Dossier Scheinevidenz

Wie wissenschaftliche Studien den Anschein objektiver Belege erzeugen können

Arbeitsversion 1.2
Erstausgabe: 15.05.2026
Letzte Korrektur: 12.06.2026

Konzept & Autorenschaft: André Wermelinger
KI-gestützte Detailausarbeitung

Für den wertvollen fachlichen Austausch und für die Ergänzungen bedanken wir und bei Herrn Prof. Dr. Hugo Bucher

Zweck: Dieses Dossier beschreibt bekannte methodische, statistische, organisatorische und kommunikative Hebel, durch die Studienergebnisse verzerrt, überinterpretiert oder selektiv dargestellt werden können. Es ist als defensive Analyse- und Prüfgrundlage formuliert.

Hinweis: Die Nennung eines Bias-Musters beweist nicht, dass eine konkrete Studie absichtlich manipuliert wurde. Sie zeigt nur, welche Punkte geprüft werden sollten.

Inhalt

1. Zusammenfassung
2. Grundprinzipien: gute und schwache Evidenz
3. Wo Verzerrungen entstehen
4. Schnelle Prüffragen für Leser
5. Warnsignale in wissenschaftlichen Studien
6. Ausführlicher Anhang: bekannte Verzerrungs- und Manipulationsansätze
7. Quellen und weiterführende Referenzen

1. Zusammenfassung

Wissenschaftliche Studien sind keine automatische Garantie für Wahrheit. Sie sind systematische Versuche, eine Frage unter bestimmten Annahmen, Methoden und Auswertungsregeln zu beantworten. Scheinevidenz entsteht dort, wo Studien, Daten oder wissenschaftliche Begriffe den Anschein belastbarer Belege erzeugen, obwohl die Aussagekraft durch methodische Schwächen, selektive Entscheidungen, verzerrte Fragestellungen oder einseitige Interpretation eingeschränkt ist.

Die Qualität eines Ergebnisses hängt deshalb nicht nur vom Resultat selbst ab, sondern vom gesamten Weg dorthin: Fragestellung, Studiendesign, Stichprobe, Messung, Datenaufbereitung, Statistik, Interpretation, Veröffentlichung und Finanzierung.

Verzerrungen können unbeabsichtigt entstehen, etwa durch schlechte Messinstrumente, kleine Stichproben, fehlende Verblindung oder unvollständige Daten. Sie können aber auch bewusst begünstigt werden, wenn Forscherinnen und Forscher, Sponsoren, Institutionen oder Publikationssysteme ein bestimmtes Ergebnis bevorzugen.

Verzerrungen entstehen nicht erst innerhalb einzelner Studien. Bereits vorgelagert kann die Forschungsagenda beeinflusst werden: durch nationale Förderprogramme, universitäre Forschungsschwerpunkte, strategische Partnerschaften, Professuren, Institutsgründungen und thematisch gebundene Mittel. Dadurch wird mitbestimmt, welche Fragen überhaupt gestellt, finanziert und sichtbar werden – und welche Fragestellungen kaum Chancen auf Ressourcen erhalten. Besonders relevant ist deshalb das Verhältnis zwischen freien Forschungsprojekten und priorisierten oder thematisch gelenkten Programmen.

Besonders problematisch sind sogenannte Freiheitsgrade im Forschungsprozess: Forscher können oft mehrere plausible Entscheidungen treffen – welche Personen eingeschlossen werden, welcher Endpunkt zählt, welche Kovariaten ins Modell kommen, welche Ausreißer entfernt werden, wann die Datenerhebung endet und welche Resultate im Abstract stehen. Wenn diese Entscheidungen nicht vorab festgelegt und transparent dokumentiert sind, kann das Ergebnis in eine gewünschte Richtung gelenkt werden.

Eine einzelne Studie sollte daher möglichst nicht isoliert bewertet werden. Aussagekräftiger sind robuste Effekte, die in unabhängigen Studien, mit transparenten Protokollen, angemessener Power, nachvollziehbaren Daten, fairen Vergleichen und externer Validierung wiederholt werden.

Kernaussage: Nicht die Existenz einer Studie ist entscheidend, sondern ob Fragestellung, Design, Datenerhebung, Analyse und Berichterstattung so angelegt sind, dass alternative Erklärungen, Zufallstreffer und Interessenkonflikte überzeugend ausgeschlossen oder zumindest transparent gemacht werden.

2. Grundprinzipien: gute und schwache Evidenz

Gute Evidenz

Gute Evidenz entsteht, wenn eine klare, vorab definierte Frage mit einem passenden Design untersucht wird, die Gruppen fair vergleichbar sind, relevante Endpunkte vollständig gemessen werden, die Analyse vorab festgelegt ist und Rohdaten, Code, Protokoll sowie Interessenkonflikte transparent sind.

