

Unfallchirurg 2021 · 124:1007–1017
<https://doi.org/10.1007/s00113-021-01101-8>
Angenommen: 6. Oktober 2021
Online publiziert: 10. November 2021
© The Author(s), under exclusive licence to
Springer Medizin Verlag GmbH, ein Teil von
Springer Nature 2021

Wissenschaftliche Leitung
Volker Alt, Regensburg
Peter Biberthaler, München
Thomas Gössling, Braunschweig
Thomas Mittlmeier, Rostock



CME

Zertifizierte Fortbildung

Klinische Studien in Unfallchirurgie und Orthopädie: lesen, interpretieren und umsetzen

Dirk Stengel¹ · Wolf Mutschler² · Luzi Dubs³ · Stephan Kirschner⁴ ·
Tobias Renkawitz⁵

¹ Forschung – Ressort Medizin, BG Kliniken – Klinikverbund der Gesetzlichen Unfallversicherung gGmbH, Berlin, Deutschland

² Klinik für Allgemeine, Unfall- und Wiederherstellungschirurgie, LMU Klinikum der Universität München, München, Deutschland

³ Orthopädische Chirurgie, Winterthur, Schweiz

⁴ Klinik für Orthopädie, St. Vincentius-Kliniken gAG, Karlsruhe, Deutschland

⁵ Klinik für Orthopädie und Unfallchirurgie, Universitätsklinikum Heidelberg, Heidelberg, Deutschland

Zusammenfassung

Für eine informierte, partizipative ärztliche Entscheidungsfindung müssen Sie neben Ihrem Erfahrungswissen aktuelle wissenschaftliche Daten berücksichtigen. Die digitale Informationsflut macht es im klinischen Alltag allerdings schwer, immer auf dem Stand des Wissens zu bleiben. Der Beitrag bietet einige Hilfen an, wie mit diesem Problem umgegangen werden kann. Ein grundlegendes Verständnis über Vor- und Nachwahrscheinlichkeiten sowie systematische Fehler (Bias) erleichtert die Abwägung zwischen Nutzen und Risiko z. B. einer (chirurgischen) Intervention im Vergleich zu einer nichtoperativen Therapie. Randomisierte Studien („randomized controlled trials“, RCT, mit allen modernen Modifikationen) liefern unverzerrte Ergebnisse, führen in Orthopädie und Unfallchirurgie evtl. zu einer stark selektierten, nichtrepräsentativen Stichprobe, und ihre Resultate müssen durch weitere, unabhängige RCT bestätigt oder widerlegt werden. Große beobachtende Datenmengen (z. B. aus Registern) können quasiexperimentell modelliert werden und RCT in der Gesundheitstechnologiebewertung flankieren.

Schlüsselwörter

Randomisierte kontrollierte Studien · Wahrscheinlichkeit · Konfidenzintervalle · Systematischer Fehler · Wissenschaftliches Fehlverhalten

Online teilnehmen unter:
www.springermedizin.de/cme

Für diese Fortbildungseinheit
werden 3 Punkte vergeben.

Kontakt

Springer Medizin Kundenservice
Tel. 0800 77 80 777
(kostenfrei in Deutschland)
E-Mail:
kundenservice@springermedizin.de

Informationen

zur Teilnahme und Zertifizierung finden
Sie im CME-Fragebogen am Ende des
Beitrags.

Medicine is a science of uncertainty and an art of probability. (William Osler, 1849–1919) [1]

Lernziele

Nach der Lektüre dieses Beitrags ...

- können Sie Studienergebnisse besser mit Ihrem Erfahrungs- und Praxiswissen kombinieren, und diese im Patient(inn)en- und Angehörigengespräch umsetzen,
- sind Sie wachsamer im Hinblick auf Datenmanipulation und Forschungsbetrug, die die Glaubwürdigkeit wissenschaftlicher Ergebnisse in der Bevölkerung schwächen,
- wissen Sie mehr über systematische Fehler (Bias), und wie diese in klinischen Studien vermieden werden können.

Einleitung

Technologischer Fortschritt und Informationsexplosion

Praxis, Forschung und Lehre in Unfallchirurgie und Orthopädie haben sich seit Beginn des neuen Jahrtausends grundlegend gewandelt. Auf der Technologie-, Struktur- und Prozessebene haben u. a. biologische und augmentierte Osteosyntheseverfahren, minimal-invasive Zugangswege, Navigation und Robotik [2, 3, 4], individualisierte Interventionen und Implantate [5], „Fast-track“-Konzepte [6], evidenzbasierte, partizipative Entscheidungsfindung [7], „value-based medicine“ [8] und Maßnahmen zur Patientensicherheit [9] Einzug in den klinischen Alltag gehalten.

Bei geschätzt etwa 3 Mio. Veröffentlichungen in der Biomedizin pro Jahr ist es nahezu unmöglich, dass alle Ärztinnen und Ärzte, Beschäftigte im Pflege- und Therapiesektor, aber auch politische Entscheider(innen) im Gesundheitswesen jederzeit auf dem aktuellen Stand des Wissens sind. Evidenz- und konsensbasierte Leitlinien der Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF), Cochrane-Reviews sowie nicht zuletzt klinische Entscheidungsunterstützungssysteme („clinical decision support systems“, CDSS) bündeln und gewichten das aktuelle Wissen. Sie können die Therapieplanung begründen und auch zu einer vertrauensvollen Kommunikation zwischen professionellen Teams, Patient(inn)en und ihren Angehörigen beitragen.

Zur Beantwortung konkreter medizinisch-wissenschaftlicher Fragen und Probleme bedarf es unverändert der Suche nach und der Interpretation von aktuellen Forschungsergebnissen in individuellen Publikationen.

Verpflichtende Vorgaben zu Planung, Durchführung und Berichterstattung klinischer Studien und Prüfungen umfassen u. a. die

- International Conference on Harmonisation – Good Clinical Practice (ICH-GCP) Guideline,
- Regelwerke
 - der US-amerikanischen Food and Drug Administration (FDA),
 - der European Medicines Agency (EMA),
 - des Bundesinstituts für Arzneimittel und Medizinprodukte (BfArM),
 - des Paul-Ehrlich-Instituts (PEI) sowie

Abstract

Clinical studies in trauma surgery and orthopedics: read, interpret and implement

Informative, participatory clinical decision-making needs to combine both skills and expertise as well as current scientific evidence. The flood of digital information makes it difficult in everyday clinical practice to keep up to date with the latest publications. This article provides assistance for coping with this problem. A basic understanding of prior and posterior probabilities as well as systematic error (bias) makes it easier to weigh up the benefits and risks, e.g. of a (surgical) intervention compared to a nonsurgical treatment. Randomized controlled trials (RCT, with all modern modifications) deliver undistorted results but in orthopedic and trauma surgery can lead to a heavily selected nonrepresentative sample and the results must be confirmed or refuted by further, independent RCTs. Large-scale observational data (e.g. from registries) can be modelled in a quasi-experimental manner and accompany RCTs in health technology assessment.

Keywords

Randomized controlled trial · Probability · Confidence intervals · Bias · Scientific misconduct

- Empfehlungen
 - des Committee on Publication Ethics (COPE),
 - des International Committee of Medical Journal Editors (ICMJE) und
 - des Enhancing the Quality and Transparency of Health Research (EQUATOR) Network.

Propädeutische Serien in Zeitschriften wie *Journal of the American Medical Association (JAMA, [10])* und *Deutsches Ärzteblatt [11]*, semiquantitative Instrumente zur Bestimmung des Risikos für systematische Fehler wie das Cochrane Risk-of-Bias Tool Version 2.0 (RoB-2) und Grading of Recommendations Assessment, Development and Evaluation (GRADE) können Kliniker(innen) und Methodiker(innen) unterstützen, um die wissenschaftliche Informationsflut zu strukturieren.

► Merke

Eine wissenschaftlich begründete klinische Entscheidungsfindung beinhaltet die gemeinsame Abstimmung zwischen Behandlern und Betroffenen und vereint ärztliche Expertise, wissenschaftliche Daten und Präferenz.

