

ÜBERSICHTSARBEIT

Ausschluss von Patienten aus der Analyse: Auswirkungen auf die Ergebnisse randomisierter kontrollierter Studien

Eine meta-epidemiologische Studie

Eveline Nüesch, Sven Trelle, Stephan Reichenbach, Anne W. S. Rutjes, Elizabeth Bürgi, Martin Scherer, Douglas G. Altman, Peter Jüni

Dieser Artikel ist die Kurzversion eines bei bmj.com publizierten Beitrags, den wir mit freundlicher Genehmigung des British Medical Journal abdrucken. Er erscheint zeitgleich in der Printversion des BMJ und des Deutschen Ärzteblattes.

Die Zitierweise lautet:
Nüesch E, Trelle S, Reichenbach S, Rutjes A, Bürgi E, Scherer M, Altman D, Jüni P: The effects of exclusions of patients from the analysis in randomised controlled trials. *BMJ* 2009; 339: b3244.

ZUSAMMENFASSUNG

Zielsetzung: Untersuchung, ob der Ausschluss von Patienten aus der statistischen Analyse in randomisierten Studien mit Fehlbewertungen der Wirksamkeit von Behandlungen sowie einer größeren Heterogenität zwischen verschiedenen Studien assoziiert ist.

Studiendesign: Meta-epidemiologische Studie auf der Basis einer Sammlung von Metaanalysen randomisierter Studien.

Datenquellen: 14 Metaanalysen, die insgesamt 167 Studien berücksichtigten. Diese verglichen die Wirksamkeit therapeutischer Interventionen bei Arthrose des Knie- oder Hüftgelenks mit Kontrollgruppen, die entweder keine Intervention oder Placebo erhielten und verwendeten jeweils Schmerz nach Angaben der Patienten als Endpunkt.

Methoden: Zur Berechnung der Effektgrößen wurden die Unterschiede in der durchschnittlichen Schmerzintensität zwischen den Gruppen am Ende durch die gepoolte Standardabweichung dividiert. Die statistische Zusammenfassung der Studien erfolgte durch eine Random-Effects-Metaanalyse. Studien mit und Studien ohne Ausschluss von Patienten aus der statistischen Analyse wurden in Bezug auf die Bewertung der Therapiewirksamkeit gegeneinander verglichen, und die Auswirkungen einer Beschränkung von Metaanalysen auf Studien ohne Patientenausschluss wurden berechnet.

Ergebnisse: In 39 Studien (23 %) waren sämtliche Patienten in die Analyse eingeschlossen. In 128 Studien (77 %) wurden Patienten von der Analyse ausgeschlossen. Die Effektgrößen waren in Studien mit Patientenausschluss ten-

denziell vorteilhafter als in Studien ohne Patientenausschluss (Differenz $-0,13$; 95-%-Konfidenzintervall $-0,29$ bis $0,04$). Allerdings variierten die Schätzungen der Verzerrung zwischen einzelnen Metaanalysen erheblich ($\tau^2 = 0,07$). Untersuchungen der Interaktion zwischen Ausschluss aus der Analyse und Bewertung der Therapiewirksamkeit waren in fünf Metaanalysen positiv. Stratifizierte Analysen zeigten, dass Unterschiede in Bezug auf Effektgrößen zwischen Studien mit versus Studien ohne Patientenausschluss stärker ausfielen in Metaanalysen mit großer Heterogenität zwischen den einzelnen Studien, in Metaanalysen mit großer geschätzter Therapiewirksamkeit und in Metaanalysen aus dem Bereich der Komplementärmedizin. Beschränkten sich die Metaanalysen auf Studien ohne Patientenausschluss, resultierte dies in geringerer geschätzter Therapiewirksamkeit, größeren p-Werten und einer beträchtlichen Minderung der Heterogenität zwischen den einzelnen Studien.

Schlussfolgerungen: Der Ausschluss von Patienten aus der statistischen Analyse in randomisierten Studien führt häufig zu Fehleinschätzungen der Therapiewirkungen. Ausmaß und Richtung dieser Verzerrung sind jedoch unvorhersehbar. Berichte randomisierter Studien sollten grundsätzlich Ergebnisse von Intention-to-treat-Analysen nennen. Systematische Übersichtsarbeiten sollten den Einfluss des Patientenausschlusses von der statistischen Analyse auf die Bewertung der Therapiewirksamkeit routinemäßig prüfen.

Schlüsselwörter: Klinische Forschung, Statistik, Datenanalyse, Verzerrung, Therapieerfolg

Abweichungen vom Studienprotokoll sowie Studienabbrüche („losses to follow-up“) führen in klinischen Studien häufig dazu, dass Patienten von der statistischen Analyse ausgeschlossen werden (1, 2). Es ist unwahrscheinlich, dass nach der Randomisierung von einer Studie ausgeschlossene Patienten repräsentativ für diejenigen sind, die in der Studie verbleiben. Das selektive Auftreten und der verzerrte Umgang mit Abweichungen vom Studienprotokoll und Studienabbrü-

che können zu Ergebnissen führen, die von den tatsächlichen Werten abweichen. Diese systematische Verzerrung nennt man Selektionsbias nach Studieneintritt oder „attrition bias“ (2). Um sicherzustellen, dass Intervention und Kontrollgruppen vergleichbar sind, und um einen Selektionsbias nach Studieneintritt zu verhindern, sollte die Analyse nach dem Intention-to-treat-Prinzip erfolgen, das heißt alle randomisierten Patienten sollten in die Auswertung eingeschlossen werden,

