

Irrungen und Wirrungen bei Metaanalysen

Andreas G. Helg

Der Artikel von Prof. Dr. med. Peter Jüni und Mitarbeitern [1] ist eine Zusammenfassung einer vor drei Jahren von den gleichen Autoren veröffentlichten Metaanalyse [2]: ein Referat der eigenen Arbeit ohne neue Daten und ohne neue Erkenntnisse. Leider gehen die Autoren in ihrem Artikel nicht auf die Irrungen und Wirrungen in ihrer eigenen Arbeit ein, auf die wir [3] und andere seinerzeit mit einem *Letter to the Editor* hingewiesen hatten. Die wichtigsten dieser Unzulänglichkeiten sind nachfolgend zusammengefasst.

Einfluss der Studiendauer nicht berücksichtigt

Reichenbach et al. [2] fanden für alle in ihre Metaanalyse eingeschlossenen Studien eine gepoolte Effektgrösse (*Effect Size*, ES) von $ES = -0,75$ (95% KI, $-0,99$ bis $-0,50$). Das entspricht einer grossen symptomatischen Wirkung. Sie stellten jedoch auch eine hohe Heterogenität unter den Studien fest. Deshalb suchten sie nach Korrelationen zwischen den methodologischen Merkmalen der Studien und den in diesen beobachteten Effektgrössen. Hierzu führten sie stratifizierte Regressionsanalysen durch. – N.B.: Eine Korrelation ist nur ein mathematischer Zusammenhang, jedoch kein Beweis für einen kausalen Zusammenhang. – Die Autoren identifizierten drei Merkmale, die mit den Effektgrössen korrelierten: die Geheimhaltung der Randomisierung (adäquat vs. unklar), die Art der Datenanalyse (*Intention-to-treat*-Analyse, ja vs. nein oder unklar), die Grösse der Studie (Anzahl Patienten >200 vs. ≤ 200). Sodann stellten sie sich auf den Standpunkt, dass die Metaanalyse auf jene Studien beschränkt werden kann, welche bezüglich der drei identifizierten Merkmale homogen sind. In der Folge schieden sie 17 der 20 anfänglich begutachteten Studien aus und berücksichtigten nur noch die verbliebenen drei. Das Merkmal *Studiendauer* wurde von Reichenbach et al. [2] jedoch nicht beachtet. Wieso? Zur Unterscheidung zwischen Kurz- und Langzeitstudien hatten die Autoren willkürlich einen Grenzwert von 6 Monaten gewählt. Für diesen Grenzwert ergab sich kein signifikanter Unterschied zwischen den Effektgrössen der Kurz- und jenen der Langzeitstudien ($p = 0,152$). Führt man die Regressionsanalyse ohne Stratifizierung in Kurz- und Langzeitstudien durch, so findet man, dass die Effektgrössen direkt mit der Studiendauer korrelieren ($p = 0,04$). Eine Korrelation zwischen den Effektgrössen und der Studiendauer findet man aber auch für eine stratifizierte Regressionsanalyse, wenn

man den Grenzwert zur Unterscheidung zwischen Kurz- und Langzeitstudien auf 12 Monate festlegt ($p = 0,04$). Dieser Grenzwert macht mehr Sinn, denn die resultierende Stratifizierung entspricht gleichzeitig der Unterscheidung zwischen kürzeren Studien mit symptomatischem primärem Endpunkt und längeren Studien mit radiologischem primärem Endpunkt.

Auswahl von 3 Studien steht auf wackligen Beinen

Die Studiendauer ist ein viertes Merkmal, das mit den Effektgrössen korreliert. Sie muss deshalb bei einer Beschränkung der Metaanalyse auf eine Auswahl von Studien berücksichtigt werden. – Das Vorliegen einer Korrelation zwischen Effektgrössen und Studiendauer ohne Stratifizierung sowie bei Stratifizierung mit einem sinnvollen Grenzwert hat mehr Gewicht als das Fehlen einer Korrelation bei Stratifizierung mit einem Grenzwert von 6 Monaten. – Zwei der drei ausgewählten Studien sind jedoch Langzeitstudien, die verbleibende Dritte ist eine Kurzzeitstudie. Das Subset von drei Studien, auf dem die Schlussfolgerungen der Autoren beruhen, ist somit gar nicht homogen – eine Beschränkung der Metaanalyse auf diese drei Studien folglich nicht zulässig.

