

Arbeitsblätter zur kritischen Bewertung wissenschaftlicher Studien

(Centre for Evidence-based Medicine 2010)*

Therapie-Studien

Welche Frage wird in der Studie gestellt? Patienten – Interventionen – Control (Vergleich) – Outcome(s)

Schritt 1: Sind die Ergebnisse der Studie valide? (Interne Validität)

1a. R-War die Zuordnung der Patienten zu Behandlungen randomisiert?	
Was ist am besten?	Wo finde ich die Informationen?
Zentrale Computer- Randomisierung ist ideal und wird oft in multizentrischen Studien eingesetzt. Kleinere Studien können eine unabhängige Person dafür einsetzen (z. B. Krankenhaus-apotheke), um die Randomisierung zu überwachen.	Die Methoden sollten darüber Auskunft geben, wie die Patienten-Gruppen zugeteilt wurden und ob die Randomisierung verborgen war.
Von diesem Paper erfüllt: Ja <input type="radio"/> Nein <input type="radio"/> Unklar <input type="radio"/> Kommentar:	
1b. R-Waren die Gruppen zu Beginn der Studie ähnlich?	
Was ist am besten?	Wo finde ich die Informationen?
Wenn der Randomisierungsprozess erfolgreich war (das heißt, eine Vergleichbarkeit der Gruppen wurde erreicht), sollten die Gruppen ähnlich sein. Je ähnlicher die Gruppen sind, desto besser ist es. Es sollte Hinweise darauf geben, ob die Unterschiede zwischen den Gruppen statistisch signifikant sind (d. h. p-Werte).	Die Ergebnisse sollten eine Tabelle der „Baseline Charakteristiken“ aufweisen, welche die randomisierten Gruppen mit einer Reihe von Variablen, die das Ergebnis beeinflussen könnten, vergleicht (z. B. Alter, Risikofaktoren etc.). Wenn das nicht der Fall ist, könnten Gruppenähnlichkeiten in den ersten Absätzen des Abschnitts Ergebnisse beschrieben werden.
Von diesem Paper erfüllt: Ja <input type="radio"/> Nein <input type="radio"/> Unklar <input type="radio"/> Kommentar:	
2a. A – Wurden Gruppen, abgesehen von der zugeteilten Behandlung, gleich behandelt?	
Was ist am besten?	Wo finde ich die Informationen?
Die Patienten der verschiedenen Gruppen sollten, abgesehen von der Intervention, gleich behandelt werden, z. B. keine zusätzliche Behandlungen oder Untersuchungen sollten gemacht werden.	Suchen Sie im Methoden-Teil nach zusätzlichen Behandlungen, etc.
Dieses Paper: Ja <input type="radio"/> Nein <input type="radio"/> Unklar <input type="radio"/> Kommentar:	
2b. A – Kann man alle Patienten, die die Studie begonnen haben, nachverfolgen? Wurden sie in den Gruppen, denen sie angehören, analysiert?	
Was ist am besten?	Wo finde ich die Informationen?
Verluste im Follow-up sollten minimal sein – vorzugsweise weniger als 20%. Wenn allerdings nur wenige Patienten ein Ergebnis von Interesse aufweisen, dann können sogar kleine Verluste des Follow-ups dazu führen, dass Ergebnisse verfälscht werden. Die Patienten sollten ebenfalls in den Gruppen, zu denen sie randomisiert wurden, analysiert werden – „Intention-to-treat-Analyse“.	Der Abschnitt Ergebnisse sollte aussagen, wie viele Patienten randomisiert (z. B. Tabelle der Baseline Charakteristika) und wie viele Patienten tatsächlich in die Analyse einbezogen wurden. Sie müssen den Abschnitt Ergebnisse lesen, um die Anzahl und den Grund für Verluste im Follow-up nachvollziehen zu können.
Dieses Paper: Ja <input type="radio"/> Nein <input type="radio"/> Unklar <input type="radio"/> Kommentar:	
3. M – Waren die Messungen objektiv und waren die Patienten und Untersucher gegenüber den Behandlungen verblindet?	
Was ist am besten?	Wo finde ich die Informationen?
Es ist ideal, wenn die Studie „doppelt verblindet“ wurde. Das heißt, Patienten und Untersucher waren sich der Behandlungs-Zuordnung nicht bewusst. Wenn das Ergebnis objektiv ist (z. B. Tod), dann ist Verblindung weniger kritisch. Wenn das Ergebnis subjektiv ist (z. B. Symptome oder Funktionen), dann ist die Verblindung des Ergebnis-Gutachters kritisch.	Sehen sie als Erstes im Methoden-Teil nach, ob eine Maskierung der Behandlungen erwähnt wird z. B. Placebos mit gleichem Aussehen oder eine Schein-Therapie. Als nächstes sollte der Methoden-Abschnitt beschreiben, wie das Ergebnis beurteilt wurde und ob der Gutachter über die Behandlung der Patienten wusste.
Dieses Paper: Ja <input type="radio"/> Nein <input type="radio"/> Unklar <input type="radio"/> Kommentar:	

