	-	-	-	-
Ref. ²⁸	+	-	-	-	-	-
Ref. ²⁹	+	-	-	-	-	-
Ref. ³⁰	+	-	-	-	-	-
Ref. ³¹	+	-	-	-	-	-
Ref. ³²	+	-	-	-	-	-
Ref. ³³	+	-	-	-	-	-
Ref. ³⁴	+	-	-	-	-	-
Ref. ³⁵	+	-	-	-	-	-
Ref. ³⁶	+	-	-	-	-	-
Ref. ³⁷	+	-	-	-	-	-
Ref. ³⁸	+	-	-	-	-	-
Ref. ³⁹	+	-	-	-	-	-
Ref. ⁴⁰	+	-	-	-	-	-
Ref. ⁴¹	+	-	-	-	-	-
Ref. ⁴²	+	-	-	-	-	-
Ref. ⁴³	+	-	-	-	-	-
Ref. ⁴⁴	+	-	-	-	-	-
Ref. ⁴⁵	+	-	-	-	-	-
Ref. ⁴⁶	+	-	-	-	-	-
Ref. ⁴⁷	+	-	-	-	-	-
Ref. ⁴⁸	+	-	-	-	-	-
Ref. ⁴⁹	+	-	-	-	-	-
Ref. ⁵⁰	+	-	-	-	-	-
Ref. ⁵¹	+	-	-	-	-	-
Ref. ⁵²	+	-	-	-	-	-
Ref. ⁵³	+	-	-	-	-	-
Ref. ⁵⁴	+	-	-	-	-	-
Ref. ⁵⁵	+	-	-	-	-	-
Ref. ⁵⁶	+	-	-	-	-	-
Ref. ⁵⁷	+	-	-	-	-	-
Ref. ⁵⁸	+	-	-	-	-	-
Ref. ⁵⁹	+	-	-	-	-	-
Ref. ⁶⁰	+	-	-	-	-	-
Ref. ⁶¹	+	-	-	-	-	-
Ref. ⁶²	+	-	-	-	-	-
Ref. ⁶³	+	-	-	-	-	-
Ref. ⁶⁴	+	-	-	-	-	-
Ref. ⁶⁵	+	-	-	-	-	-
Ref. ⁶⁶	+	-	-	-	-	-
Ref. ⁶⁷	+	-	-	-	-	-
Ref. ⁶⁸	+	-	-	-	-	-
Ref. ⁶⁹	+	-	-	-	-	-
Ref. ⁷⁰	+	-	-	-	-	-
Ref. ⁷¹	+	-	-	-	-	-
Ref. ⁷²	+	-	-	-	-	-
Ref. ⁷³	+	-	-	-	-	-
Ref. ⁷⁴	+	-	-	-	-	-
Ref. ⁷⁵	+	-	-	-	-	-
Ref. ⁷⁶	+	-	-	-	-	-
Ref. ⁷⁷	+	-	-	-	-	-
Ref. ⁷⁸	+	-	-	-	-	-
Ref. ⁷⁹	+	-	-	-	-	-
Ref. ⁸⁰	+	-	-	-	-	-
Ref. ⁸¹	+	-	-	-	-	-
Ref. ⁸²	+	-	-	-	-	-
Ref. ⁸³	+	-	-	-	-	-
Ref. ⁸⁴	+	-	-	-	-	-
Ref. ⁸⁵	+	-	-	-	-	-
Ref. ⁸⁶	+	-	-	-	-	-
Ref. ⁸⁷	+	-	-	-	-	-
Ref. ⁸⁸	+	-	-	-	-	-
Ref. ⁸⁹	+	-	-	-	-	-
Ref. ⁹⁰	+	-	-	-	-	-
Ref. ⁹¹	+	-	-	-	-	-
Ref. ⁹²	+	-	-	-	-	-
Ref. ⁹³	+	-	-	-	-	-
Ref. ⁹⁴	+	-	-	-	-	-
Ref. ⁹⁵	+	-	-	-	-	-
Ref. ⁹⁶	+	-	-	-	-	-
Ref. ⁹⁷	+	-	-	-	-	-
Ref. ⁹⁸	+	-	-	-	-	-
Ref. ⁹⁹	+	-	-	-	-	-
Ref. ¹⁰⁰	+	-	-	-	-	-

Domains:

D1: Bias arising from the randomization process.

