EvidenzWerk — Fachliches & Technisches Gesamthandbuch (Kompendium 2026)

Living Meta-Analysis Studio · Methoden-Kompendium, Studi-Modus (Voice-Lotse & Paywall-Navigator), Statistik-Rechenkern & Systemarchitektur

> ***Hinweis (IMPORTANT): Zweck dieses Handbuchs:*** *Dieses Kompendium vereint zwei Dokumente in einem durchgängigen Nachschlagewerk: 1.* ***Teil I — Fachliche Dokumentation & Methoden-Kompendium (mit Screenshots & Video):*** *Erklärt sämtliche in* ***EvidenzWerk*** *abgebildeten Verfahren der Evidenzsynthese — von klassischen Gruppenvergleichen über klinische Überlebenszeitanalysen (****Hazard Ratios nach Tierney et al. 2007****),* ***Drei-Ebenen-Meta-Analysen mit robuster Varianzschätzung (****`clubSandwich` ***CR2)****,* ***Netzwerk-Meta-Analysen (NMA)*** *und* ***Diagnostische Testgüte-Studien (DTA / Reitsma-HSROC & QUADAS-2)*** *bis hin zu* ***RoB 2****,* ***ROBINS-I****,* ***GRADE*** *und* ***RAISE-2025****. Jedes Verfahren ist so beschrieben, dass Sie auch außerhalb Ihres angestammten Fachgebiets sofort wissen, wann man es einsetzt, welche Rohzahlen man eingibt und wie man die Ergebnisse und Grafiken fachgerecht interpretiert. 2.* ***Teil II — Technische System-, Rechenkern- & API-Dokumentation:*** *Dokumentiert die gesamte Software-Architektur (* `Next.js`*,* `SQLite` *mit WAL,* `R`*-Bridge zu* `metafor` *&* `clubSandwich`*, die gläserne Rechenspur, die 100-%-paritätische* `v1`*-Agenten-API mit Idempotenz und Dry-Run, die serverseitige Mehrbenutzer-Verblindung, die nativen PDF/DOCX/SVG/Poster-Renderer ohne Fremdbibliotheken sowie das Podman-Quadlet-Deployment).*

---

0. Gesamtschau im Video & Bildergalerie

0.1 Vollständiger Durchlauf im Video (MP4)

Das folgende Video durchläuft live die beiden im Showcase aufgebauten Referenzprojekte — vom Projekt-Dashboard über Protokoll & Mehrbenutzer-Verblindung, PRISMA-S-Recherche, Doppelscreening mit Cohen's κ, KI-Extraktion mit Seitenanker, klinische Kaplan-Meier-Rekonstruktion (Tierney et al. 2007), RoB 2 & ROBINS-I, Drei-Ebenen-REML+CR2, Publikationsbias-Suite, Netzwerk-Meta-Analyse (NMA), Diagnostische Testgüte (SROC & QUADAS-2) und GRADE bis zur gläsernen Rechenspur, Ratifikation und dem druckfertigen Einreichungspaket:

![EvidenzWerk Gesamtschau-Video (2026)](/Users/franzhuetter/.gemini/antigravity-cli/brain/ddf3e4eb-130f-4bd1-a05b-29ff57ae1d90/evidenzwerk-gesamtschau-2026.mp4)

(Im Projekt-Repository liegt die Videodatei dauerhaft unter `data/demo-videos/evidenzwerk-gesamtschau-2026.mp4`, alle 16 Retina-Screenshots unter `data/handbuch-screenshots/`.)

Teil I — Fachliche Gesamtdokumentation & Methoden-Kompendium

1. Der Gesamtprozess einer systematischen Übersichtsarbeit und Meta-Analyse

EvidenzWerk bildet den gesamten Lebenszyklus einer systematischen Übersichtsarbeit (Systematic Review) und quantitativen Evidenzsynthese (Meta-Analyse) nach den Standards des **Cochrane Handbooks (Version 6.5)**, **PRISMA 2020**, **PRISMA-S**, **PRISMA-P**, **GRADE** und **RAISE 2025** ohne Medienbruch ab.

```
Prozesskette: 1. Protokoll & PICOS (PRISMA-P, Team-Verblindung, Studi-Modus) -> 2. Recherche & Import (OpenAlex, PubMed, CT.gov, RIS/BibTeX) -> 3. Doppelscreening (Reviewer A/B, Cohens Kappa) -> 4. Volltext-Extraktion & 3-Stufen-Paywall-Lotse -> 5. Codierung & Bias (10 Metriken, RoB 2, ROBINS-I, QUADAS-2) -> 6. Statistische Synthese (3-Ebenen-CR2, NMA, DTA, PubBias) -> 7. GRADE & Ratifikation (RAISE-2025) -> 8. Einreichungspaket & Klartext-Dossier (DOCX, PDF, SVG, R-Skript, ZIP)
```

1.1 Dashboard, Protokoll (PRISMA-P) und Mehrbenutzer-Verblindung

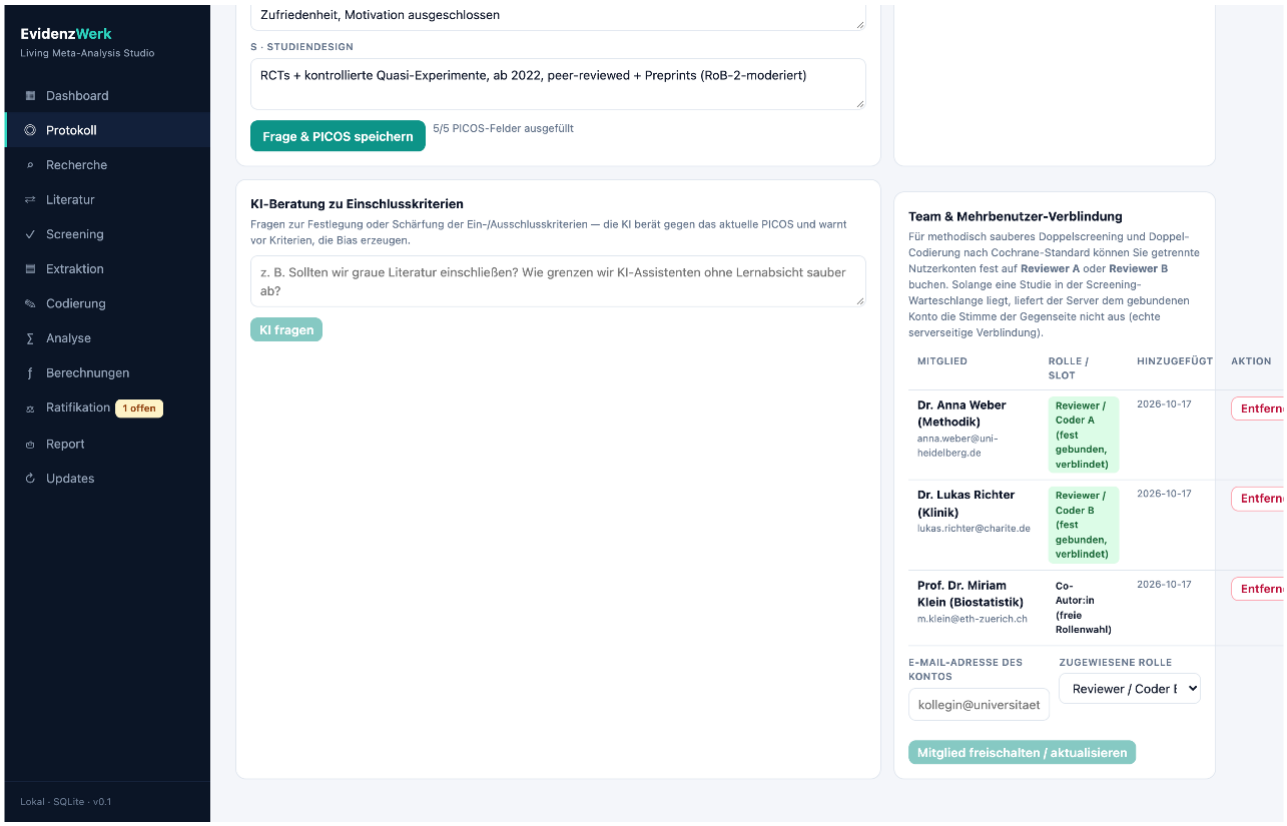


Abbildung 1: *Dashboard (01-dashboard.png)*

Jedes Projekt beginnt im Reiter **Protokoll** (`PicosEditor.tsx`, `buildPrismaP`) mit der strukturierten Festlegung der Forschungsfrage nach dem **PICOS-Schema**:

- **P (Population):** Welche Teilnehmenden oder Patientengruppen werden untersucht? (z. B. *Erwachsene Lernende ≥ 18 Jahre* oder *Patient:innen mit metastasiertem NSCLC Stadium IV, ECOG 0–1*)
- **I (Intervention / Index-Test):** Welche Maßnahme, Therapie oder diagnostische Methode wird geprüft?
- **C (Comparator / Kontrolle):** Wogegen wird verglichen (Standardtherapie, Placebo, aktive Kontrolle, Referenzstandard)?
- **O (Outcomes / Endpunkte):** Welche primären und sekundären Zielgrößen sind maßgeblich?
- **S (Study Design):** Welche Studiendesigns sind zugelassen (nur RCTs, auch quasi-experimentelle Studien oder adjustierte Registerkohorten)?

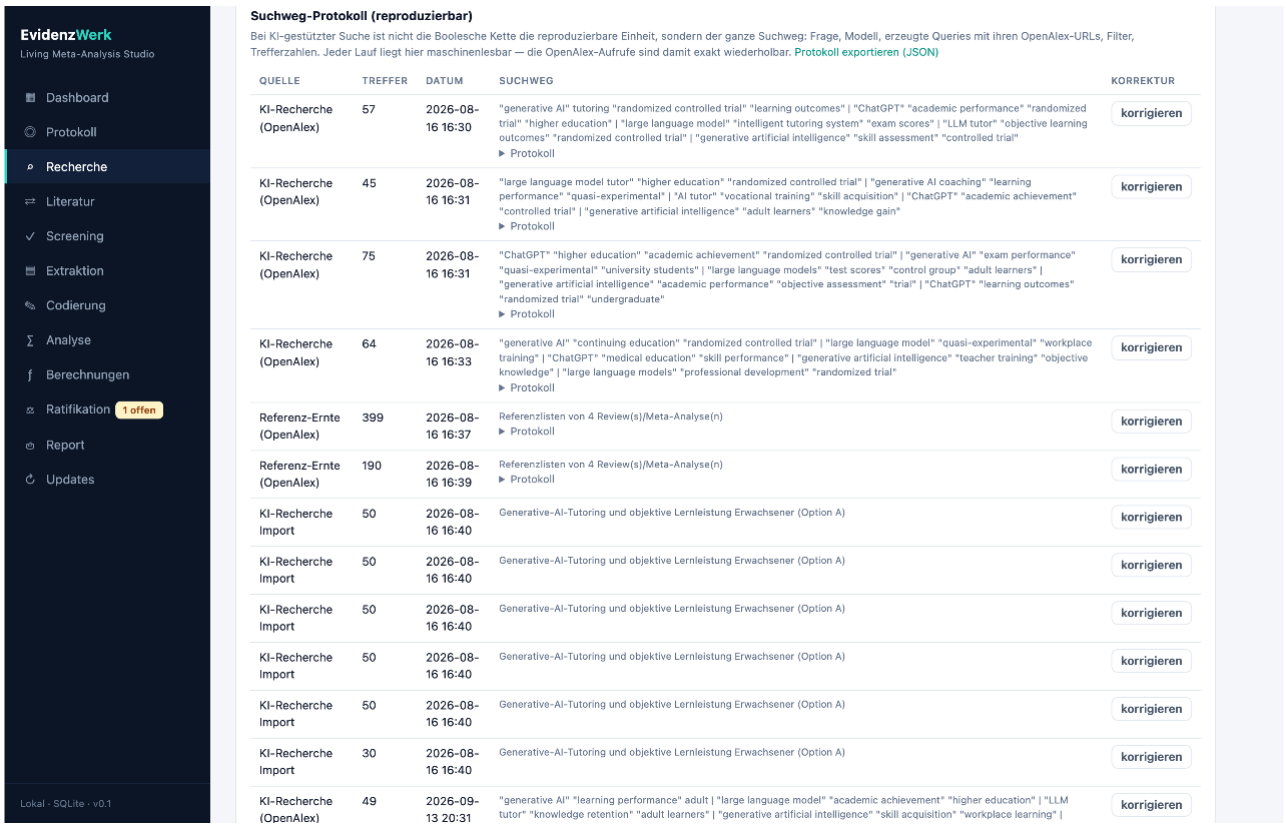


Abbildung 2: *Protokoll & Team-Verblindung (02-protokoll-team-verblindung.png)*

> ***Hinweis (TIP): Echte serverseitige Mehrbenutzer-Verblindung (***`TeamPanel.tsx`, `lib/tenancy.ts`***):*** *Nach Cochrane-Standard müssen zwei Personen (**Reviewer A** und **Reviewer B**) Studien unabhängig voneinander sichten und kodieren, ohne die Entscheidung der jeweils anderen Person vorher zu sehen. In EvidenzWerk weisen Sie im Protokoll-Reiter realen Nutzerkonten die Rolle **Reviewer / Coder A**, **Reviewer / Coder B** oder **Co-Autor:in** zu. Solange sich eine Studie in der Warteschlange (*`queued`*) befindet, filtert der Server fremde Stimmen bereits auf API- und Datenbankebene heraus; der Rollen-Umschalter im UI ist für gebundene Reviewer-Konten gesperrt.*

---

2. Literaturrecherche (PRISMA-S), Dublettenbereinigung & Doppelscreening

2.1 Viergliedrige Recherche und PRISMA-S-Protokollierung

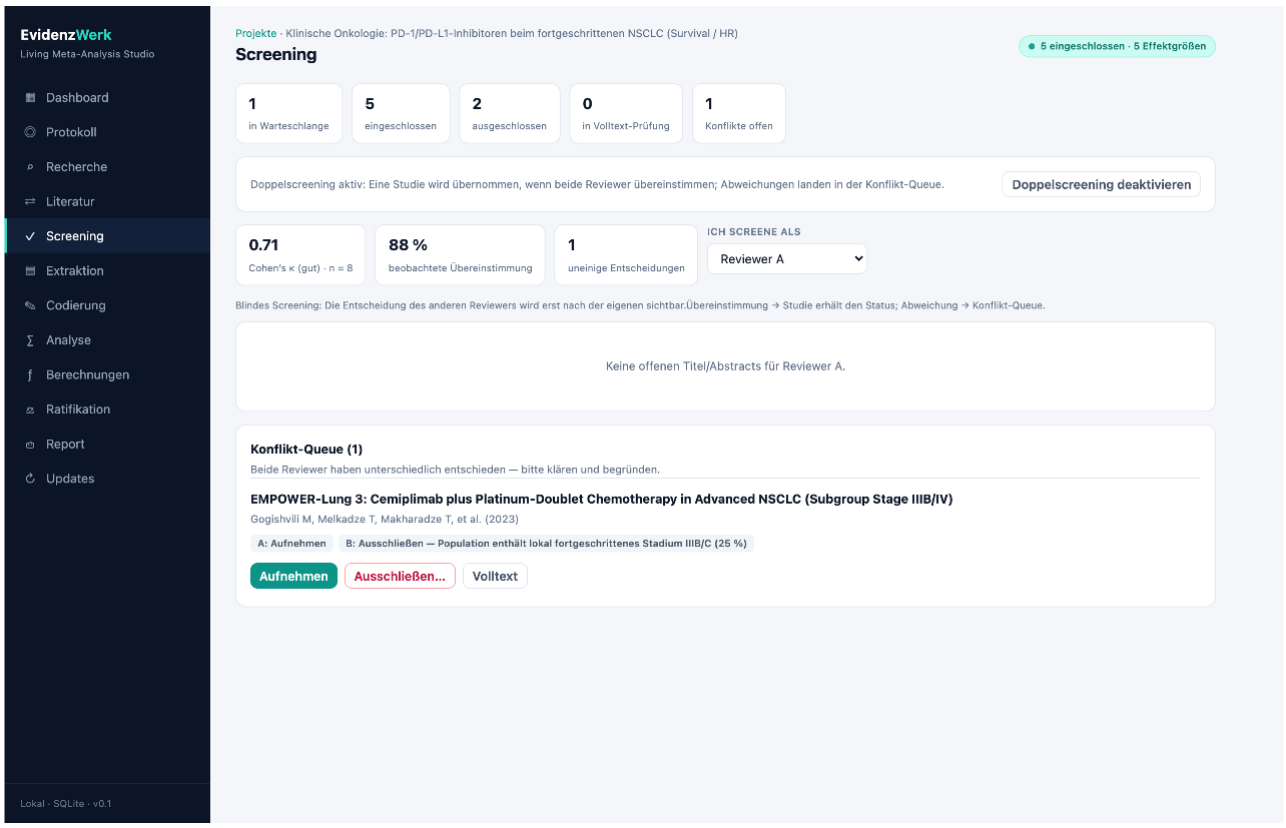


Abbildung 3: *Recherche & PRISMA-S (03-recherche-prisma-s.png)*

Im Reiter **Recherche** kombiniert EvidenzWerk vier komplementäre Zugänge, damit keine relevante Evidenz übersehen wird:

1. **KI-gestützte Mehrpfad-Suche über OpenAlex (`AiSearchPanel.tsx`, `lib/scholarly.ts`):** Zerlegt die Forschungsfrage in mehrere synonyme englischsprachige Suchpfade und speichert die exakten Abfrage-URLs, Zeitstempel und Trefferzahlen.
2. **PubMed / MEDLINE via NCBI E-Utilities (`PubmedPanel.tsx`, `lib/pubmed.ts`):** Unterstützt vollständige Boolesche Ketten inkl. `MeSH Terms` (`[mh]`), `[tiab]`-Feldern und Publikationstyp-Filtern.
3. **Studienregister ClinicalTrials.gov API v2 (`TrialsPanel.tsx`, `lib/clinicaltrials.ts`):** Essenziell gegen **Publikationsbias (Schubladeneffekt)**: Findet abgeschlossene Studien (`COMPLETED`) mit oder ohne publizierten Fachaufsatz. Unveröffentlichte Registerstudien werden mit `registry_unpublished = 1` markiert und im Analyse-Reiter gesondert als Warnhinweis ausgewiesen.
4. **Externe Datenbanken & Datei-Import (`ExternalSearchPanel.tsx`, `ImportForm.tsx`):** Direktimport von `.ris`, `.bib` (BibTeX), `.nbib` (PubMed/Medline) und `.csv` aus Lizenzdatenbanken wie **Embase (Ovid)**, **Scopus**, **Web of Science**, **PsycINFO** oder **ERIC** inklusive PRISMA-S-konformer Dokumentation der Suchoberfläche und mehrzeiligen Suchsyntax.

