

Arthroskopie 2021 · 34:299–304
<https://doi.org/10.1007/s00142-021-00469-4>
 Angenommen: 21. April 2021
 Online publiziert: 7. Juni 2021
 © Der/die Autor(en) 2021



Daniel Günther¹ · Elmar Herbst² · Brenda Laky^{3,4} · Christian Lattermann⁵ · Dominic T. Mathis⁶ · Philip Rössler^{7,8} · Karl Friedrich Schüttler⁹ · Arasch Wafaisade¹ · Sebastian Kopf¹⁰ · Research-Komitee der Arbeitsgemeinschaft für Arthroskopie (AGA)

¹ Klinik für Orthopädie, Unfallchirurgie und Sporttraumatologie, Krankenhaus Köln-Merheim, Klinikum der Universität Witten/Herdecke, Köln, Deutschland

² Klinik für Unfall-, Hand- und Wiederherstellungschirurgie, Universitätsklinikum Münster, Münster, Deutschland

³ Austrian Research group for Regenerative and Orthopedic Medicine (AURROM), Wien, Österreich

⁴ Center of Clinical Research, Universitätszahnklinik Wien, Medizinische Universität Wien, Wien, Österreich

⁵ Department of Orthopaedic Surgery, Division of Sports Medicine, Brigham and Women's Hospital, Boston, USA

⁶ Klinik für Orthopädie und Traumatologie des Bewegungsapparates, Kantonsspital Baselland (Bruderholz, Liestal, Laufen), Bruderholz, Schweiz

⁷ Orthopädische Praxisklinik, Mayen, Deutschland

⁸ Medizinische Fakultät, Rheinische Friedrich-Wilhelms-Universität, Bonn, Deutschland

⁹ Orthopaedicum, Lich, Deutschland

¹⁰ Zentrum für Orthopädie und Unfallchirurgie, Medizinische Hochschule Brandenburg Theodor Fontane der Universitätsklinik Brandenburg an der Havel, Brandenburg, Deutschland

Herausforderungen von Evidenzlevel-1-Studien in der Unfallchirurgie und Orthopädie

Randomisierte kontrollierte klinische Studien („randomized controlled trials“, RCTs) gelten als Goldstandard der evidenzbasierten Medizin [19, 20]. Auch die Verfahrensordnung des Gemeinsamen Bundesausschusses (Gba; [8]) fordert die höchstmögliche Evidenzstufe. Einschränkend ist formuliert: „Sofern es unmöglich oder unangemessen ist, Studien höchster Evidenzstufe durchzuführen oder zu fordern, sind Nachweise der bestverfügbaren Evidenz einzureichen“ [8]. Viele der Grundsätze von RCTs, die im Rahmen von pharmakologischen Medikamentenstudien zum Einsatz kommen, bergen bei ihrer Anwendung in der Chirurgie potenzielle Risiken oder Schwierigkeiten. Diese Schwierigkeiten betreffen vom Studiendesign über die Studienplanung bis hin zur Studiendurchführung und Datenauswertung viele Bereiche einer Studie.

Der Studienplanung kommt im Rahmen von RCTs eine immens wichtige Bedeutung zu. Im Rahmen dieses Manuskripts wird immer wieder deutlich

werden, dass eine akribische A-priori-Planung einer RCT die Grundlage einer validen Studiendurchführung bietet. Die Planung sollte das Studiendesign, die Erarbeitung der primären und ggf. sekundären Fragestellungen, die notwendige Fallzahl und die Festlegung geeigneter statistischer Tests enthalten. Die Anmeldung einer RCT sollte dann grundsätzlich in einem Studienregister erfolgen.

Der vorliegende Artikel soll die Ziele aufzeigen, warum RCTs für chirurgische Fachdisziplinen eine besondere Herausforderung darstellen, insbesondere, wenn eine Operation als einer der Behandlungsarme geplant ist.

