Irrungen und Wirrungen bei der Erforschung von Arthrosetherapien[1]

Peter Jüni[a, b], Sven Trelle[a, b], Eveline Nüesch[a, b], Anne W. Rutjes[a], Rebekka Sterchi[a], Stephan Reichenbach[a, c]
[a] Institut für Sozial- und Präventivmedizin, Universität Bern, [b] CTU Bern, Inselspital, [c] Klinik für Rheumatologie, Immunologie und Allergologie, Inselspital

Nationales Forschungsprogramm NFP 53 «Muskuloskelettale Gesundheit – chronische Schmerzen»

Hintergrund

Die Arthrose ist die weltweit häufigste Gelenkserkrankung. 50% der 65-jährigen und 80% der 75-jährigen Westeuropäer weisen radiologische Zeichen arthrotischer Veränderungen auf. Die damit verbundenen sozioökonomischen Probleme sind enorm. Laut Berechnungen gingen etwa in den USA der 1980er Jahre jährlich rund 68 Millionen Arbeitstage durch arthrosebedingte Arbeitsunfähigkeit verloren. Lange wurde die Arthrose in wissenschaftlichen Untersuchungen und therapeutischen Bemühungen vernachlässigt. In den letzten 15 Jahren hat die klinische Arthroseforschung nun kräftig aufgeholt. Sie ist jedoch alles andere als vor Irrtümern gefeit.

Zielsetzung

Ziel des vorliegenden Projekts war, die Wirksamkeit und Sicherheit wichtiger medikamentöser Interventionen zur Behandlung von Arthroseschmerzen in einer Serie systematischer Übersichtsarbeiten zu untersuchen. Eine systematische Übersichtsarbeit kann – muss aber nicht – eine Metaanalyse beinhalten, also eine formale statistische Kombination der Resultate der eingeschlossenen Studien. Zur statistischen Kombination verwendeten wir einerseits herkömmliche Methoden, zum Beispiel bei der Untersuchung der Sicherheit von Rofecoxib [1], der Wirksamkeit von Chondroitin [2] und beim direkten Vergleich verschiedener Hyaluronsäuren [3]. Andererseits konnten neue statistische Methoden – Netzwerk-Metaanalysen – implementiert und weiterentwickelt werden. Sie machten es möglich, in einer singulären, integrierten Analyse die kardiovaskuläre und gastrointestinale Sicherheit von sieben nicht steroidalen Antirheumatika im Vergleich zu Plazebo zu untersuchen oder die Wirksamkeit von Glucosamin, Chondroitin und deren Kombination zu analysieren. Aus diesem komplexen Teil des Projekts wurden bisher nur die Resultate einer Pilotstudie in einem anderen Indikationsgebiet publiziert [4].

In der klinischen Arthroseforschung schienen Quellen von systematischen Fehlern (engl. *Bias*) dermassen ausgeprägt, dass wir entschieden, sie in diesem NFP-53-Projekt systematisch zu untersuchen [5]. Hier illustrieren wir am Beispiel unserer systematischen Übersichtsarbeit zu Chondroitin [3] einige typische Quellen von Bias.

Methode

Wir durchsuchten das Cochrane Central Register of Controlled Trials, MEDLINE, EMBASE, CINAHL, Kongressbeiträge und Referenzlisten nach randomisierten kontrollierten Studien bei Patienten mit Knie- oder Hüftarthrose. Die Qualität der Studien wurde anhand von drei etablierten Kriterien beurteilt [5]: Geheimhaltung der Zuteilung, Verblindung der Patienten und Analyse.

