



AUS DEM NETZWERK EVIDENZBASIERTE MEDIZIN

Integritäts-Bewertungsinstrument dringend benötigt

Umgang mit problematischen Studien in systematischen Übersichtsarbeiten

**VON DR. RER. NAT. BIRGIT SCHINDLER, CLAUDIA BREUER UND PROF. DR. MED. JÖRG MEERPOHL IM
AUFTRAG DES NETZWERKS EVIDENZBASIERTE MEDIZIN E. V. (WWW.EBM-NETZWERK.DE)**



Aktuelle, methodisch hochwertige systematische Übersichtsarbeiten zu Gesundheitsinterventionen fassen alle relevante Evidenz für eine Fragestellung zusammen und spielen eine entscheidende Rolle bei der Entwicklung von Leitlinienempfehlungen und bei medizinischen Versorgungsentscheidungen. Sie können jedoch auch ungewollt Fehlinformationen verbreiten, wenn fragwürdige oder nicht authentische Studien in die Evidenzbasis einfließen. Verfügbare Instrumente zur Bewertung eingeschlossener Studien, wie beispielsweise das Risk of Bias-Tool (RoB-2) von Cochrane, gehen davon aus, dass die Studien und die berichteten Daten echt sind. Doch wenn Studien manipuliert oder komplett erfunden sind, aber eine hochwertige Methodik vorgeben, kann dies das RoB-2-Tool nicht aufdecken [1].

„Nicht authentische“ Studien werden oft erst dann als solche bekannt, wenn sie von den publizierenden

Fachzeitschriften zurückgezogen werden. Die Zahl der zurückgezogenen Forschungsartikel hat im Jahr 2023 die Marke von 10.000 überschritten und damit einen neuen jährlichen Höchstwert erreicht [2]. Die Gründe für die Rücknahme einer Publikation mögen vielfältig sein und von unbeabsichtigten Fehlern und naiven Irrtümern bis hin zu wissenschaftlichem Fehlverhalten reichen. Einer Studie zufolge liegt das Zurückziehen jedoch in den meisten Fällen tatsächlich an wissenschaftlichem Fehlverhalten wie nachgewiesenem oder vermutetem Betrug, doppelten Veröffentlichungen oder Plagiaten [3]. Und Integritätsexperten vermuten, dass die zurückgezogenen Studien nur die Spitze des Eisbergs sind [1].

PAPER MILLS – GEFÄLSCHTE WISSENSCHAFTLICHE STUDIEN ALS GESCHÄFTSMODELL

In den letzten Jahren hat sich ein besorgniserregender Trend in der wissenschaftlichen Publikationslandschaft etabliert: die systematische Produktion gefälschter Forschungsartikel durch profitorientierte Dienstleister, sogenannte „Paper Mills“. Diese haben sich darauf spezialisiert, wissenschaftliche Arbeiten gegen Bezahlung zu erstellen. Mithilfe von künstlicher Intelligenz (KI) arbeiten sie äußerst effizient und können dank eines fragwürdigen Netzwerks sogar die Veröffentlichung in Fachjournalen ermöglichen [4]. Und diese KI-generierten, gefälschten Forschungsarbeiten sind nicht mehr einfach als Plagiate oder durch schlechte Qualität zu erkennen.

Schätzungen zufolge gelangen jährlich tausende gefälschte Paper durch Paper Mills in den wissenschaftlichen Kreislauf [4].

Das Problem ist global, betrifft aber insbesondere Länder mit hohem Publikationsdruck, wie China, Indien, Russland und einige osteuropäische Staaten [4]. Aber auch in westlichen Ländern gibt es eine wachsende Zahl von Forschenden, die sich durch solche Dienste Vorteile verschaffen wollen.