Systematische Fehler – deren Quellen und Vermeidung

Wenn Sie aus einer Papierakte den Wert 3,41 in eine elektronische Datei übertragen und 3,14 eingeben, dürfte dies Ihrer Unaufmerksamkeit geschuldet sein. Dies entspricht (mit sprachlicher Unschärfe) dem englischen Fehlerbegriff des „mistake“ oder „error“, eher aber einem zufälligen Ereignis. Wenn Sie in 10/10.000 Fällen 4,13, 1,43, 3,14 oder 1,34 statt 3,41 eintragen, dürfte dies immer noch dem Zufall entsprechen.

Würde in einer Datei hingegen regelhaft 3,14 statt 3,41 erscheinen, stimmt irgendetwas nicht.

Müssten Sie eine Längenmessung mithilfe eines Zollstocks vornehmen, könnte es sein, dass dieser durch einen Produktionsfehler

In der Rettungsstelle stellt sich die resolute 69-jährige Margaret M. vor, die in Nordengland lebt und gerade ihre Tochter und Enkel in Deutschland besucht. Im Haus der Familie ist sie auf dem glatten Küchenboden ausgerutscht und auf die rechte ausgestreckte Hand gefallen. Röntgenaufnahmen zeigen eine intraartikuläre distale Radiusfraktur vom Typ AO/OTA 23C2. Das betroffene Handgelenk ist mäßig geschwollen. Margaret ist schlank, Rechtshänderin, Nichtraucherin, fährt Fahrrad, malt und bestellt einen großen Garten. Bis zu ihrer Pensionierung hat sie als Allgemeinmedizinerin („general practitioner“, GP) in Großbritannien praktiziert. Sie hat vor, noch wenigstens 3 Wochen in Deutschland bei ihrer Familie zu verbringen. Sie möchte konkret wissen, ob man ihre Fraktur operativ stabilisieren muss, ob es Alternativen gibt, und was die beste wissenschaftliche Evidenz vorschlägt. Sie möchte möglichst rasch wieder nach Hause, um die Zeit mit ihren Angehörigen vollends genießen zu können – aber natürlich auch keine Langzeitschäden davontragen.

0,27 Längeneinheiten zu kurz ist. Sie würden in diesem Fall die Distanz immer als 0,27 Längeneinheiten zu hoch notieren.

► Cave

- Systematische Fehler resultieren entweder aus einer kontinuierlichen Fehleinschätzung des/der Messenden oder einer unzureichenden Kalibrierung des Messinstruments.
- Um die Bedeutung einer wissenschaftlichen Veröffentlichung für die klinische Praxis beurteilen zu können, müssen Sie das Risiko für systematische Fehler (Bias) kennen.

Weder vor **zufälligen Fehlern** (Mistake/error) noch **bewusster Datenmanipulation** und **Forschungsbetrug** („fraud“) sind wir gefeit – hierzu später. Sie können Studienergebnissen am meisten vertrauen, wenn die *Wahrscheinlichkeit einer Verzerrung durch systematische Fehler gering ist*.

Wenn Sie einem wissenschaftlichen Manuskript auf den Grund gehen wollen, sollten Sie sich mit *zwei zentralen Fragen* auseinandersetzen:

1. Sind die Ergebnisse der Studie *valid*? Stimmt es, was die Autor(inn)en behaupten? Hierzu gehört u. a. die Frage, ob die berichteten Resultate mithilfe der gewählten Methoden generiert werden konnten, und ob sie mit den Zielstellungen bzw. Hypothesen vereinbar sind.
2. Sind die Ergebnisse der Studie *wichtig*, im Hinblick auf den medizinischen Nutzen für Patient(inn)en und die Gesellschaft? Sind die Effektstärken der untersuchten Zielparame-ter so groß (oder eben auch so minimal), dass sie klinische Entscheidungen in die eine oder andere Richtung beeinflussen können oder sollten?

Ein Studienergebnis kann wissenschaftlich bedeutsam sein, weil es ein neues Prinzip vorschlägt oder unterstützt („proof of principle“) oder dieses widerlegt. Es kann den Weg für weitere Untersuchungen bahnen und erleichtern. Diese Situation ist weitaus häufiger anzutreffen, als dass eine neue Studie unmittelbar eine Abkehr von bzw. eine Änderung bisheriger Behandlungsstandards („standard of care“) nach sich zieht. Sprunginnovationen wie etwa die Entdeckung bzw. Erfindung des Penizillins, der Röntgendiagnostik oder der Marknagelosteosynthese sind selten [12]. Wir leben von und

mit Schrittnnovationen – so wird ein neues Platten-, Nagel- oder Endoprothesendesign bei ansonsten unveränderten biologischen Rahmenbedingungen funktionelle und andere patientenzentrierte Endpunkte voraussichtlich nicht beeinflussen können.

► Merke

- **Erwarten Sie in der orthopädisch-unfallchirurgischen Literatur geringe Effektstärken.**
- **Je größer die berichteten Ergebnisunterschiede zwischen Behandlungsgruppen sind, umso intensiver sollten Sie sich mit methodischen Details auseinandersetzen, die die beobachteten Effekte beeinflusst haben könnten.**
- **Systematische Fehler führen eher zu einer Über- als Unterschätzung von Therapieeffekten.**

Redlichkeit in der Wissenschaft

Der 19 Leitlinien umfassende *Kodex zur Sicherung guter wissenschaftlicher Praxis der Deutschen Forschungsgemeinschaft* stellt das für alle Wissenschaftlerinnen und Wissenschaftler **verbindliche Regelwerk** für die klinische Forschung in Deutschland dar [13]. Das Dokument betont *redliches Denken und Handeln, Integrität und die Selbstverpflichtung aller Forschenden als Grundlage vertrauenswürdiger Wissenschaft* – und damit deren Akzeptanz in der Bevölkerung. Die durch die „coronavirus disease 2019“ (COVID-19) ausgelöste Pandemie hat eindrucksvoll gezeigt, wie stark wissenschaftliche Informationen gesellschaftlich einschneidende Maßnahmen bestimmen und auch polarisieren können.

Spektakuläre Fälle von Forschungsbetrug und Datenfabrikation in der Medizin sind selten, führten in der Vergangenheit aber zu spektakulären Konsequenzen wie z. B. den Rücktritt von Vertretern des Nobelpreis-Komitees am schwedischen Karolinska-Institut. Einen Überblick über zurückgezogene Publikationen und wissenschaftliche Hintergründe liefert z. B. die Webplattform *Retraction Watch* [14]. Datenfabrikationen lassen sich leicht aufdecken – Menschen sind schlechte Zufallsgeneratoren [15]. Wenn z. B. Mittelwerte aus kategoriellen Daten abgeleitet werden, zeigt das verblüffende Prinzip der Granularity-Related Inconsistency of Means (GRIM), dass es schlichtweg unmögliche Nachkommastellen gibt [16]. Fragen Sie z. B. 7 Proband(inn)en nach der Anzahl ihrer Kinder, können Mittelwerte wie 1,51 oder 2,02 nicht auftreten.

Die unabhängige, verblindete Begutachtung wissenschaftlicher Manuskripte („peer review“) durch internationale Expert(inn)en sichert Qualität. Die Vorabveröffentlichung von Beiträgen auf Preprint-Servern wie *medRxiv* erwies sich im Zuge der COVID-19-Pandemie als Fluch und Segen zugleich [17]. Einerseits konnten so Forschungsergebnisse frühzeitig der wissenschaftlichen Gemeinschaft barrierefrei zur Verfügung gestellt werden. Andererseits drangen auch ungesicherte Informationen zu den Medien und in die sozialen Netzwerke.

