(23) Wie funktionieren Meta-Analysen eigentlich?

Description

Meta-Analysen sind Methoden, mit denen man die Ergebnisse von verschiedenen Studien statistisch zusammenführen kann. So ähnlich wie man in einer Gruppe von Schülern die mittlere Körpergröße ausrechnen kann, indem man die Länge aller Schüler zusammenzählt und durch die Anzahl der Schüler teilt, kann man in einer Meta-Analyse den mittleren Effekt von Studien bestimmen.

Man muss sich dazu für eine Metrik entscheiden. Diese Metrik richtet sich nach den Ergebnisparametern, die die Studien verwenden.

Dichotom oder kontinuierlich?

Grundsätzlich muss man zwischen kontinuierlichen und dichotomen Maßen unterscheiden. Kontinuierliche sind solche, die ein Kontinuum abbilden. Z.B. eine Skala wie eine Schmerzskala, eine Depressionsskala, oder die Hyperaktivitätsskala von Conners, die eine Ratingskala darstellt, bei der verschiedene Fragen bewertet und am Schluss aufaddiert werden, sodass ein kontinuierlicher Wert entsteht. Das ist die bei den ADHS-Studien am meisten verwendete Skala. Auch Blutdruckwerte, überhaupt Laborwerte und Messwerte, Körpergröße, Schuhgröße, Temperatur, all das sind kontinuierliche Maße.

Davon zu unterscheiden sind die dichotomen Maße: Tot oder lebendig, krank oder gesund, rückfällig oder nicht, groß oder klein, etc.

Effektstärken von Studien mit dichotomen Ergebnisparametern

Bei diesen dichotomen Maßen berichten die originalen Studien Häufigkeiten: so viele Menschen sind in der Behandlungsgruppe gestorben, so viele in der Kontrollgruppe. Dies wird auf die Gesamtzahl der Patienten in den jeweiligen Gruppen standardisiert. Die Zahlen werden ins Verhältnis gesetzt. Und man erhält eine Kennzahl für den Unterschied zwischen den Gruppen, der je nachdem, was genau ins Verhältnis gesetzt wird als Relatives Risiko, als Odds-Ratio, als Hazard-Ratio oder sonst eine Verhältniszahl bezeichnet wird. Bei diesen dichotomen Ergebnismaßen ist ein fehlender Unterschied zwischen den Gruppen durch eine Verhältniszahl von 1 gekennzeichnet. Ist die Ergebniszahl größer – z.B. 1,5 – dann ist eine der Gruppen um die Hälfte besser dran, also um 50 %. Welche das ist, das hängt davon ab, wie die Verhältniszahl gebildet wurde.

Beispiel: Stellen wir uns eine Ergebnistabelle folgender Art vor:

	Tot	lebendig	n
Behandlung	20	30	50
Kontrolle	30	20	50
n	50	50	100

Der Begriff „Odds-Ratio“ kommt vom Englischen „Odds“, was so viel heißt, wie „Chance“, bei einer Wette beispielsweise. Hier oben in der Tabelle wäre die Chance oder die Odds unter der Behandlung zu sterben 20/30 = 0,66. Und die Chance unter der Kontrollbehandlung zu sterben wäre 30/20 = 1,5. Die Odds Ratio ist nun das

© Prof. Harald Walach

Verhältnis der beiden: 0,66/1,5 = 0,44. Damit hätte eine Person in der Behandlungsgruppe eine 44 % höhere Chance zu überleben als in der Kontrollgruppe.