D2: Bias due to deviations from intended intervention.

D3: Bias due to missing outcome data.

D4: Bias in measurement of the outcome.

D5: Bias in selection of the reported result.

Judgement

⊗ High

- Some concerns

+ Low

Schaeffele et al. (2024)

Anwendungsbeispiel

Beispielstudien

Table 1 (continued) | RCTs investigating TD-CBT for emotional disorders in individual, group and internet-based format

Authors	Country	Included diagnoses	Sample	TD-CBT protocol	Number of TD-CBT sessions	Control group(s)	Relevant features (sensitivity and depression)	Assessment times	Attrition (%) ^a
Ref. 108	Australia	MDD, GAD, SAD, PD, AG, OCD	N=156; mean age 39yr; female 85%	CBT, guided, n=39; ME-CBT, guided, n=40	6	TAU, n=39; MT, guided, n=40	GAD-7, PHQ-9	Pre, post, 3MFU	CBT: 23% post, 32% 3MFU; ME-CBT: 33% post, 28% 3MFU; TAU: 10% post, NA 3MFU; MT: 38% post, 48% 3MFU
Ref. 109	China	Anxiety disorders, depressive disorders, other emotion-related disorders	N=75; mean age 32yr; female 71%	IMED+TAU, n=37	8	TAU, n=38	BAI, BDI-II	Pre, post	IMED+TAU: 22% post, TAU: 5% post
Ref. 110	Australia	GAD, PA, SAD, MDD	N=53; mean age 28yr; female 64%	Unlabeled being course, guided (TD-CBT for students), n=30	5	WLC, n=23	GAD-7, PHQ-9	Pre, post, 3MFU	Unlabeled being: 30% post, 40% 3MFU; WLC: 5% post, NA 3MFU
Ref. 111	Australia	GAD, MDD	N=100; mean age 44yr; female 78%	Worry and sadness programme, guided (TD-CBT), n=45	6	WLC, n=54	GAD-7, PHQ-9	Pre, post, 3MFU	Worry and sadness: 7% post, 13% 3MFU; WLC: 2% post, NA 3MFU
Ref. 112	Germany	AG, GAD, PD, SAD, MDD, persistent depressive disorder, somatic symptom disorder, illness anxiety disorder	N=120; mean age 37yr; female 68%	Internet-based UP, guided, n=65	10	WLC, n=64	GAD-7, PHQ-9	Pre, post	UP: 30% post; WLC: 1% post
Ref. 113	Australia	GAD, SAD, PA	N=78; mean age 40yr; female 68%	Anxiety programme, guided (TD-CBT), n=40	6	WLC, n=38	GAD-7, PHQ-9	Pre, post, 3MFU	TD: 10% post, 20% 3MFU; WLC: 5% post, NA 3MFU
Ref. 114	Australia	GAD, PA, SAD, MDD	N=74; mean age 44yr; female 72%	Well-being programme, guided (TD-CBT), n=37	8	WLC, n=37	GAD-7, PHQ-9	Pre, post, 3MFU	TD: 8% post, 14% 3MFU; WLC: 5% post, NA 3MFU
Ref. 115	Australia	MDD	N=290; mean age 44yr; female 72%	Well-being course, guided (TD-CBT), n=149	5	DS-CBT for MDD, n=142	GAD-7, PHQ-9	Pre, post, 3MFU, 12MFU, 24MFU	TD: 5% post, 20% 3MFU, 30% 12MFU, 24% 24MFU; DS-CBT: 16% post, 16% 3MFU, 20% 12MFU, 19% 24MFU
Ref. 116	Romania	GAD, SAD, PA, PTSD, SP, OCD, MDD	N=81; mean age 34yr; female 81%	Internet-based UP, guided, n=64	9	WLC, n=33	ODIS, BDI-II	Pre, post, 6MFU	UP: 22% post, 49% 3MFU; WLC: 6% post, NA 6MFU
Ref. 117	Alghanistan	Depressive disorder, anxiety disorder	N=102; mean age 28yr; female 47%	Internet-based UP, unguided, n=51	12-14	TAU, n=51	ODIS, ODIS	Pre, post	UP: 22% post, TAU: 39% post

