



In der Flut von Studien verlässliche
Informationen finden

Evidenzbasierte Medizin

Praxisrelevante Richtlinien für Ärzte ändern sich immer rasanter. Das macht es nötig, laufende Publikationen und Neuerungen auf ihre Verlässlichkeit screenen zu können. Welche Werkzeuge die evidenzbasierte Medizin dafür bietet und wie Kennzahlen, Konfidenzintervalle und weitere statistische Werte eingeschätzt werden können, präsentierte Dr. Anna Glechner, Krems, beim **ALLGEMEINE+** Winterquartett.

Dr. Anna Glechner



© Reischer Anna

Leitung Ärzteinformationszentrum Cochrane Österreich
 Department für Evidenzbasierte Medizin und Evaluation
 Universität für Weiterbildung Krems
 Ärztin für Allgemeinmedizin
 Gesundheitszentrum Sitzenberg-Reidling
 E-Mail: anna.glechner@donau-uni.ac.at

Was ist evidenzbasierte Medizin?

Im klinischen Alltag sind Ärzte mit einer Vielzahl von Entscheidungen konfrontiert. Diese Entscheidungen basieren im Idealfall auf drei Komponenten: der klinischen Erfahrung des Arztes, den aktuellsten wissenschaftlichen Erkenntnissen und den Patientenbedürfnissen (Abb. 1). Die evidenzbasierte Medizin (EbM) vereint diese drei Säulen der klinischen Entscheidung.¹ In Situationen, in denen einer oder zwei dieser Faktoren fehlen, müssen im klinischen Alltag jedoch trotzdem Entscheidungen getroffen werden. Dies bedeutet, dass bei Fragestellungen, zu denen keine verlässliche wissenschaftliche Evidenz vorhanden ist, klinische Erfahrung und Patientenpräferenzen den Ausschlag geben werden.

„Up to date“ bleiben in der Medizin: Wie geht das?

Es ist eine große Herausforderung, auf dem aktuellen Stand der Wissenschaft zu bleiben, denn die Halbwertszeit medizinischer Erkenntnisse betrug einer Schätzung zufolge im Jahr 2010 etwa 3,5 Jahre, d. h., etwa alle 3,5 Jahre ist die Hälfte des Wissens und Handelns überholt und wird

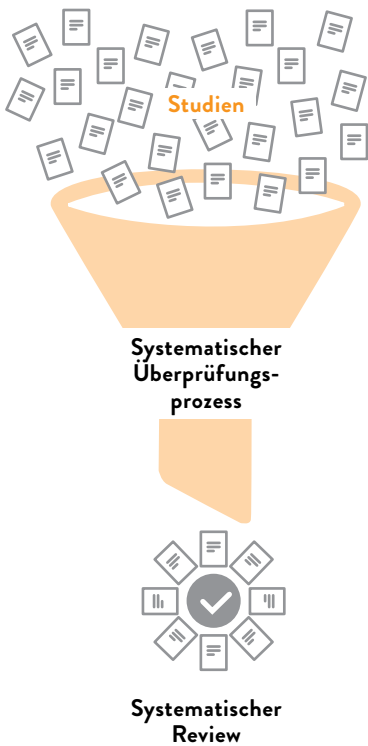
durch neue Erkenntnisse ersetzt.² Neue Medikamente und Untersuchungsgeräte kommen auf den Markt und werden von Pharmafirmen mit aktuellen Studienergebnissen angepriesen. Um Studien kritisch zu hinterfragen braucht es einiges an Übung und eine Zusatzausbildung, was im Medizinstudium häufig zu kurz kommt. Wichtig ist daher zu wissen, welche verlässlichen

Quellen es gibt. Die erste Quelle für Ärzte sind Leitlinien oder Point-of-Care Tools, wie die EbM-Guidelines für Allgemeinmedizin, „UpToDate“ oder „DynaMed“.³ Wenn Fragen in Leitlinien unbeantwortet bleiben, bieten systematische Übersichtsarbeiten eine gute Übersicht, denn sie fassen alle Studien für eine spezifische Frage zusammen.