Schwache Evidenz

Schwache Evidenz entsteht häufig bei kleinen oder selektiven Stichproben, fehlender Randomisierung oder Verblindung, vielen nachträglichen Analysen, selektiver Ergebnisberichterstattung, unklaren Ausfällen, fehlenden absoluten Risiken und überzogenen Schlussfolgerungen.

Objektiv und subjektiv

Objektive Endpunkte wie Tod, Hospitalisierung, Laborwerte oder direkt beobachtbare Ereignisse sind oft robuster als subjektive Endpunkte wie Schmerz, Zufriedenheit oder Selbstauskunft. Subjektive Endpunkte sind nicht wertlos, benötigen aber besonders gute Verblindung, validierte Instrumente und transparente Auswertung.

Abhängig und unabhängig

Viele statistische Verfahren setzen unabhängige Beobachtungen voraus. Diese Annahme ist verletzt, wenn Personen innerhalb derselben Klinik, Schule, Familie, Firma oder Region ähnlicher sind als zufällig ausgewählte Personen. Wird diese Abhängigkeit ignoriert, können Standardfehler zu klein und Resultate scheinbar signifikant werden.

3. Wo Verzerrungen entstehen

Phase	Typisches Verzerrungsrisiko
Forschungsagenda und Förderlogik	Bereits vor der einzelnen Studie wird entschieden, welche Themen, Methoden und Forschungsrichtungen Mittel, Stellen, Professuren, Institute und Sichtbarkeit erhalten. Thematisch gelenkte Programme, Forschungsschwerpunkte und strategische Partnerschaften können dadurch bestimmte Fragestellungen fördern und andere verdrängen.
Forschungsfrage	Die Frage kann so formuliert werden, dass ein bestimmtes Ergebnis wahrscheinlicher, nützlicher oder besser kommunizierbar wird.
Design	Randomisierung, Verblindung, Kontrollgruppe, Dauer und Protokoll bestimmen, welche Alternativerklärungen ausgeschlossen werden können.
Stichprobe	Wer eingeschlossen oder ausgeschlossen wird, entscheidet, ob das Resultat repräsentativ ist.
Messung	Endpunkte, Messzeitpunkte, Instrumente und Beobachter beeinflussen, was überhaupt als Resultat sichtbar wird.
Analyse	Modellwahl, Umgang mit Ausreißern, fehlenden Daten, Subgruppen und Mehrfachtests kann Ergebnisse stark verändern.
Berichterstattung	Abstract, Grafiken, Schlussfolgerungen und Publikationsentscheidungen bestimmen, welches Bild beim Leser ankommt.
Anreize	Finanzierung, Karriere, Politik, Ideologie, institutionelle Reputation und Medienlogik können die Richtung des Forschungsprozesses beeinflussen.

4. Schnelle Prüffragen für Leser

- Wurde die Studie vorab registriert und gibt es ein öffentliches Protokoll?
- Ist der primäre Endpunkt eindeutig vorab definiert?
- Wurden alle geplanten Endpunkte berichtet, auch negative oder nicht signifikante?

- Ist die Kontrollgruppe fair und entspricht sie dem relevanten Standard?
- Sind Gruppen zu Studienbeginn vergleichbar?
- Sind Randomisierung und Verblindung angemessen beschrieben?
- Wie viele Personen fielen aus, und warum?
- Wurden absolute Effekte, Konfidenzintervalle und klinische/praktische Relevanz berichtet?
- Wurden mehrere Tests, Subgruppen oder Modellvarianten korrekt behandelt?
- Gibt es Rohdaten, Code, Analyseplan und klare Interessenkonflikt-Erklärung?
- Wer hat das Forschungsprogramm, den Forschungsschwerpunkt oder die Ausschreibung definiert?
- Handelt es sich um ein freies Forschungsprojekt oder um ein thematisch priorisiertes bzw. gelenktes Programm?
- Haben Personen oder Institutionen, die das Programm mitgestaltet haben, später selbst davon profitiert?
- Welche Rolle spielen Industriepartnerschaften, institutionelle Strategien oder politische Prioritäten bei der Themenwahl?
- Passt die Schlussfolgerung zum Studiendesign, oder wird Kausalität überdehnt?
- Wurde der Befund unabhängig repliziert?