Institut für Sozial- und Präventivmedizin, Universität Bern, Schweiz:
Dipl.-Stat. Dipl.-Biol. Nüesch,
Dr. med. Trelle,
Dr. med. Reichenbach,
Dr. phil. Rutjes,
Prof. Dr. med. Jüni

CTU Bern, Universitätsspital Bern:
Dipl.-Stat. Dipl.-Biol. Nüesch,
Dr. med. Trelle,
Prof. Dr. med. Jüni

Klinik für Rheumatologie, Immunologie und Allergologie, Universitätsspital Bern:
Dr. med. Reichenbach

Abteilung für Klinische Pharmakologie und Epidemiologie, Consorzio Mario Negri Sud, Santa Maria Imbaro, Chieti, Italien:
Dr. phil. Rutjes

Klinik für Innere Medizin, Universitätsspital Bern:
Dr. phil. Bürgi

Abteilung Allgemeinmedizin der Universität Göttingen:
Prof. Dr. med. Scherer

Institut für Sozialmedizin der Universität zu Lübeck:
Prof. Dr. med. Scherer

Centre for Statistics in Medicine, University of Oxford, U.K.:
Prof. Dr. Altman

und die randomisierte Gruppenzuteilung sollte beibehalten werden ungeachtet dessen, ob sich die Patienten an das Studienprotokoll halten oder nicht (2, 3).

Vier meta-epidemiologische Studien haben den Zusammenhang zwischen Studienabbrüchen vor und nach Beginn der Intervention sowie Patientenausschlüssen nach Randomisierung und der Bewertung der Therapiewirksamkeit untersucht (1, 4–7). In Abhängigkeit von den verwendeten Methoden und Definitionen sowie von den beteiligten klinischen Bereichen variierten Ausmaß und Richtung des Selektionsbias nach Studieneintritt zwischen den Studien (1, 6, 8, 9). Randomisierte kontrollierte Studien mit subjektiven Endpunkten sind generell anfälliger für Verzerrungen als Studien, die objektiv messbare Endpunkte verwenden wie zum Beispiel die Gesamtmortalität. In Studien zur Arthrose wird die Therapiewirksamkeit häufig mithilfe subjektiver Endpunkt-Parameter gemessen, wie beispielsweise der Schmerzintensität. Metaanalysen von Studien zur Arthrose sind daher möglicherweise besonders anfällig für einen Selektionsbias nach Studieneintritt (2).

Die Autoren haben untersucht, ob der Ausschluss von Patienten aus der statistischen Auswertung in Metaanalysen von Studien zu schmerztherapeutischen Interventionen bei Arthrose mit Fehlbewertungen von Therapiewirkungen und erhöhter Heterogenität zwischen Studien assoziiert ist.

Methoden

Die Autoren führten eine elektronische Suche nach Metaanalysen randomisierter Studien (solche mit nicht voraussagbaren Zuordnungssequenzen) oder quasi-randomisierter Studien (solche mit potenziell voraussagbaren Zuordnungssequenzen) mit Patienten mit Arthrose des Knie- oder Hüftgelenks (vergleiche www.bmj.com) durch. Voraussetzung für den Einschluss in die vorliegende Studie war, dass die Metaanalysen Schmerz nach Angaben der Patienten bewerteten und eine beliebige Intervention mit Kontrollgruppen verglichen, die Placebo, Sham-Behandlung oder keinerlei Intervention erhielten.

Die Studien wurden als „ohne Patientenausschluss von der Analyse“ klassifiziert, wenn sie eine ausdrückliche Erklärung enthielten, dass alle randomisierten Patienten in die Auswertung hinsichtlich des von den Autoren gewählten Endpunktes eingeschlossen worden waren oder wenn die angegebenen Fallzahlen zu randomisierten und analysierten Patienten in Bezug auf diesen Endpunkt identisch waren. Wenn Patientenausschlüsse explizit angegeben waren, wenn die Anzahl der in der Analyse eingeschlossenen Patienten geringer war als die der randomisierten Patienten oder wenn unklar war, ob Patienten ausgeschlossen worden waren, wurden die jeweiligen Studien als Studien mit Patientenausschluss klassifiziert.

Die verdeckte Zuordnung („concealment of allocation“) stufte man als ausreichend ein, wenn den für den Einschluss von Patienten verantwortlichen Mitarbeitern die nächste Therapiezuweisung unbekannt war. Die Verblindung von Patienten wurde als hinrei-

chend angesehen, wenn die experimentelle Intervention als von der Kontroll-Intervention ununterscheidbar beschrieben war oder bei Einsatz der Double-Dummy-Technik (2).

Der primäre Endpunkt war Schmerz nach Angabe der Patienten. Wenn verschiedene Outcomes mit dem Parameter Schmerz angegeben waren, wählten die Autoren einer Hierarchie folgend ein Outcome pro Studie aus (10, 11).

Statistische Analyse

Therapiewirkungen wurden als Effektgrößen ausgedrückt (Differenz der Mittelwerte am Ende des Follow-up dividiert durch die gepoolte Standardabweichung). Negative Effektgrößen bezeichnen eine günstige Wirkung der experimentellen Intervention. Nicht verfügbare Daten wurden approximiert (11). Die Zusammenfassung der Effektgrößen verschiedener Studien erfolgte mittels Random-Effects-Metaanalysen der Varianzschätzer τ^2 diente als Maß der Heterogenität (12).