ACT, acceptance and commitment therapy; AET, autonomy enhancing therapy; AG, agoraphobia; AMT, affect regulation training; ASI, anxiety sensitivity index; AUD, alcohol use disorder; BA, behavioural activation; BDI, Beck depression inventory; BSI, brief symptom inventory; CALM, coordinated anxiety learning and management; CA-CBT, culturally adapted cognitive behavioural therapy plus problem management; CBT-F, CBT for perfectionists; CFC, common factor control; CPG, G.A., self-rating scale for effective antidepressants; CS, construct specific; DASS, short form of depression, anxiety and stress scale; DBT-ST, dialectical behaviour therapy skills training; DD, dysthymic disorder; DS, disorder-specific; DET, emotion regulation therapy; ESTEM, effective skills to empower effective men; EUC, enhanced usual care; F-SFT, false safety behaviour elimination therapy; PHQ-9, hospital anxiety and depression scale; HAM-A, Hamilton anxiety rating scale; HAM-D, Hamilton depression rating scale; ICBT, internet based CBT; IMED, internet based self-help guided intervention for emotional distress; IMT, mindfulness training; IMC, modified attention control; ME-CBT, mindfulness enhanced CBT; MFLU, month follow-up; MHS-SPIN, mini social phobia inventory; MBI, multimodal intervention; NCS, non otherwise specified; ODIS, overall anxiety severity and impairment scale; OCD, obsessive compulsive disorder; ODISG, overall depression severity and impairment scale; PA, panic disorder with/without agoraphobia; PD, panic disorder; PDSS, panic disorder severity scale; PHG, patient health questionnaire; PTSD, post-traumatic stress disorder; QDIS, quick inventory of depressive symptomatology; RNT-G, group treatment for repetitive negative thinking; RT, resistance training; SCL, symptom checklist; SP, specific phobias; SP-RAS, Sheehan patient-rated anxiety scale; ST, supportive therapy; STAI-5, state-trait anxiety inventory-state; SUB, substance use disorder; TBT, transdiagnostic behaviour therapy; OCS, transcranial direct current stimulation; TSP, transdiagnostic internet based protocol; UP, unified treatment protocol for emotional disorders; WFLU, week follow-up; WLC, waitlist control.
^aMissing data as a percentage of randomized individuals who did not provide further assessments. ^bData on depression-related outcomes (DASS-21) were provided by the authors upon request. ^cData for 12MFU were provided in ref. 116.

Schaeuffele et al. (2024)

Anwendungsbeispiel

Reformatierung des Datensatzes

```
# R Pakete
library(tidyr)
library(dplyr)

# Dateneinlesen, rearrangieren, NA Ausschluss (nach https://osf.io/ta4fg/)
load("3_Daten/td_ma.RData")
data_dep_between_post <- data_dep_between %>%
  arrange(match(comparison, c("DS-CBT", "TAU", "other", "wait-list")),
           study_name, match(setting, c("individual", "group", "internet-based")))
DS <- data_dep_between %>% drop_na(c(post_tau_m, post_tau_sd, post_td_m, post_td_sd, post_tau_n, post_tau_sd))

# Reformatierung
D = data.frame(id = DS$study_id, # Studien ID
               reference = DS$study_name, # Studienreferenz
               control = DS$comparison, # Kontrollbedingung (DS, TAU, Other, WL)
               setting = DS$setting, # Studiensetting (individual, internet-based, group)
               outcome = DS$questionnaire, # Primäres Ergebnismaß
               n_t = DS$post_td_n, # n Treatmentgruppe
               n_c = DS$post_tau_n, # n Kontrollgruppe
               m_t = DS$post_td_m, # Mittelwert Treatmentgruppe
               m_c = DS$post_tau_m, # Mittelwert Kontrollgruppe
               sd_t = DS$post_td_sd, # Standardabweichung Treatmentgruppe
               sd_c = DS$post_tau_sd) # Standardabweichung Kontrollgruppe

# Speichern des reformatierten Datensatzes
write.csv(D, "./3_Daten/TD-CBT.csv", row.names = FALSE)
```

