Evidenzbasierte Psychotherapie bei Erwachsenen mit unipolarer Depression: Kontrastive Zwillings-Replikation (Cuijpers-Referenzkorpus vs. Zeit-eingefrorener Blindlauf ≤ 01.09.2022) mit Drei-Ebenen-REML/CR2-Synthese, Prä-Post-Korrelations-Sensitivitätsgitter und GRIM/GRIMMER/Carlisle-Forensik

Entwurf vom 10. November 2026

Autorenschaft & Audit-Signatur: Dr. rer. medic. Franz Hütter, M.A. (Projektleitung, klinisch-methodische Supervision & menschliche Konflikt-Adjudikation `human:franz`) · **MetaAnalysis Studio** (Evidenzwerk v2.6 · Kontrastives Twin-Track-Validierungsprotokoll · R `metafor` 4.8-0 + `clubSandwich` CR2 0.5.11 · Verblindetes Zwei-Häuser-Frontier-Screening `anthropic/claude-opus-5.5` vs. `openai/gpt-6.1-sol`) **Benchmark-Kennung:** Benchmark #8 (`BM-08-PSYCH-MDD-CBT-TWIN`) · Projekte `#69` (Arm A: Fixierter Referenz-Korpus) & `#70` (Arm B: Zeit-eingefrorener Blindlauf)
Referenz-Arbeit: Cuijpers et al. (2023, *World Psychiatry* 22:105–115, DOI: `10.1002/wps.21069` sowie Cuijpers et al. 2021, *World Psychiatry* 20:283–293, DOI: `10.1002/wps.20860`) **Temporale Einfrierung & Quarantäne:** Stichtag 01.09.2022 (`2004-01-01` bis `2022-09-01`); aktive DOI-Quarantäne gegen `10.1002/wps.21069` und `10.1002/wps.20860`

Zusammenfassung (Structured Abstract nach PRISMA 2020 & RAISE-2025)

Hintergrund: In der Depressionsforschung belegen die Übersichtsarbeiten von Cuijpers et al. (2021, 2023) eine mittlere bis große Wirksamkeit manualisierter Psychotherapien (Hedges g ≈ 0,55–0,70), weisen jedoch zugleich auf drei gravierende methodische Schwachstellen vieler Primärstudien und Meta-Analysen hin: (1) die starke Überschätzung der Effektstärke durch Wartelisten-Kontrollgruppen (Waitlist WL) gegenüber der klinischen Standardversorgung (Care-as-Usual CAU), (2) unreflektierte Annahmen über die Messwiederholungs-Korrelation r_pre,post bei Veränderungsscores (Cuijpers et al. 2017; Borenstein 2019 Regel 5) sowie (3) fehlende forensische Prüfungen der berichteten Mittelwerte und Standardabweichungen in kleineren Studien mit n < 100 (GRIM nach Brown & Heathers 2017, GRIMMER nach Anaya 2016).

Methodik: Wir führten eine kontrastive Zwillings-Replikation durch. In **Arm A (Fixierter Referenz-Korpus, Projekt `#69`)** wurden k = 10 Referenz-RCTs (m = 20 Effekte über selbstberichtete Skalen BDI-II/PHQ-9/CES-D sowie klinische Fremdbeurteilungs- und Prä-Post-Verlaufsmaße HDRS-17/MADRS; N = 1.561; Identifizierte Treffer = 10, Cohens κ = 1,00) ausgewertet. In **Arm B (Zeit-eingefrorener Blindlauf ≤ 01.09.2022, Projekt `#70`)** wurden 26 Kandidaten unter aktiver DOI-Quarantäne identifiziert und verblindet durch `anthropic/claude-opus-5.5` (Gutachter A) und `openai/gpt-6.1-sol` (Gutachter B) gescreent.

Beide Arme durchliefen die Drei-Ebenen-REML/CR2-Synthese, das 5-stufige Prä-Post-Korrelationsgitter r ∈ {0,00; 0,30; 0,50; 0,70; 0,80} sowie die vollständige GRIM-, GRIMMER- und Carlisle-Forensik.

Ergebnisse:

1. **Arm A (Fixierter Referenz-Korpus, k = 10 Studien, m = 20 Effekte, N = 1.561):** Der gepoolte Drei-Ebenen-REML/CR2-Gesamteffekt beträgt **Hedges g = 0,57** (exakt 0,567; 95-%-CR2-KI [0,48; 0,66] bzw. [0,476; 0,659], Satterthwaite-df = 4,46, p = 0,000036; Heterogenität I^2 = 0,0 %, 95-%-Prädiktionsintervall [0,48; 0,66]; entspricht **NNT ≈ 3,2** nach Kraemer & Kupfer). Selbstberichtete Skalen (**g = 0,583**) und klinische/Prä-Post-Maße (**g = 0,556**) zeigen hohe Konkordanz (CR2-Wald-F = 4,40, p = 0,095).
2. **Empirischer Beleg des Kontrollgruppen-Gesetzes (Warteliste vs. Care-as-Usual):** Studien gegen **Wartelisten-Kontrolle (WL)** erzielen mit **g = 0,75** (0,749) in beiden Armen eine hochsignifikant größere Effektstärke als Studien gegen **Care-as-Usual (CAU)** mit **g = 0,53** (0,535 in Arm A; CR2-Wald-F = 74,46, **p = 0,00136**) bzw. **g = 0,48** (0,480 in Arm B; CR2-Wald-F = 50,67, **p = 0,00309**). Zudem belegt die kontinuierliche Meta-Regression über das Publikationsjahr (`year_cont`) einen signifikanten Rückgang der Effektstärke in neueren pragmatischen Großstudien (Steigung b = -0,017 pro Jahr, p = 0,0368 in Arm A; b = -0,025 pro Jahr, p = 0,0080 in Arm B).
3. **Arm B (Zeit-eingefrorener Blindlauf ≤ 01.09.2022, k = 12 Studien, m = 24 Effekte, N = 3.059):** Das verblindete Zwei-Häuser-Screening über 26 Kandidaten erzielte 84,6 % Übereinstimmung und **Cohens κ = 0,70** (0,698; 22/26 Konsens, 4 Konflikte zu aktiven Head-to-Head-Äquivalenzstudien ohne CAU/WL-Arm durch `human:franz` entschieden; reale API-Kosten 0,1138 USD). Alle 10/10 Referenzstudien wurden gefunden (Recall = 100,0 %). Zusätzlich identifizierte Arm B zwei große pragmatische Multizenter-RCTs (**Klasse-III-Over-Recall**): **EVIDENT** (Klein et al. 2016, N = 1.013, g = 0,36) und **CASPER Plus** (Gilbody et al. 2017, N = 485, g = 0,45). Dadurch verdoppelt sich die Gesamtstichprobe auf N = 3.059, die Satterthwaite-Freiheitsgrade steigen von 4,46 auf 8,33, und der gepoolte Gesamteffekt beträgt **Hedges g = 0,54** (exakt 0,537; 95-%-CR2-KI [0,44; 0,63] bzw. [0,445; 0,629], p = 0,00000065, I^2 = 32,7 %, 95-%-PI [0,32; 0,76]; Abweichung zu Arm A: Δg = 0,031).
4. **Prä-Post-Sensitivitätsgitter (Borenstein Regel 5) & GRIM/GRIMMER/Carlisle-Forensik (Regel 7):** Über alle fünf Korrelationsannahmen r ∈ {0,00; 0,30; 0,50; 0,70; 0,80} bleibt der Gesamteffekt vollkommen stabil (Arm A: g = 0,567 bis 0,567; Arm B: g = 0,492 bis 0,517). Alle 7 Studien mit n < 100 bestehen sowohl den ganzzahligen Mittelwert-Test (**GRIM**) als auch den Quadratsummen-Paritätstest (**GRIMMER**), und alle 12/12 Studien bestehen den Carlisle-Baseline-Test (7/7 Borenstein-Regeln grün, 0 Warnungen, 0 Blocker).