Mithilfe der Random-Effects-Metaanalyse berechneten die Autoren die Effektgrößen für Studien mit und ohne Patientenausschluss aus der Analyse separat. Anschließend wurden die Differenzen zwischen den Therapiebewertungen in Studien mit und ohne Patientenausschluss für jede Metaanalyse berechnet und mittels Random-Effects-Metaanalyse kombiniert (13). Eine negative Differenz zwischen Effektgrößen zeigt an, dass Studien mit Patientenausschluss günstigere Therapiewirkungen aufweisen. Zur Ermittlung von Interaktionen zwischen dem Ausschluss von Patienten aus der statistischen Analyse und der Schätzungen von Therapiewirkungen wurden für jede Metaanalyse einzeln z-Tests durchgeführt. Die Autoren nahmen ein stratifiziertes Analyseverfahren sowie statistische Interaktionstests für die folgenden Merkmale vor:

- Heterogenität zwischen den Studien in der gesamten Metaanalyse (niedrig: $\tau^2 < 0,06$ versus hoch: $\tau^2 \geq 0,06$)
- Therapiewirkung in der gesamten Metaanalyse (gering: Effektgrößen $> -0,5$ versus groß: Effektgrößen $\leq -0,5$) (10, 14)
- Art der in der Metaanalyse evaluierten Intervention (medikamentöse versus andere Interventionen, konventionelle versus Komplementärmedizin).

Um für Confounding durch verdeckte Zuordnung oder Patientenverblindung zu kontrollieren, haben die Autoren zusätzlich für diese Faktoren stratifiziert, um die Unterschiede zwischen Studien mit und ohne Patientenausschluss zu berechnen.

Schließlich verglichen die Autoren die gepoolten Effektgrößen, die Heterogenität zwischen den Studien, die Präzision (Reziprok des Standardfehlers), sowie die p-Werte für die gepoolten Effektgrößen von Random-Effects-Metaanalysen von sämtlichen Studien mit auf Studien ohne Patientenausschluss beschränkten Metaanalysen. Die Resultate wurden mittels Scatter-Plots (Streuplots) und Wilcoxon-Rangsummen-Tests für gepaarte Beobachtungen verglichen. Die p-Werte waren zweiseitig.

Ergebnisse

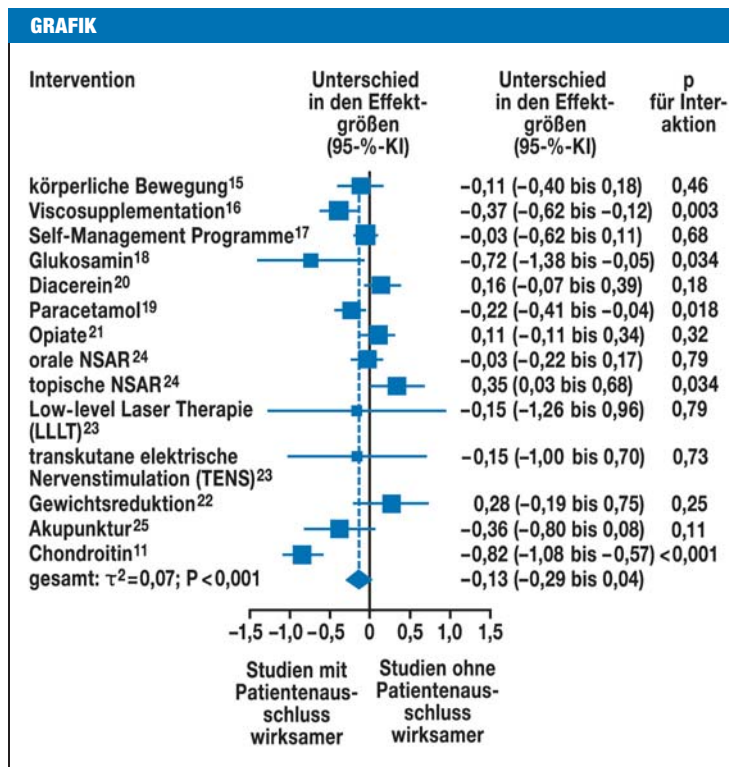
Insgesamt wurden 14 Metaanalysen (167 Studien mit 41 170 Patienten), die mindestens eine Studie mit und eine Studie ohne Patientenausschluss von der statistischen Analyse enthielten, in der vorliegenden Studie ausgewertet (11, 15–25). Acht davon untersuchten medikamentöse Interventionen, die restlichen fünf evaluierten komplementärmedizinische Verfahren. Die Anzahl der Studien pro Metaanalyse lag zwischen drei und 24 (Median 11), die Anzahl der Patienten pro Metaanalyse zwischen 278 und 13 659 (Median 1 731). Die aus den Random-Effects-Metaanalysen sämtlicher Studien gewonnenen gepoolten Effektgrößen lagen zwischen –0,07, was so gut wie keinem Therapieeffekt entspricht, und –0,88, was einem deutlichen Effekt entspricht. Die experimentelle Intervention zeigte sich in sämtlichen Metaanalysen überlegen, elf Metaanalysen fanden mit $p = 0,05$ signifikante Unterschiede zwischen der experimentellen Intervention und der Kontrollintervention. Die Varianz τ^2 schwankte zwischen 0,00 und 0,52 (Median 0,04).