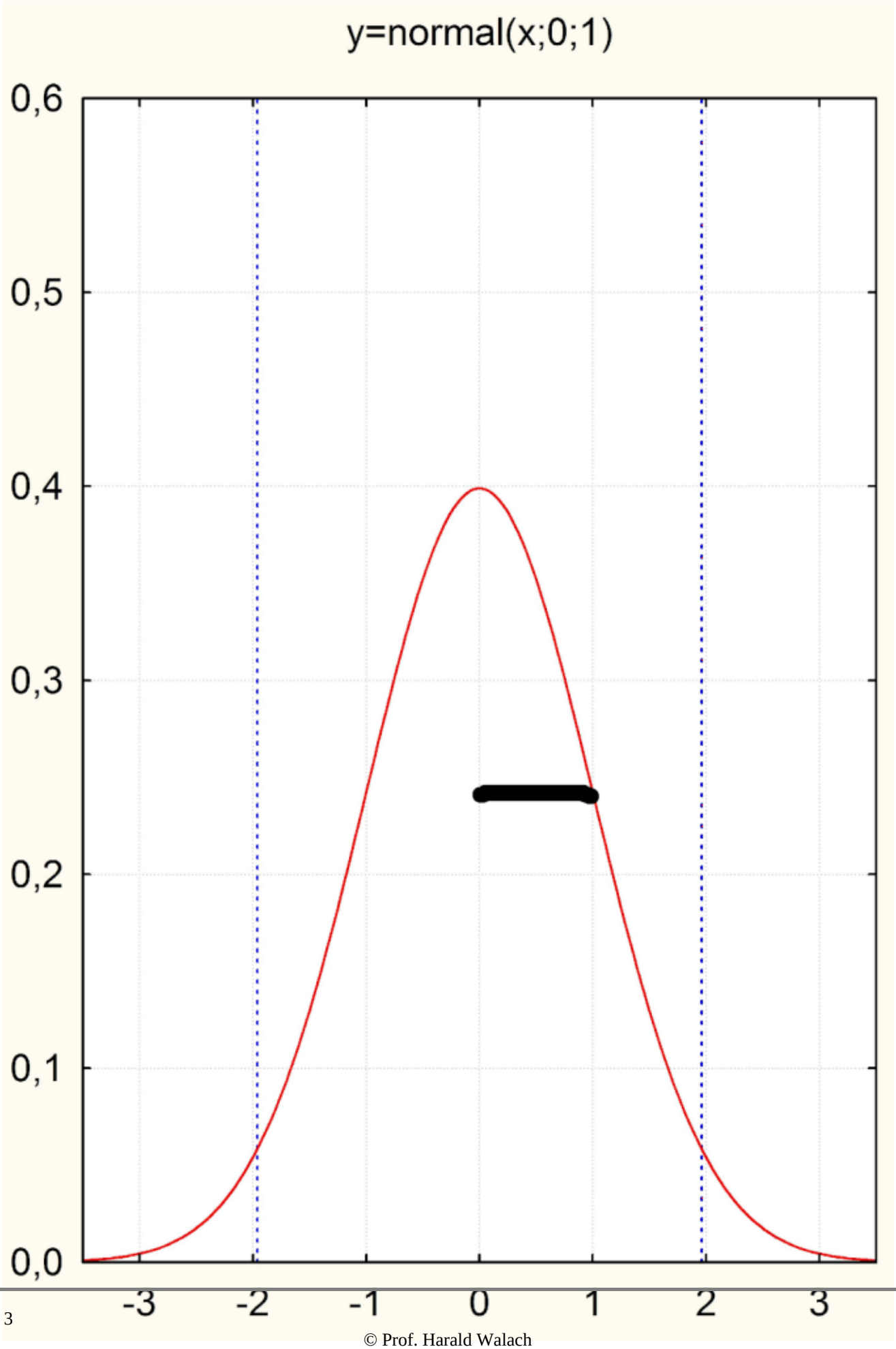
Das relative Risiko wird leicht unterschiedlich gebildet: Es wäre 20/50/30/50, würde also die Ereignisse in den beiden Gruppen ins Verhältnis zur Gesamtzahl in der Gruppe setzen und wäre damit 0,66.

Standardisierte Mittelwertdifferenz

Bei kontinuierlichen Maßen werden ebenfalls Kennzahlen gebildet, um den Unterschied zwischen den Gruppen auszudrücken. Dies waren die Kennzahlen in unseren ADHS-Meta-Analysen. In diesem Fall ist es eine Differenz zwischen den Ergebnissen der einen Gruppe minus die der anderen. Auch hier muss man darauf achten, wie die Differenz gebildet wird. Denn je nachdem kann manchmal ein positiver und manchmal ein negativer Wert eine Überlegenheit der Behandlungsgruppe ausdrücken.

Nun müssen wir aber noch das Problem lösen, dass unterschiedliche Variablen ganz unterschiedlich gemessen werden. Blutdruck etwa und Depressionsskalen haben ganz unterschiedliche Metriken. Um die Differenz zwischen Behandlungs- und Kontrollgruppe in einer Studie mit der von anderen Studien oder auch bei anderen Maßeinheiten vergleichbar zu machen, wird hier die Differenz durch die Standardabweichung oder die Streuung der Werte um den Mittelwert herum dividiert. Dies führt dazu, dass wir eine Metrik erhalten, die in den Einheiten der Standardnormalverteilung oder in Einheiten von Standardabweichungen ausgedrückt ist (Abb. 1). Die Standardabweichung gibt die Streuung der Verteilung um den Mittelwert herum an (bei einer einigermaßen normalverteilten Variable) und hat in der Standardnormalverteilung oder der Gaußkurve den Wert „1“ (siehe den dicken Strich in der Abb. 1 unten).

© Prof. Harald Walach



© Prof. Harald Walach

Abbildung 1 – Eine Normalverteilung mit Mittelwert 0 und Standardabweichung 1; der fette schwarze Strich markiert die Standardabweichung, oft als SD abgekürzt

Wenn man nun eine beliebige Differenz einer kontinuierlichen Variablen durch ihre Standardabweichung teilt, was geschieht dann? Man standardisiert diese Differenz auf die Standardnormalverteilung oder drückt den Unterschied in Einheiten von Standardabweichungen aus.

Diese Metrik, die nur für kontinuierliche Werte und damit deren Differenzen sinnvoll ist, nennt man die „standardisierte Mittelwertdifferenz (SMD)“, also den Unterschied zwischen zwei Gruppen, der mit einer metrischen Skala gemessen wurde. Am Ende ist diese SMD immer in der gleichen Metrik, egal, wie groß die Einheiten waren. Das Abkürzungssymbol für diese SMD ist in der Regel „d“ für „difference“.

