



AUS DEM NETZWERK EVIDENZBASIERTE MEDIZIN

Systematische Übersichtsarbeiten und ihre „Fallstricke“

VON DR. MED. NELE KORNDER UND DR. MED. NICOLE LINDNER IM AUFTRAG DES NETZWERKS EVIDENZBASIERTE MEDIZIN E.V. (WWW.EBM-NETZWERK.DE)



ZWEI UNTERSCHIEDLICHE ARTEN VON ÜBERSICHTSARBEITEN

Nur sehr selten reicht das Ergebnis einer einzelnen Studie, um robuste Entscheidungen, sei es in der Therapie, in der Diagnostik oder Prophylaxe, zu treffen. Aus diesem Grund spielen Übersichtsarbeiten eine herausragende Rolle in der evidenzbasierten Medizin. Sie dienen der Zusammenfassung der bestehenden Literatur zu einer spezifischen Fragestellung.

Man unterscheidet zwischen der *narrativen* und der *systematischen* Übersichtsarbeit (gleichbedeutend Reviews).

Erstere ist die schlichte Zusammenfassung eines Themas, denn bei einer narrativen Übersichtsarbeit kann allenfalls bedingt davon ausgegangen werden, dass alle verfügbaren Studien herangezogen wurden und diese auch folgerichtig bewertet wurden. Es erfolgt keine systematische Suche, was das Risiko, Arbeiten zu übersehen, mit sich bringt. Zudem fließt häufig

die persönliche Meinung der Autor:innen mit ein.

Im Gegensatz dazu ist es das Ziel einer systematischen Übersichtsarbeit, die gesamte (verfügbare) Evidenz bzw. alle einschlägigen Studien zu einer konkreten klinischen Fragestellung nach einem vorher festgelegten Protokoll zu identifizieren, auszuwählen und kritisch zu bewerten.

Heute gibt es zu sehr vielen Fragestellungen systematische Reviews. Ein einheitliches methodisches Vorgehen erleichtert das Verständnis und die kritische Interpretation.

SYSTEMATISCHE ÜBERSICHTSARBEITEN SIND GRUNDLEGENDE IN DER EVIDENZBASIERTEN MEDIZIN

Übersichtsarbeiten sind entscheidend, denn man sollte sich zur Planung einer Therapie oder auch Diagnostik auf sie verlassen können. Hierdurch lassen sich unnötige Risiken für Patient:innen und ein unnötiger Einsatz von Ressourcen verhindern. Aus diesem Grund ist es wichtig, dass Reviews systematisch durchgeführt werden, da dies dazu beiträgt, Fehler zu minimieren und eine umfassende Datenlage gewährleistet.

Ein Beispiel für die Relevanz systematischer Übersichtsarbeiten ist die vorgeburtliche Gabe von Kortikosteroiden an Schwangere, bei denen von einer Frühgeburt ausgegangen wird.

Bereits 1972 zeigte eine Studie, dass die Wahrscheinlichkeit, dass ein Neugeborenes verstirbt, reduziert wird, wenn die Mutter vor Geburt Kortikosteroide erhalten hat. In der Folge gab es weitere kleinere Studien, teils mit widersprüchlichen Ergebnissen. Erst

1989 wurden diese Studien in einer Übersichtsarbeit zusammengefasst. Hierdurch konnte die Wirkung von Kortikosteroiden erkannt werden. Wären die Ergebnisse früher systematisch zusammengetragen worden, hätte vielleicht das Leid mancher Patient:innen verhindert werden können [1].

Dieses Beispiel findet sich passenderweise in dem Logo der Cochrane Collaboration wieder [2] (siehe Kasten unten). Sie wurde 1993 mit dem Ziel gegründet, die Erkenntnisse aus randomisierten kontrollierten Studien systematisch und transparent zusammenzufassen. Heutzutage hat die Organisation eine herausragende Rolle in der Erstellung hochwertiger systematischer Übersichtsarbeiten. Ziel ist es, Ärzt:innen, Wissenschaftler:innen und politische Entscheidungsträger:innen dabei zu unterstützen, evidenzbasierte Entscheidungen zu treffen [3].

Im Folgenden skizzieren wir die methodische Vorgehensweise einer systematischen Übersichtsarbeit kurz und konkretisieren diese anhand eines Beispiels. Anschließend werden beispielhaft mögliche Verzerrungspotenziale („Biases“) erläutert.