Studiendesign

Eine Besonderheit chirurgischer klinischer Studien ist es, dass das Cross-over-Design, bei dem zwei Arzneimittel, oder ein Arzneimittel und ein Placebo, zeitlich versetzt dem gleichen Probanden verabreicht werden, schwierig implementiert werden kann. Wenn eine Operation als

Intervention bzw. Therapie im Rahmen einer RCT durchgeführt wird, so ist diese in der Regel dauerhaft. Das heißt, es gibt kein *Auswaschen*, wie es bei der Verwendung von Medikamenten auftritt. Patienten können nicht als eigene Kontrolle für zwei aufeinanderfolgende Eingriffe am gleichen Organ oder Gelenk fungieren. Bei Doppelorganen (z. B. Operation an beiden Knien oder Schultern) könnte man theoretisch die eine Seite mit Operation A und die andere mit Operation B oder einem Arzneimittel oder Physiotherapie behandeln. Die Patienten würden aber bereits eine Veränderung durch die erste Therapie erleben, die sich auch auf die zweite Therapie auswirken würde. Die Auswirkungen zwischen Doppelorganen können lediglich verglichen werden, wenn das Krankheitsniveau sehr ähnlich bleibt, was selten auftritt.

Fallzahlen, multizentrischer Ansatz und externe Validität

In RCTs ist es oft eine Herausforderung, die Mindestanzahl von Patienten einzuschließen, die erforderlich ist, um eine statistische Aussagekraft (Power) von 80 % zu erreichen. Eine zu geringe Patientenzahl birgt die Gefahr eines Typ-II-Fehlers. Bei einem Typ-II-Fehler wird die Nullhypothese fälschlicherweise angenommen, obwohl die Alternativhypothese korrekt ist. Es wird also z. B. festgestellt, dass zwischen 2 Gruppen kein Unterschied besteht, obwohl lediglich die Fallzahl zu gering ist, einen signifikanten Unterschied zu zeigen. Dies kann zu einem an einer fehlenden oder fehlerhaften Berechnung der notwendigen Fallzahl liegen oder zum anderen an der Schwierigkeit, eine ausreichende Anzahl von Probanden einzuschließen. Gründe dafür liegen in den oft hohen Kosten chirurgischer Eingriffe, dem seltenen Auftreten bestimmter Pathologien oder, insbesondere bei neuen chirurgischen Eingriffen, der validen Berechnung der idealen Stichprobengröße.

» Einen möglichen Lösungsansatz, um die Fallzahl zu erhöhen, bieten multizentrische Studien

Die Akzeptanzrate für die Teilnahme an einer chirurgischen RCT beträgt im Allgemeinen weniger als 50 %. Die Hauptgründe für die Nichtteilnahme, die Patienten angeben, sind ihre Präferenz für einen der Behandlungsarme, ihre Unzufriedenheit mit der Randomisierung und die Möglichkeit höherer Kosten [2]. Metaanalysen chirurgischer RCTs zeigen, dass sowohl in der Unfallchirurgie als auch in der Orthopädie eine adäquate Power oft nicht erreicht wird. In einer Überprüfung der Literatur zu chirurgischen Traumata von insgesamt 117 bewerteten RCTs betrug die durchschnittliche statistische Power 25 % (Spanne: 2–99 %). Generell wird in medizinischen Studien eine statistische Power von 85 % als sinnvoll angesehen. Dies erklärt, warum 91 % der evaluierten Studien Typ-II-Fehler aufwiesen [13]. Laut einer anderen Metaanalyse hatten

allerdings lediglich 28 % aller orthopädischen Studien mit negativem Ergebnis keine adäquate Power [1]. Eine A-priori-Power-Analyse könnte diese Problematik limitieren. Eine Analyse aus dem Jahr 2001 zeigte aber, dass lediglich 9 % der untersuchten orthopädischen Studien eine solche Analyse durchgeführt hatte [9].

Ein weiteres Problem bei der Rekrutierung ist die Auswahlverzerrung. Patienten, die der Teilnahme an der Studie zustimmen, können andere Faktoren aufweisen als Patienten, die die Studienteilnahme ablehnen. Die Gruppe, die der Teilnahme an der Studie nicht zugestimmt hat, sollte daher mit der Gruppe, die der Teilnahme zugestimmt hat, verglichen werden. Wenn es beim Einschluss in die Studien Unterschiede bei den Faktoren gibt, die möglicherweise das Ergebnis beeinflussen könnten, wie z. B. Alter oder Geschlecht, sollten diese in eine Kovarianzanalyse einbezogen werden. Ein weiteres Problem in diesem Zusammenhang sind strenge Ein- und Ausschlusskriterien, die dazu führen können, dass ein Großteil der Patienten mit dem zu untersuchenden Krankheitsbild aufgrund der Kriterien nicht eingeschlossen wird, wodurch das eingeschlossene Patientengut nicht mehr repräsentativ ist.

