

SERIE „STATISTIK AUS DEM IQWiG“

Das Fundament der Erkenntnis – Studiendesigns

Wer Studienergebnisse für die evidenzbasierte Entscheidung nutzen will, muss Studien nicht nur lesen, sondern auch die Vertrauenswürdigkeit der Ergebnisse einordnen können. Denn nicht jede veröffentlichte Untersuchung erlaubt belastbare Aussagen für den Praxisalltag.

PROF. DR. TIM
MATHES

Leiter des Ressorts
Gesundheitsökonomie
des Instituts für Qualität
und Wirtschaftlichkeit
im Gesundheitswesen
(IQWiG)
www.iqwig.de
tim.mathes@iqwig.de

„Vertrauenswürdig“ im statistischen Sinne bedeutet, dass die Schätzung des Effekts einer Maßnahme auf Basis einer Studie möglichst nah an den wahren Therapieeffekt herankommt. Wird der Effekt in der Studie systematisch über- oder unterschätzt, spricht man von systematischem Fehler oder Verzerrung (engl. Bias).

Eine Studie sollte so gestaltet sein, dass die Gefahr für systematische Verzerrung minimiert wird. Nur so sind kausale Aussagen – z. B. „Therapie A ist wirksamer als Therapie B“ – möglich, die als Grundlage für gute Therapieentscheidungen dienen können.

Studien ohne Kontrollgruppe: kein solides Fundament

Eines der wichtigsten Unterscheidungsmerkmale von Studiendesigns ist, ob diese „kontrolliert“ sind oder nicht. Kontrollierte Studien weisen mindestens eine Interventions- und eine Kontrollgruppe auf. Die Kontrolle ist in der Regel „keine“ oder „eine andere“ Therapie. Bei Studien ohne Kontrollgruppe, z. B. Fallserien, kann der Effekt der untersuchten therapeutischen Maßnahme (Interventionsgruppe) nicht vom natürlichen Krankheitsverlauf abgegrenzt werden. Daher kann der Effekt auch nicht auf die therapeutische Maßnahme zurückgeführt werden.

Eine kontrollierte Studie kann allerdings nur zuverlässige Ergebnisse liefern, wenn sich die Studiengruppen nicht systematisch voneinander unterscheiden. Ein systematischer Unterschied (z. B. unterschiedliches Alter) kann dazu führen, dass der Verlauf in der Interventions- und Kontroll-

gruppe auch ohne die therapeutische Maßnahme unterschiedlich ausfällt (z. B. Sterblichkeit) und dann fälschlicherweise auf die therapeutische Maßnahme zurückgeführt wird, obgleich diese eigentlich wirkungslos ist.

Randomisierte Studien: die Kraft des Zufalls

Die randomisierte kontrollierte Studie (RCT) gilt als Referenzstandard für Fragen zur Effektivität therapeutischer Maßnahmen. Der Grund für die Überlegenheit von RCTs gegenüber allen anderen Studiendesigns ist die zufällige Zuteilung (Randomisierung) der Teilnehmenden zur Interventions- und Kontrollgruppe. Hierdurch ist sichergestellt, dass Störgrößen (engl. Confounder; z. B. unterschiedliches Alter) rein zufällig auf die Interventions- und Kontrollgruppe verteilt werden. Ein großer Vorteil ist, dass dieses auch für unbekannte Störgrößen gilt. Auf diese Weise wird eine in jeder Hinsicht strukturell vergleichbare Interventions- und Kontrollgruppe geschaffen. Es können zwar auch nach der Randomisierung noch zahlenmäßige Unterschiede (z. B. beim durchschnittlichen Alter) zwischen den Gruppen bestehen, da diese aber rein zufällig sind, wird hierdurch lediglich der zufällige, nicht jedoch der systematische Fehler erhöht.

Vergleichende Beobachtungsstudien: ein unsicheres Terrain

In Beobachtungsstudien erfolgt die Auswahl der Therapie entsprechend der Versorgungspraxis. Das Hauptproblem hierbei ist, dass die Therapieauswahl und somit die Zuteilung der Patientinnen und Patienten zur Interventions- und Kontroll-

therapie in der Versorgungspraxis in der Regel von Patientenmerkmalen abhängt. Ein Beispiel hierfür wäre, dass ältere Menschen ein bestimmtes Medikament (Intervention), das genauso wirksam ist wie die bisherige Therapie (Kontrolle), aufgrund schlechter Verträglichkeit seltener bekommen. Dieses würde dazu führen, dass das Durchschnittsalter in der Kontrollgruppe höher wäre. Denn das Merkmal Alter hängt sowohl mit der Auswahl der Therapie als auch mit der Prognose (z. B. Mortalität) zusammen und ist somit eine Störgröße (engl. Confounder). Würde man nun die Differenz zwischen den Gruppen ermitteln, würde man einen Unterschied finden, der fälschlicherweise auf die Therapie zurückgeführt werden könnte. Um kausale Schlussfolgerungen aus einer Beobachtungsstudie ziehen zu können, müssen daher alle Störgrößen bekannt sein und adäquat in der Analyse berücksichtigt werden. Allerdings kann nie ganz ausgeschlossen werden, dass einige Störgrößen nicht bekannt sind, wes-

wegen im Vergleich zu einer RCT auch bei einer perfekt designten Beobachtungsstudie stets eine höhere Restunsicherheit bleibt.