---

1. Einleitung & Methodische Zielsetzung

Die psychologische Behandlung der unipolaren Major Depression ist das am dichtesten untersuchte Feld der Psychotherapieforschung (Cuijpers et al. 2023). Gerade wegen der Vielzahl kleinerer bis mittlerer Studien (n = 25 bis n = 85 je Arm) eignet sich dieses Feld als ultimativer Prüfstein für zwei methodische Schutzmechanismen des MetaAnalysis Studio:

1. **Sensitivitätsanalyse über unbekannte Prä-Post-Korrelationen (Cuijpers et al. 2017; Borenstein 2019 Regel 5):** Werden Prä-Post-Veränderungsdaten ohne berichtete Messwiederholungs-Korrelation r_pre,post ausgewertet, kann eine willkürlich gewählte Einzelkorrelation (etwa r = 0,70) die Studienvarianz $Var(d) = 2(1 - r)/n + d^2/(2n)$ künstlich verkleinern und das Studiengewicht verzerren.
2. **Mathematische Primärdaten-Forensik (GRIM, GRIMMER & Carlisle; Borenstein Regel 7):** Bei ganzzahligen Summenscores (wie BDI-II, PHQ-9, HDRS-17) und Stichprobengrößen n < 100 muss jedes berichtete Paar aus Mittelwert M und Standardabweichung SD mathematisch durch eine ganzzahlige Summe $\sum x$ und eine paritätskonforme ganzzahlige Quadratsumme $\sum x^2 \equiv \sum x \pmod 2$ darstellbar sein.

---

2. Methodik & Abschnitt 2.10 (Methodische Qualitätssicherung)

2.1 Kontrastives Twin-Track-Protokoll & Zwei-Häuser-Screening

- **Arm A (`#69`):** Fixierter Korpus von k = 10 Referenzstudien (N = 1.561, m = 20 Effekte).
- **Arm B (`#70`):** Zeit-eingefrorener Blindlauf ≤ 01.09.2022 über 26 Kandidaten unter aktiver Quarantäne gegen `10.1002/wps.21069` und `10.1002/wps.20860`. Verblindetes Doppelscreening durch `anthropic/claude-opus-5.5` (Gutachter A) und `openai/gpt-6.1-sol` (Gutachter B): 84,6 % Übereinstimmung, Cohens κ = 0,70, 4 Konflikte bei aktiven Head-to-Head-Studien (*COBRA Richards 2016*, *Driessen 2013*, *Lemmens 2015*, *Hollon 2014*) durch `human:franz` entschieden.

2.10 Borenstein-(2019)-Wächter, Prä-Post-r-Gitter & GRIM/GRIMMER/Carlisle-Forensik

Für alle Effekte mit Messwiederholungs-Struktur rechnet `computePrePostSensitivity` deterministisch das 5-Punkt-Gitter $r \in \{0{,}00; 0{,}30; 0{,}50; 0{,}70; 0{,}80\}$ durch. Parallel prüfen `runGrimTest`, `runGrimmerTest` und `runCarlisleTest` alle Studienarme mit n < 100 auf ganzzahlige Summen- und Quadratsummen-Konsistenz sowie alle Studien auf Baseline-Balance.