Wir können sehen: Ein d = 1 ist ein Unterschied zwischen zwei Gruppen, der eine Standardabweichung beträgt. Wenn wir also die obere Kurve duplizieren würden und um den Abstand des fetten Striches, der Standardabweichung nach rechts oder links verschieben würden, dann hätten wir eine Effektgröße von d = 1 visualisiert.

Eine Effektgröße von d = 0.5 ist eine halbe Standardabweichung und bereits ein klinisch bedeutsamer Unterschied. Die englische Behörde NICE hat vor Zeiten einmal gefordert, dass neue Depressionsstudien mindestens eine Effektgröße von einer halben Standardabweichung gegenüber Placebo aufweisen sollten, damit sie es wert sind, bezahlt zu werden. In Tat und Wahrheit ist der kumulierte Effekt von Antidepressiva ca. d = 0.38, also deutlich geringer [1].

Dies, etwa ein Drittel Standardabweichung, ist in der Regel die Grenze, unterhalb derer die Effekte als „klinisch klein“ eingeschätzt werden. Die bereits erwähnte Meta-Analyse von Psychotherapiestudien hat damals ein d = 0.6 zutage gefördert. Dies wird als klinisch bedeutsam gesehen.

Wenn nun viele Studien meta-analysiert werden, in denen auch kleine Studien dabei sind, dann verwendet mal als Metrik eine leicht korrigierte Fassung der SMD, die dann den Namen „Hedge’s g“ trägt, nach Larry Hedges, der sie erfunden hat. Sie ist etwas kleiner, weil sie einen Korrekturfaktor enthält, der in Rechnung stellt, dass kleinere Studien die Werte gerne überschätzen. Aber ansonsten ist es ziemlich dasselbe.

Da bei einer Meta-Analyse zwei Gruppen vorliegen, die oft unterschiedliche Standardabweichungen haben, wird meistens eine sog. gepoolte oder gemischte Standardabweichung zur Berechnung verwendet. Das ist ein Wert, bei dem die beiden Standardabweichungen in beiden Gruppen gemittelt werden.

Berechnung von Effektgrößen zum Hausgebrauch

Wenn man Effektgrößenberechnungen zum Hausgebrauch für sich selber vornehmen will, dann kann man gut auch die konservative Variante nehmen und die größere von beiden Standardabweichungen zur Standardisierung heranziehen. Das kann jeder anhand einer Ergebnistabelle mit einem Taschenrechner machen. Wenn Sie zum Beispiel in einer Ergebnistabelle einer Depressionsstudie folgende Werte sehen (willkürliche Daten):

	vorher	nachher
Behandlungsgruppe	19.5 (4.3)	16.2 (5.7)
Kontrollgruppe	20.1 (4.5)	18.7 (4.8)

Dann können Sie die Veränderungen errechnen, also 3,3 für die Behandlungsgruppe und 1,4 für die Kontrollgruppe. Oder, um es noch einfacher zu machen: Sie können einfach die Werte am Ende der Behandlung

© Prof. Harald Walach

verwenden, denn wir gehen ja davon aus, dass durch die Randomisierung, also die Zufallszuteilung, die Ausgangswerte in beiden Gruppen nur zufällig schwanken und daher ignoriert werden können. Der Unterschied zwischen beiden Gruppen am Ende der Behandlung wäre 2,5. Nun verwenden Sie entweder die größere Standardabweichung von beiden, also 5,7 zur Standardisierung, oder Sie mitteln die beiden Werte in Klammern, also 5,25. Wenn wir 2,5 durch 5,7 teilen, erhalten wir d = 0,44, also einen Unterschied von 0,44 Standardabweichungen zwischen den Gruppen. Würden wir die Differenzwerte verwenden, also 1,9 und durch die gemittelte Standardabweichung teilen, dann wäre der Effekt kleiner, d = 0,38.

Man sieht, man kann mit dieser Methode bei einzelnen Studien sehr leicht Effektgrößen ausrechnen, um sich ein Bild über die klinische Bedeutsamkeit des berichteten Effektes zu machen.

Achtung: Manchmal berichten Ergebnistabellen in Studien nicht die Standardabweichung (SD), sondern den Standardfehler des Mittelwertes (SEM = Standard Error of the Mean). Das ist ein Wert, der die statistische Schwankung der Schätzung des Mittelwertes angibt. Während die Standardabweichung die Schätzung eines Verteilungswertes ist und mit der Größe der Studie nur insofern etwas zu tun hat, als dass größere Studien diese Schätzung präziser liefern, ist der Standardfehler des Mittelwertes direkt von der Studiengröße abhängig, und zwar über die Beziehung SEM = SD/Wurzel aus n. Man sieht damit unmittelbar: Je größer die Zahl der Studienteilnehmer n in einer Studie, umso kleiner wird der Standardfehler, also der Schätzfehler des Mittelwertes. Weil in dieser Formel aber die Standardabweichung enthalten ist, kann man sie, wenn nur SEM angegeben ist, zurückrechnen, indem man die Formel arithmetisch umformuliert. Dann erhält man SD = SEM * Wurzel aus n.