Anwendungsbeispiel

n_ Gruppengröße, m_ Mittelwert, sd_ Standardabweichung für _t Treatment, _c Control

id	reference	control	setting	outcome	n_t	n_c	m_t	m_c	sd_t	sd_c
1	Barlow (2017)	WL	individual	ODSIS	65	32	3	5	4	5
1	Barlow (2017)	DS	individual	ODSIS	65	63	3	3	4	4
2	Berger (2017)	TAU	internet	BDI-II	57	63	16	23	12	13
3	de Ornelas (2017)	TAU	group	BDI	24	24	8	12	6	9
4	Dear (2015)	DS	internet	PHQ-9	142	140	6	8	4	4
5	Dear (2016)	DS	internet	PHQ-9	83	89	8	8	4	5
6	Ejeby (2014)	TAU	group	CPRS	66	65	10	14	7	9
6	Ejeby (2014)	other	group	CPRS	66	69	10	12	7	8
7	Ellard (2017)	TAU	individual	QIDS	8	10	7	12	5	4
9	Farchione (2012)	WL	individual	BDI-II	22	10	4	11	4	10
10	Fogliati (2016)	DS	internet	PHQ-9	64	58	8	8	5	5
11	Gros & Allan (2019)	other	individual	DASS	29	21	4	6	5	5
14	Johnston (2011)	WL	internet	PHQ-9	81	41	7	11	5	5
15	Lotfi (2014)	DS	individual	BDI-II	9	9	9	8	3	2
16	Mullin (2015)	WL	internet	PHQ-9	21	21	9	14	4	6
17	Neacsiu (2014)	other	group	PHQ-9	15	9	6	11	4	6
18	Newby (2013)	WL	internet	PHQ-9	43	53	6	10	4	5
20	Norton (2012)	DS	group	BDI	16	12	8	13	8	13
21	Norton (2005)	WL	group	DASS	6	9	5	23	3	3
24	Schmidt (2012)	WL	group	BDI	53	39	11	14	7	9
25	Titov (2010)	WL	internet	PHQ-9	36	34	8	11	6	7
26	Titov (2011)	WL	internet	PHQ-9	34	35	8	11	6	5
27	Titov (2015)	DS	internet	PHQ-9	142	119	8	7	5	5
28	Tulbure (2018)	WL	internet	BDI-II	50	31	11	20	11	15
30	Zemestani (2017)	WL	group	BDI-II	20	23	12	26	2	3
31	Ahmadi (2021)	other	individual	BDI	13	13	8	20	8	8
31	Ahmadi (2021)	WL	individual	BDI	13	14	8	24	8	9
32	Corpas (2021)	TAU	individual	PHQ-9	28	30	8	12	3	4
32	Corpas (2021)	TAU	group	PHQ-9	32	30	8	12	4	4
33	Corpas (2022)	TAU	group	PHQ-9	47	42	10	12	4	3

Anwendungsbeispiel

n_ Gruppengröße, m_ Mittelwert, sd_ Standardabweichung für _t Treatment, _c Control