BEDEUTUNG DES COCHRANE-LOGOS



Cochrane

Das Cochrane-Logo bezieht sich auf eine reale systematische Übersichtsarbeit, die zur Wirksamkeit der pränatalen Gabe von Kortikosteroiden auf die Sterblichkeit bei Frühgeborenen durchgeführt wurde: Die Quer-Balken stehen für die Ergebnisse einzelner Studien. Die Raute ganz unten symbolisiert das Gesamtergebnis. Sie befindet sich auf der linken Seite. Das heißt: Die Behandlung ist offenbar wirksam.

FRAGESTELLUNG EINER SYSTEMATISCHEN ÜBERSICHTSARBEIT

Ausgehend von einem klinischen Problem wird eine entsprechende Forschungsfrage abgeleitet. In unserem Fallbeispiel geht es um eine diagnostische Fragestellung. Hierbei ist die Orientierung an den so genannten „PIRT“-Kriterien sinnvoll:

P = POPULATION – die Population, der eine Patient:in mit einer bestimmten Eigenschaft angehört

I = INDEXTEST – der diagnostische Test, der untersucht werden soll

R = REFERENZTEST – der Test, mit dem der Indextest verglichen wird (im besten Falle der „Goldstandard“)

T = TARGETCONDITION – der Zustand von Interesse

BEISPIEL:

Ein Patient, der unter einem Insulinpflichtigem Diabetes mellitus Typ 2 leidet, kommt zu Ihnen in die Praxis. Er habe große Sorge, dass er eine Hypoglykämie nicht rechtzeitig mitbekommen würde. Ein Bekannter verwende ein Gerät zur kontinuierlichen Glukosemessung (CGM). Das sei eine ganz tolle Sache. Er würde von dem Gerät immer zuverlässig vor einer Unterzuckerung gewarnt, ohne sich ständig piksen zu müssen. Haben Sie schon einmal etwas davon gehört Frau Doktor? Wäre das nicht auch etwas gegen meine Ängste? Wie gut können diese Geräte denn eine Hypoglykämie erkennen?

Schlüsselt man die sich aus dem Fallbeispiel abgeleitete Fragestellung nun nach PIRT auf, ergibt sich Folgendes:

P = Erwachsene mit Insulinpflichtigem Diabetes mellitus

I = CGM

R = kapilläre / venöse Messung

T = Hypoglykämie

Sie könnten sich nun eine Vielzahl einzelner Studien zu dem Thema durchlesen. Dies wäre nicht nur sehr zeitintensiv, darüber hinaus könnte Ihr Blick auf die Ergebnisse verzerrt sein. Selbst wenn man Expert:in auf



einem bestimmten Gebiet ist, heißt das nicht, dass man auch alle Studien kennt, die sich mit eben dieser Thematik befassen. Daher machen Sie sich auf die Suche nach einem entsprechenden Review zur Hypoglykämie-detektion mittels minimalinvasiver Glukosemessung, um Ihrem Patienten die Frage evidenzbasiert beantworten zu können – und werden fündig. Anhand der gefundenen Beispielarbeit kann man das methodische Vorgehen in einem Flussdiagramm für die verschiedenen Phasen der systematischen Übersichtsarbeit (vgl. Abbildung 1) nachvollziehen [4].

WIE LASSEN SICH ALLE ARBEITEN FINDEN?

Zunächst werden relevante Suchbegriffe miteinander kombiniert. In unserem Beispiel ist dieser sogenannte Syntax aus Suchbegriffen wie „Hypoglykämie“ und „kontinuierliche Glukosemessung“ aufgebaut. Anschließend erfolgt eine Suche in allen relevanten Datenbanken (u.a. z.B. „PubMed“).

Die Auswahl der Datenbanken hängt dabei von der Fragestellung ab. Zusätzlich wird eine sogenannte Handsuche durchgeführt: Referenzlisten werden nach passenden Artikeln durchsucht, ähnliche Artikel werden in den Datenbanken gescreent. Ziel ist es, tatsächlich alle Arbeiten zu finden, die für die Beantwortung der Frage durchgeführt und publiziert worden sind.

WIE WERDEN DIE GEFUNDENEN STUDIEN DURCHGESCHAUT?