Die Kenntnis der zukünftigen Gruppenzuteilung kann die Studienärzte beim Entscheid beeinflussen, einen Patienten in die Studie aufzunehmen. Deshalb beurteilten wir, ob die Randomisierung der Patienten zum Zeitpunkt der Patientenrekrutierung geheimgehalten wurde (engl. *Concealment of allocation*). In unserer systematischen Übersichtsarbeit schlossen wir verblindete und offene Studien in die Analyse ein, beurteilten die Verblindung der Patienten in einer Studie jedoch nur dann als adäquat, falls in der Kontrollgruppe ein Plazebo verabreicht wurde und Chondroitinpräparat und Plazebo nicht zu unterscheiden waren [2]. Mitunter werden Patienten nach der Randomisierung von einer Studie ausgeschlossen, weil sie eine Verschlechterung ihrer Grundkrankheit erfahren oder gravierende Nebenwirkungen durch die Behandlung erlitten haben. Studienabbrecher sollten aber nicht nur akribisch beschrieben, sondern auch nach dem «Intention-to-Treat-Prinzip» analysiert werden. Das bedeutet: Alle Patienten, die in die Studie aufgenommen wurden, gilt es in der Analyse zu berücksichtigen – unabhängig davon, ob sie die Studie beendeten oder nicht [2, 5]. Wir beurteilten eine Analyse dann als adäquat, wenn wirklich alle randomisierten Patienten analysiert wurden [2].
Der Behandlungseffekt wurde in allen Studien standardisiert als Effektgrösse ausgedrückt, das ist die Differenz der mittleren Schmerzintensität zwischen Chondroitin- und Kontrollgruppe dividiert durch die entsprechende Standardabweichung [2, 3, 5]. Eine Effektgrösse von –0,20 Standardabweichungseinheiten entspricht einer kleinen, klinisch irrelevanten Schmerzreduktion von Chondroitin im Vergleich zu Plazebo, eine Effektgrösse von –0,50 einer mittleren und eine Effektgrösse von –0,80 einer grossen Schmerzreduktion. Zusammengefasst wurden die Studien mit Hilfe eines «Random-Effects-Modells».

1 Schmerzbehandlung bei Arthrose: Netzwerk-Meta-Analyse. Projektnummer: 405340-104762

Um die Heterogenität zwischen den Studien zu messen, bedienten wir uns der I^2-Statistik, die von 0% bis 100% reicht. Ein I^2 von 0 bis 25% bedeutet, dass zwischen den analysierten Studien keine Variabilität besteht, die über eine zufällige Streuung hinausgeht. Bei einem I^2 von 75 bis 100% ist die Variabilität sicher nicht mehr durch eine zufällige Streuung zu erklären, und es muss nach Gründen für die Variabilität gesucht werden.
Stratifizierte Analysen dienten dazu, die Quellen der Variabilität zu ermitteln. Dabei werden die Studien anhand bestimmter Merkmale gruppiert und dann analysiert [2, 5]. Kriterien für die Stratifizierung waren: Geheimhaltung der Zuteilung (adäquat vs. unklar), Verblindung der Patienten (adäquat vs. fehlend oder unklar) und Analyse (adäquat nach Intention-to-Treat-Prinzip vs. inadäquat mit Ausschluss von Patienten). Kleine, statistisch unpräzise Studien werden häufiger nicht publiziert als grosse Studien, falls sie nichtsignifikante Resultate zeigen. Sie sind häufiger mit methodischen Problemen behaftet und werden seltener nach dem Intention-to-Treat-Prinzip analysiert [5]. Deshalb stratifizierten wir unsere Analyse auch nach Studiengrösse. Wir unterschieden zwischen Studien mit mehr als 200 randomisierten Patienten – mit genügend statistischer Präzision, um eine mittlere Effektgrösse von –0,40 zu sichern – und Studien mit weniger als 200 Patienten.
Zusätzlich zeichneten wir eine Trichtergrafik (engl. *Funnel Plot*), in welcher die Effektgrössen der Studien auf der X-Achse gegen deren Standardfehler als Masszahl der statistischen Präzision auf der Y-Achse abgetragen wurde. Spielen systematische Fehler keine Rolle, hat die resultierende Grafik die Form eines auf dem Kopf stehenden Trichters. Falls das Zusammenspiel systematischer Fehler in kleinen Studien zu einer Überschätzung des Behandlungseffektes führt, wird die Grafik asymmetrisch.