Zusammenfassung von Effektstärken

Bei einer Meta-Analyse wird nun für jede Studie ein Effektstärkemaß als Differenzmaß d bzw. g zwischen Behandlungs- und Kontrollgruppe gebildet, oder eben ein Verhältnismaß, wenn man es mit dichotomen Werten zu tun hat. Wenn eine Studie mehrere Ergebnisparameter hat, kann man entweder nur die Hauptzielkriterien nehmen. Oder man mittelt die Effektstärken auf Studienebene, oder, wenn z.B. über mehrere Studien hinweg ähnliche Outcome-Maße vorliegen, dann rechnet man für unterschiedliche Outcomes eine je eigene Analyse. Das hängt sehr von der Situation und dem Ziel der Analyse ab. Dieses Ziel und die verwendete Methodik muss man sich vorher überlegen und in einem Protokoll formulieren, das man idealerweise auch vorneweg in einer Datenbank registriert oder anderweitig publiziert. Dann können Leser der Meta-Analyse prüfen, ob man sich an seine eigenen Vorgaben gehalten hat und man selber schützt sich davor, durch Herumprobieren Ergebnisse zu „erzeugen“, die eigentlich eine Zufallsschwankung darstellen. Die entsprechende Datenbank für systematische Reviews und Meta-Analysen heißt „PROSPERO“.

Wenn man nun für jede Studie ein Effektstärkemaß d/g oder eine Verhältniszahl (Odds Ratio, Risk Ratio, etc.) hat, dann muss man diese für eine Meta-Analyse zusammenfassen.

Das funktioniert im Prinzip so: Man bildet einen Mittelwert, so ähnlich wie man bei einer Klasse von Kindern die mittlere Körpergröße ausrechnet. Dadurch erhält man auch automatisch einen Streuungswert, also eine Kennzahl dafür, wie stark diese einzelnen Effektstärken um den Mittelwert streuen. Wenn wir nun meine Abb. 4 aus dem Beitrag über die ADHS-Meta-Analyse dazunehmen, dann sehen wir: Die einzelnen Effektgrößen streuen sehr stark um den Mittelwert von ca. g = 0.2.

Eine solche Situation wird als heterogen beschrieben, und man geht in der Regel davon aus, dass es Einflüsse gibt, die man nicht kennt, die diese Streuung verursachen. Daher nimmt man für die Zusammenfassung solcher heterogener Effektstärken ein statistisches Modell, das annimmt, dass es nicht nur einen wahren Mittelwert und eine unbekannte Fehler-Abweichung davon gibt, sondern einen wahren Mittelwert, eine unbekannte Fehler-

© Prof. Harald Walach

Abweichung und einen Streuungsterm, hinter dem systematische Einflussgrößen stecken. Dieser wird geschätzt. In diesem Falle geht man von einem sog. Modell „zufälliger Effekte“ aus, andernfalls ist es ein einfacheres Modell „fester Effekte“.

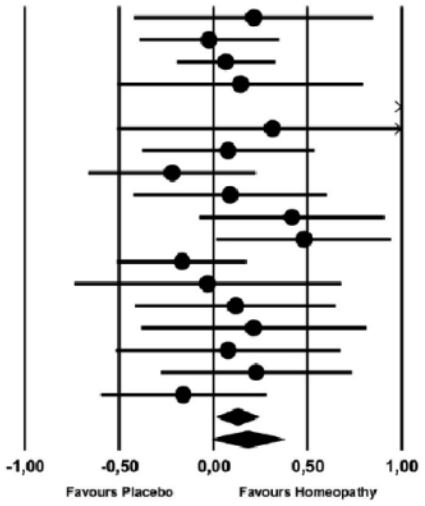
Durch die Bestimmung der Streuung kann man auch eine Signifikanzberechnung vornehmen. Diese sagt uns, ob eine gefundene Effektstärke statistisch überzufällig von null verschieden ist, und zwar völlig unabhängig von der Größe der Effektstärke.

Es gibt Meta-Analysen, die finden sehr kleine Effekte, die signifikant sind (z.B. weil alle Effekte sehr homogen sind, weil die Studien groß sind und viele Patienten untersucht haben). Es gibt Meta-Analysen, die sehr große Effekte isolieren, diese sind aber nicht signifikant (z.B. weil es nur wenige, kleine Studien gibt, die stark streuen).

Daher muss man immer auch auf die absolute Größe des Effektes blicken, nicht nur auf die Signifikanz.

Das Ergebnis einer Meta-Analyse wird dann in einem sog. Forest-Plot oder einer Baum-Grafik dargestellt. Ich drucke hier unsere Meta-Analyse der Arnica-Studien nochmals ab (Abb. 2).