Eigenschaften der Studien

39 der insgesamt 167 Studien (23 %) hatten sämtliche randomisierten Patienten in die statistische Analyse eingeschlossen. 114 Studien (69 %) hatten Patienten von der Analyse ausgeschlossen und für 14 Studien (8 %) war nicht feststellbar, ob Patientenausschlüsse stattgefunden hatten oder nicht. Der Umfang der Patientenausschlüsse bewegte sich zwischen 0,1 % und 40 % (Median 7,2 %). Berichte von Studien mit Patientenausschlüssen machten seltener Angaben zu Studienabbrüchen („losses to follow-up“) ($p = 0,002$). Datenimputationen durch Ersetzung der fehlenden Werte mit der letzten gemachten Beobachtung („last observation carried forward“) wurden von 27 % der Studien mit Patientenausschlüssen angegeben und von 49 % der Studien ohne Patientenausschlüsse; multiple Imputation wurde in 4 % beziehungsweise 15 % eingesetzt. In 68 % beziehungsweise 15 % der Studien war unklar, wie mit fehlenden Daten verfahren worden war. Studien mit Patientenausschlüssen wurden in der Regel früher publiziert als solche ohne Patientenausschluss (Mittelwert 1998 [Standardabweichung, SD 6] versus 2001 [SD 4]; $p = 0,002$) und wiesen seltener hinreichende verdeckte Zuordnung und Fallzahlkalkulation auf.

Auswirkung von Patientenausschluss auf die Bewertung von Therapiewirkungen

Im Durchschnitt wurden Therapiewirkungen in Studien mit Patientenausschlüssen vorteilhafter bewertet als in Studien ohne Patientenausschluss (Differenz zwischen den Effektgrößen –0,13; 95%-KI –0,29 bis 0,04; $p = 0,13$), allerdings variierte das Ausmaß der Verzerrung zwischen den Metaanalysen beträchtlich ($\tau^2 = 0,07$; $p < 0,001$) (Grafik). Unterschiede zwischen den Effektgrößen lagen zwischen –0,82 und 0,35. Interaktionstests zur Ermittlung des Zusammenhangs von Patientenausschlüssen aus der statistischen Analyse und der geschätzten Therapiewirkungen waren positiv ($p = 0,05$) für fünf Metaanalysen. Vier Metaanalysen ergaben vorteilhaftere Therapiewirkungen für Studien mit Patientenausschlüssen aus



Differenz zwischen den Effektgrößen in 128 Studien mit und 39 Studien ohne Patientenausschluss von der statistischen Analyse. Eine negative Differenz zwischen Effektgrößen zeigt an, dass Studien mit Patientenausschlüssen aus der Analyse günstigere Therapiewirkungen ergeben. Die p-Werte beziehen sich auf die Interaktion von Patientenausschluss und Effektgrößen. KI, Konfidenzintervall; NSAR, nicht-steroidale Antirheumatika

der Analyse, während eine Metaanalyse günstigere Therapiewirkungen für Studien ohne Patientenausschluss aufzeigte (Grafik).

In Metaanalysen mit einem hohen Maß an Heterogenität zwischen den einzelnen Studien zeigten sich deutliche Unterschiede zwischen Studien mit und solchen ohne Patientenausschluss – jedoch nicht in Metaanalysen mit geringer Heterogenität zwischen den Studien (p für Interaktion $< 0,001$). Ebenso waren die Unterschiede in Metaanalysen mit großer geschätzter Therapiewirkung in der gesamten Metaanalyse stärker ausgeprägt als in Metaanalysen mit geringer geschätzter Therapiewirkung sowie in Metaanalysen von komplementärmedizinischen Interventionen verglichen mit Metaanalysen von konventionellen medizinischen Interventionen (p für Interaktion $< 0,001$ für beide). Eine Stratifizierung nach diesen Merkmalen führte zu einem beträchtlichen Rückgang der Variabilität der Verzerrung. Wurde um die Faktoren verdeckte Zuordnung ($-0,11$; 95%-KI $-0,28$ bis $0,05$; $p = 0,18$) und Patientenverblindung ($-0,15$; 95%-KI $-0,30$ bis $0,00$; $p = 0,047$) bereinigt, blieben die durchschnittlichen Differenzen zwischen Studien mit und Studien ohne Patientenausschluss stabil. In beiden bereinigten Auswertungen war die Variabilität der Verzerrung zwischen den Metaanalysen mit Varianzeinschätzer τ^2 von $0,08$ beziehungsweise $0,06$ (beide $p < 0,001$) im Vergleich zur ursprünglichen Analyse praktisch unverändert.

Auswirkungen der Beschränkung von Metaanalysen auf Studien ohne Patientenausschluss

Wurden die Metaanalysen ausschließlich auf Studien ohne Patientenausschluss beschränkt, reduzierte sich die mediane Anzahl der in einer Metaanalyse eingeschlossenen Studien von elf auf zwei und die mediane Anzahl der Patienten von 1 731 auf 401. Bei der Schätzung der Therapiewirksamkeit nahmen die Werte in zehn Metaanalysen ab und in vier Metaanalysen zu ($p = 0,10$). Die Heterogenität zwischen den einzelnen Studien fiel in zwölf Metaanalysen ab und stieg in einer Metaanalyse an ($p = 0,006$). Eine Metaanalyse enthielt lediglich eine Studie, die keine Patienten von der statistischen Analyse ausgeschlossen hatte, sodass die Heterogenität zwischen Studien nach Beschränkung auf Studien ohne Patientenausschluss nicht berechnet werden konnte (25). Die Präzision der Schätzungen der gepoolten Effektgrößen nahm in neun Metaanalysen ab und in fünf Metaanalysen zu ($p = 0,25$). Die p-Werte wurden in zehn Metaanalysen größer, während sie in vier Metaanalysen kleiner wurden ($p = 0,016$). Nach der Beschränkung auf ausschließlich Studien ohne Patientenausschluss verloren sechs Metaanalysen statistische Signifikanz auf dem 5%-Niveau.