In der Regel führt die Suche zu einer Vielzahl von Artikeln. In unserem Beispiel wurden 6351 Artikel in allen Datenbanken und 36 zusätzliche Artikel gefunden („identification“ in der Abbildung 1). Das Filtern der gefundenen Studien erfolgt nun in mehreren Schritten: Zunächst werden alle Artikel, die beim Durchsuchen der entsprechenden Datenbanken sowie zusätzlicher Quellen gefunden wurden, zusammengetragen und Duplikate entfernt. In unserem Beispiel ergeben sich 3416 Artikel. Anschließend werden erst Titel und Abstracts von zwei Personen gescreent und diejenigen ausgeschlossen, die nicht der Fragestellung entsprechen („Screening“ in Abbildung 1). Artikel, die von einer Per-

son als möglicherweise relevant eingeschätzt wurden, werden nun von beiden Personen genauer beurteilt. In unserem Beispiel sind das 502 Artikel. Hierfür wird der Volltext des jeweiligen Artikels anhand der vorher festgelegten Kriterien beurteilt („eligibility“ in Abbildung 1). Letztlich konnten in unserem Beispiel 15 Studien eingeschlossen werden.

WAS PASSIERT MIT DEN PASSENDEN STUDIEN?

Am Ende des Screeningprozesses stehen häufig einige Studien, welche sich mit der Fragestellung beschäftigen und den zuvor festgelegten Kriterien entsprechen. Ein wichtiger Aspekt systematischer Übersichtsarbeiten ist, dass spezifische Methoden verwendet werden, um die Ergebnisse eingeschlossener Studien zusammenzufassen. Dies kann durch eine narrative Synthese geschehen, bei der die Ergebnisse in beschreibender Form zusammengefasst werden, oder durch statistische Verfahren wie die Meta-Analyse, bei der die Ergebnisse der eingeschlossenen Studien kombiniert werden, um eine quantitative Schätzung der Gesamtwirkung zu erhalten. In unserem Beispiel lässt sich so eine „gepoolte“ Sensitivität und Spezifität der Hypoglykämiedetektion angeben.

WO IST DER HAKEN? – MÖGLICHE FALLSTRICKE

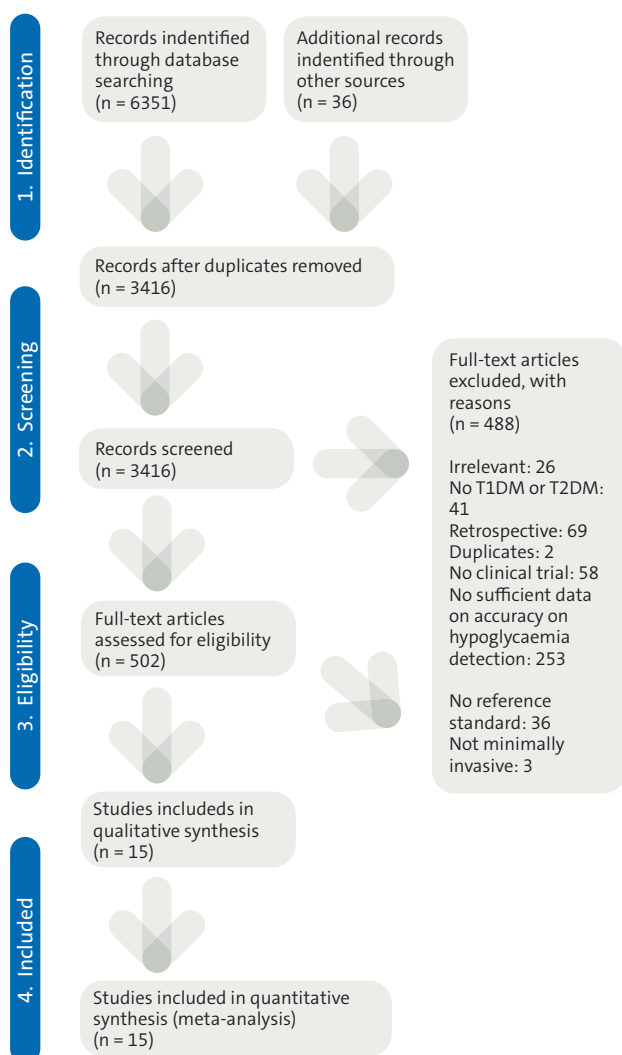
Auch systematische Übersichtsarbeit und Metaanalyse sind nicht „der Stein der Weisen“ und ihnen kann man nicht blind vertrauen. Es können Verzerrungspotenziale (auch „biases“ genannt) auftreten, die kritisch betrachtet werden sollten, bevor man die Ergebnisse in Hinblick auf eine klinische Fragestellung anwendet.