Neue Informationen entkräften oder bestärken unsere Vorannahmen

Nehmen Sie an, dass Sie nachts aufwachen und vor Ihrem Fenster ein blaues, blinkendes Licht wahrnehmen. Je nach Ihrer **individu-**

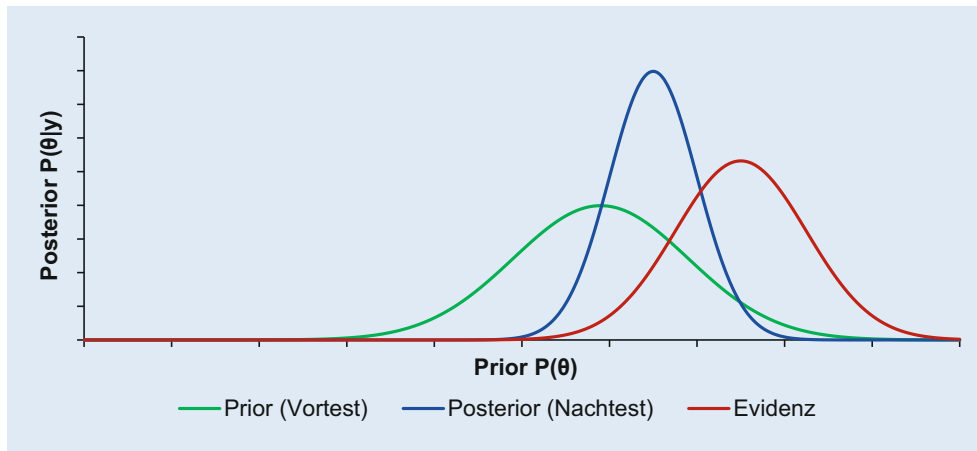


Abb. 1 ◀ Verschiebung der Vortestinformation („prior“) durch neue Evidenz (z. B. eine randomisierte Studie) hin zu einer Nachtest („Posterior“-)Information

ellen Erfahrung und Faktenkenntnis (X) würden Sie die Interpretation (D), dass das Licht durch ein außerirdisches Raumschiff hervorgerufen wird, einem bestimmten Wahrheitsgehalt zuordnen. Sie werden wahrscheinlich zuerst an einen Polizei- oder Feuerwehreinsatz und erst danach an ein unidentifiziertes Luftraumphänomen als Lichtquelle denken. Die statistische Notation wäre $P(D|X)$, entsprechend der konditionalen Wahrscheinlichkeit P für das Zutreffen eines Ereignisses, einer Beobachtung oder eines bestimmten Sachverhaltes D in Abhängigkeit von Ihrer Wahrnehmung, Ihrem Hintergrundwissen oder der bisherigen verfügbaren Evidenz X [18].

Analog zur Blinklichtsituation richtet sich auch Ihr Alltag nach der **Vorwahrscheinlichkeit** („prior probability“) über die Wirksamkeit einer medizinischen oder operativen Intervention für eine bestimmte Erkrankung $P(\theta)$. Diese ergibt sich aus der *Summe aus individuellem Wissen und bisheriger wissenschaftlicher Evidenz*. Nun tritt eine neue Studie hinzu und beeinflusst die sog. Likelihood-Funktion $P(y|\theta)$. Die **Posteriorwahrscheinlichkeit** ($P(\theta|y)$) resultiert aus der Vortestwahrscheinlichkeit (Prior) und verschobenen zusätzlichen Informationen; **Abb. 1**).

► Merke

Einfach ausgedrückt, verschiebt die Information aus einer Studie Ihr eigenes, aber auch das Vorwissen der klinisch-wissenschaftlichen Gemeinschaft über die Wirksamkeit einer Maßnahme in die eine oder andere Richtung.

Diese Form des Erkenntnisgewinns wurde durch den presbyterianischen Pfarrer, Mathematiker, Statistiker und Philosophen **Thomas Bayes** (1701–1761) vorgeschlagen und entwickelt. Wenn die Ergebnisse außerordentlich und unerwartet (im Sinne sehr starker Therapieeffekte) sind, wird sich die Posterior probability stärker verschieben als bei einem „Wir-nehmen-es-zur-Kennntnis“-Resultat. Der Extremfall würde eintreten, wenn eine bisher nichtbehandelbare oder fatale Erkrankung bzw. Verletzungsfolge (z. B. ein traumatischer Hirnschaden oder eine Querschnittlähmung) durch eine neuartige Behandlung plötzlich therapierbar würde.

In diesem Zusammenhang sollten Sie auch das von Sir Karl Raimund Popper (1902–1994) geprägte **Falsifikationsprinzip** kennen, nach dem keine Hypothese bewiesen werden kann, weil

wir nicht unendlich viele Informationen über die Vergangenheit, Gegenwart und Zukunft besitzen. Wir können Hypothesen widerlegen und uns mit einer alternativen Erklärung anfreunden. Letztere müssen wiederum durch weitere Untersuchungen belegt werden.

Was unsere klinische Entscheidungsfindung beeinflusst

Ein einzelnes noch so beeindruckendes Studienergebnis macht (wie die berühmte Schwalbe) noch keinen Sommer

Sie treffen als Ärztin oder Arzt jeden Tag Entscheidungen auf der Basis Ihrer Intuition. Dies bezeichnet man auch als **Heuristik** (Bauchentscheidung, [19]). In den chirurgischen Fächern führt die Konfrontation mit einer Akutsituation oder einem intraoperativen Situs zu einer oftmals nichterklärbaren (manuellen) Handlung [20]. Je größer die Erfahrung und das handwerkliche Geschick, umso eher wird die unbewusste Reaktion selbst auf eine unerwartete Gegebenheit zu einem günstigen Behandlungsergebnis beitragen. Erfahrungswissen unterliegt jedoch der Apoptose und wird oftmals durch neue wissenschaftliche Fakten entkräftet – mit anderen Worten: Die Expertise ist relativ und kurzlebig.

Klinisch-wissenschaftliches Beispiel

In der randomisierten Studie Open reduction and internal fixation versus casting for highly comminuted and intra-articular fractures of the distal radius (ORCHID) wurden vor einigen Jahren das funktionelle Outcome und die Lebensqualität nach der Versorgung einer intraartikulären Fraktur des distalen Radius (AO/OTA-Typ 23C) bei Patient(innen) ≥ 65 Jahren untersucht. Es wurde gezeigt, dass die in Deutschland vorwiegend stationär durchgeführte volare winkelstabile Plattenosteosynthese („open reduction and internal fixation“, ORIF) 12 Monate nach dem Unfallereignis nicht zu einem besseren Outcome als eine geschlossene Reposition und ambulante Ausbehandlung im Gipsverband führt [21]. Aufgrund der Häufigkeit der Verletzung in der alternden Bevölkerung ist es wichtig, Betroffenen und ihren Angehörigen darzulegen, dass

der Verzicht auf eine invasivere Therapie nicht notwendigerweise schlechtere Behandlungsergebnisse im Mittel- und im Langzeitverlauf mit sich bringt. Eine bewusste und gezielte nichtoperative Behandlung und Rehabilitation häufiger Verletzungen können zu ähnlich guten Ergebnissen wie eine operative Therapie führen – beide Modalitäten können sich im klinischen Alltag und in der nationalen Gesundheitsversorgung behaupten, solange sie qualitätsgesichert trainiert und umgesetzt werden.

Die Studie ORCHID musste aufgrund schleppender Rekrutierung vor Erreichung der Zielfallzahl von 252 Teilnehmer(inne)n abgebrochen werden und schloss in der „Intention-to-treat“(ITT)-Analyse (also der Population, in der Teilnehmer so ausgewertet werden, wie sie randomisiert wurden – egal, ob sie tatsächlich die per Zufall zugeloste Therapie erhielten oder nicht) 149 Patient(inn)en ein.

Die Ergebnisse wurden in der deutschen klinisch-wissenschaftlichen Gemeinschaft kontrovers aufgenommen – international lieferten sie u.a. die Grundlage für die unten erörterte Combined randomised and observational study of surgery for fractures in the distal radius in the elderly (CROSSFIRE, [22]). Eine winkelstabile Plattenosteosynthese hatte sich im britischen Distal radius acute fracture fixation (DRAFFT) trial mit 461 Teilnehmenden gegenüber der Kirschner-Draht-Osteosynthese als nichtüberlegen erwiesen [23]. Es lag also nah, die chirurgischen Extremvarianten (d.h., ORIF vs. geschlossene Reposition und Gipsstabilisierung) per Zufallszuweisung zu untersuchen – dies war ethisch gerechtfertigt, da zum Zeitpunkt der Studieninitiierung in Ermangelung wissenschaftlicher Daten **therapeutische Unsicherheit** (Equipoise) vorlag.

► Merke

Auch wenn Sie eine klare klinische Meinung vertreten und sich vielleicht wundern, warum Autor(inn)en eine vermeintlich gelöste Frage in einem aufwendigen Trial adressiert haben – seien Sie offen für die Möglichkeit, dass die beste verfügbare wissenschaftliche Evidenz vielleicht doch nicht so eindeutig ist wie vermutet.