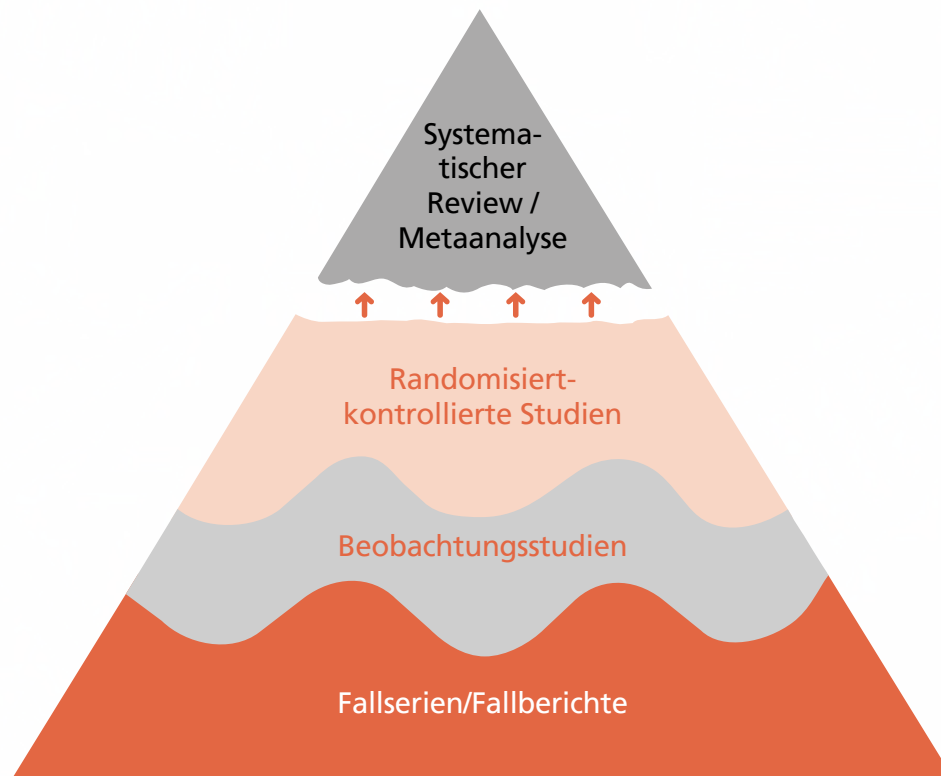
Systematische Reviews: ein breiteres Fundament

Die Anzahl medizinischer Studien wächst kontinuierlich. Für den Praxisalltag sind deshalb systematische Übersichtsarbeiten von besonderer Bedeutung. Diese bilden kein eigenes Studiendesign, sondern fassen alle relevanten Primärstudien – diese können sowohl RCTs auch als Beobachtungsstudien sein – zu einer Forschungsfrage zusammen. Falls zusätzlich eine quantitative Zusammenfassung der Ergebnisse der einzelnen eingeschlossenen Studien erfolgt, spricht man von einer Metaanalyse. Deren Vorteil liegt vor allem darin, dass nicht einzelne Studien isoliert betrachtet werden, sondern die gesamte verfügbare Evidenz. Dabei wird auch die methodische Qualität der eingeschlossenen Studien bewertet.

Quellen:

1. EBM-Netzwerk: Evidenzpyramide (2026). kvh.link/038

EVIDENZPYRAMIDE VON STUDIENDESIGNS¹



Das unabhängige Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) untersucht den Nutzen und den Schaden von medizinischen Maßnahmen für Patientinnen und Patienten. Über die Vorteile und Nachteile von Untersuchungs- und Behandlungsverfahren informiert das Institut in Form von wissenschaftlichen Berichten und allgemein verständlichen Gesundheitsinformationen.

Die Betrachtung der gesamten Evidenz und ihrer Qualität hat den Vorteil, dass die Informationsbasis vergrößert werden kann; zudem kann geprüft werden, ob sich Studienergebnisse replizieren lassen. Für den Praxisalltag können systematische Reviews einen erheblichen Zeitgewinn bringen: Statt diverse Einzelstudien lesen zu müssen, erhält man eine strukturierte Gesamteinschätzung. Systematische Reviews randomisierter Studien gelten daher in der Evidenzhierarchie als höchste Evidenzstufe für therapeutische Fragestellungen.

Evidenzpyramide: die Basis, aber nicht statisch

In der Evidenzbasierten Medizin (EbM) werden Studiendesigns traditionell in Evidenzpyramiden dargestellt (s. Abbildung). Ganz oben finden sich aus den genannten Gründen die systematischen Reviews und darunter RCTs; dann folgen Beobachtungsstudien. Ganz unten stehen Studiendesigns ohne Kontrollgruppe. Die Studiendesigns stellen eine gute Ausgangsbasis für die Beurteilung der Sicherheit der Ergebnisse dar.

Moderne Bewertungsansätze lösen sich jedoch zunehmend davon, Evidenz allein anhand des Studiendesigns zu bewerten. Sie berücksichtigen auch eine Vielzahl anderer methodischer Aspekte – wie etwa die Verblindung in RCTs oder die Berücksichtigung von Confoundern in Beobachtungsstudien –, die die Aussagesicherheit der Studien beeinflussen. Im Extremfall kann z. B. eine sehr schlecht durchgeführte RCT weniger aussagekräftig sein als eine sehr hochwertige Beobachtungsstudie. Die Übergänge bezüglich der Aussagekraft der Studiendesigns werden daher heute eher als fließend angesehen.

Mein Fazit

Das Studiendesign stellt zwar immer noch die Ausgangsbasis für die Bewertung der Studienqualität dar, dennoch ist eine tiefergehende kritische Bewertung einzelner Qualitätsaspekte unerlässlich. ■



Wenn Sie KVH aktuell auch auf Ihrem Smartphone oder Tablet lesen möchten, scannen Sie einfach den QR-Code. Sie gelangen damit direkt zum WebMagazin.

www.kvhaktuell.de

KVH WEBMAG