COCHRANE GEHT DAS PROBLEM AN

Um das Problem der gefälschten Daten bzw. komplett gefälschter Studien bei der Erstellung von systematischen Reviews in den Griff zu bekommen, entwickelt Cochrane derzeit ein „Integritätswerkzeug“ (INSPECT-SR Tool: INveStigating ProblEMatic Clinical Trials in Systematic Reviews). Durch verschiedene Informations-Checks sollen zukünftig Indizien für mögliche Integritätsprobleme gesammelt und zu einer Gesamteinschätzung zusammengeführt werden. Studien, bei denen es dann ernsthafte Zweifel an der Vertrauenswürdigkeit der Daten oder der gesamten Studie gibt, sollen dadurch früh im Reviewprozess erkannt werden. Umso besser, wenn dabei nicht nur Datenfälschungen, sondern auch unbeabsichtigte Fehler wie Flüchtigkeitsfehler und Versäumnisse im Datenmanagement bei an sich „echten“ Studien aufgedeckt werden. Gleichzeitig muss aber sichergestellt werden, dass „echte“ Studien nicht zu Unrecht unter Verdacht geraten und nicht berücksichtigt werden.

Um diese Herausforderung zu bewältigen und einen ausgewogenen Ansatz zu finden, wurde das INSPECT-SR-Projekt ins Leben gerufen.

→ Fortsetzung nächste Seite

DAS INSPECT-SR-PROJEKT

Im Rahmen des vom NIHR (National Institute for Health and Care Research) geförderten INSPECT-SR-Projekts entwickelt Cochrane derzeit gemeinsam mit einem internationalen Expertengremium ein Instrument zur Identifizierung „nicht-authentischer“ Studien. Dabei werden empirische Evidenz und Konsensmethoden kombiniert, um ein möglichst zuverlässiges Instrument zu entwickeln, das anschließend durch umfassende Tests weiter verfeinert und optimiert werden soll [5].

Das INSPECT-SR-Projekt umfasst fünf Phasen:

Sammlung von Prüfmethoden: In einer Umfrage unter 71 Experten in 16 Ländern wurde eine umfassende Liste von 102 potenziellen Prüfkriterien identifiziert [6].

Bewertung der Machbarkeit: Die gesammelten Methoden wurden auf 95 RCTs in Cochrane Reviews angewendet, um die Anwendbarkeit und die Auswirkungen zu bewerten [7].

Delphi-Umfrage und Konsensfindung: Derzeit wird durch eine Delphi-Umfrage und anschließende Konsenstreffen ermittelt, welche der Prüfkriterien in das endgültige INSPECT-SR-Werkzeug aufgenommen werden sollen.

Praxistest und Verfeinerung: Das entwickelte Werkzeug soll an neuen systematischen Übersichtsarbeiten getestet und basierend auf Nutzerfeedback optimiert werden.



→ Fortsetzung

WENN DIE STUDIENBASIS MIT DER ZEIT KLEINER STATT GRÖßER WIRD

Wie weitreichend das Phänomen „nicht-authentischer Studien“ tatsächlich ist, bleibt unklar. Es deutet sich jedoch in einigen aktualisierten Cochrane Reviews bereits an, dass nicht nur die präklinische Forschung, sondern auch die klinische Forschung von Fälschungen und Manipulationen betroffen ist.

Konkret kamen zwei unter Nutzung von Vorläuferinstrumenten von INSPECT-SR aktualisierte Cochrane Reviews zu häufig eingesetzten Mitteln – Vitamin D in der Schwangerschaft und Statine nach einer Herzoperation – in ihren aktualisierten Versionen unter Berücksichtigung einer kleineren Anzahl von Studien zu einem anderen Schluss als die vorherigen Versionen auf Basis einer größeren Anzahl von Studien.

In beiden Fällen führten die Cochrane-Autor:innen in der aktualisierten Version ein Vorab-Screening der zu ihrer Forschungsfrage gefundenen Studien auf wissenschaftliche Integrität der Studien durch und schlossen zum Teil Studien aus, die sie in der vorherigen Version berücksichtigt hatten.