Im *Einleitungsteil* einer wissenschaftlichen Veröffentlichung sollten Sie finden:

1. *eine Motivation, die Erörterung des klinischen und/oder versorgungsrelevanten Problems, unter Darlegung wissenschaftlicher Fakten und Zitation der aktuellen Literatur;*
2. *bei randomisierten Studien den Beleg, dass zu Studienbeginn 2 (oder mehr) zu vergleichende Therapieoptionen gleichwertig erschienen und keine oder nur unzureichende Vergleiche vorlagen, die eine bestimmte Behandlung favorisierten;*
3. *eine beantwortbare Zielstellung und Studienfrage („objectives“).*

Warum ist das Adjektiv „beantwortbar“ so wichtig? Kliniker(inn)en fällt es mitunter schwer, ihre zahlreichen Ideen zu fokussieren und in eine eindeutige wissenschaftliche Fragestellung zu überführen.

Eine beantwortbare Frage setzt sich aus den klassischen **W-Elementen** zusammen. Welches klinische Problem soll in welcher Population mit welcher Erkrankung oder Verletzung studiert werden? Welche Intervention soll untersucht und ggf. mit welchem Therapiestandard verglichen werden? Was sind die interessieren-

den Endpunkte, wann sollen diese erhoben werden, und welche Effektstärke, welcher Unterschied oder welche sonstige Dimension soll mit welchen Methoden bestimmt werden? Das **PICOT-Schema** („patient and problem, intervention, control, outcome, time“) ist hilfreich, um Gedanken zu sortieren und so zu gliedern, dass sie in einer Frage mit dem folgenden Muster münden:

„Führt die Behandlung A im Vergleich zum klinischen Standard oder einer Behandlung B bei einer Anzahl von C Patient(inn)en im Alter von D Jahren mit der Erkrankung/Verletzung E zu einem um F Einheiten besseren Outcome G?“

Eine klinische, insbesondere eine randomisierte Studie erfordert häufig eine **Abstraktion** sowohl bei der Definition der interessierenden Erkrankung bzw. Verletzung, den Ein- und Ausschlusskriterien, der Intervention, aber auch den Endpunkten.

► Cave

Je valider eine Studie für die nationale Gesundheitsversorgung ist, umso weniger ist sie oftmals geeignet, um fachspezifische Fragen mit hoher Auflösung und Detailschärfe zu beantworten.

Im ORCHID-Szenario bedeutete dies, dass zwar eine **generelle Aussage** über die Wirksamkeit und den Nutzen biologisch sehr unterschiedlicher Behandlungsprinzipien (und deren Effektstärken) getroffen werden konnte, diese Informationen aber nicht so tief reichten, um beispielsweise zu entscheiden, welches Implantat welches Herstellers für eine spezielle Frakturmorphologie ausgewählt werden sollte.

Im März 2021 wurden die Ergebnisse der australischen CROSSFIRE („Combined Randomised and Observational Study of Surgery for Fractures in the Distal Radius in the Elderly“-Studie veröffentlicht, die die exakt gleichen Fragen wie ORCHID untersuchte [22]. Der wesentliche Unterschied war, dass in CROSSFIRE lediglich ein Drittel der Patient(inn)en intraartikuläre Frakturen aufwies. Nichtrandomisierte Teilnehmende wurden in einer Beobachtungskohorte nachuntersucht („comprehensive cohort design“, [24]). Die randomisierte Stichprobe war mit 166 Teilnehmenden mit ORCHID vergleichbar und zeigte auch ein ähnliches Basisprofil.

Die in ORCHID beobachteten funktionellen und Lebensqualitätsmessungen nach 12 Monaten wurden durch CROSSFIRE nahezu **punktgenau bestätigt (Abb. 2)**.

Nach der derzeitigen besten wissenschaftlichen Evidenz gibt es keine Unterschiede in funktionellen und Lebensqualitätsindikatoren 3 und 12 Monate nach einer volaren Plattenosteosynthese und einer Gipsbehandlung bei distalen Radiusfrakturen bei Patient(inn)en ≥ 65 Jahren.

Sie dürfen Margaret M. unter Abwägung verschiedener Faktoren (der Prämisse der Wiederherstellung ihrer physischen Aktivität, ihrem Wunsch nach ambulanter Behandlung) und Verweis auf die Ergebnisse aus ORCHID und CROSSFIRE ein gutes funktionelles Ergebnis nach primärer nichtoperativer Ausbehandlung ihrer distalen Radiusfraktur in Aussicht stellen.

Natürlich haben Sie es nicht jeden Tag mit älteren Patient(inn)en mit der beschriebenen Verletzung zu tun, die körperlich aktiv sind, eine hohe Gesundheitskompetenz („health literacy“) aufweisen, sich der Wichtigkeit kontrollierter klinischer Studien bewusst sind oder danach fragen. Als Kliniker(in) werden Sie einwenden, dass

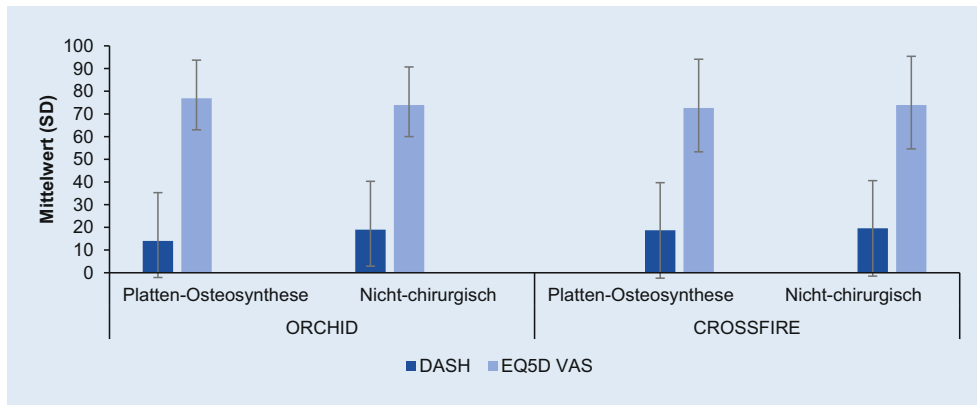


Abb. 2 ▲ Vergleich der Mittelwerte im Fragebogen Disability of the Arm, Shoulder and Hand (*DASH*) und der visuellen Analogskala (*VAS*) des EuroQol-5D Fragebogen zur gesundheitsbezogenen Lebensqualität (*EQ5D*) 12 Monate nach Einschluss in die deutsche Studie Open reduction and internal fixation versus casting for highly comminuted and intra-articular fractures of the distal radius (*ORCHID*, [21]) und die australische Combined randomised and observational study of surgery for fractures in the distal radius in the elderly (*CROSSFIRE*, [22])

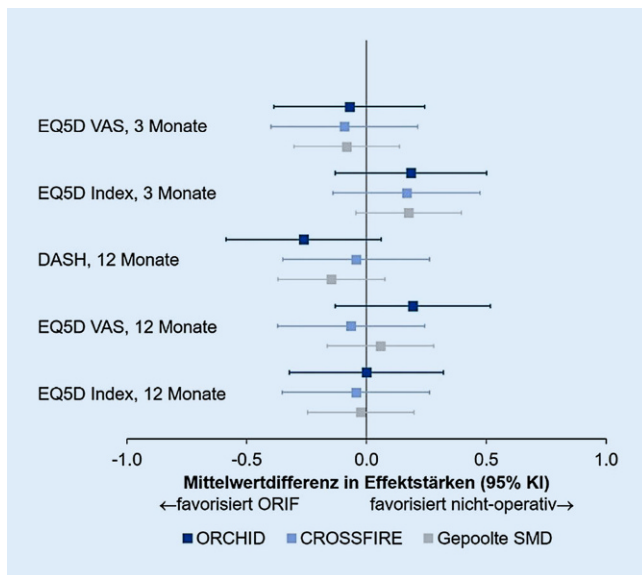


Abb. 3 ▲ Vergleich der Effektstärken (standardisierte Mittelwertdifferenz [*SMD*] = Mittelwertdifferenz/Standardabweichung) im Disability of the Arm, Shoulder and Hand (*DASH*) Instrument und EuroQol-5D Fragebogen zur gesundheitsbezogenen Lebensqualität (*EQ5D*) zwischen den Studien Open reduction and internal fixation versus casting for highly comminuted and intra-articular fractures of the distal radius (*ORCHID*, [21]) und Combined randomised and observational study of surgery for fractures in the distal radius in the elderly (*CROSSFIRE*, [22]). Die Ergebnisse der Einzelstudien wurden mithilfe der Meta-Prozedur in STATA V16.1 und einem „Fixed-effects“-Modell gepoolt. Die I^2 -Statistik zur Prüfung ergab mit Ausnahme der EQ5D-VAS-Erhebung nach 12 Monaten (21,5%) Werte von 0%. *ORIF*, „open reduction and internal fixation“, 95%-KI 95%-Konfidenzintervall