*) Die beiden Arbeitsblätter des Centre for Evidence-based Medicine der Universität Oxford „Critical Appraisal for Therapy Articles Worksheet“ und „Systematic Review Appraisal Worksheet“ können gemeinfrei heruntergeladen werden. Die deutsche Version wurde von Johannes Pohl aus dem Englischen übersetzt.

Schritt 2: Was waren die Ergebnisse?

1. Wie groß war der Effekt der Behandlung?

Meistens werden Ergebnisse als dichotome Ergebnisse präsentiert (Yes- oder No-Ergebnisse, die auftreten oder nicht auftreten) – einschließlich Ergebnissen, wie das Wiederauftreten von Krebs, Herzinfarkt oder Tod. Betrachten wir eine Studie, in der 15% (0,15) der Kontrollgruppe starben und 10% (0,10) der Behandlungsgruppe nach 2 Jahren der Behandlung starben. Die Ergebnisse können, wie unten gezeigt, in vielerlei Hinsicht zum Ausdruck gebracht werden.

Was wird gemessen?

Relatives Risiko (RR) = Risiko des Ergebnisses in der Behandlungsgruppe / Risiko des Ergebnisses in der Kontrollgruppe.

In unserem Beispiel: $RR = 0,10 / 0,15 = 0,67$

Absolute Risk Reduction (ARR) = risk of the outcome in the control group – risk of the outcome in the treatment group. This is also known as the **absolute risk difference**.

Absolute Risikoreduktion (ARR) = Risiko von den Ergebnissen in der Kontrollgruppe – Risiko des Ergebnisses in der behandelten Gruppe. Dies ist auch als die **absolute Risiko-Differenz** bekannt.

In unserem Beispiel: $ARR = 0,15 - 0,10 = 0,05$ or 5%

Relative Risikoreduktion (RRR) = absolute Risikoreduktion / Risiko des Ergebnisses in der Kontrollgruppe. Eine Alternative RRR zu berechnen, ist RR von 1 zu subtrahieren (z. B. $RRR = 1 - RR$).

In unserem Beispiel: $RRR = 0,05 / 0,15 = 0,33$ or 33%

Oder $RRR = 1 - 0,67 = 0,33$ or 33%

Number Needed to Treat (NNT) = Kehrwert der ARR wird wie folgt berechnet: $1 / ARR$.

In unserem Beispiel: $NNT = 1 / 0,05 = 20$

Was bedeutet dies?

Das relative Risiko zeigt uns, **wie viele Male wahrscheinlicher** es ist, dass ein Ereignis in der Behandlungsgruppe gegenüber der Kontrollgruppe auftritt. Ein **RR** von **1** bedeutet, dass es keinen Unterschied zwischen den beiden Gruppen gibt, die Behandlung also **keinen Effekt** hatte. Ein $RR < 1$ bedeutet, dass das Risiko das Ergebnis zu erhalten, durch die Behandlung verringert wird. Ein $RR > 1$ bedeutet, dass die Behandlung das Risiko für das Ergebnis erhöht.

Ist $RR < 1$, verringert die Behandlung das Risiko zu sterben.

Die absolute Risikoreduktion zeigt uns den absoluten Unterschied der Wahrscheinlichkeit von Ereignissen zwischen den beiden Gruppen und gibt einen Hinweis auf das grundlegende Risiko und den Behandlungseffekt. Eine **ARR** von **0** bedeutet, dass es keinen Unterschied zwischen den beiden Gruppen gibt, sodass die Behandlung keinen Effekt hatte.

Der absolute Nutzen der Behandlung entspricht einer 5%-igen Reduktion der Sterberate.