2.2 Doppelscreening, Cohen's Kappa (κ) und Konfliktlösung

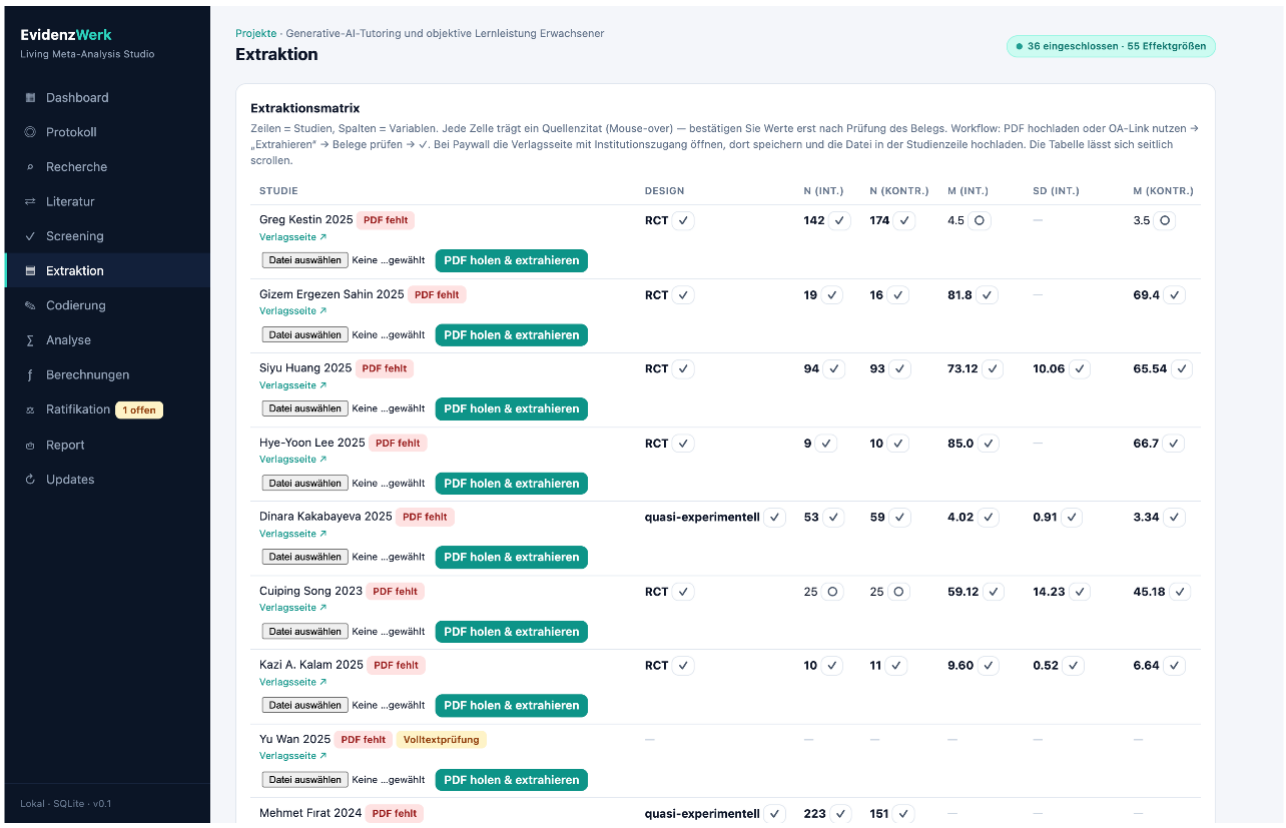


Abbildung 4: *Doppelscreening & Cohen's Kappa (04-doppelscreening-kappa.png)*

Im Reiter **Screening** (`ScreeningQueue.tsx`) wird jede importierte Arbeit zweistufig geprüft (**Titel/Abstract-Screening** arrow **Volltext-Prüfung**):

- **Inter-Rater-Reliabilität nach Cohen's κ (`screeningAgreement`):** Berechnet laufend die zufallskorrigierte Übereinstimmung zwischen Reviewer A und Reviewer B:

κ = p_o - p_e/1 - p_e wobei p_o die beobachtete Übereinstimmungsrate und p_e die bei Zufall erwartete Übereinstimmung ist.

- **Orientierungswerte nach Landis & Koch (1977):** κ < 0,40 mäßig, 0,41–0,60 moderat, 0,61–0,80 substanziell, > 0,80 fast perfekt.
- **Konflikt-Queue & Drittentscheid:** Stimmt Reviewer A für *Einschließen* und Reviewer B für *Ausschließen*, wechselt die Studie automatisch in den Status `conflict` und erscheint in der Karte **Konflikt-Lösung (Drittentscheid)**, wo eine dritte Person (oder das Konsensgespräch) die finale Entscheidung mit dokumentierter Begründung trifft.
- **PRISMA-2020-Ausschlussgründe:** Bei jedem Volltext-Ausschluss erzwingt EvidenzWerk eine strukturierte Begründung (z. B. *Population nicht PICOS-konform*, *kein kontrolliertes Design*, *Intervention nicht generativ*, *Retraktion / Datenintegrität*), da PRISMA 2020 (Item 16b) im Bericht für jeden ausgeschlossenen Volltext den exakten Grund verlangt.

---

3. Volltext-Extraktion mit Seitenanker und Belegzitat

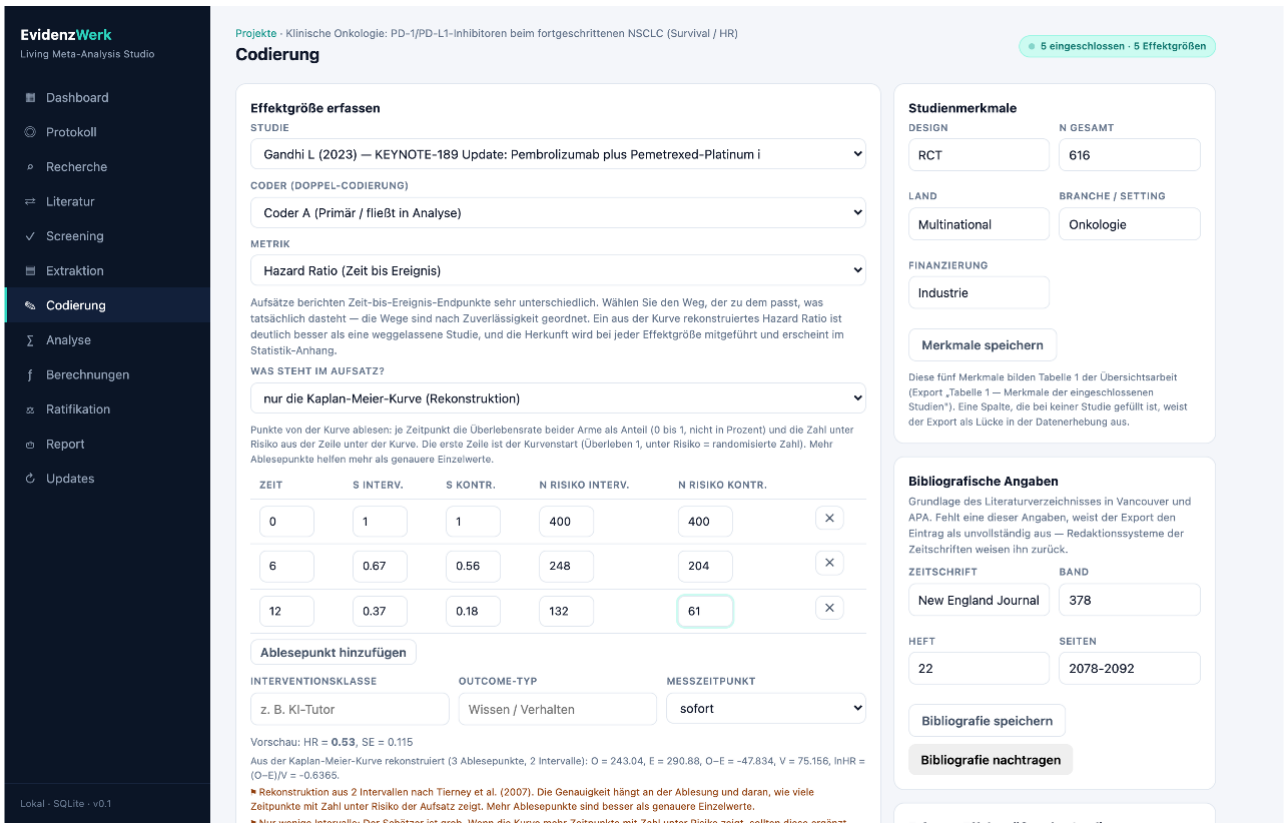


Abbildung 5: *KI-Extraktion mit Belegzitat (05-ki-extraktion-belegzitat.png)*

Im Reiter **Extraktion** (`ExtractionPanel.tsx`, `lib/extract.ts`) beschafft EvidenzWerk Open-Access-Volltexte automatisch über **Unpaywall** und **Europe PMC** (`lib/pdf-fetch.ts`), prüft anhand von DOI und Titel-Token-Abgleich, ob die PDF-Datei wirklich zur Studie gehört (`lib/pdf-identity.ts`), und extrahiert strukturierte Befunde:

- **Keine Zahl ohne wörtliches Belegzitat und Seitenzahl:** Jede extrahierte Angabe zu Stichprobengröße, Mittelwerten, Standardabweichungen, t-/F-/p-Werten oder Hazard Ratios wird zwingend mit dem **wörtlichen Zitat aus dem PDF (`quote`)** und der **Fundstelle (`page` / Abschnitt)** gespeichert.
- **1-Klick-Übernahme in die Kodiermaske:** Ein Klick auf *„Übernehmen → Codierung"* überträgt die extrahierten Zahlen direkt in die Effektgrößen-Eingabemaske, markiert den Befund als menschlich geprüft (`confirmed = 1`) und protokolliert den Vorgang im Audit-Trail.

---

4. Großes Methoden-Kompendium: Alle 10 Effektgrößen-Metriken & Studiendesigns

> ***Hinweis (NOTE): Ihr Spickzettel für fremde Fachgebiete:*** *Nicht jede Meta-Analyse nutzt dieselbe Effektgröße. Während in Psychologie und Bildungsforschung **Hedges' g** und **Korrelationen r** dominieren, arbeiten Innere Medizin, Kardiologie und Epidemiologie mit **Odds Ratios (OR)**, **Risk Ratios (RR)**, **Risikodifferenzen (RD / NNT)** und **Inzidenzratenverhältnissen (IRR)**, die Onkologie mit **Hazard Ratios (HR)** aus Überlebenszeitkurven und die Epidemiologie/Radiologie mit **Prävalenz-Anteilen** oder **Sensitivität/Spezifität**. EvidenzWerk beherrscht alle zehn Metriken nativ in `lib/stats.ts` und `lib/metrics.ts`.*

Kürzel	Name der Metrik	Typisches Fachgebiet	Wann verwenden?	Rechenskala (intern)	Berichtete Skala (Plot/Text)	Neutralwert (kein Effekt)
`smd`	**Hedges' g** (standardisierte Mittelwertdifferenz)	Psychologie, Medizin, Bildungsforschung, Sozialwissenschaften	Kontinuierliches Outcome (z. B. Schmerz, Depression, Lernleistung), das in verschiedenen Studien mit **unterschiedlichen Tests/Skalen** gemessen wurde.	Linear (g)	Linear (g)	0
`md`	**Mittelwertdifferenz (MD)**	Kardiologie, Intensivmedizin, Labormedizin	Kontinuierliches Outcome auf **exakt derselben natürlichen Einheit** in allen Studien (z. B. Blutdruck in mmHg, HbA1c in %, OP-Dauer in Minuten).	Linear (MD)	Linear (MD)	0
`r`	**Korrelation r** (über Fisher's z)	Differentielle, Arbeits- & Organisationspsychologie	Zusammenhang zweier kontinuierlicher Variablen ohne Gruppenaufteilung (z. B. Intelligenz und Berufserfolg).	Fisher's z = artanh(r)	r = tanh(z) in (-1, +1)	0
`or`	**Odds Ratio (OR)**	Epidemiologie, Fall-Kontroll-Studien, Logistische Regression	Binäres Ereignis (Ja/Nein, Tod/Überleben, Remission), insbesondere bei Fall-Kontroll-Designs oder adjustierten logistischen Modellen.	ln(OR)	OR = exp(y)	1,0
`rr`	**Risk Ratio / Relatives Risiko (RR)**	Klinische RCTs, prospektive Kohortenstudien	Binäres Ereignis in prospektiven Studien: Verhältnis der Ereigniswahrscheinlichkeit p_1 / p_2. Leichter verständlich als OR, wenn Ereignisse häufig sind (> 10 %).	ln(RR)	RR = exp(y)	1,0
`rd`	**Risikodifferenz (RD) & NNT**	Klinische Leitlinien, Versorgungsforschung, EbM	Absolute Risikosenkung p_1 - p_2 und daraus abgeleitete Number Needed to Treat (NNT = 1 /	RD	) nach Altman (1998).	Linear (RD)

Kürzel	Name der Metrik	Typisches Fachgebiet	Wann verwenden?	Rechenskala (intern)	Berichtete Skala (Plot/Text)	Neutralwert (kein Effekt)
`irr`	**Inzidenzratenverhältnis (IRR)**	Infektiologie, Pneumologie, Pharmakoepidemiologie	Zählereignisse pro **Personenzeit** (z. B. Asthma-Anfälle oder Stürze pro 1.000 Patientenjahre bei unterschiedlich langer Beobachtungsdauer).	ln(IRR)	IRR = exp(y)	1,0
`prop_logit`	**Einzelanteil / Prävalenz (Logit)**	Epidemiologie, Einarmige Phase-II-Onkologie	Zusammenfassung einer einzelnen Rate p = e / n (z. B. Komplikationsrate, Krankheitsprävalenz) ohne Vergleichsgruppe bei moderaten Raten (5–95 %).	logit(p) = lnp/1-p	Anteil p = frac11+e^(-y) in (0, 1)	—
`prop_ft`	**Einzelanteil (Freeman-Tukey Doppel-Arkussinus)**	Seltene Komplikationen, Chirurgie, Registerstudien	Einarmige Raten/Prävalenzen, wenn einzelne Studien **0 % oder 100 % Ereignisse** aufweisen — stabilisiert die Varianz ohne künstliche +0,5-Addition.	Freeman-Tukey t	Anteil p in [0, 1] (Miller 1978)	—
`hr`	**Hazard Ratio (HR, Zeit bis Ereignis)**	Onkologie, Kardiologie, Nephrologie, Intensivmedizin	**Überlebenszeitanalysen (Survival)**: Zeit bis Tod, Progression, Rückfall oder Entlassung unter Berücksichtigung **zensierter Patient:innen**.	ln(HR)	HR = exp(y)	1,0

4.1 Kontinuierliche Gruppenvergleiche (`smd` / Hedges' g und `md`) und ihre 6 Eingabewege

Oft berichten Primärstudien nicht brav Mittelwert (M), Standardabweichung (SD) und Gruppengröße (n), sondern nur einen t-Wert, einen F-Wert aus einer einfaktoriellen ANOVA, einen Mittelwertunterschied mit p-Wert oder ein fertig gerechnetes Cohen's d. EvidenzWerk rechnet in `computeEffect` alle sechs Berichtswege exakt in Hedges' g um:

1. **Rohdaten (M_1, SD_1, n_1 und M_2, SD_2, n_2):**

s (pooled) = √((n₁ - 1)SD₁² + (n₂ - 1)SD₂²/n₁ + n₂ - 2), d = fracM₁ - M₂s (pooled)

1. **Aus t-Statistik (t, n_1, n_2):** d = t · √(1/n₁ + 1/n₂)
2. **Aus einfaktoriellem F-Wert (F mit $df_1 = 1$, n_1, n_2, Wirkrichtung ± 1):** t = ±√(F), danach wie bei t.
3. **Aus Differenz Δ M und zweiseitigem p-Wert (p, n_1, n_2):** Bestimmt erst das Quantil t = F (t,df)⁻¹(1 - p/2)

über die inverse Student-t-Verteilung (df = $n_1 + n_2 - 2$), daraus SE = |Δ M| / t und s (pooled).