die spezifische Fraktur gut reponibel sein muss, um die besten Voraussetzungen für eine nichtoperative Ausbehandlung zu ermöglichen. Auch würden Sie Margaret M. natürlich empfehlen, sich nach ihrer Rückkehr in das Vereinigte Königreich in ärztliche Weiterbehandlung zu geben – aber im konkreten Szenario helfen Ihnen die Daten aus den oben genannten Studien, wissenschaftlich partizipativ zu entscheiden.

Score-basierte Endpunkte verlangen eine Gewichtung verschiedener Items und Dimensionen – eine Einschätzung von Betroffenen, ob die Therapie aus ihrer Sicht erfolgreich war, Symptome linderte oder die Funktion wiederherstellte, könnte den Nutzen einer Intervention evtl. besser beschreiben. Validierte Scores wie *DASH*, Western Ontario and McMaster Universities Osteoarthritis Index (*WOMAC*), International Knee Documentation Committee (*IKDC*) Score, Lysholm Score, Harris Hip Score u.v.a. erlauben jedoch den interkulturellen und internationalen Vergleich der Ergebnisse einzelner Studien.

Wenn Sie Mittelwerte und Mittelwertdifferenzen zwischen verschiedenen Untersuchungen aggregieren und gleichzeitig Aussagen über die klinische Relevanz von Beobachtungen treffen wollen, bietet es sich an, **Effektstärken** zu berechnen [25]. Die Effektstärke ist in etwa die *Mittelwertdifferenz in einem Score oder einer sonstigen stetigen Messgröße, geteilt durch die gemeinsame Standardabweichung* (in erster Näherung auch durch die Standardabweichung in der Kontrollgruppe). Effektstärken um 0,2 gelten als klinisch wenig relevant, um 0,5 als moderat, ab 0,8 als stark [26].

Die Ergebnisse von *ORCHID* und *CROSSFIRE* im Forest-Plot-Format illustriert **Abb. 3**. Hierbei wiederum gilt – je näher die Punktschätzer (illustriert durch Quadrate, Kreise oder Rauten) am Wert Null (bei Mittelwert- oder absoluten Risikodifferenzen) bzw. dem Wert Eins (bei relativen Risikomaßen wie „risk ratio“ bzw. relativem Risiko, „odds ratio“, „hazard ratio“ usw.) liegen, umso geringer der Therapieeffekt. Das **95%-Konfidenzintervall** (95%-KI, horizontale Fehlerindikatoren) ist wie folgt zu interpretieren: *Würden Sie die Studie 100-mal wiederholen, würde das wahre Ergebnis in 95 von 100 Fällen innerhalb des 95%-KI liegen.*

Es gibt statistische Zusammenhänge zwischen dem Konfidenzintervall, dem vor Beginn einer Studie festzulegenden **Fehler 1. Art α** (das für die individuelle Studie *akzeptable Risiko* für per Zufall auftretende Ergebnisse) und dem *nach Abschluss der Studie* zu ermittelnden **p-Wert** (dem Gradmesser der Wahrscheinlichkeit, dass die erhobenen Daten tatsächlich mit dem Zufall vereinbar sind).

Tab. 1 Verschiedene Szenarien einer Gleich- bzw. Ungleichverteilung typischer demografischer Variablen mit potenziellem Einfluss auf Behandlungsergebnisse. Überlegen Sie sich, ob im Fall einer höheren Infektionsrate in der Plattengruppe im Szenario 3 diese nicht bereits durch die Dominanz von Ausgangsrisiken hätte bedingt werden können.

Szenario 1: ideale Randomisierung			Szenario 2: zufällige Ungleichverteilung von Risikofaktoren in beiden Interventionsgruppen			Szenario 3: überzufällige Verteilung von Risikofaktoren in eine Interventionsgruppe		
Variable	Platte	Nagel	Variable	Platte	Nagel	Variable	Platte	Nagel
<i>n</i>	51	52	<i>n</i>	51	52	<i>n</i>	51	52
Geschlecht			Geschlecht			Geschlecht		
Weiblich	21 (41 %)	20 (38 %)	Weiblich	15 (29 %)	22 (42 %)	Weiblich	15 (29 %)	22 (42 %)
Männlich	29 (59 %)	30 (62 %)	Männlich	36 (71 %)	30 (58 %)	Männlich	36 (71 %)	30 (58 %)
Alter (Jahre)	45 (±9)	47 (±10)	Alter (Jahre)	42 (±8)	49 (±10)	Alter (Jahre)	49 (±10)	42 (±8)
BMI (kg/m ²)	25 (±5)	24 (±6)	BMI (kg/m ²)	29 (±8)	23 (±7)	BMI (kg/m ²)	29 (±8)	23 (±7)
Raucher	10 (20 %)	9 (17 %)	Raucher	5 (10 %)	12 (23 %)	Raucher	12 (24 %)	5 (10 %)
Diabetes	4 (8 %)	7 (15 %)	Diabetes	10 (20 %)	3 (6 %)	Diabetes	10 (20 %)	3 (6 %)

Der Statistiker, Genetiker und Evolutionstheoretiker **Sir Ronald Aylmer Fisher** (1890–1962) schlug vor, dass *ein Versuchsaufbau, der in weniger als einem von 20 Fällen (also 5 %) falsch-positive Ergebnisse liefert*, es verdient, in weiteren Experimenten bestätigt oder widerlegt zu werden [27]. Dem vorgeschlagenen Grenzwert liegen eine Normalverteilung und die Regel der 2 Standardabweichungen für die statistische Signifikanz zugrunde.

► Merke

Derzeit gibt es eine methodische Debatte darüber, ob der „klassische“ Grenzwert von 0,05 im Hinblick auf die wissenschaftliche Informations- und Manuskriptflut zur Vermeidung falsch-positiver, durch Zufall entstandener Ergebnisse auf 0,005 herabgesenkt werden sollte [28, 29].

Nehmen Sie an dieser Stelle mit, dass

1. der Begriff „statistische Signifikanz“ beschreibt, ob ein Ergebnis (noch) mit dem Zufall vereinbar ist;
2. in einem guten Manuskript mit ausführlichem Statistikeil begründet werden sollte, wie der Fehler 1. Art (α) festgelegt wurde;
3. in einem guten Manuskript auch die Wahl des p -Wertes begründet werden sollte.

► Cave

Wird lediglich eine Floskel wie „The p value was set at $<0,05$ “ in ihren verschiedenen „Copy-and-paste“-Spielarten ohne nähere Ausführung genannt, sollte das zugrunde liegende methodische Rahmenwerk kritisch hinterfragt werden.

Die Effektstärken im EQ5D-VAS und -Index nach 3 Monaten sowie dem EQ5D-Index nach 12 Monaten waren in ORCHID und CROSSFIRE nahezu identisch (Abb. 3). Strikte Rahmenbedingungen führten zu einer strengen Patientenselektion und damit einer Gesamtstichprobe von 320 Teilnehmenden. Warum sollten Sie als erfahrene(r) Unfallchirurg(in) oder Orthopäde/Orthopädin den beschriebenen Studien dennoch vertrauen und Aufmerksamkeit schenken?

Die Antwort lautet: Weil sie so frei von systematischen Fehlern sind, wie methodisch machbar.

