

## ÜBERSICHTSARBEIT

# Randomisierte kontrollierte Studien

Teil 17 der Serie zur Bewertung wissenschaftlicher Publikationen

Maria Kabisch, Christian Ruckes, Monika Seibert-Grafe, Maria Blettner

## ZUSAMMENFASSUNG

**Hintergrund:** Randomisierte kontrollierte Studien (RCTs) sind der Goldstandard in der klinischen Forschung für den Beleg von Wirksamkeit und Sicherheit einer neuen Therapie. RCTs dienen der Untersuchung von patientenrelevanten Forschungsfragestellungen, und in der Arzneimittelentwicklung sind sie die Grundlage für die Zulassungsentscheidungen der Behörden.

**Methoden:** Anhand von Literatur und einem Fallbeispiel wird eine Einführung in die Methodik randomisierter kontrollierter Studien und deren Qualitätsanforderungen gegeben, damit Publikationen zu RCTs besser verstanden und bewertet werden können. Die Ausführungen entsprechen den Inhalten zahlreicher Lehrbücher und ebenso den Erfahrungen der Autoren bei der Planung, Durchführung und Auswertung randomisierter kontrollierter Studien.

**Ergebnisse:** Methodische Aspekte hinsichtlich Fragestellung, Studiendesign, Vermeidung von systematischen Fehlern und Auswertungsstrategien sind besonders bedeutsam für die Qualität einer RCT und deshalb bei der Planung, Durchführung, Auswertung sowie Berichterstattung zu beachten. Darüber hinaus muss ethischen und gesetzlichen Anforderungen entsprochen werden.

**Schlussfolgerung:** Nur bei methodisch korrekter und der Fragestellung angemessener Planung, Durchführung und Auswertung führen randomisierte kontrollierte Studien zu belastbaren Ergebnissen. Deshalb ist die Qualität von RCTs kritisch zu prüfen, bevor die Relevanz der Studienergebnisse für die Patientenversorgung diskutiert wird.

### ► Zitierweise

Kabisch M, Ruckes C, Seibert-Grafe M, Blettner M: Randomized controlled trials: part 17 of a series on evaluation of scientific publications. Dtsch Arztebl Int 2011; 108(39): 663–8. DOI: 10.3238/arztebl.2011.0663

**K**linische Forschung ist die Grundlage für den Fortschritt in der Medizin und die Voraussetzung für evidenzbasierte Medizin. Um die Wirksamkeit und Sicherheit einer Therapie zu untersuchen, sind randomisierte und kontrollierte klinische Studien (RCTs – „randomised controlled trials“) der Goldstandard. Die Überlegenheit einer neuen Therapie im Vergleich zu einer Standardtherapie oder einer Scheintherapie (Placebo) kann in einer RCT nachgewiesen werden. In der klinischen Forschung dienen RCTs der Untersuchung von patientenrelevanten Forschungsfragestellungen und in der Arzneimittelentwicklung sind sie die Grundlage für die Zulassungsentscheidungen der Behörden. Qualitativ-hochwertige RCTs mit geringem Risiko für systematische Fehler (Bias) haben neben Metaanalysen den höchsten Evidenzgrad (1, 2).

Ziel dieses Artikels ist es, eine Einführung in die Methodik randomisierter kontrollierter Studien und deren Qualitätsanforderungen zu geben, damit Publikationen zu RCTs besser verstanden und bewertet werden können. Da RCTs per Definition interventionelle Studien sind, in denen die Intervention häufig in der Verabreichung eines Arzneimittels oder Medizinprodukts besteht, werden auch die ethischen und gesetzlichen Rahmenbedingungen angesprochen.

Die folgenden Ausführungen entsprechen den Inhalten zahlreicher Lehrbücher (3–5) und ebenso den Erfahrungen der Autoren in der Planung, Durchführung und Auswertung von RCTs. Zum besseren Verständnis werden einige methodische Aspekte anhand eines Fallbeispiels, der publizierten ALIFE-Studie (ALIFE, „Anticoagulants for Living Fetuses“), veranschaulicht. Allgemeine Grundlagen der Methodik und Statistik, die auch für randomisierte kontrollierte Studien von Bedeutung sind, sind an anderer Stelle im Rahmen der „Serie zur Bewertung wissenschaftlicher Publikationen“ im Deutschen Ärzteblatt dargelegt worden (6–11).