Die relative Risikoreduktion ist die Ergänzung des RR und wahrscheinlich das am häufigsten bekannte Maß für den Behandlungseffekt. Sie zeigt uns die Reduktion des Ergebniswertes der behandelten Gruppe gegenüber der Kontrollgruppe. Die Behandlung reduziert das Risiko des Todes um 33% bezogen auf die Kontrollgruppe.

Die number needed to treat stellt die erforderliche Anzahl von Patienten, die mit der experimentellen Therapie behandelt werden, dar, um ein schlechtes Ergebnis zu verhindern. Sie umfasst die Dauer der Behandlung. Klinische Bedeutung kann in gewissem Maße erreicht werden, indem die NNTs bestimmt werden sowie durch Aufwiegen der NNTs gegen jegliche Schäden oder Nebenwirkungen (NNHs) der Therapie. Wir müssten 20 Personen für 2 Jahre behandeln, um einen Todesfall zu verhindern.

2. Wie präzise war der Behandlungseffekt schätzungsweise?

Die wahre Gefahr von den Ergebnissen der Population ist nicht bekannt, und das Beste, was wir tun können ist es, das wahre, auf der Stichprobe von Patienten beruhende Risiko der Studie, zu schätzen. Diese Schätzung wird **Punktschätzung** genannt. Wir können messen, wie nahe diese Schätzung vom wahren Wert entfernt ist, indem man die Konfidenzintervalle (CI) für jede Schätzung untersucht. Wenn das Konfidenzintervall ziemlich schmal ist, können wir sicher sein, dass unsere Punktschätzung einem präzisen Abbild der Population entspricht. Das Konfidenzintervall gibt uns auch Informationen über die statistische Signifikanz des Ergebnisses. Wenn der Wert, der **keiner Wirkung** entspricht, außerhalb des 95%-Konfidenzintervalls liegt, dann ist das Ergebnis auf dem Niveau 0,05 statistisch signifikant. Wenn das Konfidenzintervall den Wert **keine Wirkung** enthält, dann sind die Ergebnisse nicht statistisch signifikant.

Schritt 3: Werden die Resultate mir helfen, meine Patienten zu behandeln? (Externe Validität / Anwendbarkeit)

Die Fragen, die Sie stellen, bevor Sie die Ergebnisse der Studie an den Patienten weitergeben sollten, sind:

- Unterscheidet sich mein Patient von denen in der Studie, sodass die Resultate nicht zutreffend sein könnten?
- Ist die Behandlung unter meinen Voraussetzungen möglich?
- Wird der mögliche Nutzen die möglichen Schäden der Behandlung für meinen Patienten überwiegen?

Systematische Reviews

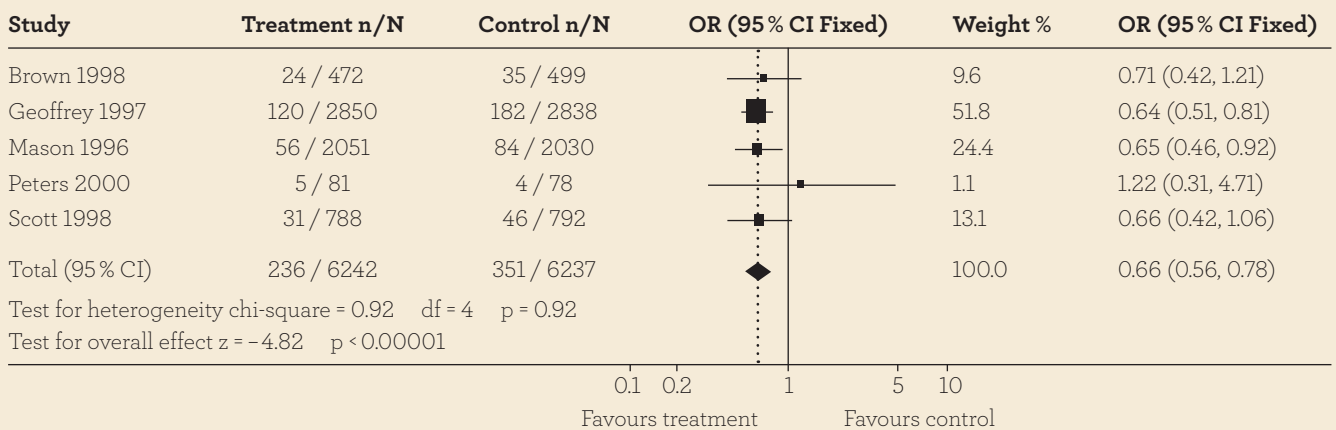
Schritt 1: Sind die Resultate des Reviews valide?