► Merke

Das Prinzip hinter der Randomisierung ist, dass durch Zufall bekannte (wie z. B. Alter, Geschlecht, Body-Mass-Index, Grunderkrankungen etc.) und damit auch unbekannte oder routinemäßig nichterfasste Risikofaktoren (z. B. bestimmte genetische Faktoren) gleichmäßig auf die Behandlungsgruppen verteilt werden.

Nach Erfahrungswerten funktioniert die zufällige Balancierung der Ausgangsrisiken ab etwa 100 Studienteilnehmern/Gruppe zuverlässig. Die meisten randomisierten Studien in Orthopädie und Unfallchirurgie weisen eher kleinere Stichprobengrößen (<100 Teilnehmende) auf – Differenzen von 10 % in kategoriellen oder binären Größen (z. B. Geschlechtsverteilung, Häufigkeit intraartikulärer oder offener Frakturen, Anteil von Rauchern oder Menschen mit Diabetes etc.) oder 10 Punkten in stetigen Größen (z. B. Alter, Body-Mass-Index etc.) sind nicht selten. Grundsätzlich ist es kein Problem, wenn derartige Unterschiede einem Zufallsmuster folgen – also z. B. Risikofaktoren in der einen oder anderen Gruppe stärker vertreten sind. Wenn sich jedoch alle Risikofaktoren in einer Therapiegruppe häufen, kann ein systematischer Fehler, der die Ergebnisse verzerren kann, nicht ausgeschlossen werden (Tab. 1).

In nichtrandomisierten bzw. Beobachtungsstudien müssen Unterschiede im Basisprofil mithilfe statistischer Verfahren (Matching, multivariate Regression etc.) korrigiert werden. Ein weit verbreitetes und etabliertes Verfahren ist das sog. **Propensity Score Matching**, das auch als „Pseudorandomisierung“ bezeichnet wird [30, 31]. Hierbei wird z. B. in einem Register zunächst die Wahrscheinlichkeit ermittelt, warum Patient(inn)en mit einem bestimmten Frakturtyp die eine oder andere Behandlung erhielten. Endpunkte werden zwischen denjenigen Patient(inn)en verglichen, die die gleiche Wahrscheinlichkeit hatten, die eine oder andere Behandlung zu erhalten. Das Propensity Score Matching kommt bei der Gleichverteilung von Ausgangsrisiken nahe an das Ideal der Randomisierung heran, wenngleich nur bekannte Risikofaktoren berücksichtigt werden können und die Stichprobe einer Stichprobe analysiert wird.

Die Methode ist daher v. a. für sehr große Datensätze geeignet [32]. So konnte z. B. in einer kombinierten Analyse von Datensätzen der ADAC Luftrettung und des TraumaRegister DGU® (AUC – Akademie der Unfallchirurgie GmbH, München, Deutschland) mithilfe des Propensity Score Matching die im Clinical randomisation of an

antifibrinolytic in significant haemorrhage 2 trial (CRASH-2, [33]) beobachtete Reduktion der Sterblichkeit Schwerverletzter durch eine präklinische Gabe von Tranexamsäure bestätigt werden [34].

Fazit für die Praxis

- Wir haben Ihnen 1. das Problem der wissenschaftlichen Informationsexplosion beschrieben, 2. Grundprinzipien von Wahrscheinlichkeit und der Bayes-Theorie der Verschiebung von Vor- hin zu Nachtestwahrscheinlichkeiten unter Einbeziehung Ihres Erfahrungswissens erläutert, und 3. illustriert, dass Randomized controlled trials (RCT) die gängige Praxis auf den Prüfstand stellen, falls ihre Ergebnisse in einer unabhängigen Studie reproduziert werden können.
- Behalten Sie beim Lesen medizinischer Veröffentlichungen immer das Risiko für systematische Fehler, die den stärksten Einfluss auf Studienergebnisse nehmen, im Auge.
- Die Randomisierung ist die derzeit einzige methodische Option, um das Risiko für systematische Fehler sicher zu minimieren.
- Insbesondere bei einer offensichtlichen Ungleichverteilung im demografischen Basisprofil sollten Sie aufmerksam sein – die Daten müssen in diesem Fall durch statistische Verfahren korrigiert werden.

Korrespondenzadresse

Prof. Dr. Dirk Stengel, MSc(Epi)

Forschung – Ressort Medizin, BG Kliniken – Klinikverbund der Gesetzlichen Unfallversicherung gGmbH
Leipziger Pl. 1, 10117 Berlin, Deutschland
Dirk.Stengel@bg-kliniken.de

Einhaltung ethischer Richtlinien

Interessenkonflikt. Gemäß den Richtlinien des Springer Medizin Verlags werden Autoren und Wissenschaftliche Leitung im Rahmen der Manuskripterstellung und Manuskriptfreigabe aufgefordert, eine vollständige Erklärung zu ihren finanziellen und nichtfinanziellen Interessen abzugeben.

Autoren. **D. Stengel:** A. Finanzielle Interessen: Referentenhonorar oder Kostenerstattung als passiver Teilnehmer: Verschiedene nationale und internationale Vorträge zu den Themen moderne Studiendesigns, klinisch-epidemiologische und statistische Methoden, Berichterstattung, Publikations- und Wissenschaftsethik, eingeladen durch akademische Institutionen und Stiftungen (Junge Akademie NOGG/Charité, AFOR, AOCID) und kommerzielle Partner (Zimmer Biomet, Stryker, Johnson & Johnson, Siemens Healthineers). – Bezahler Berater/interner Schulungsreferent/Gehaltsempfänger o. Ä.: Erstellung von „rapid reviews“ für Ethicon und CeramTec. – B. Nichtfinanzielle Interessen: ärztlicher Leiter, Zentrum für Klinische Forschung, BG Klinikum Unfallkrankenhaus Berlin gGmbH, Warener Str. 7, 12683 Berlin (bis Februar 2019), Leiter der Stabsstelle Forschung, Ressort Medizin, BG Kliniken – Klinikverbund der gesetzlichen Unfallversicherung gGmbH, Leipziger Pl. 1, 10117 Berlin (seit Februar 2019). | Rubrikherausgeber *Der Unfallchirurg*, Gutachter BMBF, klinische Studien mit hoher Relevanz für die Patientenversorgung/„systematic reviews“ | Mitgliedschaften: DGU/DGOU e. V., Deutsche Gesellschaft für Chirurgie e. V., Cochrane Injuries Group, Eastern Association for the Surgery of Trauma (EAST), Methods Board, *Journal of Bone & Joint Surgery*, Editorial Board Member *Journal of Orthopaedic Trauma*. **W. Mutschler:** A. Finanzielle Interessen: W. Mutschler gibt an, dass kein finanzieller Interessenkonflikt besteht. – B. Nichtfinanzielle Interessen: Schriftleiter *Der Unfallchirurg* | Mitgliedschaft: Deutsche Gesellschaft für Orthopädie und Unfallchirurgie. **L. Dubs:** A. Finanzielle Interessen: L. Dubs gibt an, dass kein finanzieller Interessenkonflikt besteht. – L. Dubs gibt an, dass kein nichtfinanzieller Interessen-

konflikt besteht. **S. Kirschner:** A. Finanzielle Interessen: Forschungsförderung zur persönlichen Verfügung: Aufwandsentschädigung für die Studienteilnahme an VALUE TKA Universitätsklinikum Dresden. – Bezahler Berater/interner Schulungsreferent/Gehaltsempfänger o. Ä.: Fa. Aesculap Biomechanik Studie zur Zementierung von Tibiakomponenten in Sawbones. – B. Nichtfinanzielle Interessen: angestellter Arzt, ViDiA Kliniken Karlsruhe | Mitgliedschaften: BVOU, Studiengruppe EPRD, Vizepräsident der AE (Deutsche Gesellschaft für Endoprothetik). **T. Renkawitz:** A. Finanzielle Interessen: Forschungsförderung zur persönlichen Verfügung: DePuy, Zimmer, Aesculap, Bundesministerium für Bildung und Forschung, Deutsche Arthro-Hilfe, Otto-Bock-Stiftung, Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung, Stiftung Oskar-Helene-Heim Berlin, Vielberth-Stiftung. – Kostenerstattungen: DePuy, Zimmer, Aesculap, Bundesministerium für Bildung und Forschung, Deutsche Arthro-Hilfe, Otto-Bock-Stiftung, Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung, Stiftung Oskar-Helene-Heim Berlin, Vielberth-Stiftung. – Kostenerstattungen für Schulungen/Vorträge: DePuy, Zimmer, Aesculap, Deutsche Gesellschaft für Endoprothetik (AE), Bayerischer Hausärzteverband. – B. Nichtfinanzielle Interessen: Lehrstuhlinhaber (W3) für orthopädische Chirurgie, Universität Heidelberg, Direktor der Klinik für Orthopädie und Unfallchirurgie, Universitätsklinikum Heidelberg | Mitgliedschaften: Gesamtvorstand der DGOOC, Leiter der Arbeitsgemeinschaft Evidenzbasierte Medizin der DGOU, Herausgeberband von *Der Orthopäde* und *Der Unfallchirurg* (Springer Medizin), „international advisory board“ des *Journal of the American Academy of Orthopaedic Surgeons* (AAOS).