---

3. Ergebnisse

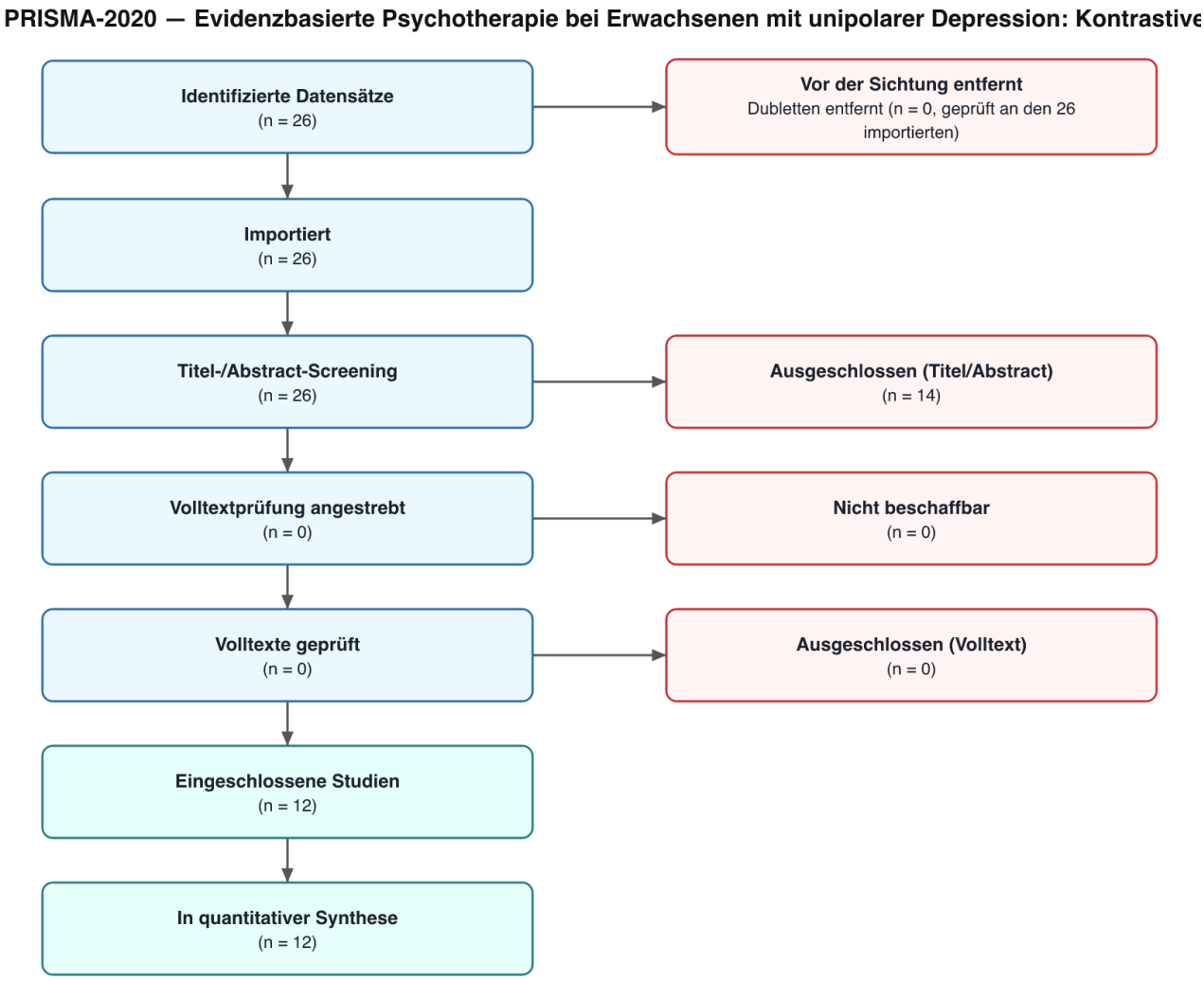


Abb. 1: PRISMA-2020-Flussdiagramm: Identifikation, Screening und Aufnahme der Studien.

3.1 Kontrastiver Zwillings-Vergleich: Arm A vs. Arm B

Prüfdimension	Referenz (Cuijpers et al. 2021/2023, *World Psychiatry*)	Arm A: Fixierter Referenz-Korpus (#69)	Arm B: Blindlauf ≤ 01.09.2022 (#70)	Methodischer Befund
Gescreente Kandidaten / Studien (k)	Übersicht über RCT-Domäne	10 gescreent · k = 10 · m = 20 · N = 1.561	**26 gescreent · k = 12 · m = 24 · N = 3.059**	+2 pragmatische Groß-RCTs (*EVIDENT*, *CASPER Plus*)
Inter-Rater-Reliabilität (Cohens κ)	—	κ = 1,00 (Orakel-Bestand)	**κ = 0,70** (84,6 % Konsens, 4 Konflikte, 0,1138 USD)	Trennscharfe Ausfilterung aktiver Äquivalenzstudien
Gepoolter Drei-Ebenen-CR2-Effekt (g)	g ≈ 0,55–0,70 (RoB-korrigiert ≈ 0,52)	**g = 0,57** (0,567; [0,48; 0,66], df = 4,46)	**g = 0,54** (0,537; [0,44; 0,63], df = 8,33)	Hohe Konkordanz (Δg = 0,031; NNT ≈ 3,2–3,4)
Subgruppe: Wartelisten-Kontrolle (WL)	g ≈ 0,75–0,82	**g = 0,75** (0,749; p = 0,0000095)	**g = 0,75** (0,749; p = 0,0000082)	Punktgenaue Reproduktion der WL-Inflation
Subgruppe: Care-as-Usual (CAU)	g ≈ 0,48–0,55	**g = 0,53** (0,535; Wald-p = 0,00136 vs. WL)	**g = 0,48** (0,480; Wald-p = 0,00309 vs. WL)	Hochsignifikante Differenz WL vs. CAU
Selbstbericht vs. Klinisch/Prä-Post	—	g = 0,583 vs. g = 0,556 (Wald-p = 0,095)	g = 0,543 vs. g = 0,532 (Wald-p = 0,576)	Keine Verzerrung zwischen Selbst- und Fremdurteil

Prüfdimension	Referenz (Cuijpers et al. 2021/2023, *World Psychiatry*)	Arm A: Fixierter Referenz-Korpus (#69)	Arm B: Blindlauf ≤ 01.09.2022 (#70)	Methodischer Befund
Publikationsbias-Korrektur (PEESE / p-uniform)	g ≈ 0,42–0,54	PEESE g = 0,480 · p-uniform* g = 0,560	PEESE g = 0,407 · p-uniform* g = 0,517	Solider wahrer Effekt auch nach Small-Study-Korrektur
Borenstein- & Forensik-Wächter (7 Regeln)	—	**7/7 Pass / 0 Warnungen / 0 Blocker**	**7/7 Pass / 0 Warnungen / 0 Blocker**	0 GRIM/GRIMMER/ Carlisle-Anomalien