Abb. 1: Komponenten der evidenzbasierten Entscheidungsfindung



Abb. 2: Systematische Übersichtsarbeiten



Leitlinien kritisch bewerten – ein Crashkurs

Leitlinien enthalten Empfehlungen für die Praxis, die auf Studienergebnissen basieren. Wenn es zu wichtigen Fragen keine Studien gibt, dann müssen im klinischen Alltag trotzdem Entscheidungen getroffen werden. In diesem Fall basieren Empfehlungen aus Leitlinien auf einem Expertenkonsensus. Leider entsprechen nicht alle Leitlinien den grundlegenden Qualitätskriterien.⁴ Einige wichtige Kriterien können Ärzten helfen zu erkennen, ob es sich um qualitativ hochwertige Leitlinien handelt.⁵

I. Wurden alle relevanten Studien zu einem Thema erfasst?

In gut gemachten Leitlinien wird systematisch in mehreren Datenbanken nach Literatur gesucht. Alternativ dazu werden in den Empfehlungen hochwertige systematische Übersichtsarbeiten zitiert. Systematische Übersichtsarbeiten fassen alle relevanten Studien zu einer bestimmten Frage zusammen (siehe unten).

II. Wurde die Qualität der Studien bewertet?

Basiert die Empfehlung auf gut gemachten Studien oder haben wir es mit Studien zu tun, die methodische Mängel aufweisen?

III. Wurden Handlungsoptionen unter Berücksichtigung von Nutzen und Risiken dargestellt?

IV. Beinhaltet die Leitlinie Empfehlungs- und Evidenzgrade?

Empfehlungen werden häufig mit einer Ziffer oder einem Buchstaben versehen, um darzustellen, wie hoch das Vertrauen in die zugrunde liegenden Studien ist. Aber auch wenn es keine Evidenz gibt oder die Studienergebnisse unsicher sind, muss eine Entscheidung getroffen werden. Meist gibt es daher zusätzlich einen Empfehlungsgrad der Experten, die die Leitlinien entwickelt haben: das heißt eine starke oder schwache Empfehlung für oder gegen eine Behandlung.

Systematische Übersichtsarbeiten – was ist zu beachten

Systematische Übersichtsarbeiten fassen alle Studien zusammen, die eine bestimmte Frage untersuchen (Abb. 2). Damit das Vorgehen für andere nachvollziehbar ist, werden systematische Übersichtsarbeiten nach einem standardisierten Schema erstellt. Damit werden Verzerrungen vermieden, die bei unsystematischem Vorgehen bei der Erfassung und Bewertung der vorhandenen Studien entstehen können.⁶ Einige der folgenden Punkte geben einen Hinweis dafür, ob eine systematische Übersichtsarbeit wesentliche Qualitätskriterien erfüllt.⁷

I. Klar definierte Frage?

Systematische Übersichtsarbeiten über medizinische Themen sind häufig mit der Frage konfrontiert, ob ein neues Medikament besser wirkt als die bisher verwendete Standard-Therapie. Um solche Fragen zu beantworten, suchen Autoren von systematischen Übersichtsarbeiten nach Studien, die zwei Gruppen miteinander vergleichen: Eine Gruppe erhält ein neues Medikament und die Kontrollgruppe die bisher verwendete Therapie. Daher hat sich für die genau definierte Fragestellung von systematischen Übersichtsarbeiten das PIKO-Schema eingebürgert. P steht dabei für Population, I für Intervention, K für die Kontrollgruppe und O für das Outcome. Für die Literatursuche in den Datenbanken müssen nicht alle Komponenten der PIKO-Frage verwendet werden. Wenn die Literatursuche zu wenige Ergebnisse ergibt, können auch nur Teile der PIKO-Fragen verwendet werden oder allgemeinere Begriffe. Wenn die Frage zu breit gestellt ist, ist sie nicht beantwortbar, und es ist daher unmöglich, alle Studien, die wichtig sein könnten, zu erfassen.