5. Warnsignale in wissenschaftlichen Studien

Warnsignal	Warum problematisch
Kein präregistriertes Protokoll	Outcome- und Analysewechsel sind schwer erkennbar.
Primärer Endpunkt unklar	Nachträgliche Auswahl eines günstigen Endpunkts möglich.
Viele Endpunkte, wenige Korrekturen	Zufallstreffer werden wahrscheinlicher.
Kleine Stichprobe, großer Effekt	Effekt kann instabil oder überschätzt sein.
Nur p-Werte, keine Effektgrößen	Praktische Relevanz bleibt unklar.
Keine absoluten Risiken	Nutzen oder Schaden können größer wirken, als sie sind.
Keine Rohdaten/kein Code	Unabhängige Reanalyse ist erschwert.
Sponsor mit direktem Interesse	Design, Analyse oder Interpretation können interessengeleitet sein.
Negative Resultate im Abstract abgeschwächt	Hinweis auf Spin.
Subgruppenbefunde prominent	Oft explorativ und nicht konfirmatorisch.
Kein sauberer Umgang mit Dropouts	Attrition Bias möglich.
Keine Sensitivitätsanalysen	Robustheit gegenüber Annahmen bleibt ungeprüft.
Keine Replikation oder externe Validierung	Befund kann zufällig oder kontextabhängig sein.

6. Ausführlicher Anhang: bekannte Verzerrungs- und Manipulationsansätze

Die folgende Taxonomie ist als Prüfkatalog zu verstehen. Sie beschreibt bekannte Muster, durch die Studienresultate in eine bestimmte Richtung gelenkt, schwächere Evidenz als stärker dargestellt oder negative Befunde unsichtbar gemacht werden können.

6.1 Fragestellung und Hypothese manipulieren

Ansatz	Typische Wirkung / Risiko
Suggestive Forschungsfrage	Die Frage ist so formuliert, dass ein bestimmtes Ergebnis plausibler erscheint.
Unfaire Vergleichsfrage	Man vergleicht eine starke Intervention mit einer schwachen Alternative statt mit dem besten verfügbaren Standard.
Zu enge Definition des Problems	Relevante Nebenwirkungen, Langzeitfolgen oder Alternativerklärungen werden ausgeblendet.
Zu breite Definition des Problems	Unterschiedliche Phänomene werden zusammengeworfen, bis ein gewünschtes Muster entsteht.
HARKing	Nachträgliche Hypothesen werden als vorher geplant dargestellt.
Politisch oder wirtschaftlich nützliche Endpunkte	Die Studie beantwortet nicht die wichtigste wissenschaftliche Frage, sondern die nützlichste Kommunikationsfrage.
Surrogatfrage statt eigentlicher Frage	Ein leicht messbarer Ersatzmarker wird statt eines relevanten klinischen, sozialen oder funktionalen Endpunkts untersucht.

6.2 Studiendesign verzerren

Ansatz	Typische Wirkung / Risiko
Nicht-randomisierte Gruppen kausal interpretieren	Gruppenunterschiede können fälschlich als Behandlungseffekt erscheinen.
Schlechte oder fehlende Randomisierung	Systematische Unterschiede zwischen Gruppen bleiben bestehen.
Allocation Bias	Die Zuteilung zu Gruppen ist vorhersagbar oder beeinflussbar.
Fehlende Verblindung	Erwartungen von Forschern, Behandlern oder Teilnehmern beeinflussen Verhalten, Messung oder Auswertung.
Ungeeignete Kontrollgruppe	Der Vergleich ist so gewählt, dass die Zielintervention besser aussieht.
Placebo oder Kontrolle nicht glaubwürdig	Teilnehmer erkennen ihre Gruppe; Erwartungseffekte verzerren Resultate.
Zu kurze Studiendauer	Späte Nebenwirkungen, Rückfälle oder abnehmende Effekte bleiben unsichtbar.
Zu langer Beobachtungszeitraum ohne Kontrolle von Änderungen	Externe Einflüsse werden als Studienwirkung fehlinterpretiert.
Unrealistische Laborbedingungen	Gute interne Validität, aber schlechte Übertragbarkeit auf reale Bedingungen.
Run-in-Phase	Nicht-Responder oder Personen mit Nebenwirkungen werden vor Studienbeginn ausgeschlossen.
Enriched Design	Nur Personen mit hoher Antwortwahrscheinlichkeit werden eingeschlossen.
Crossover-Design ohne Washout	Effekte aus der ersten Phase beeinflussen die zweite Phase.

Ansatz	Typische Wirkung / Risiko
Cluster-Effekte ignorieren	Personen innerhalb einer Klinik, Schule oder Firma sind nicht unabhängig; Signifikanz wird überschätzt.
Nicht registrierte Protokolländerungen	Design kann nach Datenlage angepasst werden.