Im April 2010 veröffentlichte das New England Journal of Medicine die Ergebnisse der ALIFE-Studie (12), die auch im Juli 2010 in der Rubrik „Studien im Fokus“ im Deutschen Ärzteblatt dargestellt wurden (13). Frauen mit zwei oder mehr Fehlgeburten wurden in der Studie entweder mit Aspirin plus Heparin, mit Aspirin allein oder mit Placebo behandelt. Die Zuteilung der Patientinnen zu den drei Behandlungsgruppen erfolgte randomisiert, das heißt nach dem Zufallsprinzip. Das primäre Ziel der Studie war es, die Wirksamkeit der Therapien anhand der Lebendgeburtenrate zu untersuchen.

Interdisziplinäres Zentrum Klinische Studien (IZKS), Universitätsmedizin der Johannes-Gutenberg-Universität Mainz: Dipl.-Biomath. Kabisch, Dipl.-Math. Ruckes, Dr. med. Seibert-Grafe

Institut für Medizinische Biometrie, Epidemiologie und Informatik (IMBEI), Universitätsmedizin der Johannes-Gutenberg-Universität Mainz: Prof. Dr. rer. nat. Blettner

## Fragestellung

Basis einer jeden RCT ist ein Prüfplan (Studienprotokoll), in dem unter anderem der medizinisch-wissenschaftliche Hintergrund, die Nutzen-Risiko-Bewertung, das Studiendesign, die Studienmethodik sowie die gesamte Planung, Durchführung und Auswertung beschrieben sind (14). Die Hauptfragestellung, also das primäre Ziel resultiert aus der medizinisch-wissenschaftlichen Begründung für die Studie.

Für die Beantwortung der Hauptfragestellung muss ein primäres Zielkriterium definiert werden. Das ist eine Mess- oder Beobachtungsgröße, die zu einem definierten Zeitpunkt erhoben wird und von der man annimmt, dass sie den Effekt einer Therapie erfasst. Dies kann auch ein klinischer Endpunkt sein wie zum Beispiel die Lebendgeburtenrate in der ALIFE-Studie.

In einer konfirmatorischen Studie werden – entsprechend der Hauptfragestellung – a priori Hypothesen formuliert. Ist das primäre Ziel der Studie, die Überlegenheit einer neuen Therapie gegenüber einer anderen Therapie oder Placebo zu zeigen, dann wird zunächst hypothetisch angenommen, dass kein Unterschied in der Wirksamkeit der Therapien besteht (Nullhypothese). Im Rahmen der Auswertung wird anhand eines statistischen Tests entschieden, ob die Nullhypothese beibehalten oder ob sie zugunsten der Alternativhypothese verworfen wird. Die Alternativhypothese wird angenommen, wenn ein statistisch signifikanter Unterschied in der Wirksamkeit der Therapien besteht (detaillierte Ausführungen zum Thema statistische Auswertung sind in dem Artikel „Auswahl statistischer Testverfahren“ [15] dieser Serie dargestellt).

Neben der Hauptfragestellung werden Nebenfragestellungen, also sekundäre Ziele untersucht. Mit den sekundären Zielkriterien werden weitere Effekte der Therapie erfasst, also zum Beispiel das Auftreten von unerwünschten Ereignissen oder der Einfluss auf Biomarker. In der ALIFE-Studie zählten die Fehlgeburtenrate, die Frühgeburtenrate und die Rate der Thrombopenien bei der Mutter zu den sekundären Zielkriterien.

Aus statistischer Sicht ist die Unterscheidung zwischen Haupt- und Nebenfragestellung unerlässlich, da sich die Bestimmung der Fallzahl ausschließlich am primären Zielkriterium orientiert (16). Bei der Fallzahlplanung wird die Zahl der Studienteilnehmer berechnet, die notwendig ist, um einen aus klinischer Sicht minimal relevanten Unterschied in der Wirksamkeit der Therapien mit einer hohen Wahrscheinlichkeit als signifikant zu erfassen, wenn dieser besteht. Die Fallzahl ist also wesentlich für die statistische Aussagekraft einer Studie (detaillierte Ausführungen zur Fallzahlplanung finden sich im Artikel „Fallzahlplanung in klinischen Studien“ [17] dieser Serie).

In der ALIFE-Studie bestand die Annahme, dass die Kombination Aspirin plus Heparin im Vergleich zu Aspirin allein oder Placebo zu einem Unterschied von 15 % in der Lebendgeburtenrate führt. Um den positiven Effekt der Kombinationstherapie nachweisen zu können, mussten 360 Frauen in die Studie eingeschlossen werden.