Verlauf der Arthrose ignoriert

Der Unterschied zwischen Verum und Placebo hinsichtlich Schmerzabnahme resp. die entsprechende Effektgrösse verändert sich mit der Zeit. Das gilt für alle medikamentösen Arthrosetherapien. Jede Therapie weist ein ihr eigenes Zeitfenster maximaler Wirkung auf. So fanden Bjordal et al. [4], dass sich dieses Zeitfenster für orale NSAIDs bis Woche 6 und für topische NSAIDs oder intra-artikulär verabreichte Steroide bis Woche 4 erstreckt. Danach nimmt die Differenz zwischen Verum und Placebo markant ab. Betrachtet man nur die symptomatische Wirkung von Chondroitinsulfat, so beginnt dieses Zeitfenster einen Monat nach Therapiebeginn und erstreckt sich bis Ende Monat 9, mit einem maximalen Unterschied zwischen Verum und Placebo Ende Monat 6 [5]. Die Unterschiede zwischen Verum und Placebo werden danach immer kleiner, weil auch die Patienten unter Placebo mit der Zeit immer weniger Schmerzen haben (Spontanverlauf der Arthrose). Diesen, dem Kliniker bekannten, charakteristischen Verlauf der Arthrose haben die Autoren

ignoriert. Unabhängig vom primären Endpunkt der Studien haben sie die Effektgrösse immer bei Studienende berechnet – bei den Langzeitstudien mit dem primären Endpunkt *Veränderung der Gelenkspaltbreite* nach zwei Jahren. Das ist ungefähr so, wie wenn man bei einer Bioverfügbarkeitsstudie mit einem Wirkstoff, der eine Halbwertszeit von einem Tag hat, die Blutprobe erst nach einer Woche nimmt. – Die sieben aktualisierten Richtlinien der Europäischen Arzneimittelbehörde zur Durchführung klinischer Studien zur Arthrose erübrigen jegliche weitere Diskussion: «The design of trials should adequately take into account the time response and duration of the action of the drug on symptoms. Primary endpoints for efficacy (change of symptom intensity) should be evaluated at a time point appropriate to show the maximum effect over placebo» [6].

Entscheidende fehlende Informationen nicht eingeholt

Die Grundlage für die stratifizierten Regressionsanalysen bildete eine Tabelle, in der die methodologischen Merkmale der Studien aufgelistet sind (Tabelle 2 im Anhang zu [2]). Diese Tabelle ist das Kernstück der Metaanalyse. Sie umfasst $22 \times 10 = 220$ Felder. Total 91 der 220 Felder darin sind mit «unklar» bezeichnet. Mit «unklar» wird ein Feld umschrieben, wenn die entsprechende Angabe in der Publikation der zugrunde liegenden Studie fehlt oder nicht eindeutig ist. Solche Informationslücken lassen sich nur füllen, indem man bei den Autoren der zugrunde liegenden Studien nachfragt.

Reichenbach et al. [2] schrieben im Abstract ihrer Publikation: «The authors ...; checked reference lists; and contacted authors», unter Literatursuche: «In addition, we ... contacted trialists and content experts» und unter Datenerfassung: «If effect sizes could not be calculated, we contacted the authors for additional data.»

Mit Ausnahme der Autoren der drei zeitlich am weitesten zurückliegenden Studien haben wir die Studienleiter resp. korrespondierenden Autoren aller Studien angeschrieben und gefragt, ob sie von den Autoren der Metaanalyse kontaktiert worden seien. Von 12 Autoren, die für 15 der 22 Studien verantwortlich waren, haben wir eine schriftliche Antwort (*Affidavit*) erhalten. Nur einer dieser Autoren hatte eine Anfrage erhalten, die jedoch nur ein einziges Feld der Tabelle betraf.

Autoren von Metaanalysen sind zwar nicht verpflichtet, fehlende oder unklare Angaben nachzufragen. Ohne Nachfrage bei den Autoren von 14 Studien ist die Bezeichnung von 61 Feldern mit «unklar» jedoch weitgehend willkürlich. Am Beispiel der Regressionsanalyse zur «Geheimhaltung der Randomisierung» (Tabelle 1 in [2]) möchten wir illustrieren, was das bedeutet: Mit einem p-Wert von 0,05 fiel dieser Test knapp positiv aus. Für 13 der 20 zu diesem Merkmal mit «unklar» bezeichneten Felder wissen wir mit Sicherheit, dass die Autoren der Metaanalyse keine weiteren Abklärungen vorgenommen hatten. Ersetzte man ein einziges «unklar» durch ein «adäquat», z.B. jenes der Studie von *Mazières et al., 2006*, so wäre

$p > 0,05$. Damit würde das Hauptargument für eine Beschränkung der Metaanalyse auf drei Studien hinfallen. Ein pikantes Detail: Die Studie von *Mazières et al., 2006* war zum Zeitpunkt der Publikation der Metaanalyse erst in Form eines Abstract veröffentlicht. Wie konnten die Autoren der Metaanalyse also wissen, ob die «Geheimhaltung der Randomisierung» adäquat war oder nicht, ohne diesen Punkt mit dem Leiter der Studie geklärt zu haben, was nachweislich nicht der Fall war.