Model	Study name	Hedges's g	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
	Jeffrey, S.L.A. et al	0,215	0,324	0,105	-0,419	0,850	0,665	0,506
	Robertson et al	-0,023	0,189	0,036	-0,393	0,347	-0,120	0,904
	Brinkhaus et al 1	0,067	0,132	0,018	-0,193	0,326	0,504	0,614
	Brinkhaus et al 2	0,145	0,332	0,110	-0,506	0,795	0,436	0,663
	Brinkhaus et al 3	2,011	0,322	0,104	1,379	2,643	6,238	0,000
	Chalet, S.R.	0,315	0,420	0,176	-0,508	1,138	0,750	0,453
	Hart et al	0,079	0,232	0,054	-0,376	0,533	0,340	0,734
	Kaziro	-0,218	0,226	0,051	-0,662	0,225	-0,965	0,334
	Kotlus et al.	0,090	0,262	0,069	-0,423	0,603	0,344	0,731
	Macedo	0,418	0,250	0,062	-0,071	0,908	1,675	0,094
	Pinsent	0,482	0,237	0,056	0,018	0,946	2,037	0,042
	Ramelet et al	-0,166	0,175	0,031	-0,509	0,176	-0,952	0,341
	Seeley	-0,030	0,361	0,130	-0,738	0,678	-0,083	0,934
	Sorrentino	0,119	0,271	0,073	-0,412	0,650	0,439	0,661
	Stevinson 1	0,214	0,304	0,092	-0,382	0,810	0,704	0,482
	Stevinson 2	0,079	0,303	0,092	-0,516	0,673	0,260	0,795
	Wolf	0,229	0,258	0,066	-0,277	0,734	0,886	0,375
	Erkan	-0,159	0,224	0,050	-0,598	0,279	-0,712	0,476
Fixed		0,129	0,056	0,003	0,019	0,239	2,297	0,022
Random		0,183	0,097	0,009	-0,007	0,373	1,891	0,059



I squared = 63,2

FIGURE 2 | Forest plot of main meta-analysis of placebo-controlled trials of *arnica* in postoperative complications.

Abb. 2 – Original Forest Plot der Arnica-Meta-Analyse (aus [2])

Jede Zeile ist eine Studie. Die Metrik ist Hedge’s g, also eine Variante der standardisierten Mittelwertdifferenz; daneben die Kennziffern (Standardfehler und Varianz), die man benötigt, um die Signifikanz der Effektgröße zu berechnen bzw. das 95%-Vertrauensintervall. Wenn dieses Vertrauensintervall sowohl bei den einzelnen Studien als auch bei der Zusammenfassung die Nullgrenze nicht mit einschließt, dann ist eine einzelne Studie bzw. der Gesamtwert signifikant.

Wir sehen etwa: Die meisten Studien clustern um den Mittelwert von g = 0.18, den die Analyse der zufälligen Effekte als gemeinsamen Mittelwert errechnet. Eine Studie, Brinkhaus et al 3, fällt komplett aus dem Rahmen: Sie hat eine riesige Effektgröße von g = 2 und passt gar nicht in das Grafikschema. Manche Studien sind sogar negativ. Das sind die, die auch in Abb. 4 aus dem Beitrag über die ADHS-Meta-Analyse unterhalb der Null-Linie landen. Individuell signifikant ist nur die Studie von Pinsent, weil hier die 95%-Vertrauensintervalle die Null-

© Prof. Harald Walach

Linie nicht schneiden.

Die Raute gibt den gemeinsamen Wert der mittleren Effektstärke an. Die Raute berührt die Null-Linie ganz leicht, weil die Irrtumswahrscheinlichkeit von p = 0.059 leicht über der konventionellen Signifikanzgrenze von 5 % liegt. Man sieht am p-Wert in der letzten Kolonne, dass der statistische Zusammenfassungswert einer Analyse fester Effekte durchaus signifikant ist. Aber man sieht am Heterogenitätsmaß, in dem Fall das I^2, das das Ausmaß der Streuung quantifiziert und signifikant ist, dass eine Analyse mit festen Effekten fehl am Platz ist.

Man erkennt also an einem solchen Forest Plot sowohl die einzelnen Effektgrößen, als auch deren Streuung und die zusammenfassende Effektstärke in der Raute. Der Abstand der Raute zur Nulllinie zeigt, wie groß der Effekt ist. Die Dicke der Raute bzw. ihre Überschneidung der Nulllinie oder nicht zeigt, wie stark dieser Summeneffekt von null verschieden ist oder nicht.

Als Beispiel für eine Meta-Analyse, die dichotome Maße zusammengefasst hat, sei die Analyse von Drouin-Chartier und Kollegen erwähnt [3]. Sie untersuchten den Einfluss des Konsums von Eiern auf Mortalität. Die Analyse ist in Abb. 3 dargestellt. Dies waren Kohortenstudien, also Beobachtungsstudien an zwei Gruppen, die teilweise über sehr lange Zeiträume Menschen, die Eier essen, mit solchen verglichen, die das nicht tun. Der Hintergrund ist die berühmt-berüchtigte Cholesterin-Hypothese der koronaren Herzkrankheit. Angeblich soll ja das Cholesterin im Ei gefährlich sein. Dies wurde in diesen Kohortenstudien untersucht: in über 32 Jahren, mit mehr als 5,5 Millionen Mannjahren. Man sieht an der Summenstatistik, des Effektmaßes „relatives Risiko“: Dieses ist RR = 0,98, also leicht unter 1. Das bedeutet: Eier essende Menschen haben sogar ein etwas geringeres Risiko, an Herzkrankheit zu sterben. Aber der Effekt ist nicht signifikant, weil das Vertrauensintervall 1, die Linie der Gleichheit oder des Nicht-Effektes, einschließt. Am Gewicht erkennt man, welche Studien größer waren; denn sie sind stärker gewichtet.