6.3 Stichprobe gezielt wählen

Ansatz	Typische Wirkung / Risiko
Selection Bias	Die untersuchte Gruppe repräsentiert nicht die Zielpopulation.
Cherry-Picking von Teilnehmern	Man nimmt Personen auf, bei denen der gewünschte Effekt wahrscheinlicher ist.
Ausschluss störender Gruppen	Ältere, Kranke, Frauen, Minderheiten, schwere Fälle oder Nicht-Responder werden ausgeschlossen.
Gesunde-Freiwillige-Bias	Freiwillige unterscheiden sich systematisch von der Zielpopulation.
Referral Bias	Spezialzentren sehen andere Fälle als Primärversorgung oder Allgemeinbevölkerung.
Survivor Bias	Nur Überlebende oder erfolgreiche Fälle werden analysiert.
Prevalence-Incidence/Neyman Bias	Fälle, die früh sterben oder schnell genesen, fehlen in der Stichprobe.
Collider Bias	Auswahl nach einem gemeinsamen Effekt zweier Ursachen erzeugt Scheinkorrelationen.
Sampling nach Verfügbarkeit	Bequeme Stichprobe statt repräsentativer Stichprobe.
Nicht-Antwort-Bias	Personen, die nicht teilnehmen oder nicht antworten, unterscheiden sich systematisch.
Über- oder Untergewichtung bestimmter Gruppen	Ergebnis wird durch Zusammensetzung der Stichprobe verschoben.

6.4 Gruppenzuteilung und Baseline manipulieren

Ansatz	Typische Wirkung / Risiko
Unbalancierte Ausgangswerte	Eine Gruppe startet günstiger.
Baseline-Werte selektiv berichten	Nur Variablen, bei denen Gruppen vergleichbar aussehen, werden gezeigt.
Regression zur Mitte ausnutzen	Extremwerte normalisieren sich ohnehin, werden aber als Interventionseffekt verkauft.
Unterschiedliche Vortherapien ignorieren	Frühere Behandlungen erklären Effekte.
Unterschiedliche Begleitbehandlungen zulassen	Zusatzbehandlungen wirken, nicht die untersuchte Intervention.
Unterschiedliche Betreuung der Gruppen	Mehr Aufmerksamkeit, Monitoring oder Beratung erzeugt bessere Ergebnisse.
Performance Bias	Gruppen werden außer der untersuchten Intervention unterschiedlich behandelt.

6.5 Intervention gezielt gestalten

Ansatz	Typische Wirkung / Risiko
Dosis des Vergleichs zu niedrig wählen	Zielbehandlung sieht wirksamer aus.
Dosis des Vergleichs zu hoch wählen	Vergleichsbehandlung erzeugt mehr Nebenwirkungen.
Suboptimale Anwendung der Konkurrenzmethode	Alternative erscheint schlechter, obwohl sie korrekt angewendet besser wäre.
Training oder Expertise ungleich verteilen	Eine Methode wird von Experten, die andere von Anfängern durchgeführt.

Ansatz	Typische Wirkung / Risiko
Adhärenz unterschiedlich fördern	Eine Gruppe wird stärker zur Einhaltung motiviert.
Co-Interventionen erlauben	Andere Maßnahmen erzeugen oder verstärken den beobachteten Effekt.
Kontamination zwischen Gruppen	Kontrollgruppe übernimmt Teile der Intervention; Effekte werden verdünnt oder verzerrt.
Intervention während Studie ändern	Anpassungen erfolgen nach Zwischenresultaten.

6.6 Messgrößen und Endpunkte manipulieren

Ansatz	Typische Wirkung / Risiko
Primären Endpunkt nachträglich ändern	Der Gewinner-Endpunkt wird nach Datenblick ausgewählt.
Viele Endpunkte testen	Irgendeiner wird statistisch signifikant.
Surrogatendpunkte verwenden	Biomarker verbessert sich, obwohl klinischer Nutzen unklar bleibt.
Kompositendpunkte	Mehrere Ereignisse werden zusammengefasst; harmlose häufige Ereignisse treiben das Ergebnis.
Subjektive Endpunkte bevorzugen	Schmerz, Wohlbefinden oder Zufriedenheit sind stärker erwartungs- und messabhängig.
Objektive Endpunkte vermeiden	Harte Resultate wie Tod, Hospitalisierung oder Funktion würden das Narrativ stören.
Unvalidierte Skalen verwenden	Messinstrument ist nicht zuverlässig oder nicht relevant.
Skalenrichtung oder Skalierung ausnutzen	Darstellung lässt kleine Effekte groß wirken.
Minimal wichtige Differenz ignorieren	Statistisch signifikant, aber praktisch bedeutungslos.
Zeitpunkt der Messung günstig wählen	Gemessen wird am Höhepunkt des Effekts, nicht langfristig.
Outcome-Switching	Geplante Outcomes werden durch andere ersetzt.
Nebenwirkungen anders definieren als Nutzen	Nutzen wird breit, Schaden eng erfasst.
Detection Bias	Outcome-Erhebung unterscheidet sich zwischen Gruppen oder ist durch Erwartungen beeinflusst.