Diskussion

Die vorliegende meta-epidemiologische Studie mit 14 Metaanalysen von insgesamt 167 Studien ergab, dass der Ausschluss von randomisierten Patienten aus der statistischen Analyse häufig zu Fehlbewertungen von Therapiewirkungen führte. Die durchschnittlich gemessene Verzerrung mag mit einer Differenz zwischen Effektgrößen von 0,13 gering erscheinen (*Grafik*), entspricht jedoch einem Viertel bis der Hälfte des typischen Therapieeffekts für Interventionen bei Arthrose (10). Die Auswirkungen von Patientenausschlüssen auf die Bewertung des Therapieeffekts schienen am ausgeprägtesten in Metaanalysen mit großen Therapieeffekten, bei Evaluation von komplementärmedizinischen Interventionen sowie in Metaanalysen mit einem hohen Maß an Heterogenität zwischen den einzelnen Studien – allerdings waren Ausmaß und Richtung der Verzerrung für eine konkrete Situation nicht voraussagbar. Tests zur Ermittlung der Interaktion zwischen Patientenausschlüssen aus der statistischen Analyse und der Bewertung der Therapiewirkungen erreichten statistische Signifikanz für fünf Metaanalysen; in vier dieser Metaanalysen waren die geschätzten Therapieeffekte in Studien ohne Patientenausschluss kleiner.

Bei der Beschränkung der Metaanalysen auf Studien ohne Patientenausschluss nahmen die p-Werte in den meisten Fällen zu; sechs Metaanalysen verloren auf dem 5%-Niveau an Signifikanz. Diese Erhöhung der p-Werte war nicht allein auf einen Verlust an statistischer Güte (Fähigkeit einer Untersuchung, einen tatsächlich vorhandenen Unterschied auch zu finden) zurückzuführen (26). Infolge der Beschränkung nahm die Heterogenität zwischen den Studien τ^2 beträchtlich ab. Daher war der durchschnittliche Verlust an statistischer Präzision in den

Random-Effects-Metaanalysen geringer als nach dem Ausschluss von über der Hälfte der Studien zu erwarten gewesen wäre. Lediglich in fünf Metaanalysen war die statistische Präzision nach Einschränkung in relevantem Ausmaß verringert; in sechs Metaanalysen blieb die Präzision weitgehend gleich, während sie in drei Metaanalysen zunahm.

Stärken und Schwächen der Studie

Die Autoren verließen sich bei der Evaluation der Studienberichte nicht auf Aussagen, ob eine Intention-to-treat-Analyse durchgeführt worden war oder nicht. Stattdessen bedurfte es detaillierter Informationen zum Patientenfluss (27, 28) oder der ausdrücklichen Erklärung, dass sämtliche randomisierten Patienten in die statistische Analyse eingeschlossen waren. Manche mögen die gewählte Abgrenzung von Studien mit gegen solche ohne Patientenausschluss von der Analyse für übermäßig strikt halten. Die Autoren nehmen an, dass jedwede mit dem Ausschluss von Patienten aus der Analyse assoziierte Verzerrung mit wachsender Anzahl ausgeschlossener Patienten zunimmt. Wieder andere mögen argumentieren, die verwendete Klassifikation sei nicht strikt genug, sondern die Autoren hätten außerdem die ausdrückliche Bestätigung zur Bedingung machen sollen, dass es keine Gruppenwechsel gegeben hatte und dass alle randomisierten Patienten bei der Auswertung denselben Gruppen zugehörten, in die sie ursprünglich randomisiert worden waren. Nur sieben der 167 Studienberichte (4 %) enthielten diese Informationen.

Die vorliegende Studie basiert auf publizierten Daten und ist somit auf die Qualität der Studienberichte angewiesen. Abgesehen von 14 Studien war es den Autoren jedoch möglich, den Berichten zu entnehmen, ob Patienten von der Analyse ausgeschlossen worden waren oder nicht, und eine Verzerrung aufgrund falscher Klassifizierungen von Studien infolge unzulänglicher Angaben in dieser Hinsicht dürfte minimal sein. Mindestens zwei Drittel der in der vorliegenden Arbeit berücksichtigten Studien wiesen unvollständige Daten bezüglich der Endpunkte auf. Aufgrund der schlechten Qualität der Studienberichte war es den Autoren nicht möglich, zu untersuchen, ob die zur Datenimputation verwendete Methode die Schätzung der Therapieeffekte beeinflusst (28–30). In der vorliegenden Studie könnte der Faktor verdeckte Zuordnung („concealment of allocation“) als Störgröße (Confounder) auf den beobachteten Zusammenhang von Patientenausschluss aus der Analyse und der Bewertung der Therapieeffekte eingewirkt und damit Scheinzusammenhänge ergeben haben. Die Sensitivitätsanalyse für den Faktor „verdeckte Zuordnung“ ergab jedoch, dass die Ergebnisse der vorliegenden Studie robust sind.