id	reference	control	setting	outcome	n_t	n_c	m_t	m_c	sd_t	sd_c
34	Diaz (2021)	WL	internet	BDI-II	45	55	12	26	9.0	14.0
34	Diaz (2021)	WL	internet	BDI-II	46	55	13	26	10.8	14.0
35	Fernández (2022)	WL	group	BDI-IA	28	26	9	14	7.0	6.6
35	Fernández (2022)	other	group	BDI-IA	28	24	9	7	7.0	4.6
35	Fernández (2022)	other	group	BDI-IA	28	21	9	9	7.0	5.0
36	González (2020)	TAU	internet	BDI-II	63	67	16	20	10.9	12.8
37	Kananian (2020)	WL	group	PHQ-9	11	12	9	15	3.7	3.6
38	Khakpoor (2019)	WL	individual	BDI-II	12	12	8	22	12.2	11.3
39	Kladnitski (2020)	TAU	internet	PHQ-9	30	33	8	12	5.0	5.0
39	Kladnitski (2020)	other	internet	PHQ-9	30	25	8	8	5.0	4.9
39	Kladnitski (2020)	TAU	internet	PHQ-9	27	33	6	12	4.9	5.0
39	Kladnitski (2020)	other	internet	PHQ-9	27	25	6	8	4.9	4.9
40	Kunst (2022)	other	group	BDI-II	50	51	11	13	9.0	10.1
41	Mahmoodi (2020)	WL	individual	BDI-II	22	19	15	24	5.7	7.4
41	Mahmoodi (2020)	other	individual	BDI-II	22	23	15	16	5.7	5.8
42	Nasiri (2020)	WL	individual	BDI-II	15	15	10	28	7.1	9.2
42	Nasiri (2020)	WL	individual	BDI-II	13	15	8	28	5.2	9.2
43	Nazari (2020b)	TAU	group	HADS	30	28	8	13	2.4	1.7
44	Nazari (2020a)	TAU	group	HADS	28	23	8	13	2.8	2.0
45	Osma (2022b)	DS	group	BDI-II	182	151	18	22	13.1	12.6
46	Osma (2022a)	TAU	group	BDI-II	87	81	18	25	12.2	13.5
48	Reinholt (2022)	DS	group	SCL	96	94	3	3	0.9	0.9
49	Roberge (2020)	TAU	group	PHQ-9	91	98	6	9	4.6	5.2
50	Rogiers (2022)	WL	group	BDI-II	42	30	13	17	10.4	9.3
51	Schaeuffele (2022)	WL	internet	PHQ-9	42	57	8	13	4.8	5.0
52	Tully (2022)	other	individual	PHQ-9	9	9	4	6	6.0	8.0
53	Gutner (2022)	TAU	individual	PHQ-9	13	11	9	12	6.0	5.5
53	Gutner (2022)	other	individual	PHQ-9	13	11	9	12	6.0	5.4
54	Yan (2022)	TAU	internet	ODSIS	40	31	8	12	3.0	2.2
58	Li (2023)	TAU	internet	BDI-II	29	36	14	20	11.2	13.7

Anwendungsbeispiel

Hedges g , asymptotischer Varianzschätzer and asymptotisches 95%-Konfidenzintervall

```
D      = read.csv("./3_Daten/TD-CBT.csv")      # Einlesen des Datensatzes
k      = nrow(D)                              # Studienanzahl
delta  = 0.05                                 # Konfidenzlevel
z_delta = qnorm((1 + delta)/2)               # Konfidenzintervallkonstante
D$hgi  = rep(NA,n,k)                          # Hedges' g
D$vgi  = rep(NA,n,k)                          # asymptotische Varianz
D$ci.lb = rep(NA,n,k)                        # untere CI Grenze
D$ci.ub = rep(NA,n,k)                       # obere CI Grenze
for(i in 1:k){
  n_t   = D$n_t[i]                            # Studieniterationen
  n_c   = D$n_c[i]                            # Umfang Treatmentgruppe
  bar_y_t = D$m_t[i]                          # Umfang Kontrollgruppe
  bar_y_c = D$m_c[i]                          # Mittelwert der Treatmentgruppe
  s2_t   = D$sd_t[i]**2                       # Mittelwert der Kontrollgruppe
  s2_c   = D$sd_c[i]**2                       # empirische Varianz der Treatmentgruppe
  s      = sqrt(((n_t-1)*s2_t + (n_c-1)*s2_c)/(n_t+n_c-2)) # empirische Varianz der Kontrollgruppe
  es     = bar_y_c - bar_y_t                  # gepoolte Standardabweichung
  d      = (bar_y_c - bar_y_t)/s              # Effektstärke
  m      = n_t + n_c - 2                      # Cohen's d
  c_m    = (gamma(m/2))/(sqrt(m/2)*gamma((m-1)/2)) # Freiheitsgradparameter
  D$hgi[i] = c_m*d                            # Biaskorrekturfaktor
  D$vgi[i] = 1/n_t + 1/n_c + (D$hgi[i]^2)/(2*(n_t+n_c)) # Hedges' g
  D$ci.lb[i] = D$hgi[i] - z_delta*sqrt(D$vgi[i]) # Geschätzte asymptotische Varianz
  D$ci.ub[i] = D$hgi[i] + z_delta*sqrt(D$vgi[i]) # untere asymptotische KI Grenze
  # obere asymptotische KI Grenze
}
write.csv(D, "./3_Daten/TD-CBT-GI.csv", row.names = FALSE) # Speichern
```