6.7 Datenerhebung beeinflussen

Ansatz	Typische Wirkung / Risiko
Unstandardisierte Messung	Messfehler können systematisch werden.
Messgeräte unterschiedlich kalibrieren	Eine Gruppe erhält systematisch andere Werte.
Beobachter nicht verblinden	Erwartungen beeinflussen Bewertung.
Interviewer-Effekt	Art der Befragung lenkt Antworten.
Suggestive Fragebögen	Antwortverhalten wird in gewünschte Richtung geschoben.
Recall Bias	Gruppen erinnern sich unterschiedlich an Expositionen oder Symptome.
Ascertainment Bias	Eine Gruppe wird intensiver untersucht; daher findet man dort mehr Ereignisse.
Hawthorne-Effekt	Teilnehmer ändern Verhalten, weil sie beobachtet werden.
Differential Misclassification	Fehlklassifikation betrifft Gruppen unterschiedlich.
Undifferenzierte Fehlklassifikation als harmlos darstellen	Auch nicht-differentielle Fehler können Effekte verzerren oder abschwächen.
Proxy-Daten verwenden	Ersatzinformationen sind ungenau oder systematisch verzerrt.
Datenqualität gruppenweise unterschiedlich	Eine Gruppe hat bessere Dokumentation als die andere.

6.8 Fehlende Daten ausnutzen

Ansatz	Typische Wirkung / Risiko
Attrition Bias	Ausfälle unterscheiden sich systematisch zwischen Gruppen.
Dropouts selektiv ignorieren	Personen mit schlechtem Verlauf verschwinden aus der Analyse.
Per-Protocol statt Intention-to-Treat verwenden	Nur regelkonforme Teilnehmer werden analysiert; reale Wirksamkeit wird überschätzt.
Intention-to-Treat falsch umsetzen	Randomisierte Personen werden trotzdem ausgeschlossen.
Imputation günstig wählen	Fehlende Werte werden so ersetzt, dass Ergebnis stabil bleibt.
Last Observation Carried Forward	Alte Werte werden fortgeschrieben, obwohl Verlauf sich geändert hätte.
Complete-Case-Analyse	Nur vollständige Fälle werden ausgewertet; Verzerrung, wenn Vollständigkeit nicht zufällig ist.
Ausfälle als nicht relevant behandeln	Fehlende Daten werden qualitativ verharmlost.
Sensitivity Analyses weglassen	Robustheit gegenüber Annahmen über fehlende Daten bleibt ungeprüft.

6.9 Statistische Analyse manipulieren

Ansatz	Typische Wirkung / Risiko
P-Hacking	Analysevarianten werden ausprobiert, bis ein gewünschter p-Wert entsteht.
Multiple Tests ohne Korrektur	Zufallstreffer werden als Befund verkauft.
Subgruppen-Fishing	Man sucht nach einer Untergruppe mit gewünschtem Effekt.
Post-hoc-Subgruppen als geplant darstellen	Explorative Ergebnisse wirken bestätigend.
Modellspezifikation variieren	Kovariaten, Interaktionen und Funktionsformen werden so gewählt, dass Ergebnis passt.
Unpassende Kovariaten kontrollieren	Kontrolle von Mediatoren, Collidern oder Folgevariablen kann Zusammenhänge verzerren.
Wichtige Confounder weglassen	Scheinkausalität bleibt bestehen.
Overadjustment	Man kontrolliert Variablen, die Teil des Effekts sind.
Underadjustment	Man kontrolliert zu wenig.
Outlier selektiv entfernen	Unbequeme Datenpunkte verschwinden.
Outlier selektiv behalten	Einzelne Extremwerte treiben den gewünschten Effekt.
Transformations-Shopping	Linear, logarithmisch, kategorisiert usw. wird ausprobiert.
Dichotomisierung kontinuierlicher Variablen	Informationsverlust; Grenzwert kann Ergebnis kippen.
Grenzwerte nachträglich wählen	Cutoff wird dort gesetzt, wo der Effekt am besten aussieht.
Einseitige Tests verwenden	Signifikanz wird leichter erreicht, wenn Gegenrichtung ignoriert wird.
Parametrische Tests trotz verletzter Annahmen	p-Werte und Konfidenzintervalle werden unzuverlässig.
Clusterung oder Abhängigkeit ignorieren	Standardfehler werden zu klein.
Zeitabhängige Verzerrungen ignorieren	Besonders bei Beobachtungsstudien können Effekte künstlich entstehen.
Immortal-Time-Bias	Personen müssen eine gewisse Zeit überleben, um in eine Gruppe zu gelangen; dadurch wirkt diese Gruppe besser.
Lead-Time-Bias	Frühere Diagnose verlängert scheinbar Überleben, ohne Tod hinauszuschieben.
Length-Time-Bias	Screening findet eher langsam verlaufende Fälle.
Simpson-Paradox ignorieren	Aggregierte Daten zeigen anderes Muster als Strata.
Bayesianische Priors gezielt wählen	Vorannahmen drücken Ergebnis in gewünschte Richtung.
Stopping Rule flexibel halten	Datenerhebung endet, wenn Ergebnis signifikant ist.