Wir wissen also nach 32 Jahren, vielen Millionen Forschungsdollars, was wir immer vermuteten: Eier sind nicht schädlich. Aber jetzt wissen wir es wirklich. Man sieht: Das Prinzip und die Darstellung ist die gleiche wie in der Analyse oben. Nur die Metrik ist eine andere, weil das Zielkriterium der Studien ein anderes war, nämlich ein dichotomes, z.B. Schlaganfall oder nicht, Herzinfarkt oder nicht, tot oder nicht.

© Prof. Harald Walach

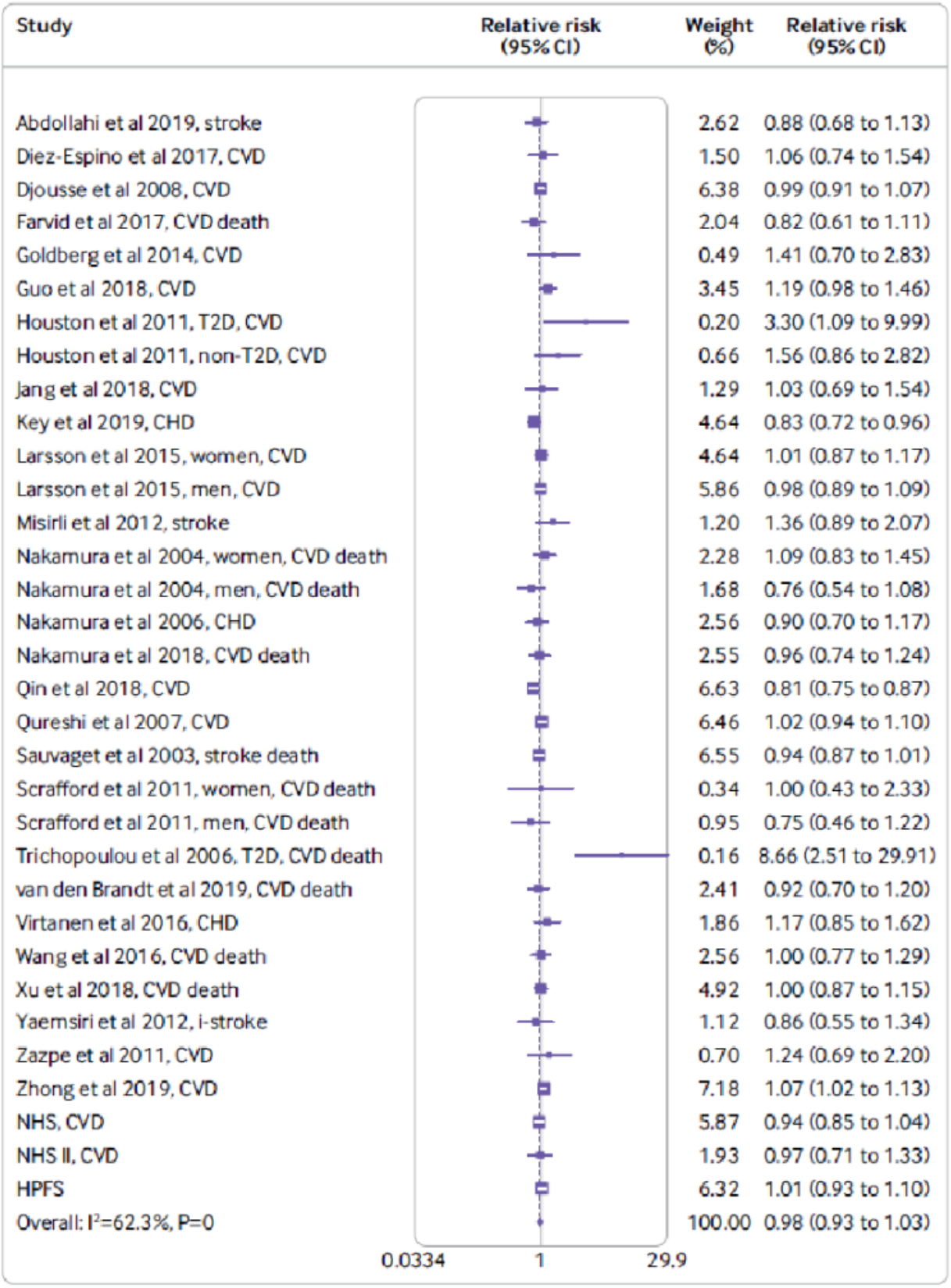


Abb. 3 – Meta-Analyse von Studien, die den Einfluss des Konsums von Eiern auf die Mortalität untersuchten, aus [3]

© Prof. Harald Walach

Sensitivitätsanalysen

Meta-Analysen haben mehrere Ziele. Eines ist, eine gemeinsame Effektstärke verschiedener Studien zu errechnen und herauszufinden, ob er signifikant von null verschieden, also statistisch bedeutsam ist. Ein anderes ist, den Effekt zu quantifizieren, also zu sehen, wie groß er ist. Ein statistisch bedeutsamer Effekt von d = 0.2 ist in der Regel klinisch nicht besonders interessant. Ein Effekt von d = 1.0 kann, auch wenn er nicht statistisch signifikant ist, trotzdem bedeutsam sein, weil das bedeuten könnte, dass man noch ein oder zwei Studien machen muss, um ihn abzusichern.