Einen möglichen Lösungsansatz, um die Fallzahl zu erhöhen, bieten multizentrische Studien. Hier müssen allerdings auch die Vor- und Nachteile eines multizentrischen Aufbaus beleuchtet werden [15]. Die chirurgischen Details der Operationen der teilnehmenden Zentren können deutlich variieren. Personal (Operateure, Anästhesisten, Pflegeteam), Verfahren (Anästhesieverfahren, Sterilisationsmethoden), Material (anästhesiologische und chirurgische Ausrüstung, Art der Spülflüssigkeit), die Nachbehandlung und Nachbehandlungcompliance der Patienten und die Umgebung (z. B. Bakterienflora) können das Endergebnis beeinflussen. Wenn eine Extrapolation der Ergebnisse einer multizentrischen RCT gewünscht wird und diese somit als allgemeingültig angesehen werden sollen, muss davon ausgegangen werden, dass die durchgeführte oder beabsichtigte Intervention an allen Standorten oder für alle Ärzte an verschiede-

nen Standorten gleich ist. Die externe Validität von RCTs in der Chirurgie ist allerdings häufig gering. Egal wie reproduzierbar eine operative Technik ist, sie ist nicht identisch. Darüber hinaus hängen neue Operationstechniken im Allgemeinen von einer Lernkurve ab. Die Lernkurve kann für jede Technik und für jeden Chirurgen variieren. Auf diese Weise können sich selbst Operationen, die zu unterschiedlichen Zeiten am selben Zentrum vom selben Operateur durchgeführt werden, erheblich unterscheiden. Auch die prä- und postoperativen Maßnahmen können stark variieren. Während das verwendete chirurgische Verfahren in einer Metaanalyse in allen Studien beschrieben wurde, wurden die prä- und postoperative Versorgung und die verwendete Anästhesie nur in 7 %, 50 % bzw. 13 % beschrieben [3]. Der Mangel an Informationen dieser Art beeinträchtigt erheblich die Verallgemeinerung der Daten und folglich die externe Validität. Es kann argumentiert werden, dass diese Verzerrungen beide Arme der Studie beeinflussen und mit zunehmender Anzahl von Personen verwässert würden. Dafür würde allerdings eine enorme Fallzahl benötigt, was jedoch aus finanzieller Sicht oft nicht praktikabel ist. RCTs mit großer Fallzahl werden zudem in der Regel oft von der Industrie gesponsert, was zu Interessenkonflikten führen kann und das Verzerrungspotenzial deutlich erhöht [4, 14].

Andererseits erhöht ein multizentrischer Aufbau mit seiner Heterogenität die externe Validität der Studie. Ausgezeichnete Bedingungen am einzelnen Studienzentrum, die an anderen Orten nur schwer reproduzierbar wären, könnten eventuell gute Ergebnisse einiger kleiner Studien in der Chirurgie erklären.

Es muss des Weiteren festgehalten werden, dass auch Studien mit geringer Power wertvoll sein können, wenn sie in Metaanalysen zusammengefasst werden. Sie können überdies vorläufige Daten zur Berechnung der notwendigen Fallzahl und zur Beurteilung der Machbarkeit der Durchführung ähnlicher Studien in der Zukunft liefern.