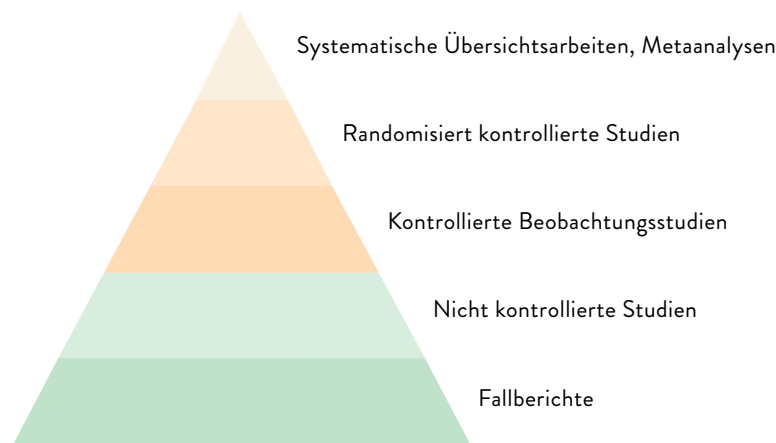
II. Literatursuche in mehreren Datenbanken?

Im nächsten Schritt wird eine umfangreiche systematische Literatursuche durchgeführt. Sie stellt die Basis einer gut durchgeführten systematischen Übersichtsarbeit dar.⁸ Eine Literaturrecherche sollte in zumindest zwei Datenbanken, wie z. B. MEDLINE, Embase, Cochrane Library, durchgeführt werden. Das Ergebnis von systematischen Übersichtsarbeiten wird jedoch genauer, wenn in mehr als zwei Datenbanken nach Studien gesucht wird und wenn zusätzlich Referenzlisten von

| | keine Müdigkeit nach dem Kaffee | Müdigkeit nach dem Kaffee | Gesamtzahl der Probanden |
|--------------------|---------------------------------|---------------------------|--------------------------|
| Kaffee mit Koffein | 70 | 30 | 100 |
| Kaffee koffeinfrei | 20 | 80 | 100 |

Tab. 1: Beispiel Koffein-Studie zur Berechnung des relativen und absoluten Risikos

Abb. 3: Hierarchie der Evidenz



ausgewählten Studien überprüft werden. Zusätzlich könnte auch „graue Literatur“, beispielsweise auf Konferenzen vorgestellte Studien, interessant sein, die nicht publiziert wurde.

III. Auswahl der Studien durch zwei Reviewer?

Die Ergebnisse der Literatursuche werden im nächsten Arbeitsschritt, im idealen Fall von zwei Personen unabhängig voneinander, nach den zuvor definierten Auswahlkriterien bewertet.⁹ Studien, die für die systematische Übersichtsarbeit relevant sind, werden im gemeinsamen Konsens ausgewählt. Dabei haben manche Studiendesigns eine höhere Wahrscheinlichkeit, falsche Ergebnisse zu liefern als andere. Randomisiert kontrollierte Studien (RCT) zum Beispiel werden höher eingestuft als Fallserien (Abb. 3).

IV. Qualität der Studien überprüft?

Studien mit methodischen Limitationen können das Resultat der systematischen Übersichtsarbeit verzerren. In der systematischen Übersichtsarbeit sollte daher die Qualität der Studien dargestellt sein, inklusive validierter Bewertungsinstrumente (z. B. Cochrane-Handbuch), die verwendet wurden, um die Studien zu überprüfen.