COCHRANE REVIEW ZU VITAMIN D IN DER SCHWANGERSCHAFT

Einer der Reviews drehte sich um die Frage, ob eine Nahrungsergänzung mit Vitamin D während der Schwangerschaft zur Vorbeugung von Komplikationen sinnvoll und verträglich ist [8].

Die vorherige Version war auf Basis von 30 Studien zu dem Schluss gekommen, dass die Nahrungsergänzung mit Vitamin D in der Schwangerschaft wahrscheinlich das Risiko für Präeklampsie, Schwangerschaftsdiabetes und geringem Geburtsgewicht reduziert und möglicherweise auch das Risiko für schwere Blutungen nach der Geburt verringert.

Die Verwendung eines Integritäts-Instruments, des sogenannten Cochrane Pregnancy and Childbirth Trust-

worthiness Screening Tools [9], führte dazu, dass 21 Studien aus der vorherigen Version des systematischen Reviews nun nicht mehr in die Analyse einfließen. Und dies hatte Auswirkungen auf die Schlussfolgerungen: Nun wird die Vertrauenswürdigkeit der Evidenz für die Wirksamkeit der Vitamin-D-Supplementation in der Schwangerschaft als niedrig oder sehr niedrig eingestuft, und die meisten berichteten Ergebnisse für Mutter und Kind stützen sich nur auf Daten aus einer, zwei oder drei Studien.

Im Bereich „Nahrungsergänzung in Schwangerschaft und Stillzeit“ scheint es möglicherweise generell ein Problem mit der Integrität von Studien zu geben. In 14 von 18 speziell mit diesem Tool untersuchten Cochrane Reviews aus diesem Bereich führte eine formale Bewertung der Integrität zum Ausschluss von Studien und bei sechs veränderten sich durch die vordefinierten Kriterien für Integrität die Schlussfolgerungen für die Praxis [9]. Das bedeutet jedoch nicht zwangsläufig, dass die ausgeschlossenen Studien tatsächlich Fälschungen sind. Vielmehr wurde ein Teil dieser Studien der Kategorie „awaiting classification“ zugeordnet.

COCHRANE REVIEW ZU STATINEN ZUR VORBEUGUNG VON KOMPLIKATIONEN NACH EINER HERZOPERATION

Der zweite Review ging der Frage nach, ob es sinnvoll ist, vor einer Herzoperation Statine einzunehmen, um beispielsweise einem Vorhofflimmern in den ersten Wochen nach der Operation vorzubeugen [10].

Ältere und neuere Studien kamen zu unterschiedlichen Ergebnissen. Ältere Studien, die weder eine Registrierung vor Studienbeginn, noch die Genehmigung einer Ethikkommission aufwiesen und vergleichsweise wenig Teilnehmende einschlossen, wurden in der aktuellen Version des Reviews nicht mehr berücksichtigt. Ausgewertet wurden jetzt nur noch 8 Studien (zuvor: 23 Studien) mit insgesamt 5592 Teilnehmenden, fünf davon neuere RCTs, die in der vorherigen Version gar nicht enthalten waren.

Man könnte hier die restriktive Vorgehensweise der Autor:innen infrage stellen und bezweifeln, ob die Grundlage für den Ausschluss einiger RCTs ausreichend war. Schließlich wurde die Registrierung erst seit den 2000er Jahren verpflichtend. Für das Vorgehen spricht aber nicht zuletzt, dass der Ausschluss kleinerer RCTs, die nicht registriert waren und keinen Nachweis über die ethische Genehmigung erbrachten, dazu beiträgt den "Small-Study-Effect" zu vermeiden. Als solcher wird in der wissenschaftlichen Forschung eine Verzerrung (Bias) bezeichnet, die dadurch entsteht, dass kleinere Studien tendenziell deutlichere Effekte berichten als größere Studien. So hatten die älteren Studien auf eine ausgeprägte Reduktion von Komplikationen nach einer Herzoperation unter Statin-Anwendung hingedeutet [11].