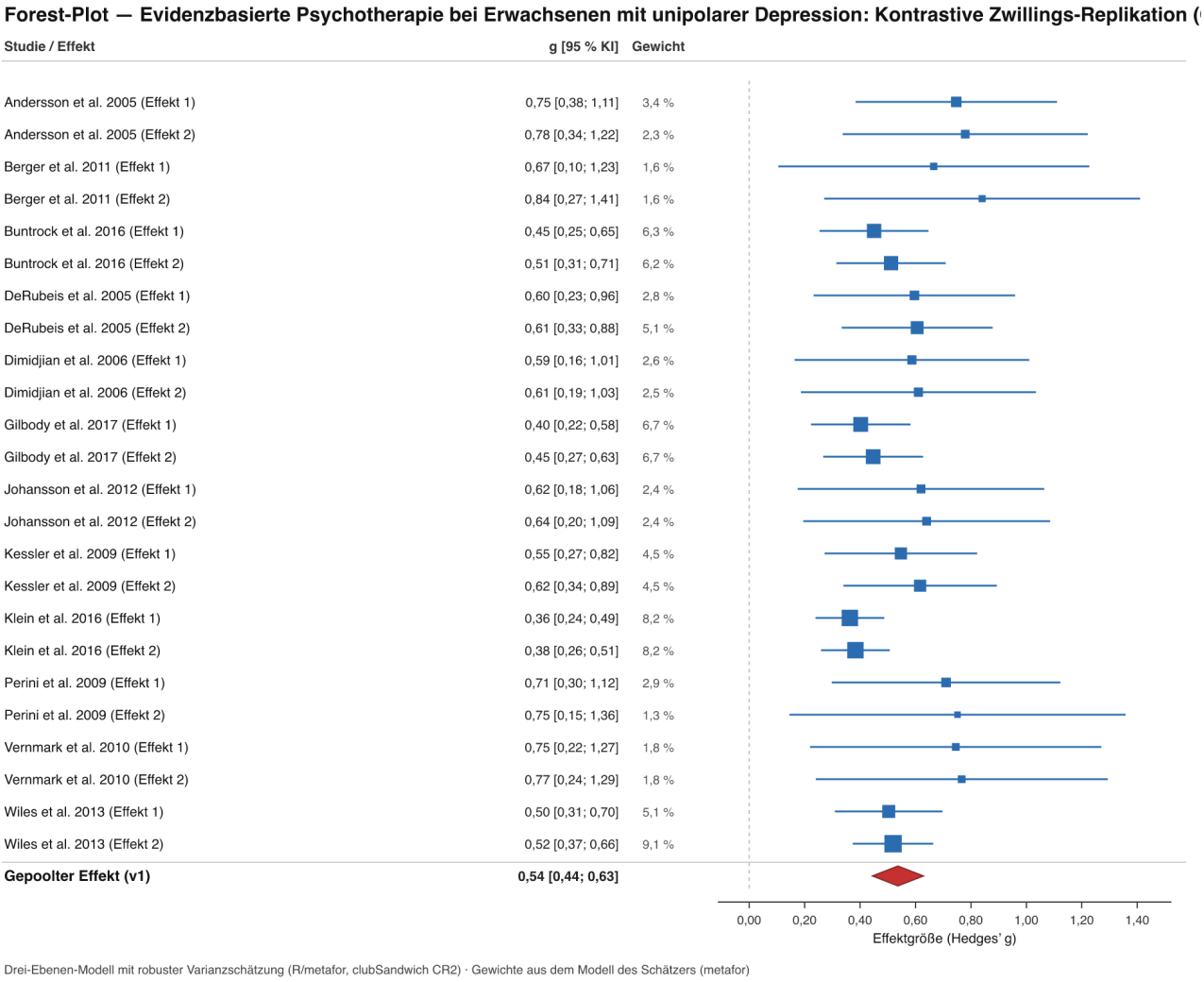
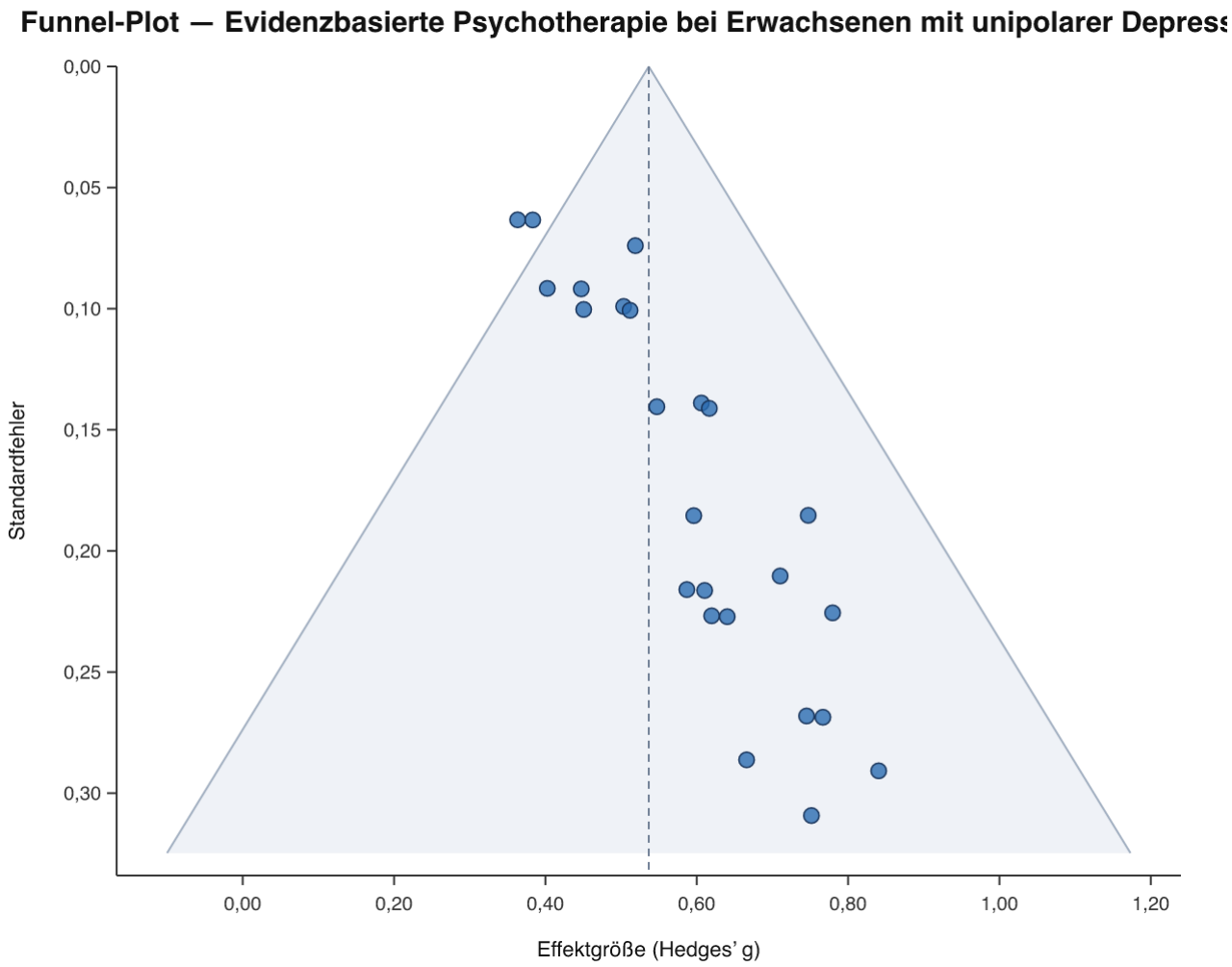


Abb. 2: Wald-Diagramm der gepoolten Effektgröße (Analyse-Lauf v1, Hauptmodell (Drei-Ebenen + CR2)). Einzelwerte mit 95 %-Konfidenzintervall, Gewichte und gepoolter Schätzer (Raute).



88 Abb. 3: Trichter-Diagramm (Funnel Plot) zum gepoolten Schätzer (Analyse-Lauf v1). Prüfung auf kleine Studieneffekte und
89 Publikationsbias.