4. **Aus fertig berichtetem Cohen's d (d, n_1, n_2):** Übernimmt d und wendet die exakte Kleinstichproben-Korrektur J(df) an.
5. **Direkteingabe (g und SE):** Falls die Studie bereits Hedges' g samt Standardfehler berichtet.

Warum immer Hedges' g statt Cohen's d? Cohen's d überschätzt in kleinen Stichproben (n < 30) systematisch den wahren Populationseffekt. EvidenzWerk multipliziert d und seine Standardabweichung deshalb mit dem exakten Gamma-Korrekturfaktor nach Hedges (1981): J(df) = fracΓ(df/2)√(df/2) Γ((df-1)/2) ≈ 1 - 3/4cdotdf - 1, g = J(df)· d, v_g = J(df)² · (n_1+n_2/n_1 n_2 + d²/2(n_1+n_2))

4.2 Studiendesigns: Prä-Post (Within) und Cluster-randomisierte Studien (ICC)

Zwei der häufigsten Methodenfehler in publizierten Meta-Analysen entstehen bei **Prä-Post-/Crossover-Designs** und **Cluster-randomisierten Studien** (Cochrane Handbook Kap. 23):

1. **Within-Group / Prä-Post / Crossover (`design = 'within'`):**

Wenn dieselben Personen vor und nach einer Intervention gemessen werden, sind beide Messungen mit r (pre,post) korreliert. Die Varianz von d beträgt nach Borenstein et al. (2009, Gl. 4.28): v_d = (1/n + d²/2n) · 2(1 - r (pre,post)) Ist r (pre,post) im Aufsatz nicht angegeben, hinterlegt EvidenzWerk einen konservativen Imputationswert (Standard r = 0,50) und protokolliert diese Annahme transparent im Statistik-Anhang.

1. **Cluster-randomisierte Studien (`design = 'cluster'`, z. B. Schulklassen, Stationen, Arztpraxen):**

Werden ganze Schulklassen oder Kliniken statt einzelner Personen randomisiert, sind die Individuen innerhalb eines Clusters einander ähnlicher (**Intraklassen-Korrelation ICC**). Wer einfach die Schülerzahl N in die Standardformel einsetzt, begeht einen **Unit-of-Analysis-Fehler** und gibt der Studie ein viel zu hohes Gewicht! EvidenzWerk multipliziert die Varianz deshalb automatisch mit dem **Design-Effekt (DE)** nach Higgins et al. (2023): DE = 1 + (barm - 1)cdotICC, v (cluster) = v (individuell) · DE wobei barm die mittlere Clustergröße ist.

4.3 Klinische Überlebenszeitanalyse: Hazard Ratios (`hr`) & alle 5 Tierney-Wege

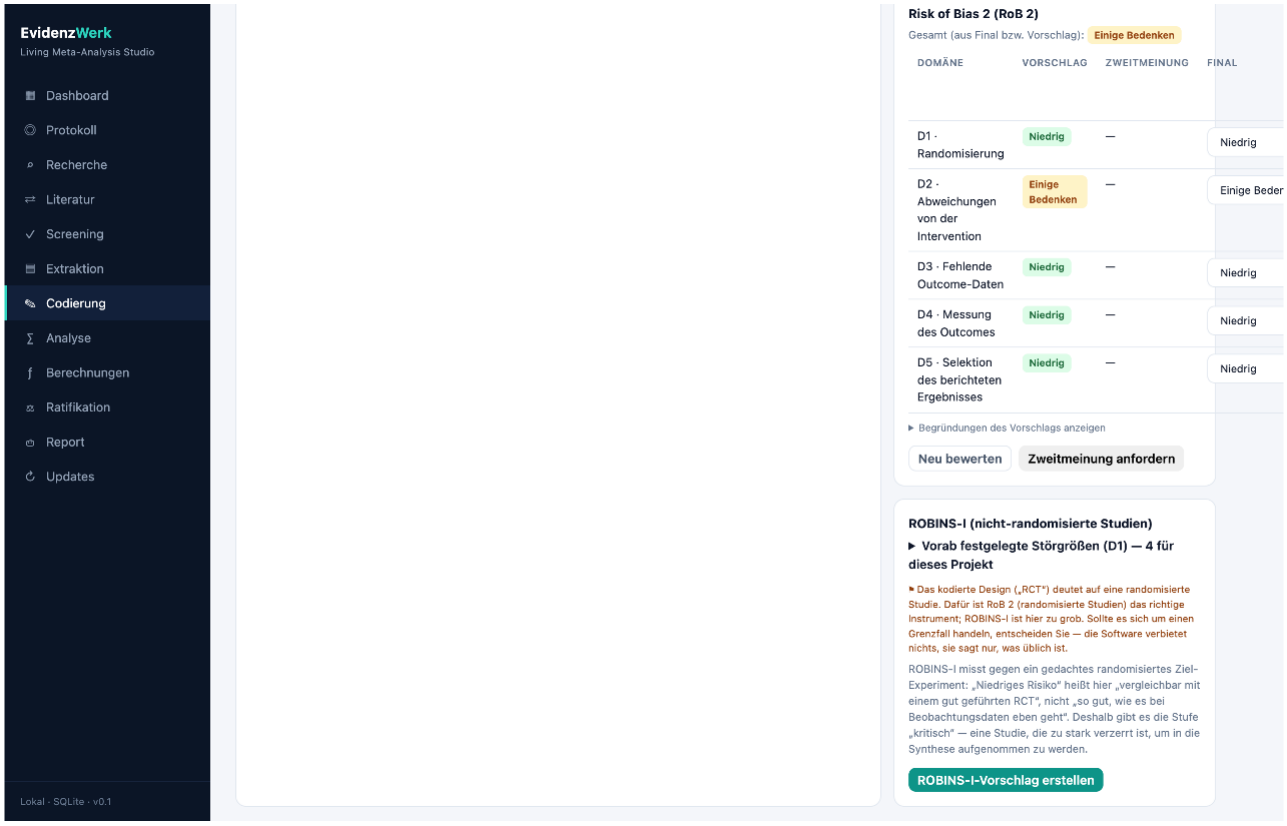


Abbildung 6: *Codierung Survival & Kaplan-Meier-Rekonstruktion (06-codierung-survival-tierney-km.png)*

In der Onkologie, Kardiologie und Transplantationsmedizin ist das zentrale Maß nicht der Anteil Überlebender zu einem willkürlichen Stichtag, sondern das **Hazard Ratio (HR)** über den gesamten Beobachtungsverlauf (z. B. *Progressionsfreies Überleben, PFS* oder *Gesamtüberleben, OS*).

- **Interpretation:** Ein HR = 0,52 [0,43; 0,64] (wie bei *Gandhi et al. / KEYNOTE-189* in unserem Onkologie-Showcase) bedeutet: Zu jedem Zeitpunkt ist das momentane Risiko für Tod oder Tumorprogression unter Pembrolizumab + Platin-Chemotherapie um **48 % niedriger** als unter alleiniger Chemotherapie. HR = 1,0 bedeutet keinen Unterschied; HR < 1,0 begünstigt die Intervention (sofern das Ereignis Tod/Progression unerwünscht ist).

Weil ältere oder chirurgisch-klinische Originalarbeiten das Hazard Ratio oft nicht direkt mit Konfidenzintervall angeben, implementiert `lib/survival.ts` **alle fünf kanonischen Herleitungswege nach Tierney et al. (2007, *Trials* 8:16)** und Parmar et al. (1998, *Stat Med* 17:2815), geordnet nach ihrer statistischen Zuverlässigkeit:

Tierney-Weg	Schlüssel (`hrSource`)	Was steht im klinischen Aufsatz?	Mathematische Herleitung in `lib/survival.ts`
Weg 1 (Beste Quelle)	`hr_ci`	Berichtetes Hazard Ratio + 95-%-Konfidenzintervall (z. B. `HR = 0,52; 95% CI 0,43–0,64`)	y = ln(HR), SE = fracln(HR (hi)) - ln(HR (lo))2 · z (0,975) (prüft zudem automatisch, ob HR symmetrisch auf der ln-Skala in [lo, hi] liegt).
Weg 2 (Exakt)	`peto`	Differenz beobachteter minus erwarteter Ereignisse ($O - E$) und Log-Rank-Varianz V	Nach Yusuf et al. (1985): y = ln(HR) = O - E/V, v = 1/V, SE = frac1√(V)

Tierney-Weg	Schlüssel (`hrSource`)	Was steht im klinischen Aufsatz?	Mathematische Herleitung in `lib/survival.ts`
Weg 3 (Sehr gut)	`logrank_p`	Zweiseitiger p-Wert des Log-Rank-Tests, Gesamtzahl der Ereignisse E, Gruppengrößen n_1, n_2 sowie Richtung (welcher Arm überlebte länger?)	Aus p folgt die Standardnormalgröße $z = \pm\Phi^{-1}(1 - p/2)$. Mit dem Allokationsverhältnis $R = n_2 / n_1$ ist die Informationsvarianz $V = E \cdot R/(1 + R)^2$ und $O - E = z \cdot \sqrt{(V)}$, woraus $\ln(HR) = O - E/V$ folgt (Tierney Gl. 7–9).
Weg 4 (Rekonstruktion)	`km_kurve`	**Nur die gedruckte Kaplan-Meier-Überlebenskurve** und die Zeile „*Number at risk*“ unter der Zeitachse	Aus mindestens zwei Zeitpunkten $t_0 < t_1 < \ldots < t_m$ mit den abgelesenen Überlebensanteilen $S_1(t_j)$, $S_2(t_j)$ in (0, 1] und den Zahlen unter Risiko rekonstruiert EvidenzWerk für jedes Intervall [t (j-1), t_j] die effektive Risikomenge, die bedingten Ereigniszahlen O (k,j) und die Intervall-Log-Rank-Größen (O_j - E_j) und V_j nach **Tierney et al. (2007, Gleichungen 10–13)** und summiert sie über alle Intervalle zu $\ln(HR) = \Sigma (O_j - E_j)/\Sigma V_j$.
Weg 5 (Näherung)	`medians`	Nur die **medianen Überlebenszeiten** beider Arme (m (Interv.), m (Kontr.)) und die Ereigniszahl E	Unter Annahme konstanter Hazards (Exponentialverteilung) gilt $HR \approx$ fracm (Kontr.)m (Interv.) mit $SE(\ln HR) \approx \sqrt{((n_1+n_2)^2/E \cdot n_1\, n_2)}$. EvidenzWerk versieht diesen Weg im Statistik-Anhang automatisch mit einem Methoden-Warnhinweis.

4.4 Binäre Endpunkte (`or`, `rr`, `rd` + Altman-NNT), Ereignisraten (`irr`) und Prävalenzen (`prop_logit`, `prop_ft`)

- **Vierfeldertafel (a, n_1, c, n_2) und Sweeting-Nullzellenkorrektur:**

Tritt in einem Studienarm kein einziges Ereignis auf (a = 0 oder c = 0), wäre ln(OR) oder ln(RR) wegen Division durch Null nicht definiert. Statt pauschal 0,5 auf jede Zelle zu addieren (was bei ungleich großen Gruppen $n_1 \neq n_2$ einen künstlichen Bias erzeugt!), verwendet EvidenzWerk die **gruppenproportionale Kontinuitätskorrektur nach Sweeting, Sutton & Lambert (2004, *Stat Med* 23:1351)**: $\delta_1 = n_1/n_1 + n_2$, $\delta_2 = n_2/n_1 + n_2$

- **Number Needed to Treat (NNT) nach Altman (1998, *BMJ* 317:1309):**

Bei der **Risikodifferenz (`rd`)** berechnet EvidenzWerk automatisch die klinisch anschauliche **Number Needed to Treat**:

- Schließt das 95-%-Konfidenzintervall der Risikodifferenz die 0 aus, wird NNTB (*Number Needed to Treat for an additional Beneficial outcome*) bzw. NNTH (*Harm*) mit sauberem Intervall ausgegeben.
- Überschreitet das Konfidenzintervall die 0 (nicht signifikant), geht das Intervall der NNT mathematisch über ∞! Viele Programme geben hier fälschlich unsinnige Zahlen wie „NNT 12 bis -25“ aus; EvidenzWerk formatiert dies streng nach **Altman (1998)** korrekt über ∞ („*NNTB 12 über ∞ bis NNTH 25*“).
- **Freeman-Tukey-Doppel-Arkussinus (`prop_ft`) für Raten mit 0 %- oder 100 %-Studien:**

Für einarmige Prävalenz- oder Komplikationsstudien transformiert EvidenzWerk nach Freeman & Tukey (1950):

$y_i = \arcsin\sqrt{(e_i/n_i + 1)} + \arcsin\sqrt{(e_i + 1/n_i + 1)}$, $v_i = 1/n_i + 0{,}5$ und transformiert den gepoolten Schätzer exakt nach **Miller (1978, *The American Statistician* 32:138)** über das harmonische Mittel tilden der Stichprobengrößen in eine Prozentzahl zurück.

---

5. Verzerrungsrisiko (Risk of Bias): RoB 2 vs. ROBINS-I vs. QUADAS-2

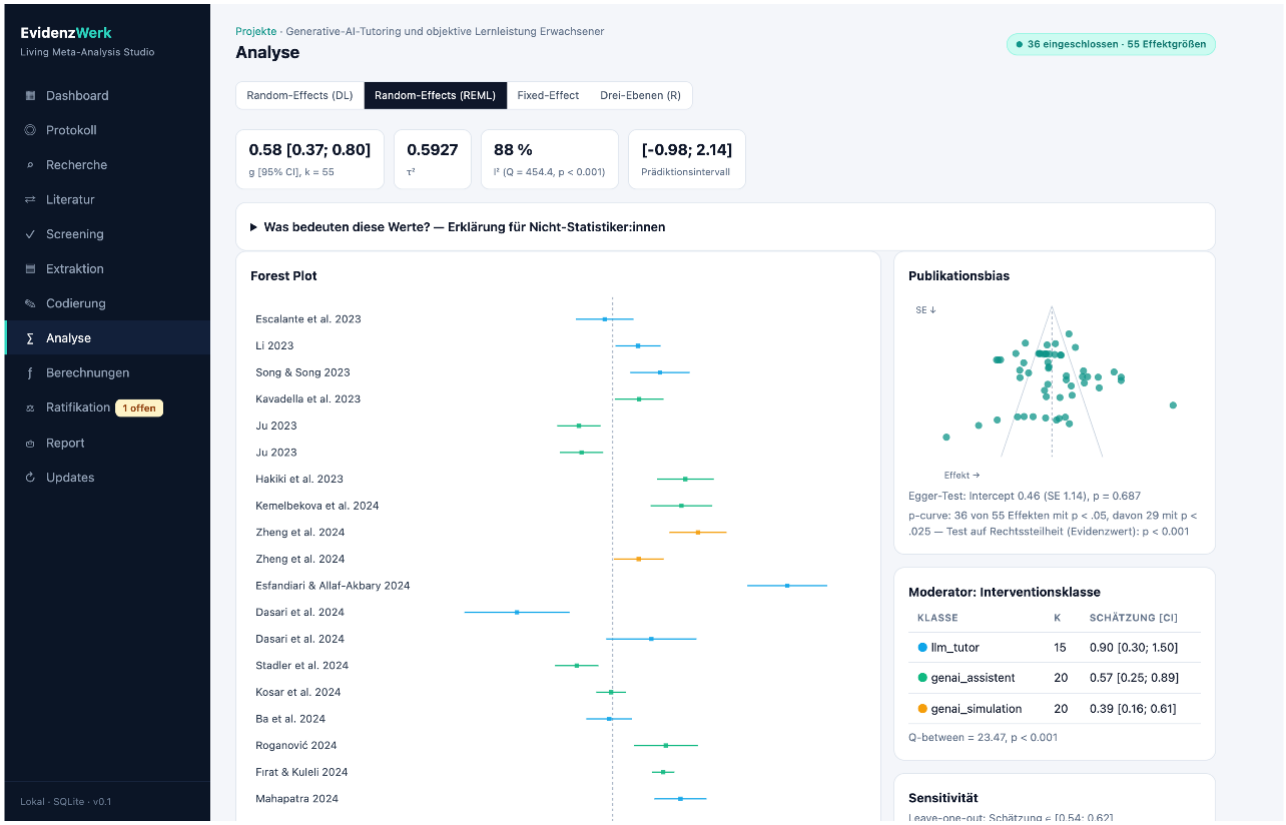


Abbildung 7: *RoB 2 & ROBINS-I Verzerrungsprüfung (07-bias-rob2-robins-i.png)*

> ***Hinweis (WARNING): Warum eine Design-Weiche zwischen RoB 2 und ROBINS-I unverzichtbar ist (*** `instrumentFuerDesign`***):*** *Ein häufiger Ablehnungsgrund bei medizinischen Fachzeitschriften ist die Bewertung von Beobachtungsstudien (Kohorten, Register, Prä-Post) mit **RoB 2**. RoB 2 gilt ausschließlich für randomisierte Studien (Domäne D1 prüft die Randomisierung!). Nicht-randomisierte Interventionsstudien müssen nach **ROBINS-I (Sterne et al. 2016, BMJ 355:i4919)** bewertet werden; diagnostische Studien nach **QUADAS-2 (Whiting et al. 2011)**. EvidenzWerk wählt anhand des kodierten Studiendesigns automatisch das methodisch richtige Instrument.*