Aber oftmals ist es interessanter herauszufinden, was nun eigentlich die Streuung der Effektstärken verursacht. Das tut man mit Sensitivitätsanalysen. Studien sind ja oft unterschiedlich: Sie haben unterschiedliche Dauer, unterschiedliche Populationen, unterschiedliche Studiendesigns und Outdcome-Maße. Man kann indessen Studien nach diesen Unterschieden getrennt untersuchen und feststellen, ob die Heterogenität dadurch abnimmt. Dann kennt man die Treiber dieser Streuung.

Wenn eine mögliche Moderatorvariable kontinuierlich ist, dann kann man diese Variable im Rahmen einer Regressionsanalyse untersuchen. Dabei wird, wie oben kurz beschrieben, die Moderatorvariable, zum Beispiel die Jahreszahl der Publikation einer Studie, dazu verwendet, um die Effektgröße vorherzusagen. Hat die Variable einen Einfluss, ergibt sich ein signifikantes Modell.

Als Beispiel für ein signifikantes Regressionsmodell im Rahmen einer Meta-Analyse zeige ich hier unten in Abbildung 4 eine Meta-Regression aus unserer Meta-Analyse von Studien zu Achtsamkeitsinterventionen bei Kindern in Schulen [4]. Die Analyse hat insgesamt eine signifikante Effektgröße von g = 0,4 erbracht, bei kognitiven Maßen sogar g = 0,8. Die Heterogenität war aber sehr groß. Sie konnte geklärt werden durch eine Regression der Meditationsintensität auf die Effektstärke (Abbildung 4):

© Prof. Harald Walach

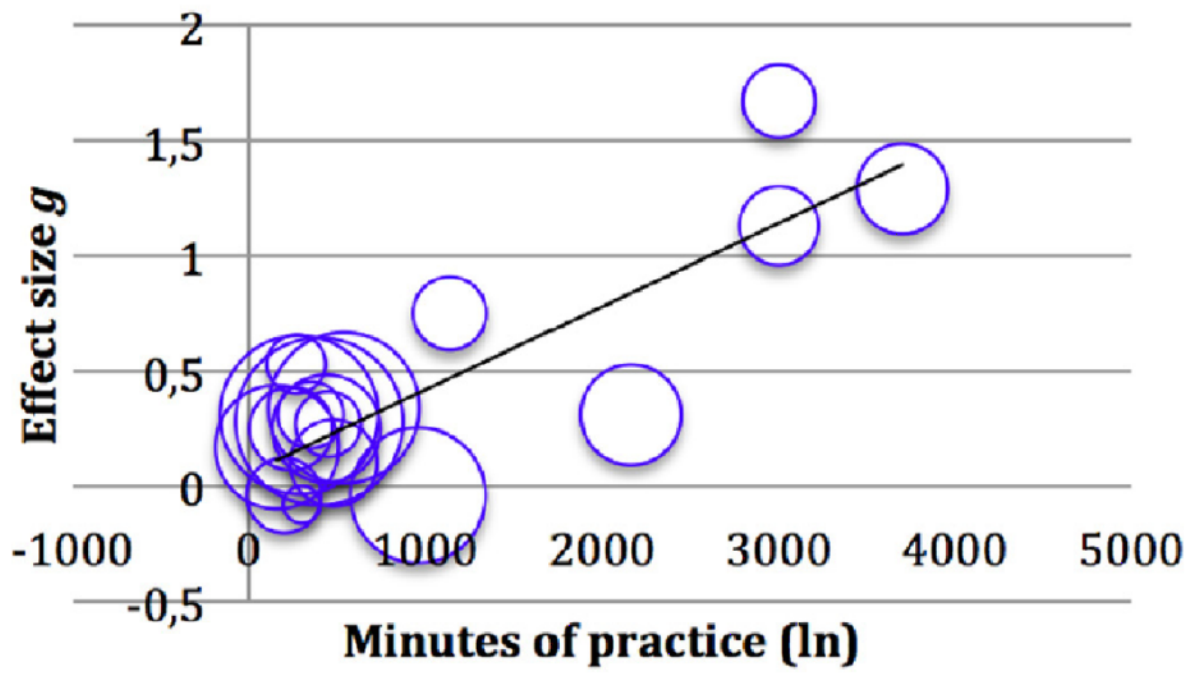


FIGURE 6 | Bubble plot of the 19 controlled effects sizes against Intensity of mindfulness training and regression line. R^2 (adjusted) = 0.52.

Abbildung 4 – Beispiel für eine Meta-Regression aus [4]: Einfluss der Meditationspraxis auf die Effektgröße bei einer Meta-Analyse von Achtsamkeitsinterventionen in der Schule – Je länger die Übungsdauer, umso größer der Effekt.