Schlussfolgerungen

Die strengste Interpretation des Intention-to-treat-Prinzips verlangt den Einschluss sämtlicher Patienten in die Auswertung, ungeachtet dessen, ob diese teilnahmeberechtigt waren, behandelt wurden oder ob sie sich an das Studienprotokoll hielten (3). In vielen Studien werden

randomisierte Patienten aus der Analyse ausgeschlossen, die nicht mindestens eine Anwendung der zugeteilten Intervention erhalten hatten, während andere Studien Patienten ausschließen, die sich nach der Rekrutierung als nicht teilnahmeberechtigt erwiesen hatten (3, 31). Beide Verfahren können potenziell unverzerrte Ergebnisse hervorbringen, falls sowohl die Patienten als auch die behandelnden Ärzte in Bezug auf die zugeteilte Intervention verblindet sind und falls die Entscheidung, Patienten von der Analyse auszuschließen, ausschließlich auf Informationen beruht, die vor der Randomisierung gesammelt wurden und diese ohne Bezug zur Gruppenzuteilung oder zum klinischen Endpunkt ist (31). Des Weiteren ist der Ausschluss von Patienten aus der Analyse aufgrund von rein zufällig fehlenden Endpunkt-Daten möglicherweise weniger problematisch als der selektive Ausschluss von Patienten aufgrund von Verstößen gegen das Studienprotokoll. Diese Annahmen lassen sich jedoch kaum jemals nachprüfen: Details zum Fluss der Studienteilnehmer durch die Studie und Beschreibungen der Verfahren, mittels derer entschieden wird, ob Patienten von der Analyse ausgeschlossen werden sollten, werden häufig aus den publizierten Berichten randomisierter Studien ausgelassen (3, 28). Es ist daher schwierig, aufgrund der publizierten Daten zu entscheiden, ob die genannten Patientenausschlüsse aus der Analyse eine Verzerrung zur Folge hatten (2) und eine strenge Einhaltung des Intention-to-treat-Prinzips sollte empfohlen werden (27, 32).

Per-Protocol-Analysen schließen nur Patienten ein, welche die im Protokoll definierten Interventionen erhielten sowie Patienten, für die Daten zu den Endpunkten vorhanden waren. Die in Per-Protocol-Analysen ausgeschlossenen Patienten unterscheiden sich jedoch mit großer Wahrscheinlichkeit von den in die Analyse eingeschlossenen Patienten (33). Studien ohne Patientenausschluss ergänzten fehlende Daten häufiger durch Imputationen als Studien mit Patientenausschlüssen. Die am häufigsten angewandte Methode war die Ersetzung eines fehlenden Werts mit der letzten gemachten Beobachtung („last observation carried forward“). Diese Methode ist in der muskuloskeletalen Forschung beliebt (34, 35), führt jedoch zu allzu präzisen Ergebnissen und potenziellen Verzerrungen (36, 37). Die Methode der multiplen Daten-Imputation ist in der Anwendung komplexer, verhindert aber solche Probleme (38). Die CONSORT-Erklärung dringt auf eine transparente Darstellung des Teilnehmerflusses durch Studien, inklusive einer Beschreibung von Patientenausschlüssen und Studienabbrüchen sowie der Gründe für Patientenausschlüsse (27, 32). Die Autoren halten darüber hinaus eine Beschreibung der verwendeten Strategien im Umgang mit fehlenden Endpunkt-Daten für entscheidend.

Fazit

Um einen potenziellen Selektionsbias nach Studieneintritt zu vermeiden, sollte gewährleistet sein, dass die Studienabbruchrate möglichst niedrig und die Compliance-Rate hoch ist, und das Fehlen von Endpunkt-Daten auf ein Minimum reduziert wird. Grundsätzlich sollten Er-

gebnisse von Intention-to-treat-Analysen genannt werden. Sensitivitätsanalysen, die auf Patienten beschränkt sind, die sich an das Studienprotokoll halten, können zusätzlich beschrieben werden. Wann immer möglich sollten die für systematische Übersichten und Metaanalysen extrahierten Daten auf Ergebnissen von Analysen beruhen, die alle randomisierten Patienten einschließen. Die Auswirkungen von Patientenausschlüssen aus der statistischen Analyse auf die Berechnung der Therapieeffekte sollte routinemäßig in stratifizierten Analysen evaluiert werden.

KASTEN

Bisherige Erkenntnisse zum Thema

- Der Ausschluss randomisierter Patienten aus der statistischen Analyse randomisierter kontrollierter Studien kann zu einem Selektionsbias nach Studieneintritt (Attrition Bias) führen.
- Richtung und Ausmaß des Selektionsbias nach Studieneintritt in verschiedenen Studien variieren in Abhängigkeit von den verwendeten Methoden und Definitionen sowie von den involvierten klinischen Bereichen.

Neue Erkenntnisse der Studie

- Der Ausschluss von Patienten aus der statistischen Analyse randomisierter Studien führte häufig zu Fehleinschätzungen im Hinblick auf Therapiewirkungen. Allerdings blieben das Ausmaß und die Richtung der Verzerrung für konkrete Situationen unvorhersehbar.
- Überbewertungen der Therapiewirkung schienen besonders ausgeprägt in Metaanalysen mit hoher Heterogenität zwischen den einzelnen Studien, in Metaanalysen mit großen geschätzten Therapievorteilen sowie in Metaanalysen aus dem Bereich der Komplementärmedizin.
- Beschränkten sich Metaanalysen auf Studien ohne Patientenausschluss, so resultierte dies in geringer geschätzten Therapieeffekten, größeren p-Werten sowie einer beträchtlichen Verminderung der Heterogenität zwischen den Studien.

Die Autoren danken Sacha Blank, Liz King, Katharina Liewald, Linda Nartey, Rebekka Sterchi und Beatrice Tschannen für ihre Mithilfe bei der Datengewinnung, den Autoren, die Studiendaten zur Verfügung gestellt haben, Malcolm Sturdy für die Entwicklung und Pflege der Datenbank und Fitore Sallahaj für die Eingabe der Daten.

Mitarbeiter: siehe www.bmj.com.