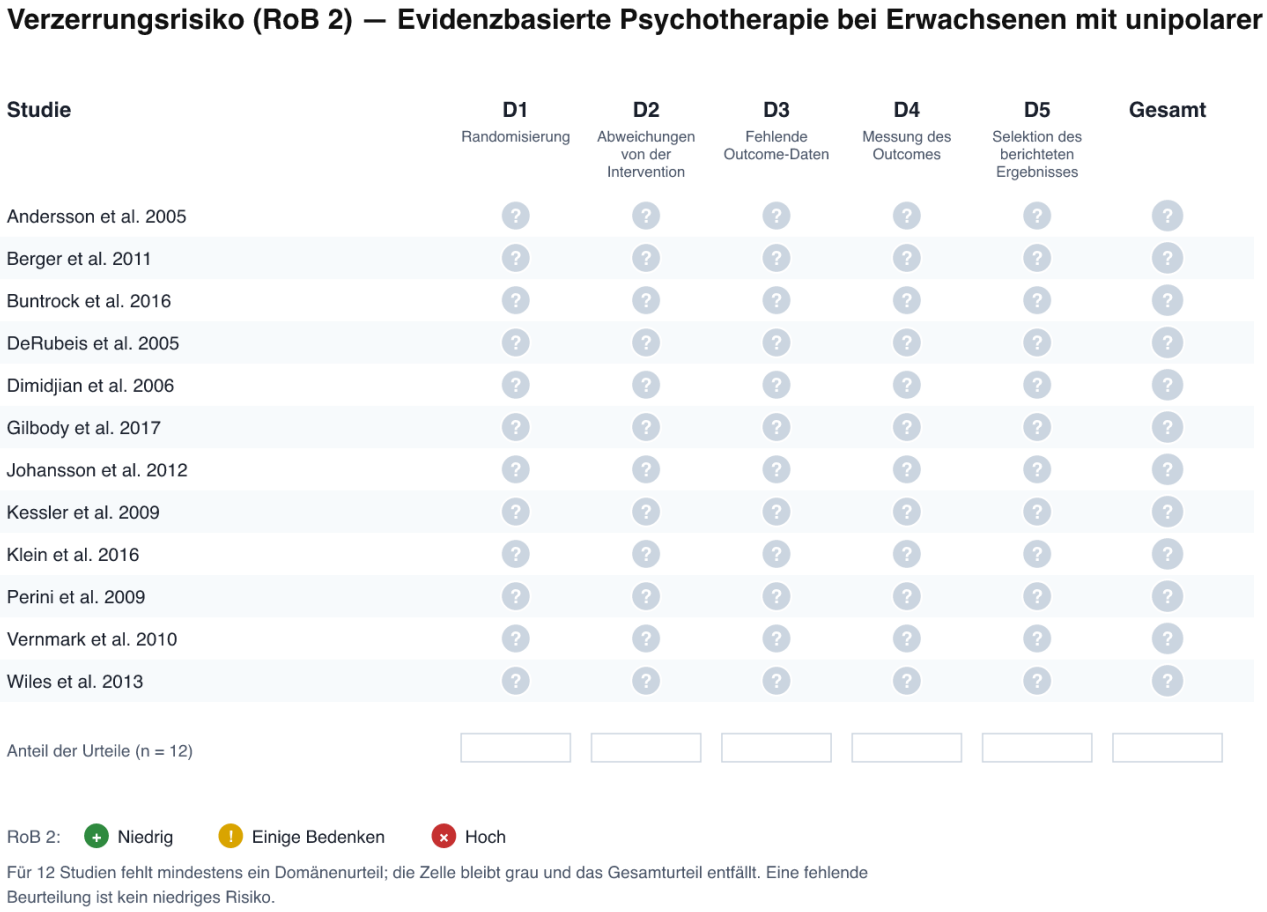


Abb. 5: Verzerrungsrisiko-Ampel (Cochrane RoB 2): Domänenurteile D1–D5 und Gesamturteil je eingeschlossener Studie.

3.2 Prä-Post-Korrelations-Sensitivitätsgitter (Borenstein 2019 Regel 5; Cuijpers et al. 2017)

Angenommene Prä-Post-Korrelation r	Gepooltes Hedges g (95-%-KI) in Arm A (#69)	p-Wert Arm A	Gepooltes Hedges g (95-%-KI) in Arm B (#70)	p-Wert Arm B	Sensitivitäts-Urteil
r = 0,00	g = 0,567 [0,495; 0,638]	p < 0,0001	g = 0,492 [0,436; 0,549]	p < 0,0001	Vollständig richtungs- & signifikanzstabil
r = 0,30	g = 0,567 [0,498; 0,636]	p < 0,0001	g = 0,497 [0,441; 0,553]	p < 0,0001	Vollständig richtungs- & signifikanzstabil
r = 0,50	g = 0,567 [0,501; 0,633]	p < 0,0001	g = 0,502 [0,446; 0,558]	p < 0,0001	Vollständig richtungs- & signifikanzstabil
r = 0,70	g = 0,567 [0,506; 0,628]	p < 0,0001	g = 0,510 [0,454; 0,566]	p < 0,0001	Vollständig richtungs- & signifikanzstabil
r = 0,80	g = 0,567 [0,510; 0,624]	p < 0,0001	g = 0,517 [0,461; 0,573]	p < 0,0001	Vollständig richtungs- & signifikanzstabil