Eigenschaft	**RoB 2 (** `lib/rob2-shared.ts`**)**	**ROBINS-I (** `lib/robins-i-shared.ts`**)**	**QUADAS-2 (**`lib/dta.ts`**)**
Studiendesign	Randomisierte kontrollierte Studien (RCT, Cluster-RCT, Crossover-RCT)	Nicht-randomisierte Interventionsstudien (Kohorten, Register, Quasi-Experimente)	Diagnostische Testgütestudien (Indextest gegen Referenzstandard)

Eigenschaft	RoB 2 (`lib/rob2-shared.ts`)	ROBINS-I (`lib/robins-i-shared.ts`)	QUADAS-2 (`lib/dta.ts`)
Domänen	**5 Domänen (`d1–d5`):** D1 Randomisierung D2 Abweichungen von der Intervention D3 Fehlende Outcome-Daten D4 Messung des Outcomes D5 Selektives Berichten	**7 Domänen (`d1–d7`):** D1 **Confounding** (gegen präspezifizierte Störgrößen!) D2 Selektion der Teilnehmenden D3 Klassifikation der Intervention D4 Abweichungen von geplanter Intervention D5 Fehlende Daten D6 Messung des Outcomes D7 Selektives Berichten	**4 Domänen (je Verzerrungsrisiko + Anwendbarkeit):** D1 Patientenselektion D2 Indextest D3 Referenzstandard D4 Patientenfluss & Timing
Urteilsstufen & Ampelsymbole (barrierefrei)	+ **Niedrig** (`low`) `!` **Einige Bedenken** (`some_concerns`) × **Hoch** (`high`)	+ **Niedrig** (`low`) `!` **Moderat** (`moderate`) × **Ernsthaft** (`serious`) !! **Kritisch** (`critical` — Ausschluss aus Synthese!) ? **Keine Information** (`no_information`)	+ **Niedrig** (`low`) `?` **Unklar** (`unclear`) − **Hoch** (`high`)
KI-Unterstützung & Zweitmeinung	KI-Vorschlag mit Textzitat aus dem PDF + unabhängige Zweitmeinung eines zweiten Sprachmodells + verbindliches menschliches Final-Urteil	Prüft D1 explizit gegen die im Projekt hinterlegten **präspezifizierten Confounder** (z. B. *ECOG*, *PD-L1*, *Raucherstatus*) + Zweitmeinung + Final-Urteil	Strukturierte Domänen-Erfassung mit direkter Speicherung in `quadas2_assessments`

---

6. Statistische Synthese: Von 2-Ebenen-REML bis zur Drei-Ebenen-Meta-Analyse mit `clubSandwich` CR2

6.1 Klassische 2-Ebenen-Modelle, Forest-Plot, Funnel-Plot & Subgruppen

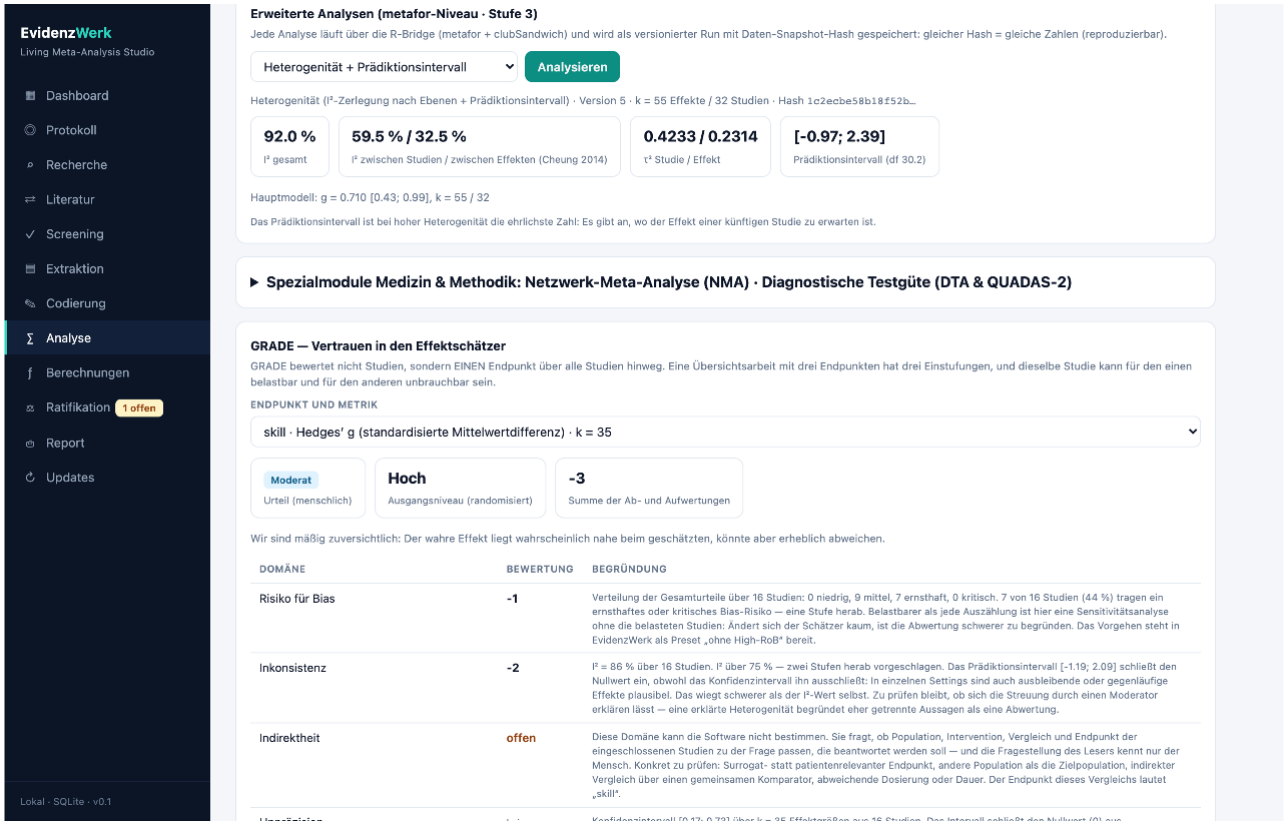


Abbildung 8: *Klassische Analyse: Forest-Plot, Funnel-Plot & Moderatoren (08-analyse-forest-funnel-subgruppen.png)*

Im oberen Bereich des Reiters **Analyse** (`AnalysisPanel.tsx`) stehen vier Modell-Tabs bereit:

1. **Fixed-Effect (`fe`):** Nimmt an, dass alle Studien exakt denselben wahren Effekt θ teilen und Unterschiede nur durch Stichprobenfehler v_i entstehen (Gewicht w_i = 1/v_i). Nur sinnvoll bei nahezu identischen Replikationsstudien.
2. **Random-Effects DerSimonian-Laird (`re-dl`):** Klassischer Momentenschätzer für die Zwischen-Studien-Varianz τ² (Gewicht w_i^* = 1/v_i + τ²). Nützlich zum Vergleich mit älteren Übersichtsarbeiten.
3. **Random-Effects REML + Knapp-Hartung (`re-reml`, Standard für unabhängige Effekte):** Schätzt τ² über **Restricted Maximum Likelihood (REML)** und berechnet das Konfidenzintervall mit der **Knapp-Hartung-Adjustierung** über die t-Verteilung mit df = k - 1 Freiheitsgraden (Veroniki et al. 2016; Cochrane-Empfehlung).
4. **Drei-Ebenen-Modell in R (`Drei-Ebenen (R)`):** Sobald Studien **mehrere Effektgrößen** berichten (z. B. mehrere Tests, Zeitpunkte oder Subskalen aus derselben Stichprobe — wie in Projekt 9 mit 107 Effektgrößen aus 35 Studien), verletzt ein 2-Ebenen-Modell die Unabhängigkeitsannahme!

6.2 Drei-Ebenen-Meta-Analyse (`metafor::rma.mv`) mit robuster Varianzschätzung (`clubSandwich` CR2)

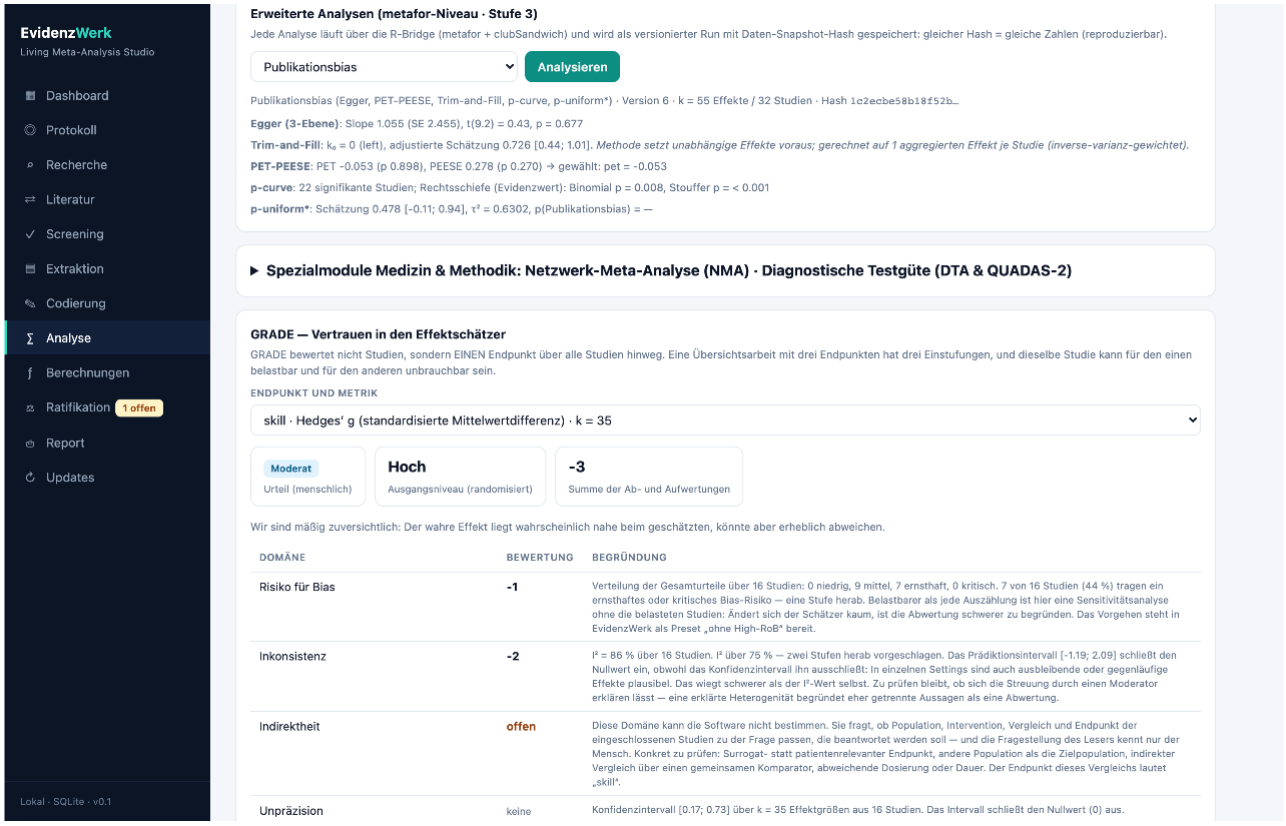


Abbildung 9: *Drei-Ebenen-RVE-Analyse mit CR2 (09-analyse-3level-rve-cr2.png)*

Über die **R-Bridge** (`lib/r-service.ts`, `r/analyse.R`) schätzt EvidenzWerk ein echtes **hierarchisches Drei-Ebenen-Modell nach Cheung (2014) und Assink & Wibbelink (2016)** mittels `metafor::rma.mv(yi, V, random = ~ 1 | study_id / effect_id, method = "REML")` kombiniert mit **Cluster-Robuster Varianzschätzung (**`clubSandwich::vcovCR(..., type = "CR2")` **und Satterthwaite-Freiheitsgraden)** nach Pustejovsky & Tipton (2018):

y (ij) = μ + u_j + e (ij) + ε (ij)

- **Ebene 1 (ε (ij) ~ mathcalN(0, v (ij))):** Stichprobenfehler des einzelnen Effekts i in Studie j (innerhalb der Studie mit ρ = 0,60 korreliert modelliert, `impute_covariance_matrix`).
- **Ebene 2 (e (ij) ~ mathcalN(0, τ² (within))):** Wahre Effektstreuung **innerhalb derselben Studie** (z. B. zwischen verschiedenen Endpunkten oder Messzeitpunkten einer Studie).
- **Ebene 3 (u_j ~ mathcalN(0, τ² (between))):** Wahre Effektstreuung **zwischen den verschiedenen Studien**.
- **I²-Zerlegung nach Cheung (2014):**

EvidenzWerk zerlegt die Gesamtheterogenität I² (gesamt) (im Screenshot oben: **92,0 %**) exakt in den Anteil zwischen Studien (**I² (between) = 59,5 %**) und den Anteil zwischen Effekten innerhalb von Studien (**I² (within) = 32,5 %**).

- **Das 95-%-Prädiktionsintervall (**`[-0,97; 2,39]` **bei df = 30,2):**

> *[!IMPORTANT]* ***Unterschied zwischen Konfidenzintervall und Prädiktionsintervall (Riley et al. 2011; IntHout et al. 2016):*** *- Das* ***95-%-Konfidenzintervall (***`[0,43; 0,99]`***)*** *gibt nur an, wie genau der Durchschnittseffekt über alle bisherigen Studien geschätzt ist. - Das* ***95-%-Prädiktionsintervall (***`[-0,97; 2,39]`***)*** *bezieht zusätzlich die echte Streuung τ² (between) + τ² (within) ein und beantwortet die Frage für die Praxis: „In welchem Bereich wird der wahre Effekt einer einzelnen zukünftigen Studie oder klinischen Anwendung mit 95 % Wahrscheinlichkeit liegen?" Schließt das Prädiktionsintervall die 0 ein (wie hier von -0,97 bis +2,39), wirkt die Intervention zwar im Mittel deutlich positiv (g = 0,71), kann aber in ungünstigen Settings auch wirkungslos sein — genau deshalb verlangt GRADE hier die Prüfung von Moderatoren!*