Finanzierung

Der Schweizerische Nationalfonds (Peter Jüni und Stephan Reichenbach, Förderungs-Nummern 4053–40–104762/3 und Nr. 3200–066378) und die Schweizerische Gesellschaft für Innere Medizin (Peter Jüni). Die Studie war ein Teilprojekt des Nationalen Forschungsprogramms NFP 53 (Muskuloskeletale Gesundheit – Chronische Schmerzen) des Schweizerischen Nationalfonds. Stephan Reichenbach erhielt ein Forschungsstipendium des Schweizerischen Nationalfonds (Förderungs-Nr. PBEB-115067). Martin Scherer erhielt einen Forschungspreis für junge Forscher des Bundesministeriums für Bildung und Forschung (Förderungs-Nr. 01 GK 0516). Peter Jüni war Teilnehmer des PROSPER Programms (Programme for Social Medicine, Preventive and Epidemiological Research) des Schweizerischen Nationalfonds (Förderungs-Nr. 3233–066377). Die Geldgeber hatten keinen Einfluss auf Studiendesign, Datenerhebung, Datenanalyse, Inter-

pretation der Daten oder auf die Abfassung des Studienberichts. Der korrespondierende Autor hatte uneingeschränkten Zugang zu sämtlichen Studiendaten sowie die ausschließliche Verantwortung für die Entscheidung, das Manuskript zur Publikation einzureichen. Keiner der Autoren erhält finanzielle Zuwendungen von einem Hersteller von Arznei-, Heil- oder Hilfsmitteln zur Behandlung der Arthrose oder ist mit einem solchen affiliert.

Interessenkonflikt

Die Autoren erklären, dass kein Interessenkonflikt im Sinne der Richtlinien des International Committee of Medical Journal Editors besteht.

Ethik-Zulassung: nicht erforderlich

LITERATUR

1. Tierney JF, Stewart LA: Investigating patient exclusion bias in meta-analysis. *Int J Epidemiol* 2005; 34: 79–87.
2. Juni P, Altman DG, Egger M: Systematic reviews in health care: assessing the quality of controlled clinical trials. *BMJ* 2001; 323: 42–6.
3. Hollis S, Campbell F: What is meant by intention to treat analysis? Survey of published randomised controlled trials. *BMJ* 1999; 319: 670–4.
4. Schulz KF, Chalmers I, Hayes RJ, Altman DG: Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995; 273: 408–12.
5. Kjaergard LL, Villumsen J, Gluud C: Reported methodologic quality and discrepancies between large and small randomized trials in meta-analyses. *Ann Intern Med* 2001; 135: 982–9.
6. Balk EM, Bonis PA, Moskowitz H, Schmid CH, Ioannidis JP, Wang C, et al.: Correlation of quality measures with estimates of treatment effect in meta-analyses of randomized controlled trials. *JAMA* 2002; 287: 2973–82.
7. Gluud LL, Thortlund K, Gluud C, Woods L, Harris R, Sterne JA: Correction: reported methodologic quality and discrepancies between large and small randomized trials in meta-analyses. *Ann Intern Med* 2008; 149: 219.
8. Juni P, Egger M: Commentary: empirical evidence of attrition bias in clinical trials. *Int J Epidemiol* 2005; 34: 87–8.
9. Wood L, Egger M, Gluud LL, et al.: Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. *BMJ* 2008; 336: 601–5.
10. Juni P, Reichenbach S, Dieppe P: Osteoarthritis: rational approach to treating the individual. *Best Pract Res Clin Rheumatol* 2006; 20: 721–40.
11. 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.
12. DerSimonian R, Laird N: Meta-analysis in clinical trials. *Control Clin Trials* 1986; 7: 177–88.
13. Sterne JA, Juni P, Schulz KF, Altman DG, Bartlett C, Egger M: Statistical methods for assessing the influence of study characteristics on treatment effects in 'meta-epidemiological' research. *Stat Med* 2002; 21: 1513–24.
14. Cohen J: *Statistical power analysis for the behavioral sciences*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum, 1988.
15. Fransen M, McConnell S, Bell M: Exercise for osteoarthritis of the hip or knee. *Cochrane Database Syst Rev* 2003(3): CD004286.
16. Lo GH, LaValley M, McAlindon T, Felson DT: Intra-articular hyaluronic acid in treatment of knee osteoarthritis: a meta-analysis. *JAMA* 2003; 290: 3115–21.
17. Chodosh J, Morton SC, Mojica W, et al.: Meta-analysis: chronic disease self-management programs for older adults. *Ann Intern Med* 2005; 143: 427–38.
18. Towheed TE, Maxwell L, Anastassiades TP, et al.: Glucosamine therapy for treating osteoarthritis. *Cochrane Database Syst Rev* 2005(2): CD002946.
19. Towheed TE, Maxwell L, Judd MG, Catton M, Hochberg MC, Wells G: Acetaminophen for osteoarthritis. *Cochrane Database Syst Rev* 2006(1): CD004257.

20. Rintelen B, Neumann K, Leeb BF: A meta-analysis of controlled clinical studies with diacerein in the treatment of osteoarthritis. *Arch Intern Med* 2006; 166: 1899–906.
21. Avouac J, Gossec L, Dougados M: Efficacy and safety of opioids for osteoarthritis: a meta-analysis of randomized controlled trials. *Osteoarthritis Cartilage* 2007; 15: 957–65.
22. Christensen R, Bartels EM, Astrup A, Bliddal H: Effect of weight reduction in obese patients diagnosed with knee osteoarthritis: a systematic review and meta-analysis. *Ann Rheum Dis* 2007; 66: 433–9.
23. Bjordal JM, Johnson MI, Lopes-Martins RA, Bogen B, Chow R, Ljunggren AE: Short-term efficacy of physical interventions in osteoarthritic knee pain. A systematic review and meta-analysis of randomized placebo-controlled trials. *BMC Musculoskelet Disord* 2007; 8: 51.
24. Bjordal JM, Klovning A, Ljunggren AE, Stordal L: Short-term efficacy of pharmacotherapeutic interventions in osteoarthritic knee pain: a meta-analysis of randomized placebo-controlled trials. *Eur J Pain* 2007; 11: 125–38.
25. Manheimer E, Linde K, Lao L, Bouter LM, Berman BM: Meta-analysis: acupuncture for osteoarthritis of the knee. *Ann Intern Med* 2007; 146: 868–77.