Ansatz	Typische Wirkung / Risiko
Zwischenanalysen ohne Korrektur	Wiederholtes Nachsehen erhöht Fehlerwahrscheinlichkeit.
Power zu niedrig planen	Instabile, überhöhte Effektschätzungen; negative Ergebnisse bleiben nicht schlüssig.
Power nachträglich schönrechnen	Schwache Evidenz wirkt geplant und solide.

6.10 Interpretation verdrehen

Ansatz	Typische Wirkung / Risiko
Korrelation als Kausalität darstellen	Beobachtungsbefunde werden überinterpretiert.
Relative statt absolute Effekte betonen	Prozentuale Effekte wirken groß, auch wenn absolute Unterschiede klein sind.
Absolute Risiken weglassen	Praktische Relevanz wird verschleiert.
Number Needed to Treat/Harm weglassen	Nutzen-Schaden-Abwägung bleibt unklar.
Konfidenzintervalle ignorieren	Unsicherheit wird nicht sichtbar.
Nicht-Signifikanz als kein Effekt darstellen	Fehlende Evidenz wird mit Evidenz für Null verwechselt.
Signifikanz als Wichtigkeit darstellen	Kleine, irrelevante Effekte wirken wichtig.
Explorativ und konfirmatorisch vermischen	Hypothesengenerierung wird als Hypothesentest verkauft.
Mechanistische Spekulation als Beweis darstellen	Plausibilität ersetzt empirische Evidenz.
Einzelstudie als endgültig darstellen	Gesamte Evidenzlage wird ignoriert.
Externe Validität überdehnen	Ergebnis aus enger Stichprobe wird auf breite Population übertragen.
Nebenwirkungen relativieren	Schaden wird als nicht kausal bewiesen abgetan, Nutzen aber kausal behauptet.
Spin in Abstract oder Conclusion	Ergebnis wird positiver formuliert als Daten es tragen.
Sekundäre positive Befunde hervorheben	Primärer negativer Befund wird in den Hintergrund gedrückt.
Grenzwertige Resultate als Trend verkaufen	Nicht signifikante Ergebnisse werden sprachlich aufgewertet.
Nicht replizierte Befunde als robust darstellen	Reproduzierbarkeit wird nicht geprüft.

6.11 Berichterstattung selektiv steuern

Ansatz	Typische Wirkung / Risiko
Publication Bias	Positive oder gewünschte Resultate werden eher veröffentlicht als negative.
File-Drawer-Problem	Negative Studien bleiben in Schubladen.
Positive Results Bias	Positive Outcomes werden bevorzugt berichtet.
Selective Reporting	Nur passende Analysen oder Endpunkte erscheinen im Paper.
Time-Lag Bias	Positive Studien erscheinen schneller.
Duplicate Publication	Gleiche oder ähnliche Daten werden mehrfach publiziert.
Salami-Slicing	Eine Studie wird in viele kleine Papers zerlegt.
Ghostwriting	Dritte schreiben oder managen Manuskript mit Interessenkonflikt.
Gift/Authorship Manipulation	Namen werden aus strategischen Gründen hinzugefügt oder entfernt.
Outcome nicht im Abstract erwähnen	Leser sehen nur die gewünschte Kurzfassung.
Supplement versteckt kritische Details	Zentrale Einschränkungen stehen nur schwer auffindbar.
Rohdaten nicht bereitstellen	Prüfung und Reanalyse werden erschwert.
Protokoll oder Statistical Analysis Plan nicht veröffentlichen	Nachträgliche Änderungen bleiben unsichtbar.

6.12 Meta-Analysen und Reviews verzerren

Ansatz	Typische Wirkung / Risiko
Suchstrategie eng wählen	Unbequeme Studien werden nicht gefunden.
Datenbanken selektiv durchsuchen	Literaturbasis wird verzerrt.
Graue Literatur ausschließen	Unveröffentlichte negative Daten fehlen.
Ein- und Ausschlusskriterien nach Ergebnisnähe setzen	Evidenzkörper wird passend gebaut.
Duplikate nicht erkennen	Einzelne Daten gehen mehrfach ein.
Heterogene Studien unkritisch poolen	Durchschnittseffekt verdeckt Unterschiede.
Fixed-Effect statt Random-Effects opportunistisch wählen	Präzision oder Effekt kann günstiger erscheinen.
Subgruppen in Meta-Analyse nachträglich wählen	Wunschmuster entsteht aus vielen Schnitten.
Risk-of-bias-Bewertung beschönigen	Schwache Studien erhalten zu viel Gewicht.
Small-Study Effects ignorieren	Kleine positive Studien dominieren Eindruck.
Funnel-Plot oder Publikationsbias nicht prüfen	Verzerrte Literaturbasis bleibt verborgen.
Narrative Review statt systematischer Review	Autor kann Studien selektiv erzählen.