## Studiendesign

In einem randomisierten und kontrollierten Studiendesign (zum Beispiel zweiarmliges Parallelgruppendesign) werden die Effekte einer Therapie (Intervention) und einer Kontrollbehandlung miteinander verglichen und die Patienten durch eine Randomisierung einer der beiden Gruppen zugeteilt. Die Patienten in der Kontrollgruppe werden entweder mit einer anderen Therapie oder mit einem Placebo behandelt. Die ALIFE-Studie ist eine dreiarmlige Parallelgruppenstudie, in der untersucht wurde, ob die Kombinationstherapie oder die Monotherapie im Vergleich zu Placebo die Lebendgeburtenrate verbessert. Der Einsatz von Placebos in klinischen Studien ist ethisch vertretbar, wenn keine Standardtherapie zur Verfügung steht. Sollte aus wissenschaftlich-methodischen Gründen der Vergleich mit Placebo zwingend sein, ist dies unter der Voraussetzung möglich, dass die Patienten keinen Schaden nehmen (18). Das ist zum Beispiel dann der Fall, wenn die Studiendauer kurz ist oder die Schwere der Erkrankung erlaubt, mit einer Therapie zu warten oder sie zu unterbrechen.

Wie für jede Studie am Menschen ist es auch für eine RCT essenziell, dass die Studienpopulation klar definiert ist. Anhand präziser Ein- und Ausschlusskriterien wird die Eignung der Patienten für die Studie sichergestellt, so dass die Studienteilnehmer bezüglich ihrer demografischen Merkmale, ihres Krankheitszustandes sowie gegebenenfalls ihrer Begleiterkrankungen und Begleitmedikation homogen sind.

Um einen möglichst „gerechten“ Vergleich zwischen den Therapien zu gewährleisten, muss die Vergleichbarkeit der Behandlungsgruppen gegeben sein. Diese kann durch Standardisierung, das heißt durch Vereinheitlichung der Studiendurchführung (zum Beispiel Zeitpunkte für die Einnahme der Prüfmedikation, Methoden für die Messung von klinischen Größen), vor allem aber durch Randomisierung der Studienteilnehmer erreicht werden.

## Randomisierung

In einer RCT erfolgt die Zuteilung der Patienten zu den Behandlungsgruppen durch das Zufallsprinzip, die Randomisierung. Dadurch soll sichergestellt werden, dass alle potenziellen Störgrößen in den zu vergleichenden Patientengruppen identische Verteilungen aufweisen (Strukturgleichheit). Störgrößen – oft mit dem Begriff Confounder bezeichnet – sind diejenigen Eigenschaften bei Patienten, die Einfluss auf das Therapieansprechen nehmen können, also zum Beispiel das Gewicht, Alter oder Geschlecht. Nur bei vorliegender Strukturgleichheit kann ein beobachteter Unterschied zwischen den Therapien tatsächlich auf den Einfluss der Behandlung, und nicht auf den Einfluss von Confoundern zurückgeführt werden. Sind diese Confounder bekannt, kann die Strukturgleichheit der Patientengruppen durch eine stratifizierte Randomisierung hergestellt werden (*Kasten*).

In der ALIFE-Studie wurden die Patientinnen in dem Randomisierungsverhältnis 1:1:1 den drei Be-

handlungsgruppen zugeteilt. Es wurde unter Berücksichtigung der prognostischen Faktoren Alter ( $< 36$  Jahre oder  $\geq 36$  Jahre) und Anzahl der erlittenen Fehlgeburten (2 oder  $\geq 3$ ) randomisiert, und da die Studie multizentrisch durchgeführt wurde, auch nach Studienzentrum stratifiziert. Würde die Zuteilung der Patienten zu den Behandlungsgruppen nicht zufällig, sondern durch eine bewusste oder unbewusste Selektion hinsichtlich Eigenschaften erfolgen, die mit der Prognose assoziiert sind, so könnte dies zu einem systematischen Fehler im Therapievergleich führen und in verzerrten Studienergebnissen resultieren (Selektionsbias).

Die Randomisierung muss nicht nur zufällig, sondern darf auch nicht vorhersagbar sein. Die Vorhersagbarkeit wird unterbunden, indem die Behandlung, zu welcher der jeweils nächste Patient zugeteilt wird, dem Studienpersonal nicht bekannt ist. Eine alternierende Behandlungszuteilung ist kein Zufallsprozess.

### Verblindung

Neben der Randomisierung trägt die Verblindung zur Vermeidung von systematischen Fehlern bei. Eine Studie kann doppelblind, einfachblind oder offen sein.