Der aktuelle Review findet im Unterschied zur vorherigen Version keine Hinweise für einen Nutzen hinsichtlich postoperativer Komplikationen. Statine beeinflussten weder die Sterblichkeit innerhalb der ersten 30 Tage nach der Herz-Operation, noch schwerwiegende Ereignisse wie Herzinfarkt, Schlaganfall, Nierenversagen, Vorhofflimmern oder die Dauer des Aufenthalts auf der Intensivstation. Werden Statine sowieso regelmäßig eingenommen, spricht nichts dagegen, sie weiter einzunehmen, denn schädliche Auswirkungen von Statinen wurden in der Phase rund um eine Herzoperation nicht berichtet.

INDIZIEN FÜR „NICHT AUTHENTISCHE“ STUDIEN SAMMELN

Um die „Verunreinigung“ der Evidenzbasis durch nicht authentische Studien zu vermeiden, soll das

BEISPIELE DES VORLÄUFIGEN INSPECT-SR-INSTRUMENTS [12]

<p>Überprüfung von Mitteilungen nach der Veröffentlichung</p>	<ul style="list-style-type: none"> ● Wurde die Studie im Nachhinein zurückgezogen? ● Wurde im Nachhinein ein „Expression of Concern“ zur Studie veröffentlicht? ● Wurde bei anderen Studien des Teams im Nachhinein ein „Expression of Concern“ zur Studie veröffentlicht oder die Studie zurückgezogen?
<p>Überprüfung der Durchführung und Transparenz</p>	<ul style="list-style-type: none"> ● Liegt die Genehmigung einer Ethikkommission vor? ● Wurde die Studie prospektiv registriert? ● Ist die Rekrutierung plausibel? (beispielsweise zu viele Betroffene einer seltenen Erkrankung innerhalb von kurzer Zeit in einem Studienzentrum)
<p>Überprüfung von Text und Abbildungen</p>	<ul style="list-style-type: none"> ● Gibt es plagierte Textstellen? ● Gibt es Duplikationen von Abbildungen?
<p>Überprüfung der Ergebnisse</p>	<ul style="list-style-type: none"> ● Sind die Baseline-Charakteristika plausibel? ● Passt die Zahl der randomisierten Personen zum Randomisierungsverfahren (Anzahl der Blöcke)? ● Keine oder kaum Studienabbrüche trotz längerer Laufzeit? ● Gibt es nicht plausible Ergebnisse bzw. statistische Auffälligkeiten?



sich derzeit in Entwicklung befindende Instrument – das INSPECT-SR-Tool (INVeStigating ProblEmatic Clinical Trials in Systematic Reviews) – eine transparente Bewertung der Integrität von RCTs in systematischen Übersichtsarbeiten ermöglichen und so zur Sicherstellung der wissenschaftlichen Validität der SR-Ergebnisse beitragen [5]. Das INSPECT-SR-Integritäts-Tool wird verschiedene Kriterien enthalten, um so Indizien für mögliche Integritätsprobleme zu sammeln (vgl. Kasten Seite 25).

Es zeichnet sich schon jetzt ab, dass es keinen eindeutigen und sicheren Referenz-Standard für verdächtige oder gefälschte Studien geben kann, und das INSPECT-SR-Tool wird auch kein eindeutiger und verlässlicher diagnostischer Test für Betrug sein. Es soll den Review-Autor:innen aber in Zukunft helfen, die Integrität der eingeschlossenen Studien besser einzuschätzen, indem Indizien für Fehler oder Hinweise für mangelnde Authentizität gesammelt werden und so zu einer Gesamteinschätzung zusammengeführt werden. ■



**DR. RER. NAT.
BIRGIT SCHINDLER**
wissenschaftliche Mitarbeiterin der
Cochrane Deutschland Stiftung



CLAUDIA BREUER
wissenschaftliche Mitarbeiterin der
Cochrane Deutschland Stiftung



**PROF. DR. MED.
JÖRG MEERPOHL**
Direktor Cochrane Deutschland
und Direktor des Instituts für
Evidenz in der Medizin des Universi-
tätsklinikums Freiburg