6.13 Finanzierung, Interessenkonflikte und institutionelle Anreize

Ansatz	Typische Wirkung / Risiko
Sponsor kontrolliert Design	Fragestellung, Endpunkte und Vergleich werden interessenfreundlich gewählt.
Sponsor kontrolliert Datenzugang	Unabhängige Prüfung ist eingeschränkt.
Publikationsrechte beim Sponsor	Negative Ergebnisse können verzögert oder verhindert werden.
Interessenkonflikte unvollständig deklarieren	Leser unterschätzen Verzerrungsrisiko.
Karriereanreize: Neuheit statt Korrektheit	Auffällige Resultate werden belohnt.
Zitations- und Impact-Faktor-Druck	Positive, überraschende und einfache Geschichten werden bevorzugt.
Regulatory Capture	Bewertende Institutionen übernehmen Perspektive der regulierten Akteure.
Think-Tank- oder Lobby-Finanzierung	Forschungsagenda und Interpretation folgen Interessen.
Nichtfinanzielle Ideologie oder Karrierebindung	Forscher haben starken Anreiz, ein Narrativ zu stützen.

6.13a Forschungsagenda, Förderprogramme und institutionelle Schwerpunktsetzung

Ein weiterer vorgelagerter Hebel liegt in der Frage, welche Forschung überhaupt ermöglicht wird. Förderlogik, Schwerpunktprogramme und institutionelle Strategien können die Forschungsagenda prägen, bevor einzelne Studien geplant, durchgeführt oder publiziert werden.

Ansatz	Typische Wirkung / Risiko
Thematisch gelenkte Förderprogramme	Forschung konzentriert sich auf politisch, wirtschaftlich oder institutionell priorisierte Themen; freie, unerwartete oder riskantere Fragestellungen geraten ins Hintertreffen.
Forschungsschwerpunkte durch Universitätsleitungen	Rektorate, Dekanate oder strategische Gremien können festlegen, welche Themen personell, finanziell und kommunikativ bevorzugt werden.
Ungleichgewicht zwischen freien und priorisierten Projekten	Je grösser der Anteil thematisch gebundener Mittel, desto stärker wird die Forschungsagenda von externen oder institutionellen Prioritäten geprägt.
Programmgestaltung durch spätere Nutzniesser	Wenn Personen oder Gruppen, die Förderprogramme mitprägen, später selbst überproportional davon profitieren, entstehen Interessenkonflikte oder Netzwerkvorteile.

Ansatz	Typische Wirkung / Risiko
Lokale Netzwerk- und Patronagestrukturen	Auf Ebene einzelner Universitäten oder Institute können kleine, gut vernetzte Gruppen Ressourcen, Stellen und Themenfelder dominieren.
Strategische Industriepartnerschaften	Private Partner können Fragestellungen, Methoden, Datenzugang, Publikationsinteressen oder institutionelle Prioritäten indirekt beeinflussen.
Modische oder reputationsstarke Themenlabels	Forschung wird auf aktuell gut finanzierbare Schlagworte ausgerichtet, auch wenn wissenschaftliche Substanz oder Langfristigkeit begrenzt sein können.
Kurzfristige Amtslogik von Leitungsfunktionen	Kurze Amtszeiten können Anreize für sichtbare, schnelle und prestigeträchtige Programme schaffen statt für langfristige, riskante oder grundlagenorientierte Forschung.
Benachteiligung freier Forschung bei Budgetkürzungen	Ungebundene, kreative oder nicht-mainstreamfähige Projekte können zuerst unter Druck geraten, während priorisierte Programme institutionell geschützt bleiben.
Erwartete Resultate bereits im Programmdesign angelegt	Förderprogramme können so formuliert sein, dass bestimmte positive Resultate, Narrative oder Anwendungsperspektiven faktisch vorausgesetzt werden.

6.14 Peer Review, Redaktion und wissenschaftliche Kommunikation

Ansatz	Typische Wirkung / Risiko
Reviewer mit gleicher Schule auswählen	Kritische Gegenperspektiven fehlen.
Reviewer mit Interessenkonflikten	Kritik wird abgeschwächt oder Konkurrenz blockiert.
Desk Reject unbequemer Replikationen	Korrekturmechanismus wird geschwächt.
Novelty Bias von Journals	Replikationen und Nullbefunde erscheinen seltener.
Press Release übertreibt Befund	Medien übernehmen zugespitzte Version.
Abstract stärker als Haupttext frisieren	Viele Leser lesen nur Kurzfassung.
Grafiken manipulativ gestalten	Achsen, Skalen, abgeschnittene y-Achsen oder relative Darstellung verändern Eindruck.
Causal Language in Überschrift	Kausalität wird suggeriert, obwohl Design sie nicht trägt.
Peer reviewed als Qualitätsersatz verwenden	Peer Review ist Filter, kein Wahrheitsbeweis.