Wissenschaftliche Leitung. Die vollständige Erklärung zum Interessenkonflikt der Wissenschaftlichen Leitung finden Sie am Kurs der zertifizierten Fortbildung auf www.springermedizin.de/cme.

Der Verlag erklärt, dass für die Publikation dieser CME-Fortbildung keine Sponsorengelder an den Verlag fließen.

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.

Literatur

1. Guileyardo JM (2015) Probability and uncertainty in clinical and forensic medicine. *Proc (Bayl Univ Med Cent)* 28(2):247–249
2. Agarwal N, To K, McDonnell S et al (2020) Clinical and radiological outcomes in robotic-assisted total knee arthroplasty: a systematic review and meta-analysis. *J Arthroplasty* 35:3393–3409.e2
3. Bauwens K, Matthes G, Wich M et al (2007) Navigated total knee replacement. A meta-analysis. *J Bone Joint Surg Am* 89:261–269
4. Gao S, Lv Z, Fang H (2018) Robot-assisted and conventional freehand pedicle screw placement: a systematic review and meta-analysis of randomized controlled trials. *eur Spine J* 27:921–930
5. Lee J-A, Koh Y-G, Kang K-T (2020) Biomechanical and clinical effect of patient-specific or customized knee implants: a review. *J Clin Med* 9:1559
6. Wainwright TW, Gill M, McDonald DA et al (2020) Consensus statement for perioperative care in total hip replacement and total knee replacement surgery: Enhanced Recovery After Surgery (ERAS[®]) Society recommendations. *Acta Orthop* 91:3–19
7. Légaré F, Adekpedjou R, Stacey D et al (2018) Interventions for increasing the use of shared decision making by healthcare professionals. *Cochrane Database Syst Rev*. <https://doi.org/10.1002/14651858.CD006732.pub4>
8. Sarkies MN, Francis-Auton E, Long JC et al (2020) Implementing large-system, value-based healthcare initiatives: a realist study protocol for seven natural experiments. *BMJ Open* 10:e44049
9. Panagioti M, Khan K, Keers RN et al (2019) Prevalence, severity, and nature of preventable patient harm across medical care settings: systematic review and meta-analysis. *BMJ* 366:l4185
10. Guyatt GH, Rennie D (1993) Users' guides to the medical literature. *JAMA* 270:2096–2097
11. Du Prel JB, Röhrig B, Blettner M (2009) Critical appraisal of scientific articles: part 1 of a series on evaluation of scientific publications. *Dtsch Arztebl Int* 106:100–105
12. Stengel D (2012) The changing landscape of product development and randomized trials. *J Bone Joint Surg Am* 94(Suppl 1):85–91
13. Deutsche Forschungsgemeinschaft (2019) Leitlinien zur Sicherung guter wissenschaftlicher Praxis. Kodex. Deutsche Forschungsgemeinschaft, Bonn
14. Retraction watch. <https://retractionwatch.com/>. Zugegriffen: 17.09.2021
15. Buyse M, George SL, Evans S et al (1999) The role of biostatistics in the prevention, detection and treatment of fraud in clinical trials. *Stat Med* 18:3435–3451

16. Brown NJL, Heathers JAJ (2017) The GRIM test: a simple technique detects numerous anomalies in the reporting of results in psychology. *Soc Psychol Person Sci* 8:363–369
17. Fraser N, Brierley L, Dey G et al (2021) The evolving role of preprints in the dissemination of COVID-19 research and their impact on the science communication landscape. *PLoS Biol* 19:e3000959
18. Kurt W (2019) Bayesian statistics the fun way. Understanding statistics and probability with star wars, LEGO, and rubber ducks. No Starch Press, San Francisco
19. Gigerenzer G, Gaissmaier W (2011) Heuristic decision making. *Annu Rev Psychol* 62:451–482
20. Patkin M (2008) Surgical heuristics. *ANZ J Surg* 78:1065–1069
21. Bartl C, Stengel D, Bruckner T et al (2014) The treatment of displaced intra-articular distal radius fractures in elderly patients. *Dtsch Arztebl Int* 111:779–787
22. Lawson A, Naylor JM, Buchbinder R et al (2021) Surgical plating vs closed reduction for fractures in the distal radius in older patients: a randomized clinical trial. *JAMA Surg* 156:229–237
23. Costa ML, Achten J, Parsons NR et al (2014) Percutaneous fixation with Kirschner wires versus volar locking plate fixation in adults with dorsally displaced fracture of distal radius: randomised controlled trial. *BMJ* 349:g4807
24. Porzolt F, Stengel D (2006) Are the results of randomized trials influenced by preference effects? Part II. Why current studies often fail to answer this question. In: Porzolt F, Kaplan RM (Hrsg) *Optimizing health: improving the value of healthcare delivery*. Springer, New York, S 292–297
25. Schäfer T, Schwarz MA (2019) The meaningfulness of effect sizes in psychological research: differences between sub-disciplines and the impact of potential biases. *Front Psychol* 10:813
26. Ferguson CJ (2016) *An effect size primer: A guide for clinicians and researchers*. American Psychological Association, Washington, DC
27. Fisher RA (1925) *Statistical methods for research workers*. Oliver and Boyd, Edinburgh
28. Ioannidis JPA (2018) The Proposal to Lower P Value Thresholds to .005. *JAMA* 319:1429–1430
29. Ioannidis JPA (2019) What have we (not) learnt from millions of scientific papers with P values? *Am Stat* 73:20–25
30. Ali MS, Prieto-Alhambra D, Lopes LC et al (2019) Propensity score methods in health technology assessment: principles, extended applications, and recent advances. *Front Pharmacol* 10:973
31. Kane LT, Fang T, Galetta MS et al (2020) Propensity score matching: a statistical method. *Clin Spine Surg* 33:120–122
32. Pincus D, Ravi B, Wasserstein D et al (2017) Association between wait time and 30-day mortality in adults undergoing hip fracture surgery. *JAMA* 318:1994–2003
33. Shakur H, Roberts I, Bautista R et al (2010) Effects of tranexamic acid on death, vascular occlusive events, and blood transfusion in trauma patients with significant haemorrhage (CRASH-2): a randomised, placebo-controlled trial. *Lancet* 376:23–32
34. Wafaisade A, Lefering R, Bouillon B et al (2016) Prehospital administration of tranexamic acid in trauma patients. *Crit Care* 20:143



Klinische Studien in Unfallchirurgie und Orthopädie: lesen, interpretieren und umsetzen

Zu den Kursen dieser Zeitschrift: Scannen Sie den QR-Code oder gehen Sie auf www.springermedizin.de/kurse-der-unfallchirurg

? Welche Elemente sollte eine wissenschaftlich begründete klinische Entscheidungsfindung beinhalten?

- Nur Empfehlungen des Klinikdirektors bzw. Chefarztes, um Patient(inn)en und ihre Angehörigen nicht zu verwirren.
- Ausschließlich die Präferenz der Patient(inn)en und ihrer Angehörigen, um die patientenzentrierte Ergebnisqualität zu verbessern.
- Nur Leitlinienempfehlungen, da diese rechtlich bindend sind.
- Eine Abstimmung zwischen Behandlern und Betroffenen, um ärztliche Expertise, wissenschaftliche Daten und Präferenz zu vereinen.
- Nur Daten aus „randomized controlled trials“ (RCT), da diese die Quelle bester biomedizinischer Evidenz darstellen.