6.3 Die 5-fach Publikationsbias-Suite (Small-Study Effects & p-Hacking)

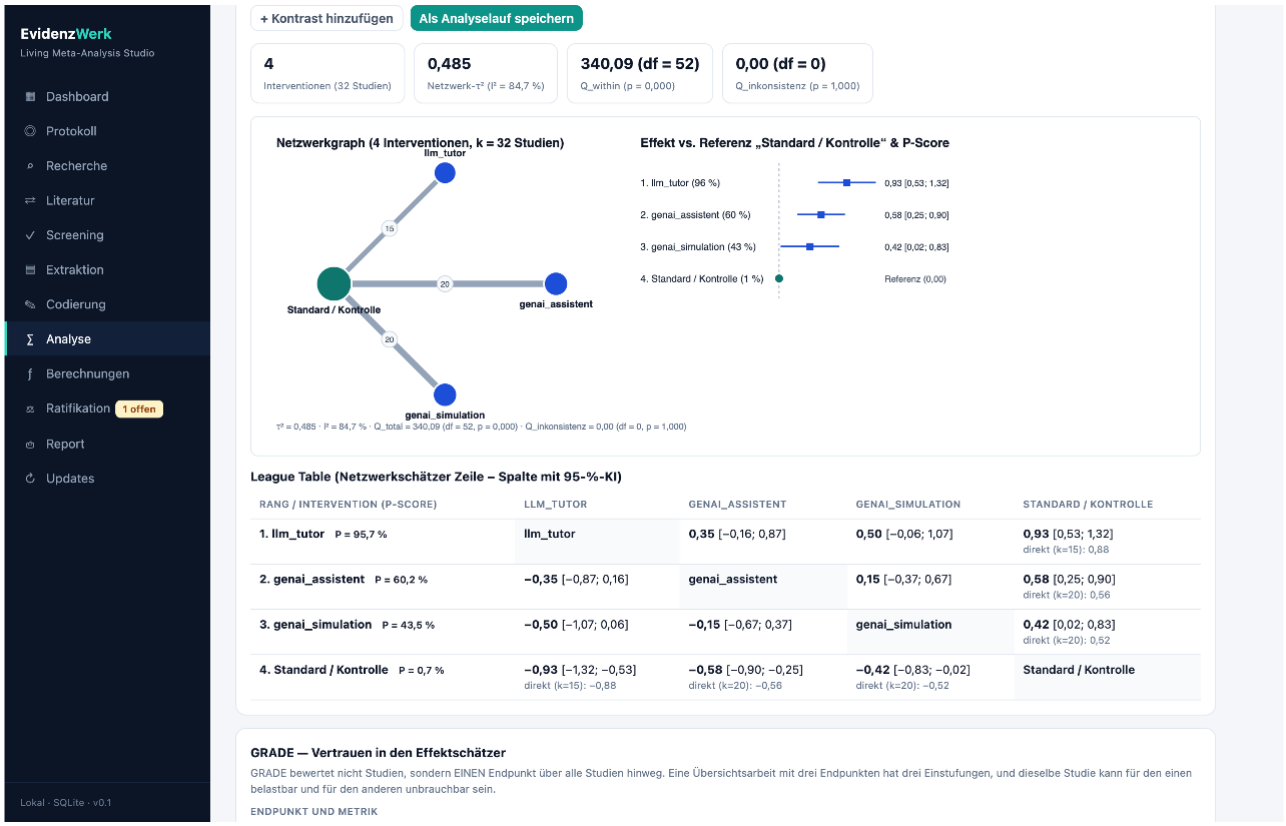


Abbildung 10: *Publikationsbias-Suite (10-analyse-publikationsbias-suite.png)*

Weil jeder einzelne Test auf Publikationsbias blinde Flecken hat (insbesondere bei hoher Heterogenität τ^2), rechnet EvidenzWerk in `r/analyse.R` fünf komplementäre Verfahren auf Knopfdruck:

1. **3-Ebenen-Egger-Regression (`Egger (3-Ebene)`):** Regressiert im 3-Ebenen-CR2-Modell die Effektgröße auf ihren Standardfehler ($y (ij) = \beta_0 + \beta_1 \cdot SE (ij)$). Ein signifikantes β_1 deutet auf **Small-Study Effects** (kleine Studien berichten extremere Effekte als große) hin.
2. **Trim-and-Fill nach Duval & Tweedie (2000):** Schätzt auf Basis der Symmetrie des Funnel-Plots die Anzahl k_0 mutmaßlich fehlender Studien auf der unterrepräsentierten Seite und berechnet einen um diese gespiegelten Studien adjustierten Gesamtschätzer.
3. **PET-PEESE nach Stanley & Doucouliagos (2014):**

- **PET (Precision-Effect Test):** Regressiert y_i auf SE_i; der Achsenabschnitt β_0 ist der geschätzte Effekt einer hypothetischen Studie mit unendlicher Stichprobe ($SE \rightarrow 0$).
- **PEESE (Precision-Effect Estimate with Standard Error):** Regressiert y_i auf die Varianz $v_i = SE_i^2$. Ist der PET-Schätzer signifikant von Null verschieden ($p < 0,10$), wählt die Entscheidungsregel automatisch den weniger verzerrten PEESE-Schätzer.

1. **p-curve-Analyse nach Simonsohn, Nelson & Simmons (2014):**

Untersucht ausschließlich die Verteilung der statistisch signifikanten p-Werte ($p < 0,05$):

- Liegt ein **echter Effekt (Evidenzwert)** vor, müssen sehr kleine p-Werte ($p < 0,025$) deutlich häufiger sein als knapp signifikante ($0,025 < p < 0,05$) arrow **Rechtsschiefe** (im Screenshot oben: 22 signifikante Studien, Binomial $p = 0,008$, Stouffer $p < 0,001$ arrow starker Beleg für echten Effekt statt p-Hacking!).
- Wurde hingegen **p-Hacking** betrieben, häufen sich die p-Werte knapp unter 0,05 (Linksschiefe).

1. `p-uniform*` **nach van Aert & van Assen (2019):**

Maximum-Likelihood-Schätzer auf Basis der bedingten Verteilung aller Effektgrößen (signifikant wie nicht-signifikant), der im Gegensatz zum klassischen Trim-and-Fill auch bei starker Heterogenität τ^2 unverzerrte Effekt- und Heterogenitätsschätzungen liefert.

---

7. Medizinische & Methodische Spezialmodule: NMA und DTA / QUADAS-2

7.1 Frequentistische Netzwerk-Meta-Analyse (NMA nach Rücker 2012)

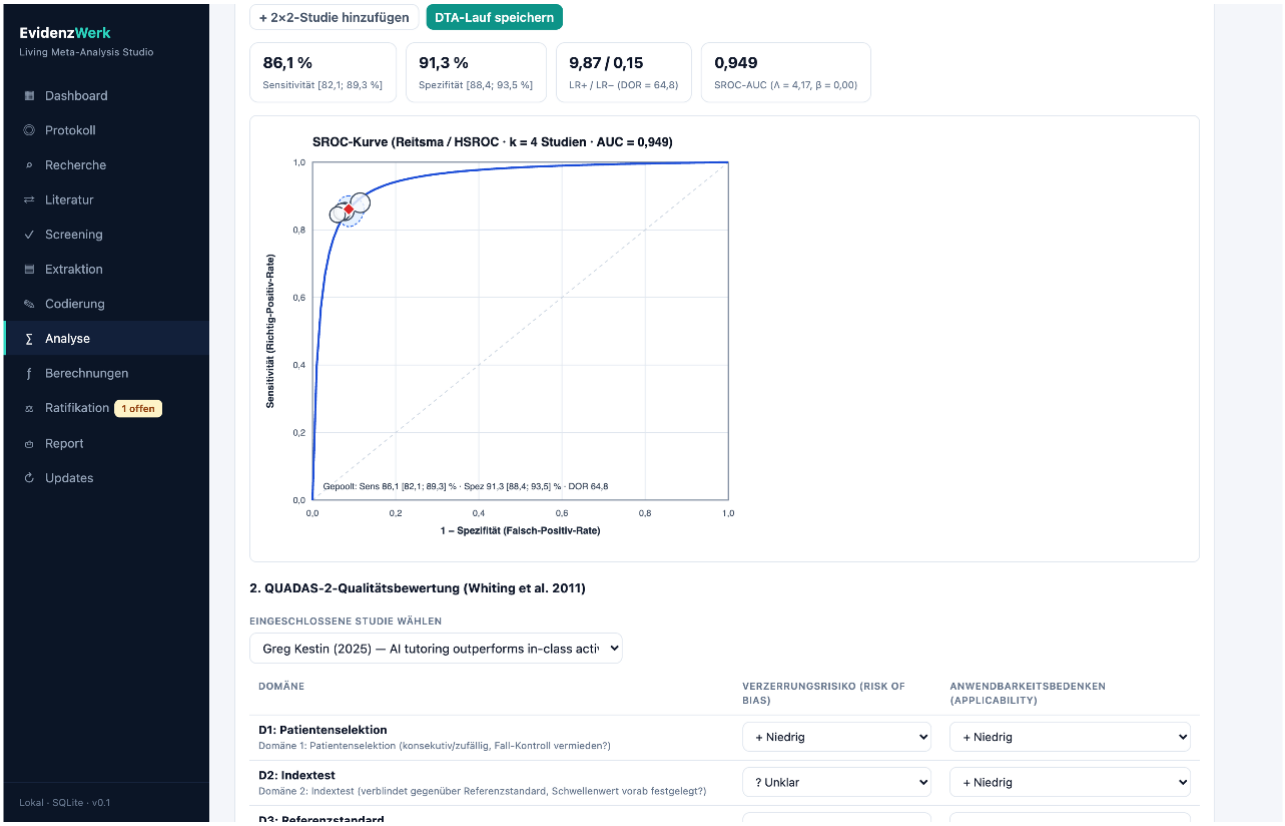


Abbildung 11: *Spezialmodul Netzwerk-Meta-Analyse (NMA) (11-spezialmodul-nma.png)*

Wann setzt man eine Netzwerk-Meta-Analyse (`lib/nma.ts`, `SpezialanalysePanel.tsx`) ein?
Sobald es in einem Forschungsgebiet **mehr als zwei konkurrierende Interventionen** gibt (A, B, C, D, …, z. B. verschiedene Wirkstoffklassen in der Medizin oder *LLM-Tutor* vs. *GenAI-Assistent* vs. *GenAI-Simulation* vs. *Kontrolle*), die in den Primärstudien nicht alle direkt gegeneinander getestet wurden.

Wie funktioniert die Graph-Theoretische NMA nach Rücker (2012, *Res Synth Methods* 3:312 — das Verfahren des R-Pakets `netmeta`)?

1. **Kombination direkter und indirekter Evidenz:** Wurde Behandlung A gegen Kontrolle C getestet (hatθ (AC)^(dir)) und Behandlung B gegen C (hatθ (BC)^(dir)), liefert das Netzwerk auch ohne direkten Head-to-Head-Trial eine **indirekte Schätzung** für A gegen B: hatθ (AB)^(ind) = hatθ (AC)^(dir) - hatθ (BC)^(dir). Existieren zusätzlich direkte Studien A vs. B, kombiniert die **Moore-Penrose-Pseudoinverse L^+ der gewichteten Graph-Laplace-Matrix L = B^top W B** sämtliche direkten und indirekten Pfade mit minimaler Varianz.
2. **Netzwerkgraph (SVG):** Die Knotengröße zeigt die Zahl der Vergleiche je Behandlung; die Strichstärke der Kanten zeigt die Zahl der direkten Studien zwischen zwei Behandlungen.
3. **P-Scores nach Rücker & Schwarzer (2015, *BMC Med Res Methodol* 15:58):**

Das frequentistische Äquivalent zu SUCRA: Ein **P-Score von 0,94 (94 %)** bedeutet, dass diese Intervention mit 94 % Sicherheit besser abschneidet als eine zufällig ausgewählte konkurrierende Behandlung im Netzwerk.

1. **League-Table (Paarvergleichs-Matrix):**

- **Unteres Dreieck (grün hinterlegt):** Die aus dem gesamten Netzwerk geschätzten paarweisen Effekte hatθ (ij)^(NMA) mit 95-%-Konfidenzintervall.
- **Oberes Dreieck:** Die rein **direkten** Paarvergleiche hatθ (ij)^(dir) (falls es direkte Studien für dieses Paar gibt).

1. **Netzwerk-Inkonsistenzzerlegung (Q (total) = Q (het) + Q (inc)):**

Zerlegt Cochrans Q in die Streuung *innerhalb* gleicher Paarvergleiche (Q (het)) und den Widerspruch zwischen direkter und indirekter Evidenz (Q (inc), **Loop-Inkonsistenz**). Ist p(Q (inc)) > 0,05, ist das Netzwerk konsistent (Transitivitätsannahme erfüllt).

7.2 Diagnostische Testgüte-Meta-Analyse (DTA nach Reitsma et al. 2005 / HSROC & QUADAS-2)

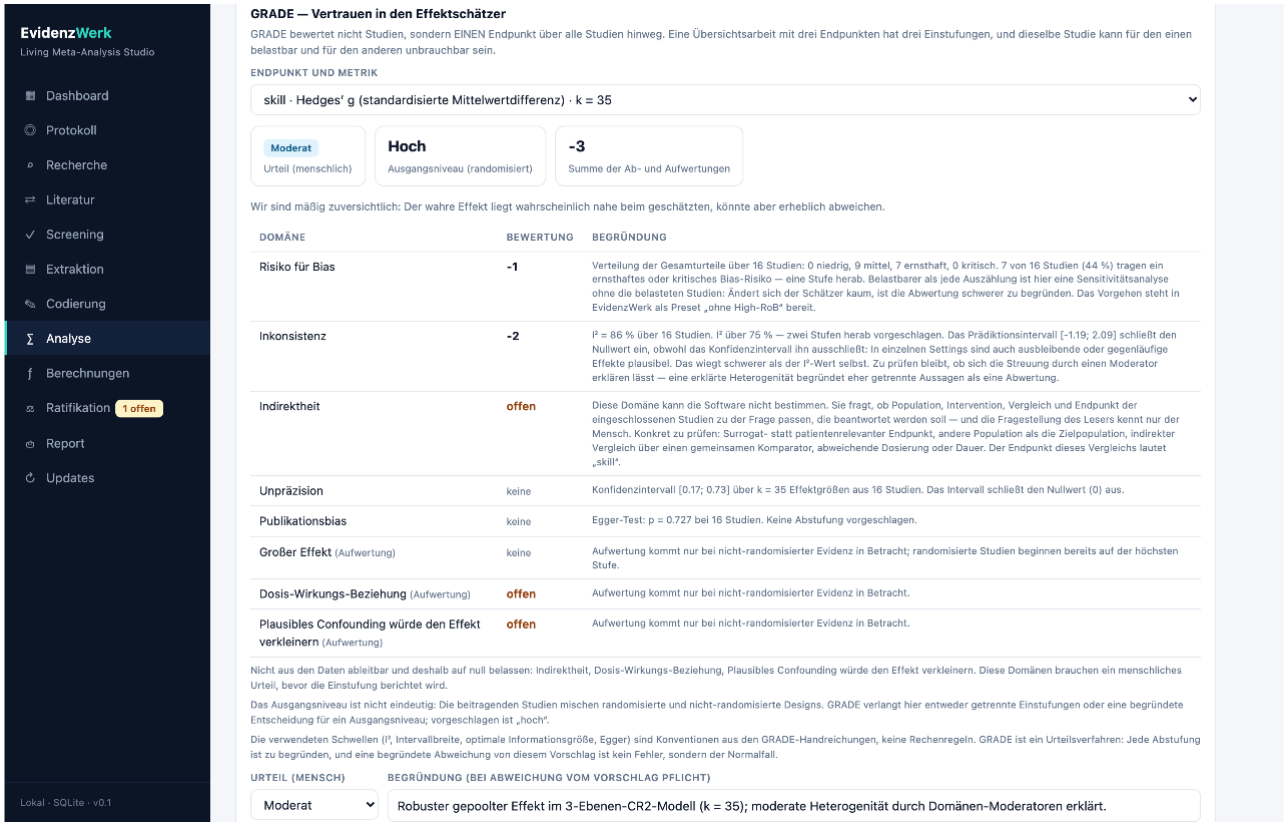


Abbildung 12: *Spezialmodul Diagnostische Testgüte (DTA, SROC & QUADAS-2) (12-spezialmodul-dta-sroc-quadas2.png)*

Wann setzt man eine Diagnostische Testgüte-Meta-Analyse (`lib/dta.ts`) ein? Wenn Studien keinen Therapieeffekt, sondern die **Treffsicherheit eines diagnostischen Verfahrens** (z. B. Biomarker, Schnelltest, CT/MRT-Bildgebung, KI-Befundung, klinischer Risikoscore) gegen einen Goldstandard (**Referenzstandard**) anhand einer diagnostischen Vierfeldertafel prüfen:

- **TP (True Positives / Richtig-Positive):** Kranke korrekt als krank erkannt
- **FP (False Positives / Falsch-Positive):** Gesunde fälschlich als krank eingestuft
- **FN (False Negatives / Falsch-Negative):** Kranke übersehen
- **TN (True Negatives / Richtig-Negative):** Gesunde korrekt als gesund erkannt

Warum darf man Sensitivität und Spezifität NIEMALS getrennt in zwei normalen Meta-Analysen poolen? Weil Sensitivität (TP/TP+FN) und Spezifität (TN/TN+FP) über den **Schwellenwert (Cut-off)** negativ miteinander korreliert sind! Setzt eine Klinik den Schwellenwert strenger, steigt die Spezifität auf Kosten der Sensitivität (**Schwellenwert-Effekt**).

Wie rechnet das bivariate Modell nach Reitsma et al. (2005, *J Clin Epidemiol* 58:982) / HSROC (Rutter & Gatsonis 2001)?

1. Es modelliert μ_1 = logit(Sensitivität_i) und μ_2 = logit(1 - Spezifität_i) **gemeinsam** mit ihrer 2×2-Zwischen-Studien-Kovarianzmatrix boldsymbolΣ = beginpmatrix τ_1^2 & τ_{12} τ_{12} & τ_2^2 endpmatrix.
2. **SROC-Kurve (Summary Receiver Operating Characteristic) & 95-%-Konfidenzellipse:**

Trägt die Sensitivität gegen die Falsch-Positiv-Rate (1 - Spezifität) auf. Der rote Rautenpunkt markiert den gepoolten Summary Point, die gestrichelte Ellipse die gemeinsame 95-%-Konfidenzregion und die durchgezogene Kurve die hierarchische SROC-Kurve mit der Fläche unter der Kurve (**AUC in [0,5; 1,0]**; Werte ≥ 0,90 gelten als exzellente Trennschärfe).