6.15 Datenintegrität: harte Formen wissenschaftlichen Fehlverhaltens

Ansatz	Typische Wirkung / Risiko
Datenfälschung	Werte werden erfunden.
Datenfabrikation	Ganze Datensätze oder Fälle existieren nicht.
Bildmanipulation	Mikroskopie, Western Blots, Gelbilder usw. werden verändert.
Selektives Entfernen von Messungen	Unbequeme Daten verschwinden.
Backdating von Protokollen	Nachträgliche Entscheidungen wirken geplant.
Manipulation von Laborbüchern	Audit-Trail wird verfälscht.
Plagiat oder Selbstplagiat	Wissenschaftliche Herkunft und Neuheit werden falsch dargestellt.
Unethische Nachrekrutierung	Teilnehmer werden ergänzt, bis Ergebnis passt.
Datenzugang verweigern mit Vorwand	Externe Prüfung wird verhindert.

6.16 KI-, Big-Data- und Modellstudien-spezifische Verzerrungen

Ansatz	Typische Wirkung / Risiko
Train/Test Leakage	Testdaten fließen ins Training; Modell wirkt besser.
Benchmark-Shopping	Nur Datensätze mit guten Resultaten werden gezeigt.

Ansatz	Typische Wirkung / Risiko
Metrik-Shopping	Accuracy, F1, AUC, RMSE usw. werden nach Vorteil gewählt.
Unfaire Baselines	Vergleichsmodelle sind schlecht getunt.
Hyperparameter auf Testset optimieren	Testset wird faktisch Validierungsset.
Nicht repräsentative Trainingsdaten	Modell funktioniert nur für bestimmte Gruppen.
Label Bias	Zielvariable enthält bereits soziale oder institutionelle Verzerrungen.
Datenbereinigung selektiv	Fälle werden entfernt, bis Performance steigt.
Ablationsstudien weglassen	Unklar, welcher Modellteil wirklich wirkt.
Keine externe Validierung	Modell generalisiert möglicherweise nicht.
Proprietäre Daten oder Modelle	Reproduktion unmöglich.

7. Quellen und weiterführende Referenzen

Die folgenden Quellen wurden zur begrifflichen und methodischen Einordnung verwendet. Die Taxonomie im Anhang ist eine strukturierte Zusammenfassung bekannter Bias- und Reporting-Muster; sie ist keine Aussage über eine konkrete Studie.

[1] Cochrane: Chapter 8 - Assessing risk of bias in a randomized trial / RoB 2. https://www.cochrane.org/authors/handbooks-and-manuals/handbook/current/chapter-08 - Beschreibt Risk-of-bias-Domänen für randomisierte Studien: Randomisierung, Abweichungen von Interventionen, fehlende Outcome-Daten, Outcome-Messung und selektive Ergebnisberichterstattung.

[2] Cochrane: About Risk of Bias 2 (RoB 2). https://www.cochrane.org/learn/courses-and-resources/cochrane-methodology/risk-bias/about-risk-bias-2-rob-2 - Erklärt die RoB-2-Systematik und die Unterscheidung verschiedener Bias-Domänen.

[3] Catalogue of Bias. https://catalogofbias.org/biases/ - Katalog zahlreicher Biasformen in Gesundheits- und Evidenzforschung, z. B. Selection Bias, Attrition Bias und Ascertainment Bias.

[4] Catalogue of Bias: Reporting biases. https://catalogofbias.org/biases/reporting-biases/ - Beschreibt Reporting-Bias-Kategorien wie Publication Bias, Time-Lag Bias, Duplicate Publication Bias, Location Bias, Citation Bias, Language Bias und Outcome Reporting Bias.

[5] ROBINS-I / NCBI Bookshelf: Bias domains in non-randomised studies. https://www.ncbi.nlm.nih.gov/books/NBK553695/ - Beschreibt Bias-Domänen für nicht-randomisierte Interventionsstudien, u. a. Confounding, Selection, Classification of Interventions, Missing Data, Measurement of Outcomes und Selection of Reported Result.

[6] BMJ: ROBINS-I tool for assessing risk of bias in non-randomised studies of interventions. https://www.bmj.com/content/355/bmj.i4919 - Fachartikel zur Entwicklung und Begründung des ROBINS-I-Werkzeugs.

[7] Stanford Encyclopedia of Philosophy: Reproducibility of Scientific Results. https://plato.stanford.edu/entries/scientific-reproducibility/ - Überblick über Reproduzierbarkeit, Replikation und die metawissenschaftliche Diskussion zur Replikationskrise.

[8] Stanford Encyclopedia of Philosophy: Scientific Objectivity. https://plato.stanford.edu/entries/scientific-objectivity/ - Diskutiert wissenschaftliche Objektivität, p-hacking, selektive Berichterstattung und fragwürdige Forschungspraktiken im Kontext wissenschaftlicher Anreize.