Resultate

Insgesamt schlossen wir 20 Studien mit 3846 Patienten in unsere systematische Übersichtsarbeit ein (Tab. 1). Die Metaanalyse aller Studien ergab eine grosse, klinisch klar relevante Effektgrösse von –0,75 Standardabweichungen (95%-Vertrauensintervall –0,99 bis –0,50). Die Studien waren mit einer I^2-Statistik von 92% jedoch extrem heterogen. Ältere, vor 1995 publizierte Studien kamen auf einen Effekt, der grösser war als der geschätzte Effekt einer Knie- oder Hüftgelenkstotalprothese (sic!). 2005 oder später publizierte Studien hingegen zeigten einen Nulleffekt. Je länger man nach einer Wirkung von Chondroitin suchte, desto weniger war sie zu finden – weshalb?

Die Resultate der stratifizierten Analyse liefern die wahrscheinlichste Erklärung. Studien mit adäquater Geheimhaltung der Randomisierung, Studien, die nach dem Intention-to-Treat-Prinzip analysiert wurden sowie adäquat gepowerte, grosse Studien mit mehr als 200 Patienten weisen klinisch irrelevante, kleine Behandlungseffekte oder klare Nulleffekte nach. Da die Studien in den letzten zehn Jahren immer grösser und besser wurden, ist der beobachtete zeitliche Trend wenig überraschend.
In der Trichtergrafik markiert die graue Schattierung den Bereich, in welchem die Resultate mit einem zweiseitigen P-Wert von ≤0,05 signifikant ausfallen – links

Tabelle 1. Resultate der Metaanalyse aller 20 Studien und der stratifizierten Analysen.

	Anzahl Studien	Anzahl randomisierter Patienten	Behandlungseffekt (95% CI)	P für Interaktion
Alle Studien	20	3846	–0,75 (–0,99 bis –0,50)	–
Geheimhaltung der Randomisierung				0,05
Adäquat	2	1253	–0,01 (–0,12 bis 0,10)	
Unklar	18	2593	–0,84 (–1,08 bis –0,59)	
Plazebokontrolliert				0,64
Ja	17	3091	–0,78 (–1,06 bis –0,50)	
Nein	3	755	–0,62 (–0,94 bis –0,30)	
Patientenverblindung				0,45
Adäquat	14	2260	–0,83 (–1,16 bis –0,50)	
Fehlend oder unklar	6	1586	–0,59 (–0,99 bis –0,19)	
Intention-to-Treat-Analyse				0,02
Ja	3	1553	–0,03 (–0,13 bis 0,07)	
Nein oder unklar	17	2293	–0,88 (–1,13 bis –0,64)	
Anzahl randomisierter Patienten				0,02
>200	5	2419	–0,26 (–0,56 bis 0,04)	
≤200	15	1427	–0,93 (–1,22 bis –0,65)	

Falls der P-Wert für Interaktion ≤0,05 ausfällt, gilt der gefundene Unterschied zwischen zwei analysierten Straten als statistisch gesichert. Dies ist der Fall für Geheimhaltung der Randomisierung, Intention-to-Treat-Analyse und Anzahl randomisierter Patienten.