Einfluss multipler Faktoren auf das Krankheitsbild

Bei Krankheitsbildern, bei denen das Ergebnis wesentlich von verschiedenen Faktoren abhängt, wird es mit zunehmender Anzahl der Faktoren schwieriger, diese zu kontrollieren. So ist es z. B. bekannt, dass in der Therapie von Knorpeldefekten die Ausdehnung des Defekts, die genaue Lokalisation des Defekts im Knie, die Beinachse des Patienten, das Gewicht und das biologische Alter des Patienten wichtige Faktoren sind. Weitere Faktoren wie Ernährung, hormonelle Ausgangslage, Arbeitsumfeld oder supportive Injektionen bestimmter Medikamente können ebenfalls einen Einfluss haben. An diesem Beispiel sieht man, dass eine RCT, welche alle Faktoren kontrolliert, kaum durchführbar ist. Die Fallzahl wäre dann insgesamt sehr klein. Daten werden in solchen Fällen z. B. häufig über große Registerstudien generiert. Selbst eine methodisch völlig korrekt durchgeführte RCT mit adäquater statistischer Power kann daher zu einer unrealistischen Aussage kommen, wenn ein wichtiger Einflussfaktor keine Berücksichtigung findet. Dies ist schwierig, da es oft besonderer Sachkenntnisse bedarf, um diesen fehlenden Einflussfaktor überhaupt als solchen zu erkennen. Eine gute RCT benötigt daher eine adäquate Studienplanung im Expertenkreis. Gelingt es, die einzelnen Einflussfaktoren im Rahmen der Planung zu berücksichtigen und gezielt auszuwählen, welche Faktoren im Rahmen der RCT priorisiert werden sollen, wird dies letztendlich auch zu einer wissenschaftlich hochwertigen RCT führen.

Randomisierung

Die Randomisierung der Patienten erfordert, dass bei zwei zu vergleichenden Verfahren beide Operationstechniken unter Beachtung der Randomisierung durchgeführt werden. Jede Einschränkung bei der Durchführung einer der Techniken (z. B. nur an bestimmten Wochentagen, nur durch einen bestimmten Chirurgen) birgt das Risiko, dieses Prinzip zu verletzen. Beispielsweise würde in Notfällen

oder Traumata, in denen die Therapie komplex ist und von einem bestimmten Chirurgen durchgeführt wird, die Randomisierung von der Verfügbarkeit des Arztes abhängig gemacht, und daher wäre eine Zuordnung nach Zweckmäßigkeit erforderlich, wodurch die Randomisierung beeinträchtigt wird. Idealerweise sollte die Auslosung zum Zeitpunkt der Operation oder gar in der Operation erfolgen. In manchen Fällen ist dies sogar notwendig. Beispielsweise kann in einer Studie über Meniskusnähte der Patient nur dann in die Studie aufgenommen werden, wenn die Naht möglich ist [5].

Verblindung

Das Verblinden ist ein wichtiger Bestandteil von RCTs. Die Doppelblindstudie (sowohl der Patient als auch der Arzt oder der Forscher, der die Studie durchführt, sind verblindet) ist der am häufigsten verwendete Typ. Im Rahmen von chirurgischen Studien kann es dabei zu den im Folgenden genannten Schwierigkeiten kommen [7].

Verblindung des Chirurgen

Für den Chirurgen, der eine Operation durchführt, ist eine Verblindung schwierig, wenn sich die Operationstechniken oder Implantate unterscheiden [18]. Es könnte argumentiert werden, dass, wenn die postoperative Datenerfassung verblindet wäre, die Verzerrungsgefahr minimiert würde. Wenn jedoch der Chirurg an eine der Techniken mehr glaubt, könnte er größere Anstrengungen in diese Operation stecken und so die Ergebnisse der Studie beeinflussen [17]. Es ist nicht zu verhindern, dass der Chirurg weiß, dass die Studie stattfindet und Patienten eingeschlossen werden. Es gibt jedoch Möglichkeiten diesen Bias zu reduzieren. Beispielsweise muss der Chirurg nicht wissen, welche Patienten dies im Einzelnen sind. Der Studienleiter könnte ihm vor jeder Operation die Methodik mitteilen, es werden aber z. B. nur 20 % der Patienten eingeschlossen. Der Operateur wüsste in dem Fall nicht, um welche Patienten es sich konkret handeln würde.

Arthroskopie 2021 · 34:299–304
<https://doi.org/10.1007/s00142-021-00469-4>
 © Der/die Autor(en) 2021

D. Günther · E. Herbst · B. Laky ·
 C. Lattermann · D. T. Mathis · P. Rössler ·
 K. F. Schüttler · A. Wafaisade · S. Kopf ·
 Research-Komitee der
 Arbeitsgemeinschaft für Arthroskopie
 (AGA)

Herausforderungen von Evidenzlevel-1-Studien in der Unfallchirurgie und Orthopädie