Anschrift für die Verfasser

Prof. Dr. med. Peter Jüni
 Institut für Sozial- und Präventivmedizin, Universität Bern
 Finkenhubelweg 11, CH-3012 Bern
 juni@ispm.unibe.ch

SUMMARY

The Effects of Exclusions of Patients From the Analysis in Randomised Controlled Trials: Meta-Epidemiological Study

Objective: To examine whether exclusions of patients from the analysis of randomized trials are associated with biased estimates of treatment effects and higher heterogeneity between trials.

Design: Meta-epidemiological study based on a collection of meta-analyses of randomised trials.

Data sources: 14 meta-analyses including 167 trials that compared therapeutic interventions with placebo or non-intervention control in patients with osteoarthritis of the hip or knee and used patient reported pain as an outcome.

Methods: Effect sizes were calculated from differences in means of pain intensity between groups at the end of follow-up, divided by the pooled standard deviation. Trials were combined by using random effects meta-analysis. Estimates of treatment effects were compared between trials with and trials without exclusions from the analysis, and the impact of restricting meta-analyses to trials without exclusions was assessed.

Results: 39 trials (23 %) had included all patients in the analysis. In 128 trials (77 %) some patients were excluded from the analysis. Effect sizes from trials with exclusions tended to be more beneficial than those from trials without exclusions (difference –0.13, 95 %-confidence interval –0.29 to 0.04). However, estimates of bias between individual meta-analyses varied considerably (2=0.07). Tests of interaction between exclusions from the analysis and estimates of treatment effects were positive in five meta-analyses. Stratified analyses indicated that differences in effect sizes between trials with and trials without exclusions were more pronounced in meta-analyses with high between trial heterogeneity, in meta-analyses with large estimated treatment benefits, and in meta-analyses of complementary medicine. Restriction of meta-analyses to trials without exclusions resulted in smaller estimated treatment benefits, larger P values, and considerable decreases in between trial heterogeneity.

Conclusion: Exclusions of patients from the analysis in randomised trials often result in biased estimates of treatment effects, but the extent and direction of bias is unpredictable. Results from intention to treat analyses should always be described in reports of randomised trials. In systematic reviews, the influence of exclusions from the analysis on estimated treatment effects should routinely be assessed.

Key words: clinical research, statistics, data analysis, bias, treatment effect

 Mit „e“ gekennzeichnete Literatur:
www.aerzteblatt.de/lit3909

Ausschluss von Patienten aus der Analyse: Auswirkungen auf die Ergebnisse randomisierter kontrollierter Studien

Eine meta-epidemiologische Studie

Eveline Nüesch, Sven Trelle, Stephan Reichenbach, Anne W. S. Rutjes, Elizabeth Bürgi,
Martin Scherer, Douglas G. Altman, Peter Jüni

e-LITERATUR

1. Pildal J, Hrobjartsson A, Jorgensen K, Hilden J, Altman D, Gotzsche P: Impact of allocation concealment on conclusions drawn from meta-analyses of randomized trials. *Int J Epidemiol* 2007; 36: 847–57.
2. Altman DG, Schulz KF, Moher D, et al.: The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med* 2001; 134: 663–94.
3. Egger M, Juni P, Bartlett C: Value of flow diagrams in reports of randomized controlled trials. *JAMA* 2001; 285: 1996–9.
4. Gravel J, Opatrný L, Shapiro S: The intention-to-treat approach in randomized controlled trials: are authors saying what they do and doing what they say? *Clin Trials* 2007; 4: 350–6.
5. Huwiler-Muntener K, Juni P, Junker C, Egger M: Quality of reporting of randomized trials as a measure of methodologic quality. *JAMA* 2002; 287: 2801–4.
6. Fergusson D, Aaron SD, Guyatt G, Hébert P: Postrandomization exclusions: the intention to treat principle and excluding patients from analysis. *BMJ* 2002; 325: 652–4.
7. Moher D, Schulz KF, Altman DG: The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *Ann Intern Med* 2001; 134: 657–62.
8. Sackett DL, Gent M: Controversy in counting and attributing events in clinical trials. *N Engl J Med* 1979; 301: 1410–2.
9. Baron G, Boutron I, Giraudeau B, Ravaud P: Violation of the intent-to-treat principle and rate of missing data in superiority trials assessing structural outcomes in rheumatic diseases. *Arthritis Rheum* 2005; 52: 1858–65.
10. Kim M: Statistical methods in Arthritis & Rheumatism: current trends. *Arthritis Rheum* 2006; 54: 3741–9.
11. Julious SA, Mullee MA: Issues with using baseline in last observation carried forward analysis. *Pharm Stat* 2008; 7: 142–6.
12. Streiner DL: The case of the missing data: methods of dealing with dropouts and other research vagaries. *Can J Psychiatry* 2002; 47: 68–75.
13. Baron G, Ravaud P, Samson A, Giraudeau B: Missing data in randomized controlled trials of rheumatoid arthritis with radiographic outcomes: a simulation study. *Arthritis Rheum* 2008; 59: 25–31.