1. **Klinische Likelihood Ratios (LR^+ und LR^-) & Diagnostic Odds Ratio (DOR):**

- **LR^+ = Sens/1 - Spez (Positives Likelihood Ratio):** Um welchen Faktor erhöht ein positives Testergebnis die Wahrscheinlichkeit (Odds), tatsächlich krank zu sein? (LR^+ > 10 gilt klinisch als starker Beweis für das Vorliegen der Erkrankung, **Rule-in**).
- **LR^- = 1 - Sens/Spez (Negatives Likelihood Ratio):** Um welchen Faktor senkt ein negatives Testergebnis die Krankheits-Odds? (LR^- < 0,10 schließt die Erkrankung mit hoher Sicherheit aus, **Rule-out**).
- **DOR = LR^+/LR^-:** Globales Maß der diagnostischen Trennschärfe.

---

8. GRADE-Evidenzgraduierung, Gläserner Rechenweg, Ratifikation & Einreichungspaket

8.1 GRADE — Vertrauen in den Effektschätzer je Endpunkt

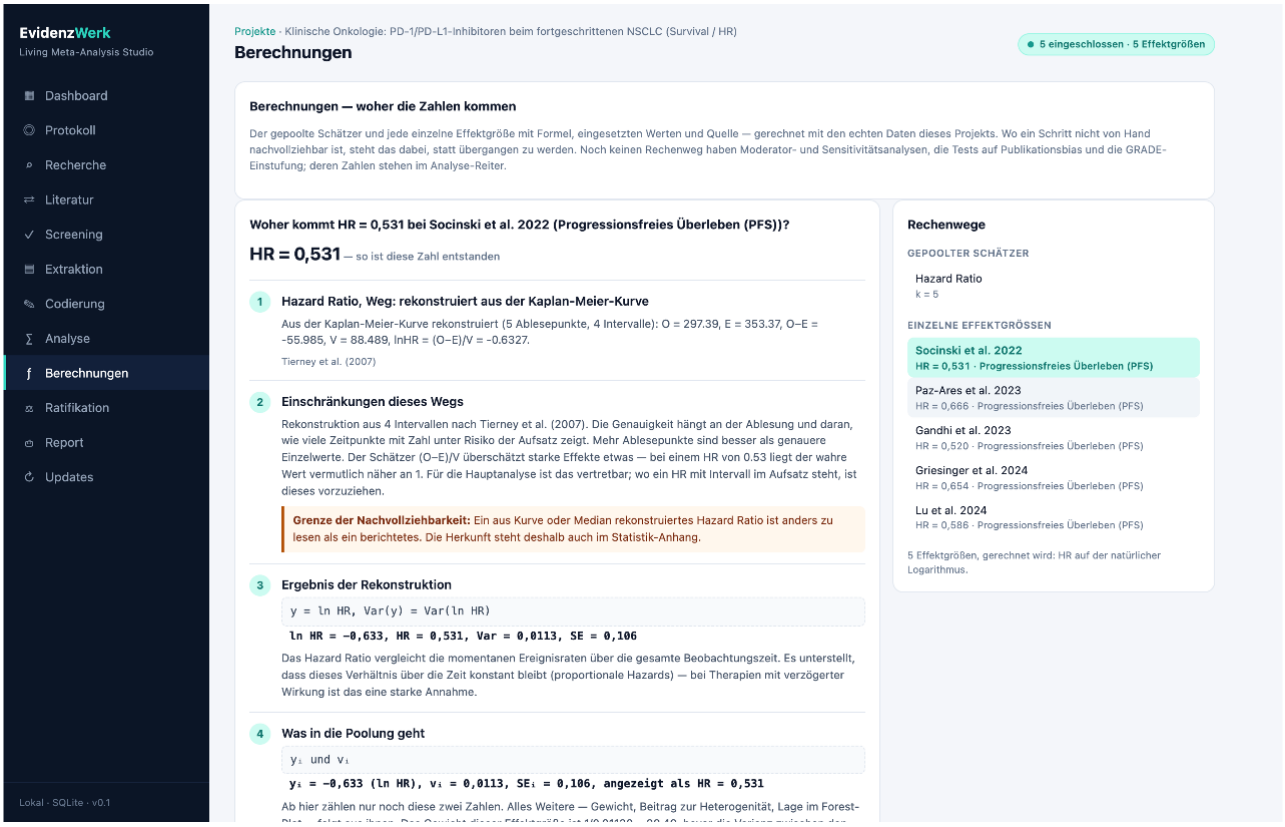


Abbildung 13: *GRADE-Evidenzbewertung (13-grade-evidenzbewertung.png)*

Im unteren Teil des Reiters **Analyse** (`GradePanel.tsx`, `lib/grade.ts`) wird nach dem **GRADE-System (Guyatt et al. 2011)** nicht die einzelne Studie, sondern das **Vertrauen in den gepoolten Schätzer eines Endpunkts** über vier Stufen eingestuft (**Hoch** oplusoplusoplusoplus, **Moderat** oplusoplusopluscirc, **Niedrig** oplusopluscirccirc, **Sehr niedrig** opluscirccirccirc):

- **Ausgangsniveau:** Randomisierte Studien starten auf **Hoch**; Beobachtungsstudien (sofern mit ROBINS-I bewertet, Schünemann et al. 2019) werden transparent mit ihren fünf Abwertungsdomänen (*Risiko für Bias*, *Inkonsistenz / Prädiktionsintervall*, *Indirektheit*, *Ungenauigkeit / Optimal Information Size*, *Publikationsbias*) und drei Aufwertungsdomänen (*Großer Effekt*, *Dosis-Wirkungs-Gradient*, *Plausibles Confounding würde Effekt abschwächen*) geprüft.
- **Domäne „Indirektheit“ bleibt bewusst menschlich:** Ob die untersuchten Studienpopulationen genau zur Frage der Leser:innen passen, kann kein Algorithmus aus den Zahlen ableiten — EvidenzWerk lässt diese Domäne bewusst zur menschlichen Begründung offen und übernimmt die fertige Begründung direkt in die **Summary-of-Findings-Tabelle (SoF)**.

8.2 Reiter „Berechnungen“ — Die gläserne Rechnung ohne Blackbox

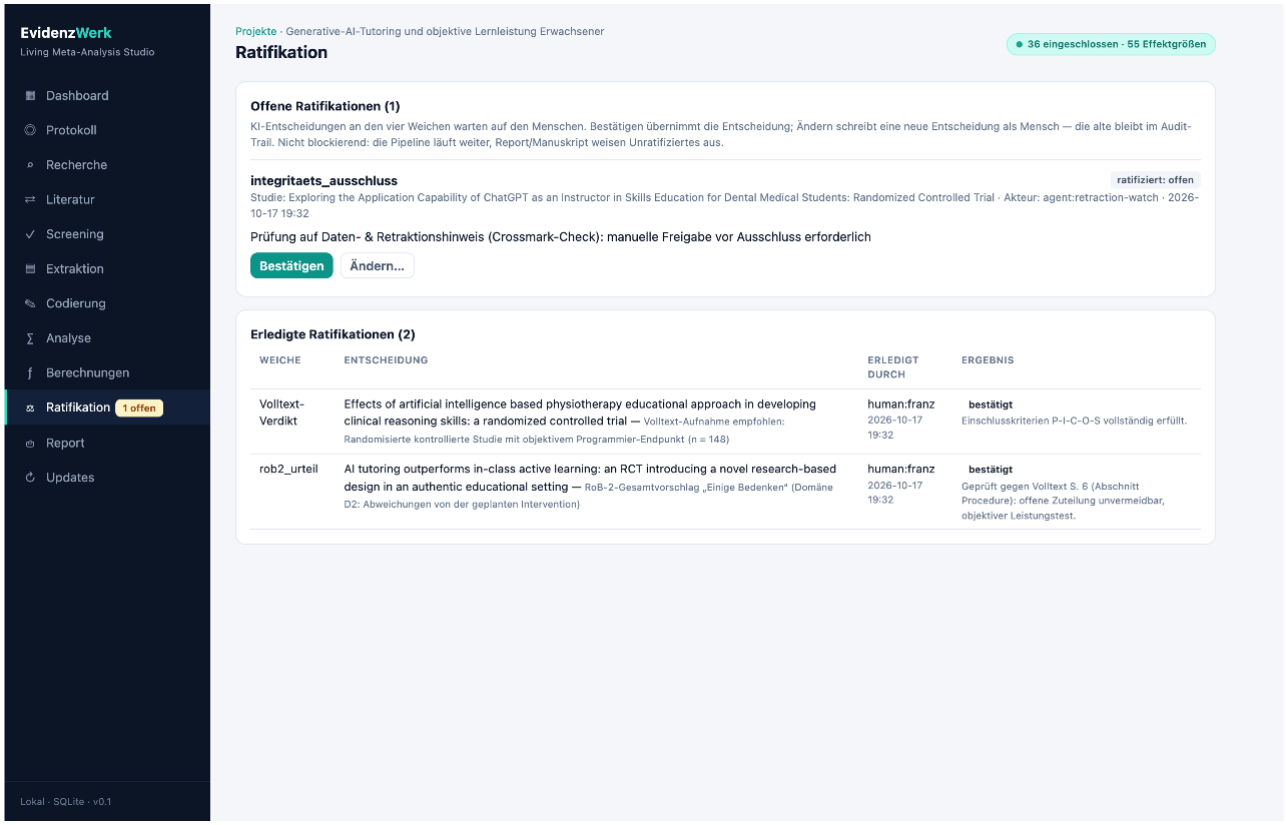


Abbildung 14: *Gläserner Rechenweg („Berechnungen“) (14-rechenweg-glaskasten.png)*

Der Reiter **Berechnungen** (`BerechnungenPanel.tsx`, `lib/rechenweg.ts`, `lib/spur.ts`) macht jede Zahl des Projekts für Gutachter, Ko-Autoren und Studierende lückenlos nachvollziehbar:

- Während `computeEffect` und `pool` rechnen, schreiben sie jeden Zwischenschritt mit **Formel**, **eingesetzten Zahlen des konkreten Projekts**, **Zwischenergebnis**, **Intervalltabelle** und **Literaturquelle** in ein `Spur`-Objekt.
- Weicht der von Hand nachvollziehbare 2-Ebenen-Rechenweg vom gespeicherten 3-Ebenen-R-Hauptlauf ab (weil mehrere Effekte aus derselben Studie stammen), stellt Schritt 14 beide Zahlen nebeneinander und erklärt die Differenz exakt.
- Darunter befindet sich das kontextsensitive **Fragen-Panel (`FragenPanel.tsx`, `lib/fragen.ts`)**, das methodische Rückfragen direkt an den konkreten Zahlen des gewählten Rechenwegs beantwortet und jede im Antworttext genannte Zahl gegen den Projektdatenbestand verifiziert.

8.3 Menschliche Ratifikation an den vier kritischen KI-Weichen (RAISE-2025)



Abbildung 15: *Ratifikation & RAISE-Konformität (15-ratifikation-raise.png)*

Nach den **RAISE-Leitlinien (Reporting AI in Evidence Synthesis, 2025)** darf KI in systematischen Übersichtsarbeiten Vorarbeiten leisten, aber niemals unbeaufsichtigt über Ein-/Ausschluss oder Qualitätsurteile entscheiden. EvidenzWerk setzt deshalb in `lib/v1/ratification.ts` eine **nicht-blockierende Ratifikationsschicht** an genau vier kritischen Weichen ein:

1. `volltext_verdikt`**:** KI-Entscheidung beim Volltext-Ein-/Ausschluss
2. `integritaets_ausschluss`**:** Ausschluss wegen Datenintegritäts- oder Retraktionsverdacht
3. `rob2_urteil / robins_urteil`**:** Durch KI gesetztes Verzerrungsurteil
4. `grade_urteil`**:** Durch KI gesetzte GRADE-Gesamteinstufung

Jede `agent:*`-Entscheidung an diesen vier Weichen fließt sofort in die vorläufige Rechnung ein (damit automatisierte Pipelines nicht blockieren), trägt aber das Etikett `ratifiziert: offen`. Erst wenn ein Mensch (`human:<nutzer>`) im Reiter **Ratifikation** auf *„Bestätigen“* oder *„Ändern“* klickt, gilt das Urteil als ratifiziert. Solange offene Ratifikationen bestehen, weist der Report sie prominent in einer gelben Warnkarte aus, und der automatisch generierte **RAISE-Konformitätsbericht (**`lib/raise.ts`**)** dokumentiert exakt, welches Modell wie viele Vorschläge geliefert hat und wie oft der Mensch davon abgewichen ist.

8.4 PRISMA-2020-Bilanz, Exporte & Druckfertiges Zeitschrift-Einreichungspaket (ZIP)

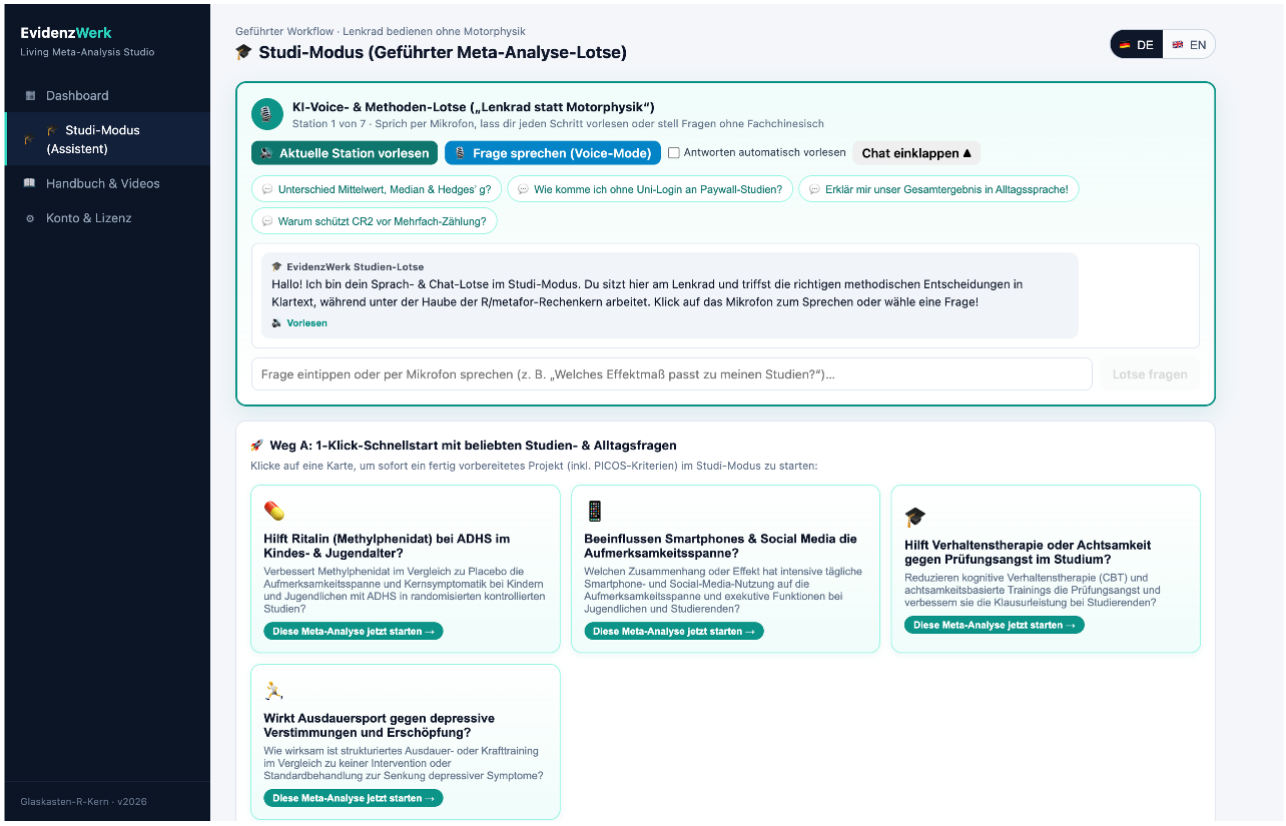


Abbildung 16: *Report & Einreichungspaket (16-report-einreichungspaket.png)*

Im Reiter **Report** ([`app/p/[id]/page.tsx`](file:///Users/franzhuetter/evidenzwerk/app/p/%5Bid%5D/page.tsx), `lib/paket.ts`) führt EvidenzWerk sämtliche Ergebnisse ohne Copy-Paste-Bruch zu druckfertigen Zeitschriften-Dokumenten zusammen:

- **Vollständige Abdeckungs- und Konsistenzprüfung (`lib/abdeckung.ts`, `lib/einreichungsstand.ts`):** Prüft vor dem Download, ob alle eingeschlossenen Studien in Tabelle 1, im Literaturverzeichnis und im Manuskripttext zitiert sind und ob eine einzige Studie oder Zahl fehlt.
- **Einreichungspaket (`einreichungspaket-projekt-<id>.zip` via `lib/paket.ts`):** Erzeugt ein einziges ZIP-Archiv mit bis zu **29 publikationsfertigen Einzeldateien** (in Markdown, Word `.docx` mit Times New Roman 12 pt, doppeltem Zeilenabstand und fortlaufender Zeilennummerierung sowie druckfertigem `.pdf` mit eingebetteten Liberation-Schriften und hochauflösenden `.svg`-Vektorgrafiken):