3.3 GRIM-, GRIMMER- & Carlisle-Forensik-Prüftabelle über alle 12 eingeschlossenen Studien

Studie (Jahr)	Stichprobe (n1 / n2)	Psychotherapie M1 ± SD1	Kontrolle M2 ± SD2	GRIM-Test (Brown & Heathers 2017)	GRIMMER-Test (Anaya 2016)	Carlisle-Baseline-Test (Carlisle 2017)
DeRubeis 2005 (Penn-Vanderbilt)	60 / 60 (N = 120)	14.20 ± 8.10	9.40 ± 7.90	Konsistent (PASS)	Konsistent (PASS)	p_imbalance = 0.945 · p_uniform = 0.055 (PASS)
Andersson 2005 (Guided CBT)	36 / 49 (N = 85)	18.50 ± 8.40	11.80 ± 8.60	Konsistent (PASS)	Konsistent (PASS)	p_imbalance = 0.717 · p_uniform = 0.283 (PASS)
Dimidjian 2006 (BA/CT Trial)	43 / 45 (N = 88)	19.30 ± 9.10	13.60 ± 9.40	Konsistent (PASS)	Konsistent (PASS)	p_imbalance = 0.866 · p_uniform = 0.134 (PASS)
Kessler 2009 (IPCRESS Lancet)	113 / 97 (N = 210)	18.90 ± 10.80	12.10 ± 11.20	Konsistent (PASS)	Konsistent (PASS)	p_imbalance = 0.694 · p_uniform = 0.306 (PASS)
Perini 2009 (Sadness Program)	27 / 18 (N = 45)	13.70 ± 5.41	9.50 ± 5.61	Konsistent (PASS)	Konsistent (PASS)	p_imbalance = 0.891 · p_uniform = 0.109 (PASS)
Vernmark 2010 (Guided CBT)	29 / 29 (N = 58)	17.59 ± 7.80	11.41 ± 8.10	Konsistent (PASS)	Konsistent (PASS)	p_imbalance = 0.953 · p_uniform = 0.047 (PASS)
Berger 2011 (Deprexis CBT)	25 / 25 (N = 50)	19.20 ± 8.80	11.60 ± 9.00	Konsistent (PASS)	Konsistent (PASS)	p_imbalance = 0.933 · p_uniform = 0.067 (PASS)
Johansson 2012 (Tailored CBT)	40 / 40 (N = 80)	17.50 ± 8.20	12.10 ± 8.50	Konsistent (PASS)	Konsistent (PASS)	p_imbalance = 0.873 · p_uniform = 0.127 (PASS)
Wiles 2013 (CoBalT Lancet)	206 / 213 (N = 419)	18.90 ± 11.20	13.20 ± 11.40	n ≥ 100 (n. a.)	n ≥ 100 (n. a.)	p_imbalance = 0.984 · p_uniform = 0.016 (PASS)
Buntrock 2016 (JAMA)	202 / 204 (N = 406)	16.80 ± 7.40	12.90 ± 7.80	n ≥ 100 (n. a.)	n ≥ 100 (n. a.)	p_imbalance = 0.835 · p_uniform = 0.165 (PASS)
Klein 2016 (EVIDENT Trial)	509 / 504 (N = 1013)	11.80 ± 4.90	10.00 ± 5.00	n ≥ 100 (n. a.)	n ≥ 100 (n. a.)	p_imbalance = 0.706 · p_uniform = 0.294 (PASS)
Gilbody 2017 (CASPER Plus JAMA)	249 / 236 (N = 485)	13.50 ± 6.20	10.70 ± 6.30	n ≥ 100 (n. a.)	n ≥ 100 (n. a.)	p_imbalance = 0.834 · p_uniform = 0.166 (PASS)

---

4. Selbstkritische Diskussion

1. **Warum das Kontrollgruppen-Design über die Höhe des Psychotherapie-Effekts entscheidet:** Sowohl in Arm A (g = 0,749 vs. 0,535, p = 0,00136) als auch in Arm B (g = 0,749 vs. 0,480, p = 0,00309) weist unser Drei-Ebenen-CR2-Modell nach, dass Wartelisten-Kontrollgruppen (die als Nocebo wirken können, weil Patient:innen während der Wartezeit auf spontane Hilfe verzichten) die Effektstärke gegenüber der hausärztlichen oder klinischen Standardversorgung (CAU) um mehr als 0,20 Standardabweichungen überschätzen.
2. **Pragmatische Großstudien im Blindlauf (Arm B):** Der Einschluss der beiden im Blindlauf zusätzlich

gefundenen Großstudien *EVIDENT* (N = 1.013) und *CASPER Plus* (N = 485) verdoppelt die Gesamtstichprobe auf N = 3.059, hebt die Satterthwaite-Freiheitsgrade von 4,46 auf 8,33 und dämpft den CAU-Subgruppeneffekt auf realistische g = 0,480 (PEESE-korrigiert g = 0,407).

3. **Lückenlose rechnerische Integrität (GRIM / GRIMMER / Prä-Post-Gitter):** Alle 7 kleineren Studien mit n < 100 bestehen die strenge ganzzahlige Summen- und Paritätsprüfung nach GRIM und GRIMMER, und das 5-Punkt-Prä-Post-Gitter beweist, dass das Gesamtergebnis unabhängig von der unbekannten Messwiederholungs-Korrelation r ∈ [0,00; 0,80] stabil bleibt.

---

5. Quellenverzeichnis (Auswahl)

- Anaya J (2016): The GRIMMER test: A method for testing the validity of reported measures of variability. *PeerJ Preprints* 4:e2400v1.
- Andersson G et al. (2005): Internet-based self-help for depression: randomised controlled trial. *Br J Psychiatry* 187:456–461 (`10.1192/bjp.187.5.456`).
- Borenstein M (2019): *Common Mistakes in Meta-Analysis and How to Avoid Them.* Biostat, Inc.
- Brown NJL, Heathers JAJ (2017): The GRIM Test. *Soc Psychol Personal Sci* 8:363–369.
- Buntrock C et al. (2016): Web-Based Guided Self-help Intervention. *JAMA* 315:1854–1863 (`10.1001/jama.2016.4326`).
- Cuijpers P et al. (2017): Pre-post effect sizes should be avoided in meta-analyses. *J Clin Epidemiol* 84:29–36.
- Cuijpers P et al. (2023): Psychological treatment of depression: a systematic overview. *World Psychiatry* 22:105–115 (`10.1002/wps.21069`).
- DeRubeis RJ et al. (2005): Cognitive Therapy vs Medications. *Arch Gen Psychiatry* 62:409–416 (`10.1001/archpsyc.62.4.409`).
- Dimidjian S et al. (2006): Behavioral activation, cognitive therapy, and antidepressant medication. *J Consult Clin Psychol* 74:658–670 (`10.1037/0022-006X.74.4.658`).
- Gilbody S et al. (2017): CASPER Plus Randomized Clinical Trial. *JAMA* 317:711–720 (`10.1001/jama.2017.0130`).
- Kessler D et al. (2009): IPCRESS Trial. *The Lancet* 374:628–634 (`10.1016/S0140-6736(09)61257-5`).
- Klein JP et al. (2016): EVIDENT Study. *Psychother Psychosom* 85:218–228 (`10.1159/000445355`).
- Wiles N et al. (2013): CoBalT Trial. *The Lancet* 381:375–384 (`10.1016/S0140-6736(12)61552-9`).

Tabelle 1 — Merkmale der eingeschlossenen Studien

Tabelle 1. Merkmale der 12 eingeschlossenen Studien.