In einer doppelblinden Studie wissen weder Patient noch Prüfarzt, zu welcher Behandlung der Patient randomisiert wurde. Doppelblinde Studien sind von Vorteil, wenn die Kenntnis der zugeteilten Behandlung bewusst oder unbewusst den Studienverlauf und somit das Studienergebnis beeinflussen würde. So ist besonders bei der Erhebung subjektiver Zielkriterien die Verblindung des beurteilenden Prüfarztes sinnvoll. Die Verblindung des Patienten ist zum Beispiel dann wichtig, wenn dessen Einstellung zur Behandlung Einfluss auf seine Zuverlässigkeit bei der Einnahme der Prüfmedikation (Compliance) oder sogar auf das Therapieansprechen haben könnte.

Ist nur der Patient oder nur der Prüfarzt hinsichtlich der zugeteilten Behandlung verblindet, dann heißt eine Studie einfachblind. Liegt keine Verblindung vor, dann spricht man von einer offenen Studie. Um systematische Fehler zu minimieren, sollte der höchstmögliche Verblindungsgrad gewählt werden.

### Auswertungskollektiv

Die statistische Auswertung der Studiendaten einer RCT erfolgt auf bereits im Prüfplan definierten Patientenkollektiven. Das sogenannte Intention-to-Treat-Kollektiv (ITT) ist das primäre Auswertungskollektiv. Es besteht aus allen randomisierten Patienten. Bei der Auswertung nach dem ITT-Prinzip werden die Patienten der Behandlungsgruppe zugeordnet, zu der sie randomisiert wurden. Die Vorteile einer Randomisierung, wie die Strukturgleichheit, können so für die Auswertung beibehalten werden. Da das ITT-Kollektiv alle Patienten umfasst, für die eine Behandlung vorgesehen war, fließen auch Daten von Patienten in die Auswertung ein, die zum Beispiel die Studienbehandlung nicht erhielten, unterbrachen oder vorzeitig beendeten. Somit ist die Auswertungsstrategie

nach dem ITT-Prinzip konservativ, das heißt tendenziell wird der Therapieeffekt unterschätzt (19), unabhängig davon, ob das primäre Zielkriterium eine Verbesserung oder eine Verschlechterung beschreibt. In vielen Studien wird ein modifiziertes ITT-Kollektiv (mITT) definiert, das zum Beispiel alle Patienten umfassen kann, die ein Minimum der Studienbehandlung erhielten.

Eine weitere Auswertungsstrategie ist die Auswertung basierend auf dem Per-Protocol-Kollektiv (PP). Dabei werden Patienten von der Auswertung ausgeschlossen, bei denen die Studiendurchführung vom Prüfplan abgewichen ist. Diese sogenannten Protokollverletzungen können zum Beispiel die Nicht-Einhaltung von Ein- und Ausschlusskriterien oder die nicht korrekte Anwendung der Studienbehandlung sein. Bei der Auswertung nach dem PP-Prinzip werden die Patienten, abhängig von der Behandlung, die sie tatsächlich erhielten, den Behandlungsgruppen zugeordnet. Da das PP-Kollektiv nur diejenigen Patienten umfasst, die die Studie prüfplankonform beendeten, kann es zu einer Verzerrung der Studienergebnisse zugunsten der untersuchten Intervention kommen (19).

Um die Robustheit der Studienergebnisse zu überprüfen, wird die Auswertung nach dem PP-Prinzip als Sensitivitätsanalyse durchgeführt, wenn das ITT-Kollektiv das Patientenkollektiv für die primäre Wirksamkeitsanalyse ist (16). Liefert die Auswertung des primären Zielkriteriums nach dem ITT- und dem PP-Prinzip annähernd die gleichen Ergebnisse, dann können diese als verlässlich angesehen werden. Ist dies nicht der Fall, dann müssen die möglichen Gründe für die Diskrepanz zwischen den Ergebnissen der ITT- und PP-

### KASTEN

#### Stratifizierte Randomisierung

Betrachtet man die Stratifizierungsfaktoren Geschlecht (männlich, weiblich) und Alter ( $< 18$  Jahre,  $\geq 18$  Jahre), und sollen 150 Patienten in dem Verhältnis 1:1 (2 x 75 Patienten) zu den Behandlungsgruppen Verum und Placebo randomisiert werden, dann muss die Randomisierung für jede Untergruppe (Stratum) getrennt erfolgen. Für 2 Stratifizierungsfaktoren mit je 2 Ausprägungen entstehen 4 Strata (männlich und  $< 18$  Jahre, männlich und  $\geq 18$  Jahre, weiblich und  $< 18$  Jahre, weiblich und  $\geq 18$  Jahre).