Auf welche Fragestellung bezieht sich das Systematic Review?	
Was ist am besten?	Wo finde ich die Informationen?
Die wichtigste Frage sollte anfänglich klargestellt werden. Die Exposition, wie eine Therapie oder ein Diagnose-Test, und das Ergebnis, welches von Interesse ist, werden oft bezüglich einfacher Zusammenhänge ausgedrückt.	Aus dem Titel , dem Abstract oder dem letzte Absatz der Einleitung sollte die Frage klar hervorgehen. Falls Sie nach dem Lesen dieser Abschnitte noch immer nicht feststellen können, was die zentrale Frage ist, sollten Sie nach anderen Studien suchen!
Von diesem Paper erfüllt: Ja <input type="radio"/> Nein <input type="radio"/> Unklar <input type="radio"/> Kommentar:	
Ist es unwahrscheinlich, dass wichtige, relevante Studien weggelassen wurden?	
Was ist am besten?	Wo finde ich die Informationen?
Der Ausgangspunkt für eine umfassende Suche nach allen relevanten Studien besteht in den größten bibliographischen Datenbanken (z. B. Medline, Cochrane, EMBASE etc). Es sollte aber auch eine Suche in Referenzlisten relevanter Studien sowie der Kontakt mit Experten einbezogen werden, vor allem um sich über unveröffentlichte Studien zu erkundigen.	Der Abschnitt Methoden sollte die Suchstrategie, einschließlich der detaillierten Suchbegriffe beschreiben. Der Abschnitt Ergebnisse wird die Anzahl der behandelten Titel und Abstracts, der abgerufenen Volltexte, die Anzahl der ausgeschlossenen Studien sowie den Grund des Ausschlusses erläutern. Diese Informationen können in einer Abbildung oder einem Flussdiagramm dargestellt werden.
Dieses Paper: Ja <input type="radio"/> Nein <input type="radio"/> Unklar <input type="radio"/> Kommentar:	
Waren die Einschlusskriterien angemessen?	
Was ist am besten?	Wo finde ich die Informationen?
Die Aufnahme oder der Ausschluss von Studien in einem systematischen Review sollten a priori klar definiert sein. Die verwendeten Einschlusskriterien sollten die Patienten, Eingriffe oder Exposition und Ergebnisse von Interesse spezifizieren. In vielen Fällen ist auch die Art des Studiendesigns ein wesentlicher Bestandteil der Auswahlkriterien.	Der Abschnitt Methoden sollte die Ein- und Ausschlusskriterien im Detail beschreiben. Normalerweise wird dieser auch das Studiendesign beinhalten.
Dieses Paper: Ja <input type="radio"/> Nein <input type="radio"/> Unklar <input type="radio"/> Kommentar:	
Waren die eingeschlossenen Studien genügend valide, um die Fragestellung zu beantworten?	
Was ist am besten?	Wo finde ich die Informationen?
Der Artikel sollte beschreiben, wie die Qualität der einzelnen Studien, anhand vorgegebener Qualitätskriterien sowie an die Art der klinischen Fragestellung angepasst, beurteilt wurde (z. B. Randomisierung, Verblindung und Vollständigkeit der Follow-up).	Der Abschnitt Methoden sollte die Beurteilung der Qualität und die angewandten Kriterien beschreiben. Der Abschnitt Ergebnisse sollte Informationen über die Qualität der einzelnen Studien bieten.
Dieses Paper: Ja <input type="radio"/> Nein <input type="radio"/> Unklar <input type="radio"/> Kommentar:	
Waren die Ergebnisse zwischen den Studien ähnlich?	
Was ist am besten?	Wo finde ich die Informationen?
Idealerweise sollten die Ergebnisse der verschiedenen Studien ähnlich oder homogen sein. Wenn Heterogenität besteht, können die Autoren einschätzen, ob die Unterschiede signifikant sind (Chi-Quadrat-Test). Mögliche Gründe für die Heterogenität sollten untersucht werden.	Der Abschnitt Ergebnisse sollte angeben, ob die Ergebnisse heterogen sind und mögliche Gründe dafür diskutieren. Das „Forest Plot“ sollte die Ergebnisse des Chi-Quadrat-Tests für Heterogenität zeigen und, falls vorhanden, die Gründe diskutieren, die für Heterogenität sprechen.
Dieses Paper: Ja <input type="radio"/> Nein <input type="radio"/> Unklar <input type="radio"/> Kommentar:	

Schritt 2: Was waren die Ergebnisse?