V. Einschätzung über das Vertrauen in die Ergebnisse?

Studien mit guter methodischer Qualität sollten bei der Zusammenfassung der Ergebnisse mehr Gewicht erhalten als Studien mit methodischen Problemen. Zusätzlich sind Studienergebnisse häufig unpräzise. Die Autoren der systematischen Übersichtsarbeit sollten eine Einschätzung darüber abgeben, wie hoch das Vertrauen in die Studienergebnisse ist: das heißt, wie hoch die Wahrscheinlichkeit ist, dass zukünftige Studien die Ergebnisse ändern werden.

Relative Kennzahlen in Studien – Vorsicht bei der Interpretation

Häufig werden auf Kongressen Studienergebnisse vorgestellt. Für dichotome Endpunkte gibt es immer zwei Möglichkeiten: zum Beispiel entweder krank oder gesund, tot oder lebend. Eine sehr anschauliche Kennzahl, um ein Studienergebnis darzustellen, ist das absolute Risiko. Das absolute Risiko ist die Wahrscheinlichkeit, dass ein bestimmtes Ereignis in einem bestimmten Zeitraum auftritt. Stellen Sie sich vor, Sie führen eine Studie durch, um herauszufinden, wie viele Personen 30 Minuten nach einem Kaffee wieder fit sind, und teilen 100 Personen der Gruppe zu, die koffeinhaltigen Kaffee bekommt, und weitere 100 Personen der Gruppe, die

koffeinfreien Kaffee trinkt (Tab. 1). Nach 30 Minuten sind von 100 Personen, die nach dem Mittagessen einen koffeinhaltigen Kaffee getrunken haben, 70 Personen wieder fit. Das absolute Risiko, nach 30 Minuten mit einem koffeinhaltigen Kaffee wieder fit zu sein, beträgt demnach 70 Prozent. Für die 100 Personen, die einen koffeinfreien Kaffee erhielten, sind nach 30 Minuten 20 ohne Koffein wieder fit, also nur 20 Prozent.

Häufig wird in den Studien das relative Risiko präsentiert, das ist ein Verhältnis der beiden Risiken.¹⁰ Dividiert man nun 70 Prozent durch 20 Prozent, ergibt das 3,5 – das relative Risiko. Ein relatives Risiko von 3,5 bedeutet, dass das Risiko, 30 Minuten nach einem koffeinhaltigen Kaffee wieder fit zu sein, um das 3,5-Fache höher ist als nach einem koffeinfreien Kaffee. Ähnlich wie das relative Risiko wird auch die Odds-Ratio interpretiert. Dabei sollte allerdings beachtet werden, dass die Odds-Ratio das relative Risiko bei hohen Ereigniszahlen überschätzt.

Häufig wird in den Medien oder in Werbematerialien die relative Risikozunahme oder -reduktion präsentiert, die einer prozentuellen Änderung entspricht. Angewandt auf unser Beispiel mit dem Kaffee könnte man behaupten, dass mit koffeinhaltigem Kaffee die Wahrscheinlichkeit, fit zu sein, um 250 % zunimmt (= relative Risikozunahme). Denn nur 20 von 100 Personen waren nach 30 Minuten wieder fit, und mit koffeinhaltigem Kaffee waren es um 250 % mehr, nämlich 70 von 100. Die relative Risikozunahme wird tatsächlich häufig verwendet, daher empfiehlt es sich für Ärzte, immer auf die tatsächlichen Ereigniszahlen zu achten.

Das Konfidenzintervall – ein Maß für die Genauigkeit

Für Studien gilt, dass Ergebnisse genauer sind, wenn mehr Personen an einer Studie teilnehmen. Ein Maß für die Genauigkeit ist das Konfidenzintervall, das umso breiter ist, je weniger Personen an einer Studie teilnehmen. Meist wird das 95 %-Konfidenzintervall dargestellt, das sich aus

einer statistischen Berechnung ergibt. Ein 95 %-Konfidenzintervall bedeutet, dass für 100 Stichproben mit derselben Teilnehmerzahl ein Konfidenzintervall berechnet wird und der wahre Wert bei 95 der 100 Konfidenzintervalle enthalten ist.¹¹

P-Wert: Ist das Ergebnis Zufall oder nicht?