	Verum	Placebo
männlich und $< 18$ Jahre	10	10
männlich und $\geq 18$ Jahre	16	17
weiblich und $< 18$ Jahre	24	23
weiblich und $\geq 18$ Jahre	25	25
Summe	75	75

**TABELLE 1**

**Ergebnisse der ALIFE-Studie (nach [12])**

	Aspirin plus Heparin	Aspirin allein	Placebo	p-Wert
Intention-to-Treat-Kollektiv n	123	120	121	
Lebendgeburten n (%)	67 (54,5)	61 (50,8)	69 (57,0)	0,63
relatives Risiko (95%-KI)	0,96 (0,76–1,19)	0,89 (0,71–1,13)	1,00	
absolute Differenz der Lebendgeburtenraten (95%-KI) %	-2,6 (-15,0–9,9)	-6,2 (-18,8–6,4)		

Relatives Risiko und absolute Differenz wurden berechnet für den Vergleich zwischen den Behandlungsgruppen Aspirin plus Heparin und Placebo und zwischen Aspirin allein und Placebo. P-Wert gilt für alle Behandlungsgruppenvergleiche. 95%-KI bezeichnet das 95%-Konfidenzintervall.

Auswertung im Ergebnisteil der Publikation diskutiert werden.

Die Daten der ALIFE-Studie, insbesondere das primäre Zielkriterium, wurden auf der Basis des ITT-Kollektivs statistisch ausgewertet. Die Raten der Lebendgeburten in den drei Behandlungsgruppen unterschieden sich nicht signifikant voneinander (*Tabelle 1*). Eine Auswertung nach dem PP-Prinzip bestätigte dieses Ergebnis. Weder die Kombination Aspirin plus Heparin, noch Aspirin allein hatten eine nachweislich bessere Wirkung auf die Rate der Lebendgeburten als Placebo.

### Qualitätsstandards und gesetzliche Anforderungen

Klinische Studien unterliegen nationalen und internationalen wissenschaftlichen und ethischen Anforderungen. 1964 hat der Weltärztebund in der inzwischen mehrfach revidierten Deklaration von Helsinki (20) ethische Grundsätze für die Forschung am Menschen formuliert. Arzneimittel- und Medizinprodukte-Studien unterliegen in Deutschland dem Arzneimittelgesetz (AMG, [BGBl. I S. 2262]) und der GCP-Verordnung (21) einerseits und dem im März 2010 novellierten Medizinproduktegesetz (MPG, [BGBl. I S. 983]) andererseits. Mit der GCP-Verordnung wurde 2004 in Deutschland die Einhaltung von GCP („Good Clinical Practice“/Gute Klinische Praxis) erstmals gesetzlich verankert (21). Für die der GCP-Verordnung zugrundeliegenden europäischen Richtlinien 2001/20/EG und 2005/28/EG stellt die GCP-Guideline ICH-E6 von 1997 die Basis dar (14). Die GCP-Verordnung hat den Schutz der Studienteilnehmer sowie die Sicherung der Datenqualität der Daten zum Ziel.

Im Jahr 2004 vereinbarten die Herausgeber medizinischer Fachzeitschriften als Voraussetzung für die Publikation klinischer Studien, die Registrierung der Studie in einer öffentlich zugänglichen Datenbank (22). Jede Studie am Menschen ist von Ärzten gemäß der ärztlichen Berufsordnung der zuständigen Ethikkommission zur Begutachtung vorzulegen. Bei Arzneimittel- und den meisten Medizinprodukte-Studien ist eine zustimmende Bewertung durch die Ethikkommission sowie eine Genehmigung durch die Bundesoberbehörde (Bundesinstitut für Arzneimittel und Medizinprodukte [BfArM]) oder das Bundesinstitut für Impfstoffe und biomedizinische Arzneimittel (Paul-Ehrlich-Institut [PEI]) notwendig. Neben den Anträgen ist insbesondere der Prüfplan, die Patienteninformation, die Einwilligungserklärung sowie eine Versicherungsbestätigung einzureichen.