Wie werden die Resultate präsentiert?
Ein systematischer Review enthält eine Zusammenfassung der Daten aus den Ergebnissen einer Reihe von einzelnen Studien. Wenn die Ergebnisse der einzelnen Studien ähnlich sind, wird eine statistische Methode (Meta-Analyse genannt) verwendet, um die Ergebnisse der einzelnen Studien zu kombinieren. Außerdem wird ein Gesamtüberblick über die Schätzung berechnet. Die Meta-Analyse liefert die gewichteten Werte zu jeder der einzelnen Studien, die sich nach ihrer Größe richten. Die einzelnen Ergebnisse der Studien, wie relatives Risiko, Chancenverhältnis oder die mittlere Differenz zwischen den Gruppen, werden auf standardisierte Weise zum Ausdruck gebracht. Ergebnisse werden für gewöhnlich in einer Darstellung (Forest Plot) wiedergegeben.

Comparison: O3 Treatment versus Placebo
Outcome: O1 Effect of treatment on mortality



Der oben dargestellte Forest Plot ist eine Meta-Analyse von 5 Studien, welche die Auswirkungen einer hypothetischen Behandlung auf die Mortalität untersuchen. Die einzelnen Studien werden durch ein schwarzes Quadrat und eine horizontale Linie dargestellt, was der Stichprobe sowie dem 95%-Konfidenzintervall des Chancenverhältnisses entspricht. Die Größe des schwarzen Quadrats spiegelt die Gewichtung der Studie in der Meta-Analyse wider. Die vertikale Linie entspricht einer Trennlinie von Wirkung (hier links) und keiner Wirkung (hier rechts) - also einem Chancenverhältnis von 1,0. Wenn das Konfidenzintervall den Wert 1 einschließt, weist dies darauf hin, dass das Ergebnis auf dem üblichen Niveau ($p > 0,05$) nicht signifikant ist.

Der Diamant an der Unterseite stellt das kombinierte und gepoolte Chancenverhältnis der 5 Studien mit ihrem 95%-Konfidenzintervall dar. In diesem Fall zeigt er an, dass die Behandlung die Sterblichkeit um 34% (bzw. 0,66 95% CI 0,56 bis 0,78) reduziert. Beachten Sie, dass der Diamant und die „kein Effekt“-Linie sich nicht überlappen (der Wert 1 ist im Konfidenzintervall nicht enthalten). Also können wir davon ausgehen, dass der gepoolte OR statistisch signifikant ist. Der Test für den Gesamteffekt zeigt ebenfalls die statistische Signifikanz ($p < 0,0001$).

Untersuchung auf Heterogenität

Heterogenität kann anhand des „Eyeball-Tests“ oder formeller mithilfe statistischer Tests, wie dem Cochran Q-Test, ermittelt werden. Mit dem „Eyeball-Test“ sucht man nach Überlappung der Konfidenzintervalle der Versuche mit der zusammengefassten Schätzung. Beachten Sie im obigen Beispiel, dass die senkrechte gepunktete Linie, welche durch die Kombination der Chancenverhältnisse (OR) der einzelnen Studien entsteht, sich mit den Konfidenzintervallen (durchgezogene horizontale Linien) der einzelnen Studien kreuzt. Dies weist darauf hin, dass die Studien homogen sind. Heterogenität kann auch anhand des Cochran-Chi-Quadrat (Cochran Q) ermittelt werden. Wenn Cochran Q statistisch signifikant ist, besteht Heterogenität. Ist der Cochran Q nicht statistisch signifikant, aber das Verhältnis von Cochran Q und den Freiheitsgraden (Q / df) > 1 , so besteht möglicherweise Heterogenität. Wenn Cochran Q statistisch nicht signifikant ist und $Q / df < 1$, dann ist Heterogenität sehr unwahrscheinlich. Im obigen Beispiel ist $Q / df < 1$ ($0,92 / 4 = 0,23$) und der p-Wert nicht signifikant (0,92), d. h. dass keine Heterogenität besteht.

Hinweis: Um Heterogenität erkennen zu können, wird das Signifikanzniveau für Cochran Q, aufgrund der geringen Teststärke oft auf 0,1 gesetzt.