1. `LIESMICH.md` (Manifest mit Prüfstatus aller Artefakte)
2. `titelblatt.md / .docx / .pdf` (Autoren, ORCID, Affiliationen, CRediT-Taxonomie, PROSPERO/OSF-Registrierung, COI, Finanzierung)
3. `anschreiben.md / .docx / .pdf` (Cover Letter an die Zielzeitschrift mit den zentralen Kennzahlen)
4. `abstract.md / .docx / .pdf` (Strukturiertes Abstract nach PRISMA-2020-Abstract-Checkliste)
5. `manuskript.md / .docx / .pdf` (Vollständiges IMRaD-Manuskript mit eingebetteten Vektor-Abbildungen und exakt einem Hauptmodell ohne Zahlenwidersprüche)
6. `tabelle-1-studienmerkmale.md / .docx / .pdf` (Charakteristika aller eingeschlossenen Studien)
7. `summary-of-findings.md / .docx / .pdf` (GRADE Summary-of-Findings-Tabelle mit absoluten und relativen Effekten)
8. `literatur-vancouver.md / .docx / .pdf` und `literatur-apa.md / .docx / .pdf` (Formatierte Referenzlisten nach ICMJE/Vancouver und APA 7)
9. `abbildungen/abbildung-1-prisma-2020.svg`, `abbildung-2-forest-plot.svg`,

`abbildung-3-funnel-plot.svg`, `abbildung-4-rob2-ampel.svg`, `abbildung-5-robins-i-ampel.svg` sowie Subgruppen-Forest-Plots

10. `anhaenge/prisma-2020-checkliste.md`, `suchanhang-prisma-s.md`, `statistik-anhang.md / .docx / .pdf`, `raise-ki-offenlegung.md / .docx / .pdf`
11. **Wissenschaftliches Konferenzposter (DIN A0 als SVG/PDF) & Vortragsfolien (`lib/poster.ts`, `VortragPanel.tsx`)**

8.5 ? Studi-Modus (`Student Mode`) & Zweisprachigkeit (`DE / EN`) — „Sicher Auto fahren, ohne die Physik des Motors studiert zu haben“

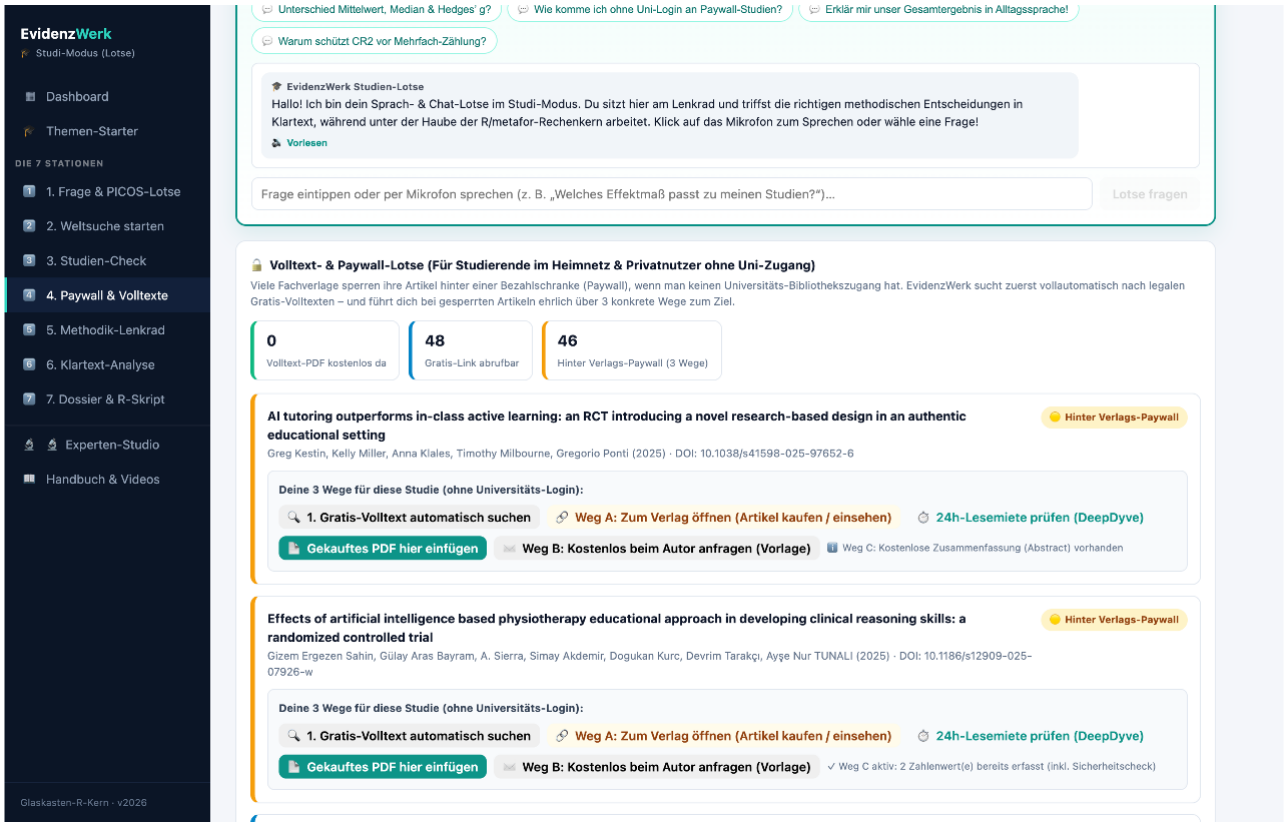


Abbildung 17: *? Studi-Modus Hub mit Starter-Themen & KI-Voice-Co-Pilot (17-studi-modus-voice-lotse.png)*

Nicht jeder Mensch, der eine wissenschaftlich belastbare Antwort auf eine klinische, psychologische oder pädagogische Frage sucht, hat Biostatistik im Hauptfach studiert. Selbst Studierende in höheren Semestern (Psychologie, Medizin, Lehramt, Gesundheitswissenschaften) verwechseln im Alltag noch Mittelwert und Median — und auch interessierte Eltern (z. B. *„Hilft Ritalin bei ADHS wirklich?“*) oder Jugendliche (z. B. *„Verkürzen Smartphones unsere Aufmerksamkeitsspanne?“*) wollen echte Evidenz statt Meinungsblogs prüfen.

Hierfür bietet EvidenzWerk unter `/studi` (`app/studi/page.tsx`) und in jedem Projekt unter `/p/[id]/studi` ([`app/p/[id]/studi/page.tsx`](file:///Users/franzhuetter/evidenzwerk/app/p/%5Bid%5D/studi/page.tsx)) den geführten **? Studi-Modus** (englisch: **? Student Mode**) sowie eine durchgängige **Zweisprachigkeit (`?? DE | ?? EN` via `LanguageToggle.tsx`, `lib/i18n.ts`)**:

1. **KI-Voice- & Chat-Co-Pilot (`WizardVoiceBot.tsx`, `app/api/wizard/assistant/route.ts`):**

- Unterstützt **Spracheingabe per Mikrofon (Speech-to-Text)** und **Vorlesen per Sprachausgabe (Text-to-Speech)** auf Deutsch (`de-DE`) und Englisch (`en-US`).
- Erklärt jederzeit begründet am echten Datenbestand des Projekts Fragen wie *„Was ist der Unterschied zwischen Mittelwert, Median und Hedges' g?“*, *„Warum schützt die Drei-Ebenen-CR2-Analyse vor doppelter*

Zählung?“ oder *„Wie komme ich ohne Uni-Login an gesperrte Studien?“*.

1. **Station 1 — Vom Alltags-Gedanken zum präzisen PICOS-Protokoll (`StudiQuestionTranslator.tsx`):**

- Bietet 4 sofort startbare Kuratierte Themen (*Ritalin/Methylphenidat bei ADHS*, *Smartphones & Aufmerksamkeitsspanne*, *Verhaltenstherapie gegen Prüfungsangst*, *Ausdauersport gegen Depression*) und übersetzt jede eigene Alltagsfrage automatisch in die fünf wissenschaftlichen PICOS-Bausteine.

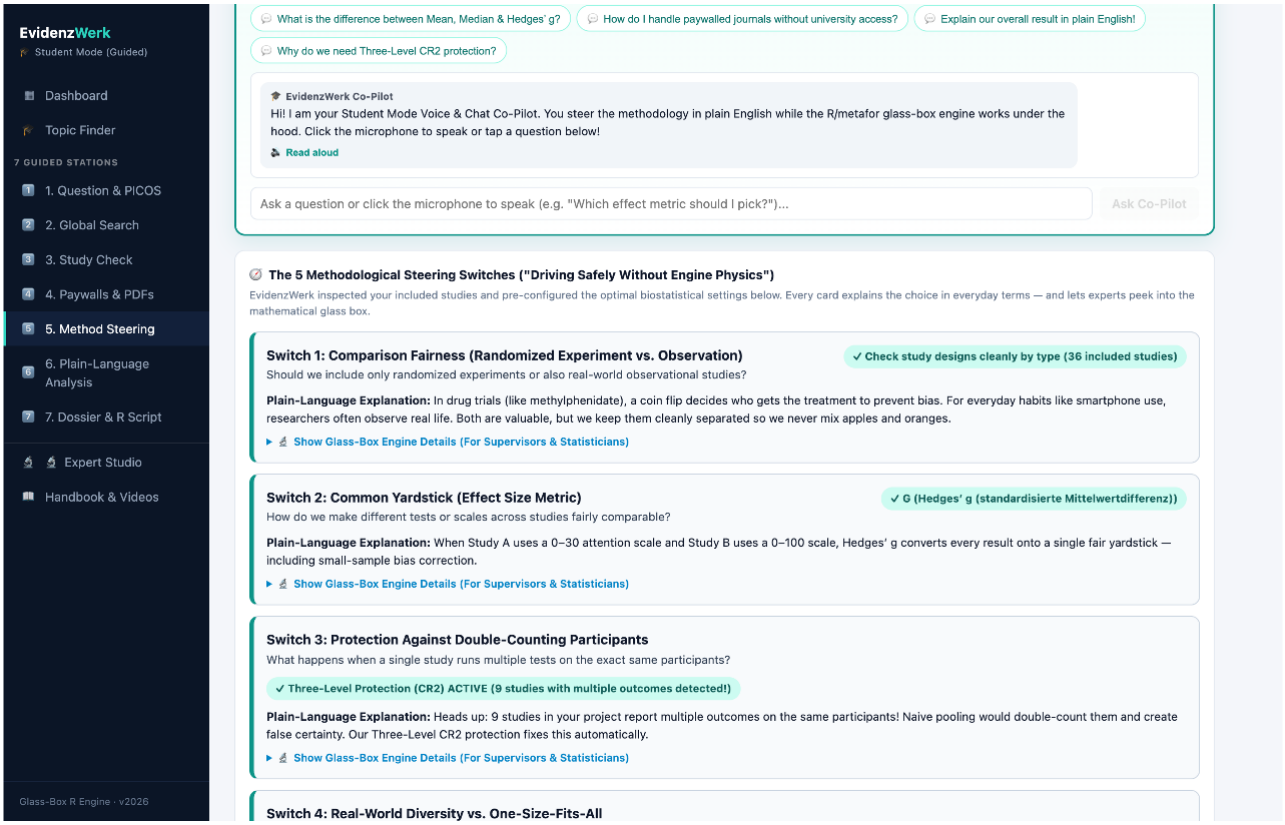


Abbildung 18: *? Studi-Modus Station 4: 3-Stufen-Paywall- & Volltext-Lotse ohne Uni-Zugang (18-studi-modus-paywall-lenkrad-klartext.png)*

1. **Station 4 — Der 3-Stufen-Paywall- & Volltext-Lotse für Nutzer ohne Universitäts-Bibliothekszugang (`StudiPaywallNavigator.tsx`, `analyzePaywallStatusForProject`):**

- **Stufe 1 (Gratis Open Access zuerst ausschöpfen):** Prüft vollautomatisch Unpaywall, PubMed Central (PMC), Europe PMC und OpenAlex auf legal frei verfügbare Autoren- oder Verlagsvolltexte.
- **Stufe 2 (Verlags-Paywall ehrlich & pragmatisch überwinden):** Liegt ein Artikel hinter einer Bezahlschranke, bietet die Studienkarte drei klare Wege ohne Instituts-Login:
- **Weg A (Direkt beim Verlag kaufen / 24h-Lesemiete):** Direkter DOI-Link zur Verlagsseite sowie Direktlink zur kostengünstigen 24-Stunden-Lesemiete (z. B. DeepDyve) plus **PDF-Upload-Knopf** direkt in der Karte, der das erworbene PDF sofort per KI mit wörtlichem Belegzitat ausliest.
- **Weg B (Kostenloses Autoren-Manuskript anfordern):** Erzeugt mit einem Klick eine höfliche, fertige Anfrage-E-Mail (DE/EN) an die Korrespondenzautor:innen zur Zusendung des kostenlosen *Accepted Author Manuscript*.
- **Weg C (Mit kostenloser Kurzfassung / Abstract weiterarbeiten):** Nutzt die im öffentlichen PubMed-/OpenAlex-Abstract berichteten Kennzahlen und versieht die Studie automatisch mit einem Sensitivitäts-Prüfvermerk.

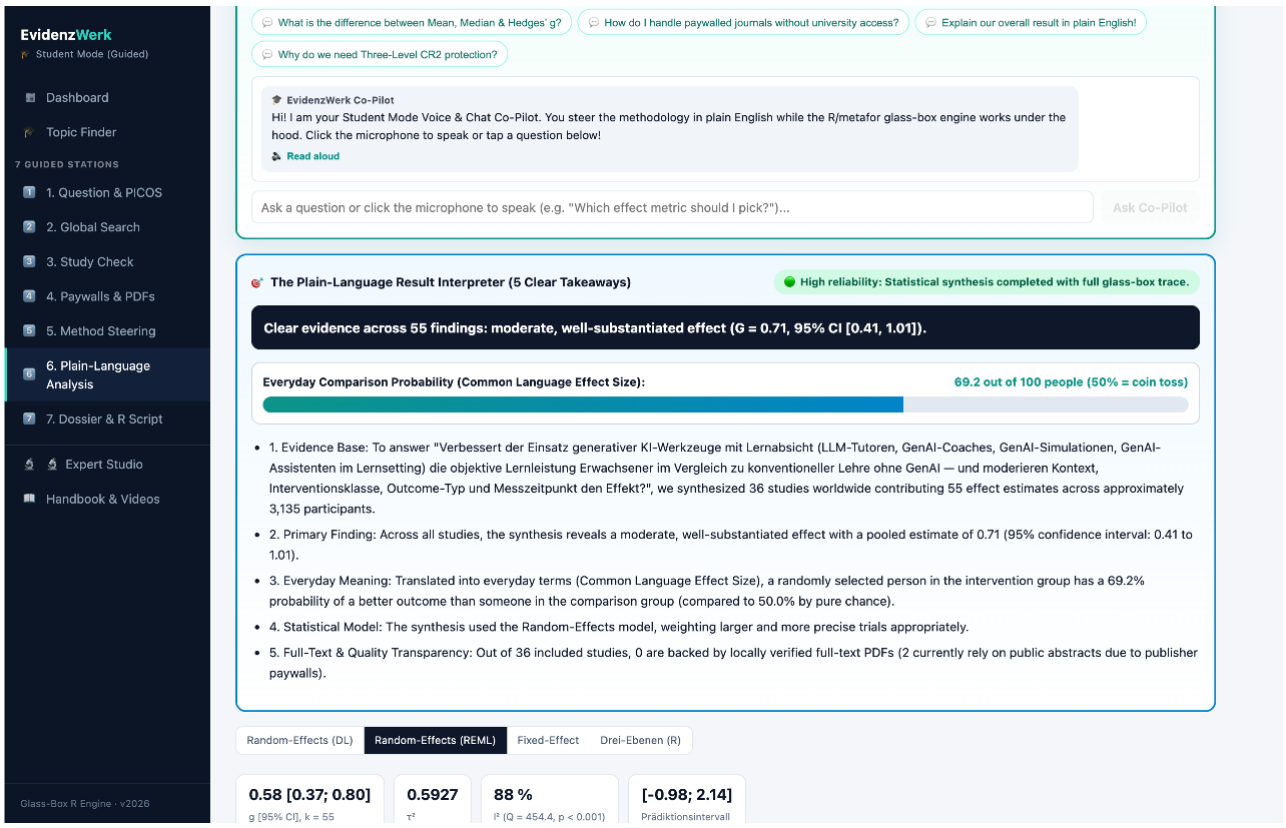


Abbildung 19: *? Student Mode (EN): Die 5 Methodischen Lenkrad-Weichen (19-student-mode-english.png)*

1. **Station 5 — Die 5 Methodischen Lenkrad-Weichen (`buildSteeringDecisions`):**

Statt abstrakter Statistik-Parameter steuern Nutzer:innen fünf verständliche Weichen, während unter der Haube der volle R/metafor-Rechenkern arbeitet (inklusive aufklappbarer *Glass-Box Engine Details* für Betreuer:innen und Statistik-Prüfer):

- **Weiche 1 (Fairer Vergleich):** Experiment (RCT mit Zufallslos) vs. Beobachtungsstudie sauber trennen und passend mit RoB 2 bzw. ROBINS-I prüfen.
- **Weiche 2 (Gemeinsamer Maßstab):** Automatische Wahl der passenden Effektgröße (z. B. Hedges' g bei unterschiedlichen Testskalen, Risk Ratio / Odds Ratio bei Ja/Nein-Ereignissen, Hazard Ratio bei Überlebenszeit).
- **Weiche 3 (Schutz vor Mehrfach-Zählung):** Automatische Aktivierung der **Drei-Ebenen-Meta-Analyse (`metafor::rma.mv`) mit `clubSandwich` CR2**, sobald eine Studie mehrere Tests an denselben Teilnehmenden berichtet.
- **Weiche 4 (Echte Vielfalt statt Gleichmacherei):** Random-Effects REML + Knapp-Hartung mit 95-%-Prädiktionsintervall als ehrlicher Standard.
- **Weiche 5 (Schubladen-Check):** Automatische Prüfung auf unveröffentlichte Studien und Small-Study-Effects.