Der p-Wert gibt die Wahrscheinlichkeit dafür an, dass ein Resultat nur durch Zufall entstanden ist.¹¹ Ein p-Wert von $\leq 0,05$ gilt als statistisch signifikant. Das heißt, die Wahrscheinlichkeit eines falsch positiven Resultats ist 5 %.

Metaanalysen

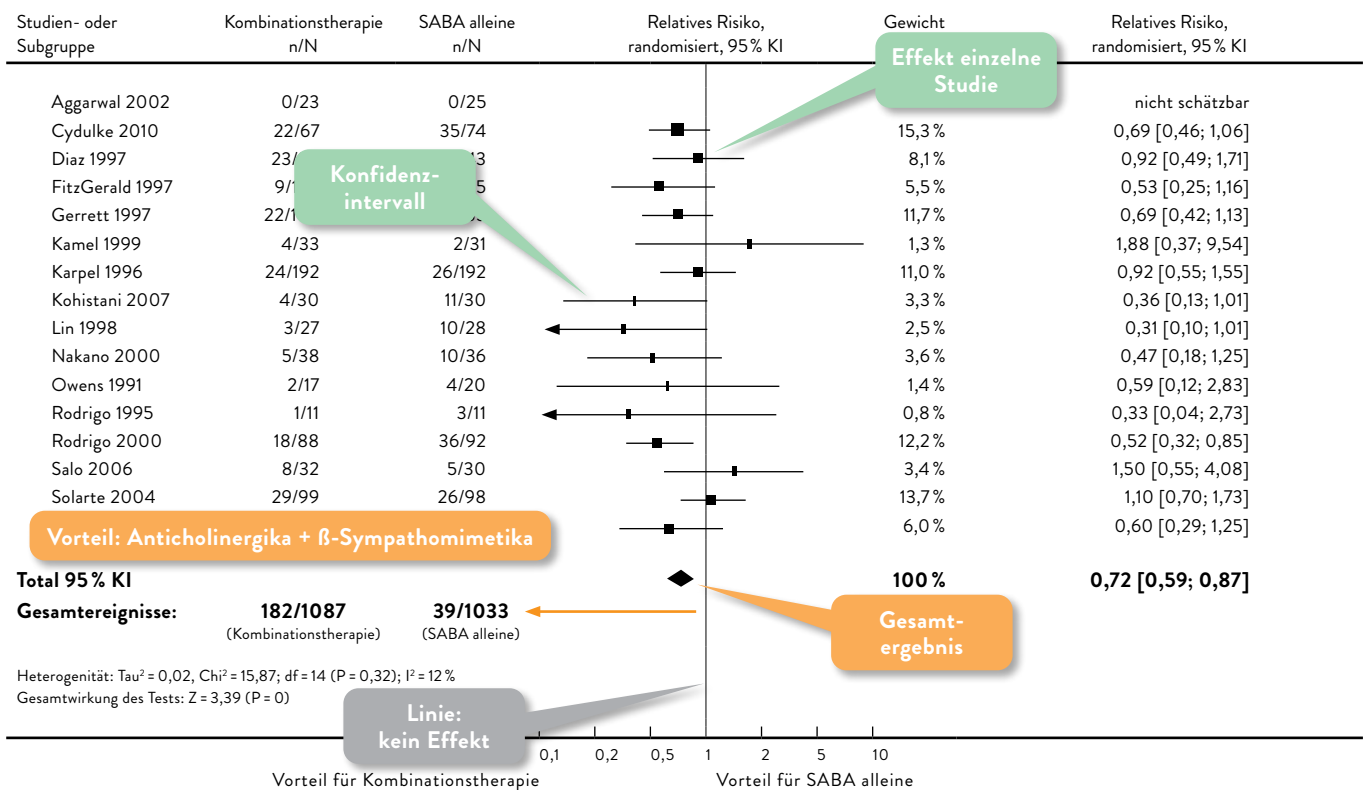
In einer Metaanalyse werden Ergebnisse von verschiedenen, jedoch vergleichbaren Studien zusammengefasst.¹⁰ In einer Metaanalyse sollten alle für eine bestimmte Fragestellung relevanten Studien berücksichtigt werden. Daher wird eine