Schwächen zu Beginn können einen großen Einfluss auf das Ergebnis eines systematischen Reviews haben: Wenn Einschlusskriterien, Suchbegriffe oder auch die Quellenauswahl unangemessen gewählt wurden, kann dies dazu führen, dass nicht alle einschlägigen Studien berücksichtigt werden und das Ergebnis der Übersichtsarbeit nicht der tatsächlichen Datenlage entspricht.

Besonders relevant ist dies für kleinere Studien, welche ein negatives nicht-signifikantes Ergebnis zeigen.

ABBILDUNG 1: FLUSSDIAGRAMM EINES REVIEWS ZUR HYPOGLYKÄMIEDETEKTION MITTELS MINIMALINVASIVER GLUKOSEMESSUNG [4]

1. Identifikation („Identification“) der in der Datenbank und zusätzlichen Quellen gefundenen Datensätze („records“)
2. In der Vorauswahl („Screening“) werden nach Entfernen der Duplikate 3416 Studien aufgenommen.
3. Die Volltexte werden auf „Eignung“ („Eligibility“) geprüft.
4. 14 Studien werden eingeschlossen („Included“).



Studien mit einem positiven signifikanten Ergebnis werden schon ca. 2,5 Mal so oft eingereicht [5]. Studien mit negativen Ergebnissen werden zudem seltener zitiert („citation bias“) und oft zeitverzögert berichtet („time-lag bias“). Die Chance, dass eine kleine Studie mit unspektakulärem Ergebnis in eine Übersichtsarbeit eingeschlossen wird, ist deutlich geringer. Dies kann einen großen Einfluss auf die Übersichtsarbeit haben. Die tatsächliche Wirksamkeit einer Therapie könnte überschätzt werden. Dieser Publikationsbias wurde im Jahr 1986 von Simes anhand zweier Krebstherapien näher beschrieben. Er verglich die Daten der publizierten Literatur zur Auswirkung auf das Überleben mit den Daten eines Registers für Krebsstudien. Es zeigte sich, dass in beiden Fällen die Auswirkungen der Therapien auf das Überleben wesentlich geringer waren oder sogar komplett verschwanden [6].

Im Zusammenhang mit systematischen Reviews lohnt sich ein Blick auf die Autor:innen und mögliche Interessenskonflikte, die zu einem „confirmation bias“ führen können. Dies ist der Fall, wenn Autor:innen eine bestimmte Meinung über die Wirksamkeit einer Maßnahme vertreten und nur Studien in die Überprüfung einbeziehen, die diese Meinung unterstützen [7].

Es gibt verschiedene Checklisten, die zur Bewertung der Qualität einer systematischen Übersichtsarbeit verwendet werden können. Geläufig sind beispielsweise das AMSTAR (Assessing the Methodological Quality of Systematic Reviews) tool [8] oder das ROBIS (Risk of Bias in Systematic Reviews) tool [9]. Etwaige biases können so identifiziert werden.

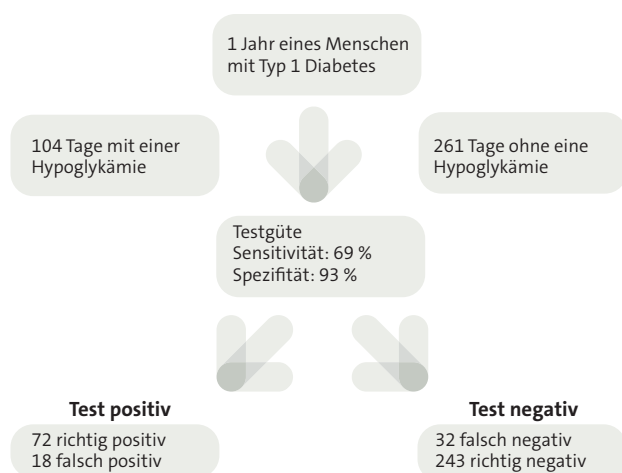
UND WAS IST NUN MIT MIR, FRAU DOKTOR?