Zusammenfassung

Randomisierte kontrollierte klinische Studien stellen für chirurgische Fachdisziplinen eine Herausforderung dar. Dieser Artikel beschreibt die Besonderheiten und Schwierigkeiten solcher Studien. Wann immer möglich, sollte das höchste Evidenzlevel angestrebt werden. Ist ein hohes Evidenzlevel allerdings methodisch nicht korrekt erreichbar, sollte das bestmögliche nächste Evidenzlevel erzielt werden. Die Bewertung, auch im Hinblick auf die Kostenerstattung, sollte daher gerade in chirurgischen Disziplinen nicht ausschließlich am Evidenzlevel festgemacht werden. Vielmehr sollte im Rahmen von Expertengremien und unter Zuhilfenahme der besten, verfügbaren Evidenz beurteilt werden.

Wenn die Operationstechniken identisch sind und die Intervention darin besteht, einen zusätzlichen Faktor einzuführen (Injektion eines Wachstumsfaktors, postoperative Medikation), kann eine Verblindung auch für den Chirurgen einfacher möglich sein. Es muss jedoch vorausgesetzt sein, dass die Intervention eine Verblindung mit einem Placebo ermöglicht oder die gleichen physikalischen Eigenschaften wie die Kontrolle hat.

Verblindung der Patienten

Wenn das Ziel der Studie der Vergleich zwischen zwei verschiedenen Operationstechniken ist, birgt es eine Schwierigkeit, wenn die Zugangswege unterschiedlich sind. Wenn der Patient während der Nachsorge Zugang zu den Röntgenuntersuchungen hat, was normalerweise der Fall ist, können auch

Unterschiede zwischen den verwendeten Implantaten festgestellt werden. In der Literatur findet man im Rahmen von RCTs durchaus auch Placebo-Operationen (Scheinoperationen; [16]). Die ethischen Implikationen solcher Studien sind jedoch offensichtlich, und es ist selten, dass solche Studien von Ethikkommissionen genehmigt und von Patienten akzeptiert werden [10, 12]. Darüber hinaus muss angemerkt werden, dass der Terminus „Scheinoperation“ in der Literatur nicht klar definiert ist. Das klassische Beispiel ist das Anlegen der Hautschnitte wie bei der eigentlichen Operation. Bei einer kürzlich veröffentlichten RCT zum Vergleich der konservativen Therapie mit der arthroskopischen Teilresektion von degenerativen Innenmeniskusläsionen wurde gar die arthroskopische Lavage als Scheinoperation bezeichnet. Bereits die Anlage der Hautschnitte kann zu Hämatomen, Schmerzen und einer Infektion führen. Gerade das zweite Beispiel zeigt, dass hier der Begriff der Scheinoperation sehr weit gefasst wurde, da eine Lavage durchaus bereits eine therapeutische Intervention ist. Des Weiteren können postoperative Symptome wie Schmerzen, Schwellung oder Hämatome bei Scheinoperationen geringer ausgeprägt sein als bei der eigentlichen Operation, so dass eventuell durch den Patienten aber auch den Nachuntersucher auf die Gruppenzugehörigkeit geschlossen werden kann.

Verblindung der Nachuntersucher

Der unabhängige Nachuntersucher, der in der Regel ein Arzt oder Physiotherapeut ist, der ansonsten nicht weiter in die Studie involviert ist, spielt eine wichtige Rolle bei RCTs. Er kann einen Verlust der Verblindung erfahren, wenn die zu bewertenden Verfahren über unterschiedliche Zugangswege durchgeführt wurden, was zu chirurgischen Narben führt, die die Identifizierung der Gruppe zur Folge hat. Das Maskieren der Narbe bei allen Bewertungen durch die Verwendung geeigneter Kleidung ist eine Möglichkeit, die Verblindung sicherzustellen. Auch wenn unterschiedliche Operationstechniken

unterschiedliche Rehabilitationsprotokolle erfordern, kann die Verblindung verloren gehen. Des Weiteren kann ggf. in der direkten postoperativen Phase, wie bereits im vorherigen Absatz angesprochen, auf Grund des Erscheinungsbildes des Operationsgebiets mit Ausprägung der Hämatome oder Weichteilschwellung oder aufgrund der Beschwerden des Patienten auf die Gruppenzugehörigkeit geschlussfolgert werden.