Metaanalyse im Idealfall im Rahmen einer systematischen Übersichtsarbeit durchgeführt. Bei einer Metaanalyse werden die Ergebnisse der Einzelstudien zusammengefasst und ein gemeinsamer Effekt wird errechnet. Das Gesamtergebnis ist genauer und die statistische Aussagekraft ist erhöht, da mehrere Studien mit einer insgesamt höheren Anzahl an Studienteilnehmern berücksichtigt werden.^{12, 13} In einer Metaanalyse kann ein Vorteil einer Behandlung nachgewiesen werden, auch wenn die Einzelstudien, die in der Metaanalyse zusammengefasst werden, aufgrund von zu geringen Teilnehmerzahlen keine ausreichende statistische Signifikanz erreichen. In einer Metaanalyse wird gewichtet summiert, d. h., Studien mit einer größeren Zahl an Teilnehmern erhalten mehr Gewicht als kleinere Studien. Größere Studien liefern genauere Ergebnisse, da Zufallsergebnisse („random error“) eine geringere Rolle spielen als bei kleineren Studien.

Die Resultate der Metaanalyse werden zu meist in einem „Forest Plot“ grafisch dargestellt. Abb. 4 zeigt den „Forest Plot“ einer Metaanalyse von 16 Studien, die untersuchen, ob eine Kombination von Anticholinergika und β -Sympathomimetika besser wirkt als β -Sympathomimetika alleine, um die Hospitalisierungsrate bei Patienten mit akutem Asthma bronchiale zu reduzieren.¹⁴

In einem „Forest Plot“ werden die Ergebnisse der Einzelstudien und der Gesamteffekt mit dem zugehörigen Konfidenzintervall abgebildet. Die schwarzen Vierecke mit den horizontalen Linien stellen die Einzelstudien dar. Die schwarzen Vierecke sind umso größer, je mehr Personen an einer Studie teilnehmen. Die Konfidenzintervalle sind die horizontalen Linien, auf denen die schwarzen Vierecke liegen. Das Konfidenzintervall ist ein Maß für die Streuung und umso breiter, je größer die Streuung ist. Der schwarze Diamant symbolisiert den

Abb. 4: Metaanalyse: Hospitalisierung bei akutem Asthma bronchiale (modifiziert nach Kirkland SW et al. 2017)¹⁴



Cochrane Österreich Workshops 2022

Kritische Bewertung medizinischer Studien

10.–11.05.2022, Universität für Weiterbildung Krems

Kritische Bewertung von randomisiert kontrollierten Studien – Onlinekurs

11.10.2022

Systematische Literaturrecherche (Grund- und Aufbaukurs)

17.–18.10.2022, Universität für Weiterbildung Krems

Medizinische Statistik für Nicht-StatistikerInnen

08.–09.11.2022, Universität für Weiterbildung Krems

Ursache-Wirkung aus Big Data ableiten – geht das?

28.–29.11.2022, Universität für Weiterbildung Krems



statistisch zusammengefassten Gesamteffekt aller in die Metaanalyse inkludierten Studien.

In unserem Beispiel wurde das relative Risiko („Risk-Ratio“) als Effektmaß verwendet und es beschreibt, um wie viel geringer das Risiko für eine Hospitalisierung ist, wenn Patienten mit akutem Asthma mit einer Kombination aus Anticholinergika und β -Sympathomimetika behandelt werden, verglichen mit β -Sympathomimetika alleine.