Wie sieht es nun mit der Messgenauigkeit der kontinuierlichen Glukosemessung aus? Mit der ist es leider, laut dem Ergebnis unseres Beispiels, nicht so gut bestellt. Mit einer mittleren Sensitivität von 69% werden Hypoglykämien zu häufig verpasst. Am besten lassen sich diese Ergebnisse mit einer „Bierdeckelrechnung“ veranschaulichen (siehe nächste Seite, Abbildung 2).

In einem Jahr mit kontinuierlicher Glukosemessung erlebt ein Mensch mit Typ 1 Diabetes etwa 32 Tage mit



ABBILDUNG 2: "BIERDECKELRECHNUNG" ZUR VERANSCHAULICHUNG DER TESTGÜTE



In einem Jahr erlebt ein Mensch mit Typ 1 Diabetes etwa 18 Tage mit falsch positiven Alarmen und 32 Tage mit falsch negativen Messungen. Dieser Schätzung liegt eine wöchentliche Inzidenz von zwei Episoden symptomatischer Hypoglykämie pro Woche und Patient zugrunde. [10] Diese Häufigkeit wurde nicht in Hausarztpraxen in Deutschland nachgewiesen.

falsch-negativen Messungen (dieser Schätzung liegt eine wöchentliche Inzidenz von zwei Episoden symptomatischer Hypoglykämie pro Woche und Patient zugrunde [10]). Die Werte lassen sich nicht vollständig auf den Patienten übertragen, da die zugrundeliegende Anzahl von Hypoglykämien aus einem anderen Kontext stammen. Trotzdem müssen wir den Patienten enttäuschen. Seine Ängste vor einer Unterzuckerung lassen sich leider auch mit einer kontinuierlichen Glukosemessung nicht sicher ausräumen. Wenn er sich dennoch dazu entschließt, muss er trotzdem auf Symptome achten und ggf. zusätzlich messen. ■



DR. MED. NELE KORNDER
Wissenschaftliche Mitarbeiterin
Philipps-Universität Marburg
Institut für Allgemeinmedizin



DR. MED. NICOLE LINDNER
Wissenschaftliche Mitarbeiterin
Philipps-Universität Marburg
Institut für Allgemeinmedizin

References

- [1] Evans I, Thornton H, Chalmers I, Glasziou P. Testing Treatments: Better Research for Better Healthcare. 2nd. London 2011.
- [2] Das Cochrane Logo | Cochrane Schweiz; 2023 [cited 2023 January 15] Available from: URL: <https://swiss.cochrane.org/de/unsere-logo>
- [3] Shah HM, Chung KC. Archie Cochrane and his vision for evidence-based medicine. *Plast Reconstr Surg* 2009; 124(3): 982–8 <https://doi.org/10.1097/PRS.0b013e3181b03928> [PMID: 19730323]
- [4] Lindner N, Kuwabara A, Holt T. Non-invasive and minimally invasive glucose monitoring devices: a systematic review and meta-analysis on diagnostic accuracy of hypoglycaemia detection. *Syst Rev* 2021; 10(1): 145 <https://doi.org/10.1186/s13643-021-01644-2> [PMID: 33971958]
- [5] Dickersin K. The Existence of Publication Bias and Risk Factors for Its Occurrence. *JAMA* 1990; 263(10): 1385 <https://doi.org/10.1001/jama.1990.03440100097014>
- [6] Simes RJ. Publication bias: the case for an international registry of clinical trials. *J Clin Oncol* 1986; 4(10): 1529–41 <https://doi.org/10.1200/JCO.1986.4.10.1529> [PMID: 3760920]
- [7] Catalogue of Bias Collaboration. Spencer, E. A., Heneghan C. Confirmation bias in: Catalogue Of Bias; 2018 [cited 2023 January 15] Available from: URL: <https://catalogofbias.org/biases/confirmation-bias/>
- [8] Shea BJ, Reeves BC, Wells G, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ* 2017; 358: j4008 <https://doi.org/10.1136/bmj.j4008> [PMID: 28935701]
- [9] Whiting P, Savović J, Higgins JPT, et al. ROBIS: A new tool to assess risk of bias in systematic reviews was developed. *J Clin Epidemiol* 2016; 69: 225–34 <https://doi.org/10.1016/j.jclinepi.2015.06.005> [PMID: 26092286]
- [10] Frier BM. Defining hypoglycaemia: what level has clinical relevance? *Diabetologia* 2009; 52(1): 31–4 <https://doi.org/10.1007/s00125-008-1209-3> [PMID: 19018507]