Verblindung des Statistikers

Die Verblindung des Statistikers kann durchaus sinnvoll sein, um einen möglichen Einfluss auf die Datenauswertung zu verhindern. Der Statistiker erhält hierbei lediglich die Daten der einzelnen Studiengruppen, welche allerdings mit einem Code verblindet sind. Hierzu ist es natürlich notwendig, dass die Statistik nicht, wie häufig in klinischen Studien üblich, durch einen der Operateure oder Nachuntersucher durchgeführt wird.

Nachuntersuchung

Chirurgische Eingriffe sind häufig kurativ. Dies unterscheidet sie von Pathologien wie arteriellem Hypertonus oder Diabetes mellitus, bei denen über lange Zeiträume und möglicherweise während des gesamten Patientenlebens regelmäßige Kontrollen erforderlich sind. Patienten mit akuten Zuständen, wie Frakturen, oder auch chronischen Zuständen, wie einer Arthrose, die endoprothetisch versorgt wird, zeigen kurz- und mittelfristig eine signifikante Verbesserung der Symptome. Wenn es den Patienten gut geht, kann es daher schwierig sein, das Follow-up über einen langen Zeitraum aufrechtzuerhalten.

Eine weitere Schwierigkeit im Zusammenhang mit der Einhaltung chirurgischer Protokolle durch den Patienten ist die Notwendigkeit einer Rehabilitation. Die Einnahme eines Medikaments zu Hause erfordert weniger persönlichen, körperlichen und oft auch finanziellen Aufwand von einem Patienten als das Verlassen des Hauses und der Weg zum Physiotherapieort. Da ein Mangel an angemessener Rehabilitation oder ein Abbruch der Rehabilitation häufig

ein Ausschlusskriterium der Studie ist (wenn es festgestellt wird), kann dies das Endergebnis deutlich beeinträchtigen. Gerade in chirurgischen Studien sollten die Nachuntersuchungsintervalle daher regelmäßig erfolgen, um Nachuntersuchungsdaten an Zeitpunkten zu haben, bevor ein Follow-up abgebrochen oder in einen anderen Behandlungsarm gekreuzt wird. Interessanterweise haben nur 16,4% der RCTs in der orthopädischen Chirurgie zwischen 2005 und 2008 die Nachuntersuchungsdaten von Zeitpunkten, bevor ein Follow-up abgebrochen oder in einen anderen Behandlungsarm gekreuzt wurde, inkludiert [11]. Bei den meisten dieser Studien wurden die während des Follow-ups verlorenen Patienten von der endgültigen statistischen Analyse ausgeschlossen. Das Auslassen dieser Daten kann zu Verzerrungen führen, da dies die Integrität der Randomisierung beeinträchtigen kann [11]. Zu kleine Nachuntersuchungsintervalle wiederum können die Compliance der Studienteilnehmer senken. Nachuntersuchungszeitpunkte sollten daher wohlüberlegt gesetzt werden. Eine routinemäßige Nachuntersuchung nach 6 Wochen, 12 Wochen, 6 Monaten und 12 Monaten, wie es im klinischen Vorgehen oft Tradition ist, macht daher nur Sinn, wenn die Nachuntersuchungszeitpunkte im Einzelnen auch eine sinnvolle Aussagekraft haben.

Datenanalyse

Es wird empfohlen, die Datenanalyse einer RCT anhand der Gruppe durchzuführen, für die der Patient ausgewählt wurde, selbst wenn dieser im Verlauf der Studie die Gruppe wechselt (*Intention-to-treat*-Analyse). In chirurgischen Studien kann das Befolgen dieses Prinzips seltsame oder sogar inkongruente Ergebnisse hervorrufen, insbesondere, wenn man das Wechseln von einem Studienarm in den anderen erlaubt (z. B. beim Vergleich zwischen einer konservativen und einer operativen Therapie). Ein Patient, der für eine konservative Behandlung ausgewählt wurde und sich im Verlauf der operativen Therapie unterzieht, wird bei einer *Intention-to-treat*-Analyse weiterhin als konservativer Fall analy-

siert. Wenn dieser Patient eine Infektion an der Operationsstelle präsentiert, würden die Ergebnisse einer solchen Studie das „Auftreten einer Infektion der Operationsstelle“ als „Komplikation durch konservative Behandlung zeigen“ [15]. Weitere mögliche Analysen stellen die *As-treated*-Analyse und *Per-protocol*-Analyse dar. Die *As-treated*-Analyse inkludiert alle Patienten, welche die im Rahmen der Studie zu prüfende Intervention wirklich erhalten haben, unabhängig davon, welcher Behandlung sie ursprünglich laut Studienplan zugeordnet waren. Die *Per-protocol*-Analyse inkludiert alle Patienten, die die laut Studienplan zugewiesene Behandlung auch tatsächlich erhalten haben.