? Wie entstehen systematische Fehler (Bias)?

- Sie entstehen als reines Zufallsprodukt.
- Sie resultieren u. a. aus einer Ungleichverteilung im demografischen Basisprofil und einer unzureichenden Kalibrierung des Messinstruments.
- Sie entwickeln sich aus zu langen Untersuchungszeiträumen.
- Systematische Fehler resultieren ausschließlich aus gefälschten Studienergebnissen.
- Sie begründen sich durch fehlende chirurgische Expertise.

? Wie ist die Vorwahrscheinlichkeit („prior probability“) zugunsten eines therapeutischen Verfahrens vollständig definiert?

- Aus der Summe individuellen Wissens und bisheriger wissenschaftlicher Evidenz
- Aus chirurgischer Erfahrung im Klinikalltag
- Aus der Posteriorwahrscheinlichkeit
- Durch das Falsifikationsprinzip von Sir Karl Raimund Popper.
- Durch die Mittelwertdifferenz und die gemeinsame Standardabweichung.

? Welchen Vorteil haben randomisierte Studien („randomized controlled trials“, RCT) in Orthopädie und Unfallchirurgie?

- Randomized controlled trials reduzieren die digitale Informationsflut.
- Randomized controlled trials minimieren das Risiko systematischer Fehler.
- Randomized controlled trials sind die einzige Möglichkeit, um Nachwahrscheinlichkeiten zu berechnen.
- Randomized controlled trials sind immer kostengünstig und gewinnbringend.
- Randomized controlled trials schützen Leser(innen) vor bewusster Datenmanipulation.

? Was versteht man unter dem Begriff der „Effektstärke“?

- Sie ist mit der statistischen Signifikanz von Studienergebnissen gleichzusetzen.
- Sie ist durch einen p -Wert $< 0,05$ definiert.

- Sie stellt einen Weg dar, um aus Mittelwerten und Standardabweichungen die klinische Relevanz von Studienergebnissen abzuleiten.
- Sie stellt die Summe aus Mittelwertdifferenzen dar.
- Sie besteht aus der gemeinsamen Standardabweichung.

? Wie ist das 95 %-Konfidenzintervall (95 %-KI) zu interpretieren?

- Es gibt an, wie häufig ein Ergebnis bei 100-maliger Wiederholung des Experiments in 95 Fällen innerhalb des angegebenen Konfidenzintervalls liegt.
- Es gibt an, in welchem Wertebereich ein Studienergebnis zu 95 % wahr ist.
- Es gibt an, wenn ein zu erwartendes Studienergebnis zu 5 % falsch ist.
- Es gibt an, wann 95 % aller Studien nicht zutreffen.
- Es gibt an, wie sich der p -Wert und die Standardabweichung beeinflussen.

? Mithilfe welches Untersuchungsverfahrens könnte ein Forschungsbetrug aufgedeckt werden?

- „Prior probability“
- Falsifikationsprinzip
- Granularity-related inconsistency of means (GRIM) Test
- Heuristik
- Pseudorandomisierung

Informationen zur zertifizierten Fortbildung

Diese Fortbildung wurde von der Ärztekammer Nordrhein für das „Fortbildungszertifikat der Ärztekammer“ gemäß § 5 ihrer Fortbildungsordnung mit **3 Punkten** (Kategorie D) anerkannt und ist damit auch für andere Ärztekammern anerkennungsfähig.

Anerkennung in Österreich: Für das Diplom-Fortbildungs-Programm (DFP) werden die von deutschen Landesärztekammern anerkannten Fortbildungspunkte aufgrund der Gleichwertigkeit im gleichen Umfang als DFP-Punkte anerkannt (§ 14, Abschnitt 1, Verordnung über ärztliche Fortbildung, Österreichische Ärztekammer (ÖÄK) 2013).

Hinweise zur Teilnahme:

- Die Teilnahme an dem zertifizierten Kurs ist nur online auf www.springermedizin.de/cme möglich.
- Der Teilnahmezeitraum beträgt 12 Monate. Den Teilnahmeschluss finden Sie online beim Kurs.
- Die Fragen und ihre zugehörigen Antwortmöglichkeiten werden online in zufälliger Reihenfolge zusammengestellt.

- Pro Frage ist jeweils nur eine Antwort zutreffend.
- Für eine erfolgreiche Teilnahme müssen 70% der Fragen richtig beantwortet werden.
- Teilnehmen können Abonnenten dieser Fachzeitschrift und e.Med-Abonnenten.
- Abonnenten von „Der Orthopäde“ oder „Der Unfallchirurg“ können kostenlos an CME-Kursen beider Zeitschriften teilnehmen.

? Wofür steht das „PICOT-Schema“?

- „Prior probability, incidence, computation, outcome, time“
- „Preference, intellectual property, compliance, testing“
- „Particular interests, contracts and testimonies“
- „Patients, incidence, co-morbidity, treatment“
- „Patient and problem, intervention, control, outcome, time“

? Welche nationale und internationale Vorgabe spielt bei der Planung, Durchführung und Berichterstattung klinischer Studien und Prüfungen die geringste Rolle?

- Leitlinien zur Sicherung guter wissenschaftlicher Praxis der Deutschen Forschungsgemeinschaft (DFG)

- International Conference on Harmonisation – Good Clinical Practice (ICH-GCP) Guideline
- Enhancing the Quality and Transparency of Health Research (EQUATOR) Network
- Regularien des Bundesinstituts für Arzneimittel und Medizinprodukte (BfArM) bzw. Paul-Ehrlich-Instituts (PEI)
- Verordnung zur Verhütung übertragbarer Krankheiten (IfSG)

? Welche Rolle nimmt das „Propensity Score Matching“ in orthopädisch-unfallchirurgischen Studien ein?

- Propensity Score Matching erlaubt den Vergleich verschiedener Scores wie z. B. Disabilities Of Arm, Shoulder And Hand (DASH) Score, Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC), International Knee Documentation Committee (IKDC) Score oder Harris Hip Score.

- Propensity Score Matching kann zu einer Gleichverteilung bekannter Risikofaktoren führen und damit das Risiko für systematische Fehler reduzieren.
- Propensity Score Matching stellt die einzige Möglichkeit dar, um unbekannte Risikofaktoren gleichmäßig auf Studiengruppen zu verteilen.
- Propensity Score Matching ist lediglich ein anderer Begriff für multivariante Adjustierung.
- Propensity Score Matching ist rein experimentell und sollte in orthopädisch-unfallchirurgischen Studien nicht zum Einsatz kommen.



www.SpringerMedizin.de/cme

Automatische Übermittlung Ihrer CME-Punkte an die Ärztekammer

Die auf www.SpringerMedizin.de/cme erworbenen CME-Punkte können auf Ihren Wunsch hin elektronisch an die Ärztekammer übermittelt werden.

So einfach geht's:

> Einheitliche Fortbildungsnummer (EFN) hinterlegen

Möchten Sie Ihre auf CME.SpringerMedizin.de gesammelten CME-Punkte direkt an Ihre Ärztekammer übermitteln, hinterlegen Sie Ihre EFN bitte bei der Registrierung. Wenn Sie bereits registriert sind, können Sie Ihre EFN jederzeit unter dem Punkt *Meine Daten* nachtragen. Ihre CME-Punkte werden ab sofort automatisch an Ihre Ärztekammer übermittelt.

Weitere Informationen zur elektronischen Punkteübermittlung der Bundesärztekammer finden Sie unter www.eiv-fobi.de

Teilnehmen und weitere Informationen unter: www.SpringerMedizin.de/cme

Unser Tipp: Mit den **e.Med-Kombi-Abos** stehen Ihnen die CME-Kurse der Fachzeitschriften von Springer Medizin in elektronischer Form zur Verfügung. Auf Wunsch erhalten sie mit den e.Med-Kombi-Abos darüber hinaus eine gedruckte Fachzeitschrift Ihrer Wahl.

Testen Sie **e.Med kostenlos und unverbindlich!**

Jetzt informieren unter www.springermedizin.de → „Abo-Shop“ oder telefonisch unter 0800-77 80 777 (Montag bis Freitag, 10 bis 17 Uhr)