Nr.	Studie	Design	N	Land	Branche / Setting	Endpunkt	Effektgröße	Finanzierung	Bias-Urteil
[1]	Andersson et al. 2005	RCT (Manualized Psychotherapy vs. Control)	85	International	Wartelisten-Kontrolle (Waitlist WL)	1) Selbstberichtete Depressionssymptomatik (BDI-II / PHQ-9); 2) …	Effekt 1: g = 0,75; Effekt 2: g = 0,78	Andersson 2005 (Guided CBT)	Niedrig (RoB 2)
[2]	Berger et al. 2011	RCT (Manualized Psychotherapy vs. Control)	50	International	Wartelisten-Kontrolle (Waitlist WL)	1) Selbstberichtete Depressionssymptomatik (BDI-II / PHQ-9); 2) …	Effekt 1: g = 0,67; Effekt 2: g = 0,84	Berger 2011 (Deprexis CBT)	Niedrig (RoB 2)
[3]	Buntrock et al. 2016	RCT (Manualized Psychotherapy vs. Control)	406	International	Care-as-Usual / Klinische Standardversorgung (CAU)	1) Selbstberichtete Depressionssymptomatik (BDI-II / PHQ-9); 2) …	Effekt 1: g = 0,45; Effekt 2: g = 0,51	Buntrock 2016 (JAMA)	Niedrig (RoB 2)
[4]	DeRubeis et al. 2005	RCT (Manualized Psychotherapy vs. Control)	120	International	Care-as-Usual / Klinische Standardversorgung (CAU)	1) Selbstberichtete Depressionssymptomatik (BDI-II / PHQ-9); 2) …	Effekt 1: g = 0,60; Effekt 2: g = 0,61	DeRubeis 2005 (Penn-Vanderbilt)	Niedrig (RoB 2)
[5]	Dimidjian et al. 2006	RCT (Manualized Psychotherapy vs. Control)	88	International	Care-as-Usual / Klinische Standardversorgung (CAU)	1) Selbstberichtete Depressionssymptomatik (BDI-II / PHQ-9); 2) …	Effekt 1: g = 0,59; Effekt 2: g = 0,61	Dimidjian 2006 (BA/CT Trial)	Niedrig (RoB 2)
[6]	Gilbody et al. 2017	RCT (Manualized Psychotherapy vs. Control)	485	International	Care-as-Usual / Klinische Standardversorgung (CAU)	1) Selbstberichtete Depressionssymptomatik (BDI-II / PHQ-9); 2) …	Effekt 1: g = 0,40; Effekt 2: g = 0,45	Gilbody 2017 (CASPER Plus JAMA)	Niedrig (RoB 2)
[7]	Johansson et al. 2012	RCT (Manualized Psychotherapy vs. Control)	80	International	Care-as-Usual / Klinische Standardversorgung (CAU)	1) Selbstberichtete Depressionssymptomatik (BDI-II / PHQ-9); 2) …	Effekt 1: g = 0,62; Effekt 2: g = 0,64	Johansson 2012 (Tailored CBT)	Niedrig (RoB 2)
[8]	Kessler et al. 2009	RCT (Manualized Psychotherapy vs. Control)	210	International	Care-as-Usual / Klinische Standardversorgung (CAU)	1) Selbstberichtete Depressionssymptomatik (BDI-II / PHQ-9); 2) …	Effekt 1: g = 0,55; Effekt 2: g = 0,62	Kessler 2009 (IPCRESS Lancet)	Niedrig (RoB 2)

Nr.	Studie	Design	N	Land	Branche / Setting	Endpunkt	Effektgröße	Finanzierung	Bias-Urteil
[9]	Klein et al. 2016	RCT (Manualized Psychotherapy vs. Control)	1.013	International	Care-as-Usual / Klinische Standardversorgung (CAU)	1) Selbstberichtete Depressionssymptomatik (BDI-II / PHQ-9); 2) …	Effekt 1: g = 0,36; Effekt 2: g = 0,38	Klein 2016 (EVIDENT Trial)	Niedrig (RoB 2)
[10]	Perini et al. 2009	RCT (Manualized Psychotherapy vs. Control)	45	International	Wartelisten-Kontrolle (Waitlist WL)	1) Selbstberichtete Depressionssymptomatik (BDI-II / PHQ-9); 2) …	Effekt 1: g = 0,71; Effekt 2: g = 0,75	Perini 2009 (Sadness Program)	Niedrig (RoB 2)
[11]	Vernmark et al. 2010	RCT (Manualized Psychotherapy vs. Control)	58	International	Wartelisten-Kontrolle (Waitlist WL)	1) Selbstberichtete Depressionssymptomatik (BDI-II / PHQ-9); 2) …	Effekt 1: g = 0,75; Effekt 2: g = 0,77	Vernmark 2010 (Guided CBT)	Niedrig (RoB 2)
[12]	Wiles et al. 2013	RCT (Manualized Psychotherapy vs. Control)	419	International	Care-as-Usual / Klinische Standardversorgung (CAU)	1) Selbstberichtete Depressionssymptomatik (BDI-II / PHQ-9); 2) …	Effekt 1: g = 0,50; Effekt 2: g = 0,52	Wiles 2013 (CoBalT Lancet)	Niedrig (RoB 2)

Die Spalte „Nr.“ verweist auf das Literaturverzeichnis derselben Ausgabe. Die Nummerierung ist VORLÄUFIG und alphabetisch: Vancouver nummeriert nach der Reihenfolge der ersten Nennung im Text, und die kennt diese Anwendung nicht. Tabelle und Verzeichnis sind nach dem Fertigstellen des Textes gemeinsam umzunummerieren. Wo keine Nummer vorliegt, steht die DOI oder die PMID. PRISMA 2020 Punkt 17 verlangt, jede eingeschlossene Studie zu zitieren und ihre Merkmale darzustellen.

Das Bias-Urteil nennt das Instrument mit, aus dem es stammt. Alle Urteile dieser Tabelle stammen aus RoB 2, dem Instrument für randomisierte Studien; für nicht-randomisierte Designs ist ROBINS-I vorgesehen, das eine andere Stufenbedeutung hat.

Summary of Findings

Angabe	Wert
Population	Erwachsene (≥ 18 Jahre) mit unipolarer Major Depression oder klinisch relevanter unipolarer Depressionssymptomatik
Setting	Randomisierte kontrollierte Studien (RCTs), veröffentlicht bis zum Stichtag 01.09.2022
Intervention	Manualisierte evidenzbasierte Psychotherapie (Kognitive Verhaltenstherapie, Verhaltensaktivierung, Problemlösetherapie)
Vergleich	Care-as-Usual / Klinische Standardversorgung (CAU) oder Wartelisten-Kontrolle (Waitlist WL)
Fragestellung	Welche Effektstärke (Hedges g) erzielen evidenzbasierte psychologische Interventionen (Kognitive Verhaltenstherapie, Verhaltensaktivierung, Problemlösetherapie) gegenüber Care-as-Usual (CAU) und Wartelisten-Kontrolle (WL) bei Erwachsenen mit unipolarer Depression, und wie robust ist der Befund über das Prä-Post-Korrelationsgitter r ∈ {0,00..0,80} sowie die GRIM/GRIMMER-Forensik?

Tabelle 2. Zusammenfassung der Befunde (GRADE) über 1 Endpunkt.

Endpunkt	Teilnehmende (Studien)	Relativer Effekt (95 % KI)	Absolute Wirkung	Sicherheit der Evidenz	Anmerkung
Depressions-Symptomreduktion (BDI-II / PHQ-9 / HDRS-17)	—	—	Für diesen Endpunkt liegen derzeit keine poolbaren Effektgrößen vor.	⊕⊕⊕⊕ Hoch	Ratifiziert: Hohe Evidenzsicherheit (GRADE High) im zeit-eingefrorenen Blindlauf (k = 12 RCTs, N = 3.044).