Fragen und Kommentare zu Schlüsselaussagen

In ihrer Publikation umschrieben die Autoren das Ziel ihrer Arbeit so [2]: «We performed a systematic review and meta-analysis of all available randomized, controlled trials to determine the effects of chondroitin on pain and joint space width and to explore whether reported beneficial effects could be explained by biases affecting individual trials or publication bias.» Wir fragen uns, ob die Autoren ihre Analysen unbefangen genug angingen.

Bei ihrer Erklärung der Trichtergrafik [1] suggerieren die Autoren, dass die Punkte (Studien) symmetrisch zur Kegelachse angeordnet sein sollten und dass die erkennbare Asymmetrie ein Indiz für schlechte Methodologie sei. Symmetrisch zur Kegelachse liegen die einzelnen Studien jedoch nur bei einem Nulleffekt (den Jüni et al. [1] unterstellen). Verfügt die untersuchte Therapie über eine Wirkung, so liegen die Punkte im Idealfall auf einer Geraden links der Kegelachse und parallel zu dieser.

Zu den stratifizierten Regressionsanalysen hielten Reichenbach et al. [2] fest: «These analyses should be viewed as hypothesis-generating. They are observational in nature and have the same disadvantages as do other observational studies (ref.). In addition, the multiplicity of analyses has increased the probability of identifying spurious associations.» Wir fragen uns, wie man als Wissenschaftler auf einer solchen Basis ein finales Verdikt über Chondroitinsulfat aussprechen kann. Der Ansicht, dass die Schlussfolgerungen von Reichenbach et al. nicht auf solidem Fundament stehen, sind auch Harley A. Goldberg und Andrew L. Avins, zwei unabhängige Autoren einer der zahlreichen *Letter to the Editor* [7].

Fazit

Die Metaanalyse kann nicht auf drei von 20 Studien beschränkt werden. Langzeitstudien mit radiologischem primärem Endpunkt hätten in dieser Metaanalyse zur symptomatischen Wirkung separat betrachtet oder ausgeschlossen werden müssen. Eine mögliche Alternative hierzu hätte in der konsequenten Berechnung der Effektgrössen für alle Studien nach 6 Monaten bestanden. Fehlende oder unklare Informationen hätten eingeholt oder klargestellt werden müssen, insbesondere für neuere Studien mit über 200 Patienten sowie für Studien, die bis zum Zeitpunkt der Veröffentlichung der Metaanalyse nur in Form von Abstracts vorlagen.

Die wichtigste Botschaft

Die Metaanalyse von Reichenbach et al. [2] hält einer sorgfältigen Prüfung nicht stand. Die Schlussfolgerungen, Empfehlungen und Forderungen der Autoren sind vermessen – sie stehen auf zu wenig solidem Fundament und decken sich nicht mit den klinischen Erfahrungen von tausenden von Ärztinnen und Ärzten in der Schweiz und weltweit.

Korrespondenz:

IBSA Institut Biochimique SA
 Andreas G. Helg, Dr. sc. nat. ETH
 Product Manager
 Via del Piano
 CH-6915 Pambio-Noranco
andreas.helg@ibsa.ch

Literatur

- 1 Jüni P, Trelle S, Nüesch E, et al. Irrungen und Wirrungen bei der Erforschung von Arthrosetherapien. *Schweiz Med Forum*. 2010;10(26–27):461–3.
- 2 Reichenbach S, Sterchi R, Scherer M, et al. Meta-analysis: chondroitin for osteoarthritis of the knee or hip. *Ann Intern Med*. 2007;146:580–90.
- 3 Helg AG, DeVathaire F. How solid are the results of the meta-analysis by Reichenbach et al. and its conclusions. *Ann Intern Med*. Rapid Responses. http://www.annals.org/content/146/8/580.abstract/reply#annintmed_el_44993.
- 4 Bjordal JM, Klovning A, Ljunggren AE, et al. Short-term efficacy of pharmacotherapeutic interventions in osteoarthritic knee pain: A meta-analysis of randomised placebo-controlled trials. *Eur J Pain*. 2007;11:125–38.
- 5 Kahan A, Uebelhart D, De Vathaire F, et al. Long-Term Effects of Chondroitins 4 and 6 Sulfate on Knee Osteoarthritis: The Study on Osteoarthritis Progression Prevention, a Two-Year, Randomized, Double-Blind, Placebo-Controlled Trial. *Arthritis Rheum*. 2009;60(2):524–33.
- 6 CPMP/EWP/784/97 Rev. 1 of January 20, 2010, Seite 10.
- 7 Goldberg HA, Avins AL. Interpretation of chondroitin meta-analysis. *Ann Intern Med*. Rapid Responses. http://www.annals.org/content/146/8/580.abstract/reply#annintmed_el_44993.