Hedges' g und asymptotischer Varianzschätzer mit metafor

```
library(metafor)
D      = read.csv("./3_Daten/TD-CBT-GI.csv")           # Einlesen des Datensatzes
D      = escalc()                                     # escalc() aus metafor
data   = D,                                           # Datensatz
measure = "SMD",                                     # Hedges g
m1i    = m_c,                                        # Mittelwerte Treatmentgruppe
sd1i   = sd_c,                                       # Standardabweichungen Treatmentgruppe
m2i    = m_t,                                        # Mittelwerte Kontrollgruppe
sd2i   = sd_t,                                       # Standardabweichungen Kontrollgruppe
n1i    = n_c,                                        # Umfang Treatmentgruppe
n2i    = n_t,                                        # Umfang Kontrollgruppe
slab   = id,                                         # Studien ID
write.csv(D, "./3_Daten/TD-CBT-GI.csv", row.names = FALSE) # Speichern
```

Anwendungsbeispiel

Hedges' g , asymptotischer Varianzschätzer und asymptotisches 95%-Konfidenzintervall: hgi, vgi, ci.lb, ci.ub

	reference	hgi	vgi	ci.lb	ci.ub
1	Barlow et al. (2017)	0.44777333	0.04766813	0.43408254	0.46146412
2	Barlow et al. (2017)	0.03980061	0.03126382	0.02871306	0.05088815
3	Berger et al. (2017)	0.56413341	0.03474290	0.55244521	0.57582160
4	de Ornelas Maia et al. (2017)	0.44134228	0.08536232	0.42302135	0.45966322
5	Dear et al. (2015)	0.37763087	0.01443796	0.37009615	0.38516560
6	Dear et al. (2016)	0.08734206	0.02330632	0.07776899	0.09691513
7	Ejeby et al. (2014)	0.54383594	0.03166498	0.53267749	0.55499440
8	Ejeby et al. (2014)	0.29770488	0.02997252	0.28684872	0.30856103
9	Ellard et al. (2017)	1.12500710	0.26015669	1.09302316	1.15699104
10	Farchione et al. (2012)	1.09396134	0.16415379	1.06855513	1.11936756
11	Fogliati et al. (2016)	0.02448452	0.03286884	0.01311593	0.03585311
12	Gros & Allan (2019)	0.39731311	0.08368038	0.37917357	0.41545266
13	Johnston et al. (2011)	0.83737896	0.03960971	0.82489894	0.84985898
14	Lotfi et al. (2014)	-0.30485836	0.22480385	-0.33458983	-0.27512689
15	Mullin et al. (2015)	0.84216853	0.10368152	0.82197719	0.86235987
16	Neacsiu et al. (2014)	0.91959924	0.19539575	0.89188059	0.94731789
17	Newby et al. (2013)	1.00167975	0.04734958	0.98803478	1.01532472
18	Norton & Barrera (2012)	0.49271022	0.15016839	0.46841036	0.51701008
19	Norton & Hope (2005)	4.97163638	1.10168339	4.90581865	5.03745411
20	Schmidt et al. (2012)	0.37350274	0.04526713	0.36016120	0.38684428
21	Titov et al. (2010)	0.48894292	0.05889715	0.47372478	0.50416107
22	Titov et al. (2011)	0.56822095	0.06032287	0.55281972	0.58362218
23	Titov et al. (2015)	-0.20254139	0.01552420	-0.21035442	-0.19472837
24	Tulbure et al. (2018)	0.68492930	0.05515392	0.67020269	0.69965590
25	Zemestani et al. (2017)	5.74821096	0.47768674	5.70487124	5.79155067
26	Ahmadi et al. (2021)	1.44410631	0.19395083	1.41649034	1.47172228
27	Ahmadi et al. (2021)	1.73672104	0.20420720	1.70838430	1.76505779
28	Corpas et al. (2021)	1.05514936	0.07864538	1.03756401	1.07273471
29	Corpas et al. (2021)	0.97379716	0.07223076	0.95694422	0.99065009
30	Corpas et al. (2022)	0.61192636	0.04718979	0.59830443	0.62554829