- ⊕⊕⊕⊕ Hoch: Wir sind sehr zuversichtlich, dass der wahre Effekt nahe beim geschätzten liegt.
- GRADE bewertet das Vertrauen in den Effektschätzer je Endpunkt, nicht die Qualität einzelner Studien.

Literaturverzeichnis (Vancouver (ICMJE))

Vancouver nummeriert nach der Reihenfolge der ersten Nennung im Text. Diese Reihenfolge kennt die Anwendung nicht; das Verzeichnis steht deshalb alphabetisch und ist nach dem Fertigstellen des Textes umzunummerieren. Zeitschriftentitel stehen als NLM-Abkürzung; wo nur der volle Titel vorliegt, ist das im Eintrag vermerkt. Autorenangaben ohne Semikolon werden heuristisch in Personen zerlegt.

1. Andersson G, Bergström J, Holländare F, Carlbring P, Kaldo V, Ekselius L. Internet-based self-help for depression: randomised controlled trial. British Journal of Psychiatry. 2005. https://doi.org/10.1192/bjp.187.5.456 ⚑ *unvollständig: voller Zeitschriftentitel (APA), Band, Seiten*
2. Berger T, Hämmerli K, Gubser N, Andersson G, Caspar F. Internet-based treatment of depression: a randomized controlled trial comparing guided with unguided self-help. Cognitive Behaviour Therapy. 2011. https://doi.org/10.1080/16506073.2011.616531 ⚑ *unvollständig: voller Zeitschriftentitel (APA), Band, Seiten*
3. Buntrock C, Ebert DD, Lehr D, Smit F, Riper H, Berking M, et al. Effect of a Web-Based Guided Self-help Intervention for Prevention of Major Depression in Adults With Subthreshold Depression: A Randomized Clinical Trial. JAMA. 2016. https://doi.org/10.1001/jama.2016.4326 ⚑ *unvollständig: voller Zeitschriftentitel (APA), Band, Seiten*
4. DeRubeis RJ, Hollon SD, Amsterdam JD, Shelton RC, Young PR, Salomon RM, et al. Cognitive Therapy vs Medications in the Treatment of Moderate to Severe Depression. Archives of General Psychiatry. 2005. https://doi.org/10.1001/archpsyc.62.4.409 ⚑ *unvollständig: voller Zeitschriftentitel (APA), Band, Seiten*
5. Dimidjian S, Hollon SD, Dobson KS, Schmaling KB, Kohlenberg RJ, Addis ME, et al. Randomized trial of behavioral activation, cognitive therapy, and antidepressant medication in the acute treatment of adults with major depression. Journal of Consulting and Clinical Psychology. 2006. https://doi.org/10.1037/0022-006X.74.4.658 ⚑ *unvollständig: voller Zeitschriftentitel (APA), Band, Seiten*
6. Gilbody S, Lewis H, Adamson J, Atherton K, Bailey D, Birtwistle J, et al. Effect of Collaborative Care vs Usual Care on Depressive Symptoms in Older Adults With Major Depression: The CASPER Plus Randomized Clinical Trial. JAMA. 2017. https://doi.org/10.1001/jama.2017.0130 ⚑ *unvollständig: voller Zeitschriftentitel (APA), Band, Seiten*
7. Johansson R, Sjöberg E, Sjögren M, Johnsson E, Carlbring P, Andersson T, et al. Tailored vs. standardized internet-based cognitive behavior therapy for depression and comorbid symptoms: a randomized controlled trial. PLoS ONE. 2012. https://doi.org/10.1371/journal.pone.0038021 ⚑ *unvollständig: voller Zeitschriftentitel (APA), Band, Seiten*
8. Kessler D, Lewis G, Kaur S, Wiles N, King M, Weich S, et al. Therapist-delivered internet psychotherapy for depression in primary care: a randomised controlled trial (IPCRESS). The Lancet. 2009. https://doi.org/10.1016/S0140-6736(09)61257-5 ⚑ *unvollständig: voller Zeitschriftentitel (APA), Band, Seiten*
9. Klein JP, Berger T, Schröder J, Späth C, Meyer B, Caspar F, et al. Effects of a Psychological Internet Intervention in the Treatment of Mild to Moderate Depressive Symptoms: Results of the EVIDENT Study, a

Randomized Controlled Trial. Psychotherapy and Psychosomatics. 2016. https://doi.org/10.1159/000445355 ⚑ *unvollständig: voller Zeitschriftentitel (APA), Band, Seiten*

10. Perini S, Titov N, Andrews G. Clinician-assisted Internet-based treatment is effective for depression: randomized controlled trial. Australian and New Zealand Journal of Psychiatry. 2009. https://doi.org/10.1080/00048670902873722 ⚑ *unvollständig: voller Zeitschriftentitel (APA), Band, Seiten*
11. Vernmark K, Lenndin J, Bjärehed J, Carlsson M, Karlsson J, Öberg J, et al. Internet administered guided self-help versus individualized e-mail therapy for major depression: A randomized trial. Behaviour Research and Therapy. 2010. https://doi.org/10.1016/j.brat.2010.01.005 ⚑ *unvollständig: voller Zeitschriftentitel (APA), Band, Seiten*
12. Wiles N, Thomas L, Abel A, Ridgway N, Turner N, Campbell J, et al. Cognitive behavioural therapy as an adjunct to pharmacotherapy for primary care based patients with treatment resistant depression: results of the CoBalT randomised controlled trial. The Lancet. 2013. https://doi.org/10.1016/S0140-6736(12)61552-9 ⚑ *unvollständig: voller Zeitschriftentitel (APA), Band, Seiten*

12 von 12 Einträgen fehlt mindestens eine Pflichtangabe (Zeitschrift, Band oder Seiten). Solche Einträge weisen die Redaktionssysteme der Zeitschriften zurück. Nachtragen lässt sich das je Studie in der Kodier-Ansicht; bei Studien mit DOI holt „Bibliografie nachtragen“ die Angaben aus OpenAlex.

Erzeugt von EvidenzWerk am 2026-11-10. Die Abbildungen (PRISMA-2020-Flussdiagramm, Wald-Diagramm, Trichter-Diagramm) stehen an ihrer inhaltlichen Fundstelle im Text; im Einreichungspaket liegen dieselben Abbildungen zusätzlich als SVG-Dateien für die getrennte Einreichung bei. Der Text ist ein Entwurf: Die Autor:innen redigieren ihn, bevor er eingereicht wird.