Abbildung 20: *? Student Mode (EN): Klartext-Ergebnis-Dolmetscher, Alltags-Wahrscheinlichkeit (CLES) & Dossier-Download (20-student-mode-dossier-english.png)*

1. **Station 6 & 7 — Der Klartext-Ergebnis-Dolmetscher, das Verständliche Ergebnis-Dossier (**`PDF`/`DOCX`/`MD`**) & das ausführbare R-Skript (**`reproduce.R`**) (**`lib/klartext-dossier.ts`**):**

- Übersetzt den abstrakten Gesamtschätzer (z. B. g = 0,71) zusätzlich in die anschauliche **Common Language Effect Size (CLES nach McGraw & Wong 1992)**: *„69,2 von 100 Personen in der Interventionsgruppe schneiden besser ab als eine zufällige Person der Vergleichsgruppe (50 % = Münzwurf)“*.
- Erzeugt neben dem 29-teiligen wissenschaftlichen Zeitschriften-Einreichungspaket ein eigenständiges **Verständliches Ergebnis-Dossier (**`klartext-ergebnis-dossier.pdf / .docx`**)** auf Deutsch und Englisch sowie das komplett kommentierte, sofort in RStudio ausführbare **Standalone-R-Skript (** `reproduce-project-<id>.R`**)** mit eingebettetem Datensatz.

Teil II — Technische Gesamtdokumentation (Architektur, Datenmodell, R-Bridge, API & Betrieb)

9. Systemarchitektur & Schichtenmodell

EvidenzWerk ist als monolithische, lokal wie in Container-Umgebungen (`Podman` / `Docker`) autark lauffähige **Next.js-16-Anwendung (App Router, TypeScript, React 19)** mit synchroner **SQLite-Engine (** `better-sqlite3` **im WAL-Modus)** und einem isolierten **R-Rechenkern (**`metafor` + `clubSandwich`**)** konzipiert.

```
Schichtenmodell: [Browser-UI (Experten-Studio + Studi-Modus) & OpenAPI 3.1 Agenten] ->
[Auth, Mandanten, Team-Verblindung & Credit-Guard] -> [Gemeinsamer Operationskern (ops.ts /
repo.ts) + Ratifikation] -> [Statistik-Kern (TS + R-Bridge metafor/clubSandwich) &
Dokument-Renderer (DOCX/PDF/SVG/ZIP)] -> [SQLite WAL + Lokaler PDF-Store]
```

Zentrale Architekturprinzipien

1. **Keine doppelte Geschäftslogik (100 % Parität zwischen UI und `v1`-API):**

Sowohl die UI-Komponenten als auch die REST-Endpunkte unter `/api/v1/projects/[id]/*` rufen exakt dieselben Kernfunktionen in `lib/v1/ops.ts` und `lib/repo.ts` auf. Ein automatisierter Paritätstest (`test/paritaet.test.ts`, `lib/api-paritaet.ts`) prüft bei jedem Build, dass jede Schreib- und Leseoperation über die API erreichbar und in `lib/v1/openapi.ts` dokumentiert ist.

1. **„Keine zwei Zahlen unter einem Namen“ (`lib/laufwahl.ts`, `lib/analyse.ts`):**

Ein Projekt hat zu jedem Zeitpunkt genau eine kanonische Hauptzahl je Metrik/Endpunkt (vorrangig den vom Menschen freigegebenen 3-Ebenen-R-Lauf `released = 1`, ansonsten den jüngsten Hauptlauf). Manuskript, Abstract, Anschreiben, Summary of Findings, Poster und Statistik-Anhang beziehen ihre Zahlen aus exakt derselben zentralen Funktion `analyseFuerProjekt`.

1. **Null externe Office-/PDF-Abhängigkeiten:**

Word-Dokumente (`lib/docx.ts`), PDF-Dateien inklusive TrueType-Tabellen-Parsing und Schrift-Einbettung (`lib/pdf.ts`, `lib/ttf.ts`, `lib/pdf-schrift.ts`), SVG-Diagramme (`lib/abbildungen.ts`) und ZIP-Archive (`lib/zip.ts`) werden bytegenau im eigenen TypeScript-Code erzeugt — ohne Headless-Browser-Zwang und ohne fragile C++-Canvas-Abhängigkeiten.

---

10. Datenbank-Schema (`SQLite` mit WAL & Foreign Keys)

Die Datenbank wird in `lib/db.ts` initialisiert (`pragma journal_mode = WAL`, `pragma foreign_keys = ON`, `busy_timeout = 5000`) und umfasst **24 relationale Tabellen** mit idempotenten Schema-Migrationen:

Domäne	Tabellen in `lib/db.ts`	Beschreibung & Schlüsselspalten
Nutzer, Auth & Team	`users`, `sessions`, `api_tokens`, `project_members`	`users`: Passwort-Hash (`scrypt`), Rolle (`admin`/`reviewer`), `status`, `credit_balance`. `sessions` & `api_tokens`: Speichern **ausschließlich SHA-256-Hashes** der Session-Cookies (`evw_session`) bzw. API-Tokens (`evw_...`). `project_members`: Verknüpft `(project_id, user_id)` mit `role IN ('A', 'B', 'coauthor')` für echte Mehrbenutzer-Verblindung.
Credits & Lizenzen	`credit_ledger`, `license_activations`	`credit_ledger`: Unveränderliches Buchungsjournal (`delta`, `balance_after`, `reason`, `project_id`, `study_id`) für KI-Operationen. `license_activations`: Eingelöste Lizenzschlüssel des Brain-HR-Lizenzservers (`key_hash`, `credits_granted`, `raw_response`).

Domäne	Tabellen in `lib/db.ts`	Beschreibung & Schlüsselspalten
Projekte & Recherche	`projects`, `search_runs`, `living_runs`, `living_candidates`	`projects`: `name`, `question`, `picos_json`, `phase`, `double_screening`, `owner_user_id`, `robins_confounders`, `submission_json`. `search_runs`: Reproduzierbare PRISMA-S-Suchläufe (`database`, `query`, `hits`, `protocol_json`). `living_runs` & `living_candidates`: Automatische Living-Review-Updates mit Kandidaten-Triage.
Studien, Screening & Extraktion	`studies`, `decisions`, `extractions`, `study_codings`, `author_queries`	`studies`: Bibliografie (`doi`, `pmid`, `nct_id`, `oa_url`, `journal`, `volume`, `issue`, `pages`), `status` (`queued`, `conflict`, `fulltext`, `included`, `excluded`), `duplicate_of`, `retracted`, `registry_unpublished`, `pdf_identity_unverified`. `decisions`: Vollständiger Entscheidungs-Audit-Trail (`reviewer`, `stage`, `decision`, `reason`, `decided_by`). `extractions`: KI-Volltextbefunde inkl. `findings_json` (mit `quote` & `page`) und `confirmed`. `study_codings`: Kodierte Merkmale für Tabelle 1 je Coder (A/B).
Effektgrößen & Synthese	`effects`, `analysis_runs`	`effects`: `study_id`, `coder` (A/B), `metric`, `input_json`, `y`, `v`, `se`, `display`, `intervention_class`, `outcome_type`, `timepoint`, `decided_by`. `analysis_runs`: Versionierte R- und TS-Analyseläufe (`version`, `model`, `metric`, `k`, `data_hash`, `results_json`, `released`).
Qualitäts- & Evidenzbewertung	`rob2_assessments`, `robins_assessments`, `quadas2_assessments`, `grade_assessments`	`rob2_assessments` (`d1`–`d5`), `robins_assessments` (`d1`–`d7`), `quadas2_assessments` (`patient_selection`, `index_test`, `reference_standard`, `flow_timing` mit `risk_of_bias` + `applicability`), `grade_assessments` (je `(project_id, outcome, metric)`).
Governance & API-Sicherheit	`ratifications`, `idempotency_keys`, `llm_calls`	`ratifications`: Status (`offen`, `bestaetigt`, `geaendert`) für `agent:*`-Entscheidungen an den 4 Weichen. `idempotency_keys`: Scoped auf `(key, actor, method, path)` mit `request_hash` (SHA-256), verhindert doppelte Ausführung und Cross-Tenant-Replays. `llm_calls`: Token- und USD-Kostenprotokoll je Modell und Aufgabe.

11. R-Bridge (`metafor + clubSandwich`) & Snapshot-Hashing

Die Datei `lib/r-service.ts` bildet die Brücke zwischen dem TypeScript-Server und der statistischen Referenzimplementierung in `r/analyse.R`:

1. **Prozess-Isolation & Sicherheit:**

- Der Aufruf erfolgt ausschließlich über `execFile('Rscript', ['--vanilla', 'r/analyse.R', inputPath, outputPath])` ohne Shell-Interpolation (`shell: false`), mit hartem Timeout (`60 000 ms`) und automatischer Bereinigung des temporären Arbeitsverzeichnisses (`rmSync(..., { force: true, recursive: true })` im `finally`-Block).
- Retrahierte Studien (`study_retracted = 1`) werden vor der Übergabe an R zentral herausgefiltert.

1. **Deterministischer Snapshot-Hash (`computeDataHash`):**

- Vor jedem Analyselauf berechnet `computeDataHash` einen kanonischen **SHA-256-Hash** über alle sortierten Tupel `(id, study_id, metric, y, v, intervention_class, outcome_type, timepoint)`.
- Jeder Lauf in `analysis_runs` speichert diesen `data_hash`. Sobald eine Effektgröße nachträglich geändert, gelöscht oder ergänzt wird, ändert sich der aktuelle Projektdaten-Hash, und EvidenzWerk kennzeichnet ältere gespeicherte Analyseläufe im UI und im Export sofort als **„veraltet gegenüber aktuellem Kodierstand“** (`lib/vergleichsbasis.ts`).

---

12. OpenAPI 3.1 (`v1`-API), Idempotenz, Dry-Run & Akteurs-Audit

Unter `/api/v1/openapi.json` (`lib/v1/openapi.ts`) stellt EvidenzWerk eine vollständige **OpenAPI-3.1-Spezifikation** für externe KI-Agenten, R/Python-Pipelines und CI-Workflows bereit.

12.1 Authentifizierung & Akteurs-Kennzeichnung (`actor`)

- **Session-Cookie (`evw_session`)** für Browser-Aufrufe arrow Akteur wird automatisch als `human:<email-prefix>` gestempelt.
- **Bearer-Token (`Authorization: Bearer evw_...`)** oder Header `X-Agent-Actor: <name>` arrow Akteur wird als `agent:<token-name>` gestempelt und löst an den vier methodischen Weichen automatisch Ratifikations-Einträge aus.

12.2 Transaktionale Garantien für Agenten (`?dry_run=true` & `Idempotency-Key`)

- **Dry-Run-Modus (`?dry_run=true`):** Jede mutierende `POST`/`PUT`/`PATCH`/`DELETE`-Route unter `/api/v1/*` unterstützt den Query-Parameter `?dry_run=true`. Dabei führt `withDryRun` die echte Datenbankoperation innerhalb eines `SAVEPOINT dry_run` aus, berechnet das reale Ergebnis (inklusive Validierung, Effektberechnung und Folgezustand), rollt den Savepoint danach vollständig zurück (`ROLLBACK TO SAVEPOINT dry_run`) und liefert das Ergebnis mit `"dry_run": true` aus.
- **Idempotenz (`Idempotency-Key`):** Übergibt ein Client den Header `Idempotency-Key: <uuid>`, speichert `lib/v1/idempotency.ts` Statuscode und Antwort-JSON gebunden an `(key, actor, method, path)` und `request_hash`. Wiederholte Netzwerk-Retries liefern exakt dieselbe Antwort ohne doppelte Datensätze; ein abweichender Payload unter demselben Key wird mit `409 Conflict` (`IDEMPOTENCY_KEY_REUSED_DIFFERENT_BODY`) abgewiesen.

---

13. Verzeichnisstruktur des Quellcodes

Pfad	Aufgabe & wichtigste Module
[app/p/[id]/page.tsx](file:///Users/franzhuetter/evidenzwerk/app/p/%5Bid%5D/page.tsx)	Zentrale Arbeitsfläche eines Projekts mit allen 11 Reitern (`protokoll`, `recherche`, `literatur`, `screening`, `extraktion`, `coding`, `analyse`, `berechnungen`, `ratifikation`, `report`, `updates`).
`app/api/v1/`	46 versionierte REST-Endpunkte mit OpenAPI-3.1-Spezifikation, Idempotenz, Dry-Run und standardisiertem Fehlerformat.
`components/`	28 React-Komponenten, darunter `CodingForm.tsx`, `AnalysisPanel.tsx`, `AdvancedAnalysisPanel.tsx`, `SpezialanalysePanel.tsx`, `BerechnungenPanel.tsx`, `GradePanel.tsx`, `TeamPanel.tsx`.
`lib/stats.ts` & `lib/metrics.ts`	Mathematischer Kern für alle 10 Metriken (`smd`, `md`, `r`, `or`, `rr`, `rd`, `irr`, `prop_logit`, `prop_ft`, `hr`), Poolung (FE, DL, REML + Knapp-Hartung), Moderatoren und Sensitivitätsanalysen.
`lib/survival.ts`	Klinische Überlebenszeitanalyse: Alle 5 Tierney-Wege (2007) inkl. Kaplan-Meier-Kurvenrekonstruktion.
`lib/nma.ts` & `lib/dta.ts`	Frequentistische Netzwerk-Meta-Analyse (Rücker 2012, League-Table, P-Scores) und bivariate Diagnostische Testgüte-Synthese (Reitsma 2005 / HSROC, SROC-Ellipse, QUADAS-2).
`lib/rob2-shared.ts`, `lib/robins-i-shared.ts`, `lib/grade.ts`	Regelwerke für RoB 2, ROBINS-I (mit Design-Weiche) und GRADE.
`lib/spur.ts` & `lib/rechenweg.ts`	Gläserne Rechnung: Schritt-für-Schritt-Protokollierung jeder Formel mit eingesetzten Projektwerten.
`lib/docx.ts`, `lib/pdf.ts`, `lib/abbildungen.ts`, `lib/paket.ts`	Eigene Dokument-, Grafik-, Poster- und ZIP-Einreichungspaket-Generatoren.
`r/analyse.R`	R-Skript für `metafor::rma.mv` (3-Ebenen-REML), `clubSandwich` CR2, I²-Zerlegung und 5-fach Publikationsbias-Suite.
`test/`	40 automatisierte Testsuiten (`node --test`) mit über 680 Einzelprüfungen gegen analytische Referenzwerte, R-Gegenproben und UI/API-Parität.

---

14. Betrieb, Tests & Deployment (`Franz2`)

14.1 Automatisierte Qualitätssicherung lokal ausführen

```
# 1. Vollständige Testsuite (40 Testdateien, >680 Assertions inkl. R-Bridge & Parität)
npm test

# 2. Produktions-Build prüfen (Next.js TypeScript & Route-Compiler)
npm run build
```

14.2 Produktivbetrieb auf `Franz2` (`https://metaanalysis.brain-hr.de`)

Die Produktionsinstanz läuft als **rootless Podman-Quadlet-Service** (`evidenzwerk.service`) hinter Caddy mit automatischem TLS:

- **Container-Definition:** `deploy/evidenzwerk.container` und `Containerfile` (Multi-Stage-Build auf Basis von `node:22-bookworm-slim` inklusive `r-base`, `metafor`, `clubSandwich`, `meta`, `puniform` und Liberation-Schriften).
- **Persistentes Datenverzeichnis auf dem Host:** `/srv/evidenzwerk/data` (gemountet nach `/data` im Container; enthält `evidenzwerk.db` sowie `/data/pdfs/`).
- **Ein-Befehl-Deployment (`deploy/deploy.sh`):**

Synchronisiert den aktuellen Git-Stand auf `Franz2`, baut das Container-Image `localhost/evidenzwerk:latest`, startet `systemctl --user restart evidenzwerk.service` und prüft anschließend den Health-Endpunkt `/api/health/live`.