Anwendungsbeispiel

Forestplot mit metafor

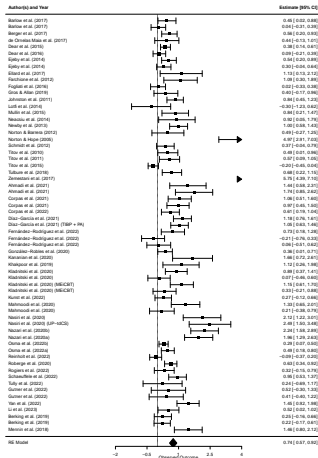
```
# figure setup
library(latex2exp)
library(metafor)
pdf(file = "./3_Abbildungen/eva_3_forest.pdf", width = 12, height = 18)
par(
  family      = "sans",
  mfcol       = c(1,1),
  pty         = "m",
  bty         = "l",
  lwd         = 1,
  las         = 1,
  mgp         = c(2,1,0),
  xaxs        = "i",
  yaxs        = "i",
  font.main   = 1)

# forest plot
D      = read.csv("./3_Daten/TD-CBT-GI.csv")           # Einlesen des Datensatzes
res    = rma(yi=yi, vi=vi, data = D)                 # Schätzen des Random Effects Modells
forest(
  res,
  slab  = paste(D$reference),                        # metafor forest() Funktion
  alim  = c(-2,4),                                  # Random-effect-model object
  header = paste0("Author(s) and Year"),            # Studienreferenzen
  dev.off())                                         # Hedges g Spanne
                                                    # header for study label and SMD
```

Anwendungsbeispiel

Forestplot mit metafor

- Die 95%-Konfidenzintervalle entsprechen nicht denen nach Hedges & Olkin (1985) (S. 86) vorgeschlagenen.



Motivation

Standardisierte Effektstärkeschätzung in Einzelstudien

Anwendungsbeispiel

Selbstkontrollfragen

Selbstkontrollfragen

1. Geben Sie die Definition von Cohen's d im Treatment- und Kontrollgruppen-Design wieder.
2. Geben Sie die Definition des Einzelstudienmodells (ESMs) wieder.
3. Geben Sie die Definition des Effektstärkeschätzers im ESM wieder.
4. Geben Sie Aussage (1) und (2) des Theorems zu den Eigenschaften des Effektstärkeschätzers im ESM wieder.
5. Geben Sie das Theorem zum Unverzerrten Effektstärkeschätzer im ESM - Hedges' g wieder.
6. Geben Sie das Theorem zur Asymptotischen Normalverteilung von Hedges' g wieder.

Definition (Nichtzentrale t -Zufallsvariable)

T sei eine Zufallsvariable mit Ergebnisraum \mathbb{R} und WDF

$$p : \mathbb{R} \rightarrow \mathbb{R}_{>0}, t \mapsto p(t) := \frac{1}{2^{\frac{\nu-1}{2}} \Gamma(\frac{\nu}{2}) (\nu\pi)^{\frac{1}{2}}} \int_0^\infty \tau^{\frac{\nu-1}{2}} \exp\left(-\frac{\tau}{2}\right) \exp\left(-\frac{1}{2} \left(t \left(\frac{\tau}{\nu}\right)^{\frac{1}{2}} - \mu\right)^2\right) d\tau. \quad (30)$$