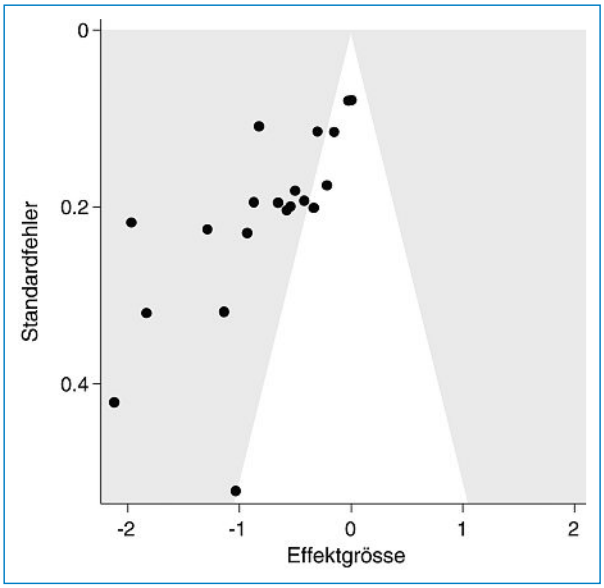

Abbildung 1
Trichtergrafik, in welcher die Effektgrössen der Studien auf der X-Achse gegen deren Standardfehler als Masszahl der statistischen Präzision auf der Y-Achse abgetragen wurden. Spielen systematische Fehler keine Rolle, hat die resultierende Grafik die Form eines auf dem Kopf stehenden Trichters. Falls das Zusammenspiel systematischer Fehler in kleinen Studien kumulativ zu einer Überschätzung des Behandlungseffektes führt, wird die Grafik asymmetrisch. Die grau schattierten Bereiche markieren den Bereich, in welchem die Resultate mit einem zweiseitigen P-Wert von ≤0,05 signifikant ausfallen, links zugunsten von Chondroitin, rechts zugunsten von Plazebo.

zugunsten von Chondroitin, rechts zugunsten von Plazebo (Abb. 1). Es fällt auf, dass sich die Resultate sehr kleiner Studien mit Standardfehlern von mehr als 0,2 alle im grauen Bereich links befinden, d.h. signifikant zugunsten von Chondroitin sind. In der weissen, nicht signifikanten Mitte und im grauen Bereich rechts herrscht unten in der Trichtergrafik vollkommene Leere. Diese Asymmetrie ist wahrscheinlich bedingt durch das Zusammenspiel von schlechter Studienmethodologie, selektiven Ausschlüssen von Patienten aus der Analyse und Publikationsbias in kleinen, unprofessionell durchgeführten Studien.

Praktische Schlussfolgerung

Die sorgfältige Analyse einer systematischen Übersichtsarbeit zu Chondroitin bei Knie- und Hüftgelenksarthrose zeigt, dass Chondroitinpräparate bezüglich Schmerzhemmung wahrscheinlich nicht besser wirksam sind als Plazebo. Die Anwendung von Chondroitin in der täglichen Praxis wird daher nicht empfohlen.

Ausblick

Behandlungen mit Chondroitin sollten künftig nur im Rahmen randomisierter Studien erfolgen. Die Herstellerfirmen sind aufgefordert, industrieunabhängigen Forschern Einsicht in die individuellen Patientendaten der bisherigen Studien zu gewähren, um Patientengruppen zu identifizieren, bei denen Chondroitin eventuell wirksam ist – zum Beispiel diejenigen mit leichtgradiger Arthrose.

Die wichtigste Botschaft

Das Zusammenspiel systematischer Fehler kann zu einer massgeblichen Überschätzung von Behandlungseffekten führen. Die kritische Analyse randomisierter Studien in systematischen Übersichtsarbeiten ist deswegen unerlässlich.

Korrespondenz:
Prof. Peter Jüni
Abteilungsleiter
Institut für Sozial- und Präventivmedizin
Universität Bern
Finkenhubelweg 11
CH-3012 Bern
juni@ispm.unibe.ch

Literatur

1 Jüni P, Nartey L, Reichenbach S, Sterchi R, Dieppe PA, Egger M. Risk of cardiovascular events and rofecoxib: cumulative meta-analysis. Lancet. 2004;364:2021–9.
2 Reichenbach S, Sterchi R, Scherer M, Trelle S, Bürgi E, Bürgi U, et al. Meta-analysis: chondroitin for osteoarthritis of the knee or hip. Ann Intern Med. 2007;146:580–90.
3 Reichenbach S, Blank S, Rutjes WS, Shang A, King EA, Dieppe PA, et al. Hylan versus hyaluronic acid for osteoarthritis of the knee: a systematic review and meta-analysis. Arthritis Care Res. 2007;57:1410–8.
4 Stettler C, Wandel S, Allemann S, Kastrati A, Morice MC, Schömig A, et al. Outcomes associated with drug-eluting and bare-metal stents: A collaborative network meta-analysis. Lancet. 2007;370:937–48.
5 Nüesch E, Trelle S, Reichenbach S, Rutjes AWS, Bürgi E, Scherer M, et al. The effects of the exclusion of patients from the analysis in randomised controlled trials: meta-epidemiological study. BMJ. 2009;339:b3244.