Diskussion

RCTs stellen in der klinischen Forschung den Goldstandard dar. Insbesondere die Kostenträger fordern die Durchführung von RCTs, um Behandlungen in die Vergütung einzuschließen. Im Rahmen dieses Artikels wird deutlich, dass die Durchführung valider RCTs in der Chirurgie nicht immer einfach ist. Trotzdem sollte die Durchführung hochwertiger RCTs in der Unfallchirurgie und Orthopädie höchste Priorität haben. Dieser Artikel kann helfen, sich der Schwierigkeiten einer RCT bewusst zu werden. Er soll aber auch Anreiz geben, diese Kenntnisse zu nutzen, um auf Basis einer soliden Planung das höchstmögliche Evidenzlevel zu erreichen.

Insbesondere bei selten auftretenden Pathologien oder solchen, die multifaktoriell beeinflusst werden, kann es allerdings durchaus sein, dass methodisch korrekt durchgeführte Studien auf niedrigerem Evidenzlevel einen höheren Stellenwert erhalten müssen, da eine RCT nicht adäquat durchgeführt werden kann. Evidenzlevel 1 ist nicht immer gleichzusetzen mit qualitativ hochwertig, und in bestimmten Situationen ist die Durchführung von RCTs ggf. sogar unmöglich, obwohl die Therapie erfolgversprechend oder sogar allen anderen verfügbaren Therapien überlegen ist [6]. Gerade in solchen Fällen können Registerstudien durch ihre oft hohen Fallzahlen Aufschlüsse über die

Qualität einzelner Interventionen geben. Auch ein Expertenkonsensus bietet die Möglichkeit einer Bewertung und kann letztendlich sogar die Basis für folgende RCTs legen, da wichtige Einflussfaktoren gewichtet werden.

» Es sollte immer die höchste Evidenz zur Bewertung herangezogen werden

Liegen zur Bewertung einer Therapie methodisch korrekte Studien verschiedener Evidenzlevel vor, sollte jedoch immer die höchste Evidenz zur Bewertung herangezogen werden. Man kann argumentieren, dass die externe Validität beim Vorliegen einer monozentrischen RCT ggf. nicht gegeben ist. Allerdings müssen auch die Ergebnisse von monozentrischen Studien mit geringerer Evidenz, wie z. B. retrospektive Fallserien, nicht unbedingt auf andere Kliniken übertragbar sein. Das Vorliegen mehrerer ähnlicher RCTs unterschiedlicher Arbeitsgruppen mit gleichen Ergebnissen erhöht die externe Validität. Die Durchführung additiver RCTs bei bereits vorliegenden methodisch korrekten RCTs muss allerdings ethisch jeweils detailliert abgewogen werden.

Es bleibt festzuhalten, dass methodisch gute Level-1-RCTs angestrebt werden sollten. Im Ausnahmefall, bei sehr seltenen Pathologien, sollte das bestmögliche nächste Evidenzlevel genutzt werden. Registerstudien stellen eine sehr gute Ergänzung zu RCTs dar. Die Bewertung, auch im Hinblick auf die Kostenersatzung, sollte daher gerade in chirurgischen Disziplinen nicht ausschließlich anhand des Evidenzlevels erfolgen. Vielmehr sollte im Rahmen von Expertengremien unter Zuhilfenahme der bestverfügbaren Evidenz beurteilt werden.

Korrespondenzadresse

PD Dr. med. Daniel Günther, MHBA
Klinik für Orthopädie, Unfallchirurgie und Sporttraumatologie, Krankenhaus Köln-Merheim, Klinikum der Universität Witten/Herdecke
Ostmerheimer Str. 200, 51109 Köln, Deutschland
GuentherD@kliniken-koeln.de

Funding. Open Access funding enabled and organized by Projekt DEAL.