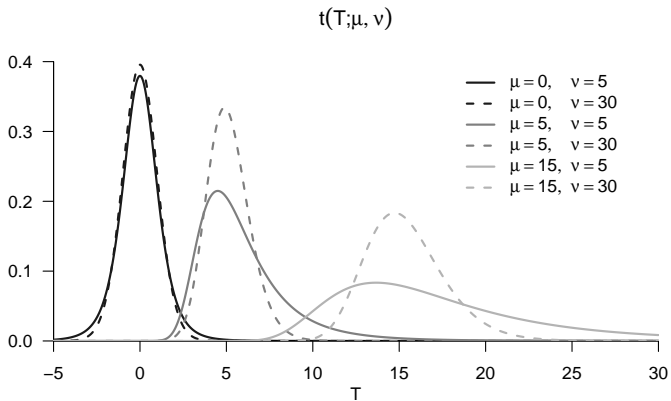
Dann sagen wir, dass T einer nichtzentralen t -Verteilung mit Nichtzentralitätsparameter μ und Freiheitsgradparameter ν unterliegt und nennen T eine *nichtzentrale t -Zufallsvariable mit Nichtzentralitätsparameter μ und Freiheitsgradparameter ν* . Wir kürzen dies mit $T \sim t(\mu, \nu)$ ab. Erwartungswert und Varianz einer nichtzentral t -verteilten Zufallsvariable T mit Nichtzentralitätsparameter μ und Freiheitsgradparameter $\nu > 1$ ergeben sich zu

$$\mathbb{E}(T) = \mu \sqrt{\frac{\nu}{2}} \frac{\Gamma((\nu-1)/2)}{\Gamma(\nu/2)} \quad \text{und} \quad \mathbb{V}(T) = \frac{\nu(1+\mu^2)}{\nu-2} - \frac{\mu^2\nu}{2} \left(\frac{\Gamma((\nu-1)/2)}{\Gamma(\nu/2)}\right)^2. \quad (31)$$

Bemerkung

- Nichtzentrale t -Zufallsvariablen sind für die Testgütefunktion des T-Tests essentiell.
- Für die Herleitung von Erwartungswert und Varianz verweisen wir auf Hogben et al. (1961).

Wahrscheinlichkeitsdichtefunktionen spezieller nichtzentralen t -Zufallsvariablen.



Referenzen

- Borenstein, M. (Ed.). (2009). *Introduction to meta-analysis*. John Wiley & Sons.
- Cuijpers, P., Karyotaki, E., Reijnders, M., & Ebert, D. D. (2019). Was Eysenck right after all? A reassessment of the effects of psychotherapy for adult depression. *Epidemiology and Psychiatric Sciences*, *28*(1), 21–30. <https://doi.org/10.1017/S2045796018000057>
- Hedges, L. V. (1981). *Distribution Theory for Glass's Estimator of Effect Size and Related Estimators*. 23.
- Hedges, L. V., & Olkin, I. (1985). *Statistical Methods for Meta-Analysis*. Academic Press.
- Hogben, D., Pinkham, R. S., & Wilk, M. B. (1961). *The Moments of the Non-Central t-Distribution*. 5.
- Johnson, N. L., & Welch, B. L. (1940). Applications of the Non-Central t-Distribution. *Biometrika*, *31*(3/4), 362. <https://doi.org/10.2307/2332616>
- Lin, L., & Aloe, A. M. (2021). Evaluation of various estimators for standardized mean difference in meta-analysis. *Statistics in Medicine*, *40*(2), 403–426. <https://doi.org/10.1002/sim.8781>
- Munder, T., Flückiger, C., Leichenring, F., Abbass, A. A., Hilsenroth, M. J., Luyten, P., Rabung, S., Steinert, C., & Wampold, B. E. (2019). Is psychotherapy effective? A re-analysis of treatments for depression. *Epidemiology and Psychiatric Sciences*, *28*(03), 268–274. <https://doi.org/10.1017/S2045796018000355>
- Schaeuffele, C., Meine, L. E., Schulz, A., Weber, M. C., Moser, A., Paersch, C., Recher, D., Boettcher, J., Renneberg, B., Flückiger, C., & Kleim, B. (2024). A systematic review and meta-analysis of transdiagnostic cognitive behavioural therapies for emotional disorders. *Nature Human Behaviour*, *8*(3), 493–509. <https://doi.org/10.1038/s41562-023-01787-3>