Wir wissen damit: Je länger (in der Tendenz) die Kinder meditierten, umso größer war der Effekt in der Studie. Die Größe der Blasen gibt die Größe der Studien an. Links, auf der y-Achse, ist die Effektgröße abgetragen. Unten auf der x-Achse die Dauer der Praxis. Man erkennt in der Tendenz, dass die Effektstärken ansteigen, je länger die Übungsdauer war. Natürlich gibt es auch Ausreißer: eine Studie mit sehr hohen Effekten und kurzer Praxis, und welche mit längerer Dauer und trotzdem kleinen Effekten. Aber in der Tendenz erkennt man einen Anstieg.

Solche Sensitivitätsanalysen helfen, um herauszufinden, was bei weiteren Studien beachtet werden muss. In dem Falle würde man etwa davon ausgehen, dass es nützlich ist, die Übungsdauer zu verlängern, wenn man größere Effekte sehen will.

In unserer ADHS-Meta-Analyse war zu sehen, dass der größte Effekt von einer Studie mit über einem Jahr Laufzeit kam. Also würde man bei einer weiteren Studie versuchen, die Behandlungsdauer zu verlängern.

© Prof. Harald Walach

Man kann Sensitivitätsanalysen auch dazu verwenden, um zu sehen, wie anfällig die Analyse gegenüber Annahmen ist. Dann würde man etwa Studien mit einem bestimmten Design herausnehmen. Oder man könntebei nicht-signifikanten Analysen ausrechnen, wie viele weitere Studien dieser Größe es bräuchte, um signifikante Effekte zu erhalten. Oder man könnte die Studien nach unterschiedlichen Typen von Interventionen herunterbrechen. Das hängt alles von der Fragestellung und dem Interesse der Forscherin ab.

Meta-Analysen sind Momentaufnahmen

Meta-Analysen sind nicht für Dauer. Man kann mit einer geschickten Fotografie beweisen, dass Pferde fliegen können. Nämlich dann, wenn man justament in dem Moment auf den Auslöser drückt, wenn das Pferd beim Galopp alle 4 Beine in der Luft hat. Pferde können aber natürlich nicht fliegen. So ähnlich ist das auch bei Meta-Analysen. Wenn man zu einem bestimmten Zeitpunkt alle Studien zusammennimmt, könnte sich durchaus ein signifikanter – oder nicht-signifikanter – Effekt ergeben. Kommt eine weitere gut gemachte Studie hinzu, kann sich das Bild wieder ändern.

Daher ist gerade bei älteren Analysen Vorsicht geboten. Und daher ist vor allem bei solchen Meta-Analysen Vorsicht geboten, bei denen nicht wirklich alle Studien eingeschlossen sind. Denn oftmals werden negativ ausgegangene Studien nicht publiziert. Das verzerrt natürlich das Bild. Daher ist es wichtig, bei publizierten Meta-Analysen zu kontrollieren, wie die Suchstrategie war. Sind auch nicht-publizierte Arbeiten erfasst? Ist auch die graue Literatur – Diplomarbeiten, Doktorarbeiten, akademische Abschlussarbeiten, in denen „schlechte“ Ergebnisse oft versteckt werden – mit erfasst?

Man erreicht diese Vollerfassung, indem man Forscher auf dem Gebiet kontaktiert, Firmen anschreibt etc. Bei Arzneimittelstudien ist es mittlerweile auch durchaus üblich, sich die Dokumente der Zulassungsbehörden kommen zu lassen. Das haben Peter Doshi, Peter Gøtzsche und Kollegen damals bei der EMA, der europäischen Zulassungsbehörde, erreicht [5]. Aber damit kommen tausende von Seiten Papier ins Haus geflattert.

Daher ist es nicht nur wichtig, beim Lesen von Meta-Analysen auf die Ergebnisse zu schauen, sondern auch darauf, wie die Literatursuche vonstattenging.

Auf jeden Fall sind Meta-Analysen nützlich, um den Stand einer Disziplin zusammenzufassen. Im Fall unserer Homöopathie-Analysen sieht man dadurch: Manchmal ist Homöopathie wirksam und manchmal sogar besser als Placebo. Auf jeden Fall bei ADHS.

Quellen und Literatur

1. Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R. Selective publication of antidepressant trials and its influence on apparent efficacy. New England Journal of Medicine. 2008;358:252-60.
2. Gaertner K, Baumgartner S, Walach H. Is homeopathic arnica effective for postoperative recovery? A meta-analysis of placebo-controlled and active comparator trials. Frontiers in Surgery. 2021;8:680930. doi: https://doi.org/10.3389/fsurg.2021.680930
3. Drouin-Chartier J-P, Chen S, Li Y, et al. Egg consumption and risk of cardiovascular disease: three large prospective US cohort studies, systematic review, and updated meta-analysis. BMJ. 2020;368:m513. doi: https://doi.org/10.1136/bmj.m513

© Prof. Harald Walach

4. Zenner C, Herrnleben-Kurz S, Walach H. Mindfulness-based interventions in schools – a systematicreview and meta-analysis. Frontiers in Psychology. 2014;5:art 603; doi: https://doi.org/10.3389/fpsyg.2014.00603
5. Doshi P, Jefferson T. The first 2 years of the European Medicines Agency's policy on access to documents: secret no longer. Archives of Internal Medicine. 2013;doi: https://doi.org/10.1001/jamainternmed.2013.3838.

Date Created
15. Juni 2022

© Prof. Harald Walach