Einhaltung ethischer Richtlinien

Interessenkonflikt. D. Günther, E. Herbst, B. Laky, C. Lattermann, D. T. Mathis, P. Rössler, K.F. Schüttler, A. Wafaisade und S. Kopf geben an, dass kein Interessenkonflikt besteht.

Für diesen Beitrag wurden von den Autoren keine Studien an Menschen oder Tieren durchgeführt. Für die aufgeführten Studien gelten die jeweils dort angegebenen ethischen Richtlinien.

Open Access. Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>.

Literatur

1. Abdullah L, Davis DE, Fabricant PD et al (2015) Is there truly „no significant difference“? Underpowered randomized controlled trials in the orthopaedic literature. *J Bone Joint Surg Am* 97:2068–2073
2. Abraham NS, Young JM, Solomon MJ (2006) A systematic review of reasons for nonentry of eligible patients into surgical randomized controlled trials. *Surgery* 139:469–483
3. Ahmad N, Boutron I, Moher D et al (2009) Neglected external validity in reports of randomized trials: the example of hip and knee osteoarthritis. *Arthritis Rheum* 61:361–369
4. Bhandari M, Jonsson A, Buhren V (2006) Conducting industry-partnered trials in orthopaedic surgery. *Injury* 37:361–366
5. Bhandari M, Richards RR, Sprague S et al (2002) The quality of reporting of randomized trials in the *Journal of Bone and Joint Surgery* from 1988 through 2000. *J Bone Joint Surg Am* 84:388–396
6. Bothwell LE, Greene JA, Podolsky SH et al (2016) Assessing the Gold standard—lessons from the History of RCTs. *N Engl J Med* 374:2175–2181
7. Boutron I, Tubach F, Giraudeau B et al (2004) Blinding was judged more difficult to achieve

-
- ve and maintain in nonpharmacologic than pharmacologic trials. *J Clin Epidemiol* 57:543–550
8. Bundesausschuss G (2020) Verfahrensordnung des Gemeinsamen Bundesausschusses. *Bundesanzeiger BAnz AT B2*
 9. Freedman KB, Back S, Bernstein J (2001) Sample size and statistical power of randomised, controlled trials in orthopaedics. *J Bone Joint Surg Br* 83:397–402
 10. Heckerling PS (2006) Placebo surgery research: a blinding imperative. *J Clin Epidemiol* 59:876–880
 11. Herman A, Botser IB, Tenenbaum S et al (2009) Intention-to-treat analysis and accounting for missing data in orthopaedic randomized clinical trials. *J Bone Joint Surg Am* 91:2137–2143
 12. Horng S, Miller FG (2002) Is placebo surgery unethical? *N Engl J Med* 347:137–139
 13. Lochner HV, Bhandari M, Tornetta P 3rd (2001) Type-II error rates (beta errors) of randomized trials in orthopaedic trauma. *J Bone Joint Surg Am* 83:1650–1655
 14. Lynch JR, Cunningham MR, Warme WJ et al (2007) Commercially funded and United States-based research is more likely to be published; good-quality studies with negative outcomes are not. *J Bone Joint Surg Am* 89:1010–1018
 15. Malavolta EA, Demange MK, Gobbi RG et al (2011) Randomized controlled clinical trials in orthopedics: difficulties and limitations. *Rev Bras Ortop* 46:452–459
 16. Moseley JB, O'malley K, Petersen NJ et al (2002) A controlled trial of arthroscopic surgery for osteoarthritis of the knee. *N Engl J Med* 347:81–88
 17. Roman H, Marpeau L, Hulsey TC (2008) Surgeons' experience and interaction effect in randomized controlled trials regarding new surgical procedures. *Am J Obstet Gynecol* 199:108.e1–108.e6
 18. Simunovic N, Devereaux PJ, Bhandari M (2008) Design considerations for randomised trials in orthopaedic fracture surgery. *Injury* 39:696–704
 19. Soucacos PN, Johnson EO, Babis G (2008) Randomised controlled trials in orthopaedic surgery and traumatology: overview of parameters and pitfalls. *Injury* 39:636–642
 20. Zlowodzki M, Jonsson A, Bhandari M (2006) Common pitfalls in the conduct of clinical research. *Med Princ Pract* 15:1–8

Hier steht eine Anzeige.