Das Ergebnis von 16 Studien mit 2120 Teilnehmenden zeigt, dass das Risiko für eine Hospitalisierung mit der Kombinationstherapie um 28 % geringer ist als mit β -Sympathomimetika alleine (RR; relatives Risiko: 0,72; 95 % CI: 0,59–0,87). In der Gruppe, die mit Anticholinergika und β -Sympathomimetika behandelt wurde, wurden 17 % (182 von 1087) im Spital aufgenommen verglichen mit 23 % (239 von 1033), die mit β -Sympathomimetika alleine behandelt wurden.

Metaanalysen können nur mit Studien durchgeführt werden, die ähnlich sind in Bezug auf Population, Studiendesign und Intervention.¹⁵ In einem „Forest Plot“ ist von relevanten Ungleichheiten auszugehen, wenn sich die Einzelstudien mit den zugehörigen Konfidenzintervallen nicht oder nur wenig überschneiden. Das Ausmaß der

Heterogenität wird mit dem I²-Test errechnet und sollte in jeder Metaanalyse erhoben werden.¹⁶

Literatur:

- 1 Sackett DL et al.: Evidence based medicine: what it is and what it isn't. *BMJ* 1996; 312(7023): 71
- 2 Densen P: Challenges and opportunities facing medical education. *Trans Am Clin Climatol Assoc* 2011; 122: 48-58
- 3 Haynes RB: Of studies, syntheses, synopses, and systems: the “4S” evolution of services for finding current best evidence. *Evid Based Med* 2001; 6(2): 36
- 4 Qaseem A et al.: Guidelines International Network: toward international standards for clinical practice guidelines. *Ann Intern Med* 2012; 156(7): 525-31
- 5 Semlitsch T et al.: Development of a workable mini checklist to assess guideline quality. *Z Evid Fortbild Qual Gesundheitswes* 2014; 108(5-6): 299-312
- 6 Pae CU: Why systematic review rather than narrative review? *Psychiatry Investig* 2015; 12(3): 417-9
- 7 AMSTAR 2: Assessing the methodological quality of systematic reviews: Bruyère Research Institute. https://amstar.ca/Amstar_Checklist.php; zuletzt aufgerufen am 11.3.2022
- 8 Egger M et al.: How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? *Empirical study. Health Technol Assess* 2003; 7(1): 1-76
- 9 Buscemi N et al.: Single data extraction generated more errors than double data extraction in systematic reviews. *J Clin Epidemiol* 2006; 59(7): 697-703
- 10 Fletcher RH et al.: *Clinical Epidemiology: The Essentials*. Philadelphia/Baltimore/New York/London/Buenos Aires/Hong Kong/Sydney/Tokyo: Wolters Kluwer, Lippincott Williams & Wilkins, 2012
- 11 Magnusson K: Interpreting confidence intervals: an interactive visualization. <https://rpsychologist.com/d3/ci/>; zuletzt aufgerufen am 11.3.2022
- 12 Egger M et al.: Meta-analysis: principles and procedures. *BMJ* 1997; 315(7121): 1533-7
- 13 Cohn LD, Becker BJ: How meta-analysis increases statistical power. *Psychol Methods* 2003; 8(3): 243-53

- 14 Kirkland SW et al.: Combined inhaled beta-agonist and anticholinergic agents for emergency management in adults with asthma. *Cochrane Database Syst Rev* 2017; 1(1): Cd001284
- 15 Glasziou PP et al.: Investigating causes of heterogeneity in systematic reviews. *Stat Med* 2002; 21(11): 1503-11
- 16 Higgins JP et al.: Measuring inconsistency in meta-analyses. *BMJ* 2003; 327(7414): 557-60

Fortbildungsanbieter:

Gesellschaft der Ärzte in Wien – Billrothhaus

Lecture Board:

Dr. Brigitte Piso, MPH
Department für Evidenzbasierte Medizin und Evaluation
Universität für Weiterbildung Krems

Dr. Reinhard Dörflinger
Arzt für Allgemeinmedizin
Wien



+ 1 DFP-Punkt



<https://allgemeineplus.at/article/137476>