Darüber hinaus sind Arzneimittel- und Medizinprodukte-Studien den zuständigen Landesbehörden anzuzeigen. Für Verdachtsfälle unerwarteter schwerwiegender Nebenwirkungen, den vorzeitigen Abbruch einer Studie sowie den Abschlussbericht bestehen gesetzlich definierte Meldeverpflichtungen. Das Bundesdatenschutzgesetz (BDSG, [BGBl. I S. 2814]) und das AMG verpflichten zur Pseudonymisierung von personenbezogenen Daten, die im Rahmen einer klini-

**TABELLE 2**

**Mindestangaben zu einer randomisierten kontrollierten Studie (nach [23])**

Studiendesign	Beschreibung des Studiendesigns (z. B. Parallelgruppenvergleich)
Studienpopulation	Nennung der Ein- und Ausschlusskriterien für Patienten
Behandlungen	präzise Angaben zu den Behandlungen und deren Durchführung in jeder Interventionsgruppe und Kontrollgruppe
Ziele	genaue Formulierung der primären und sekundären Ziele/Fragestellungen
Zielkriterien	klare Definition der primären und sekundären Zielkriterien
Fallzahl	Beschreibung der Fallzahlbestimmung
Randomisierung	Beschreibung der Methode zur Generierung der zufälligen Behandlungszuteilung (z. B. stratifizierte Randomisierung)
Verblindung	Nennung des Verblindungsgrades (z. B. doppelblind)
Auswertungskollektive	Anzahl der in der Auswertung eingeschlossenen Patienten je Behandlungsgruppe und Definition der Auswertungskollektive (z. B. ITT)
Ergebnisse	Darstellung der Ergebnisse aller primären und sekundären Zielkriterien je Behandlungsgruppe
unerwünschte Ereignisse	Angabe aller wichtigen unerwünschten Ereignisse je Behandlungsgruppe
Interpretation	Interpretation der Ergebnisse unter Berücksichtigung der Fragestellung, möglicher Ursachen für systematische Fehler, des aktuellen Forschungsstandes und Publikationen anderer Forscher zur untersuchten Fragestellung
Generalisierbarkeit	Angaben zur Übertragbarkeit der Studienergebnisse auf die Patientenversorgung

schen Studie erhoben, dokumentiert, gespeichert und ausgewertet werden. Das heißt, dass die Angaben zur Identität eines Patienten (Name oder Initialen) durch einen Code ersetzt werden. Die Teilnahme an einer klinischen Studie ist nur zulässig, wenn der Patient vor Studienaufnahme in die Aufzeichnung, Aufbewahrung, Verarbeitung und Weitergabe seiner Daten einwilligt.

## Diskussion

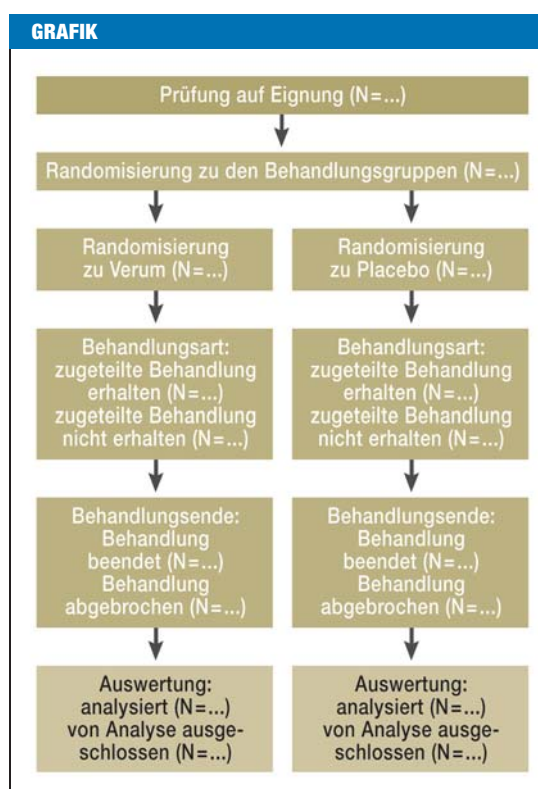
Bei der Veröffentlichung einer RCT sind Studienplanung, -durchführung und -auswertung nachvollziehbar zu beschreiben. Im CONSORT-Statement sind alle Inhalte, die beim Publizieren einer RCT angegeben werden sollten, dargestellt (23). In Anlehnung an CONSORT sind die wichtigsten Angaben, die in einer Publikation enthalten sein müssen, in *Tabelle 2* zusammengefasst. Der Patientenfluss in einer RCT und die Anzahl der Patienten, deren Daten statistisch ausgewertet wurden, können wie in *Grafik* gezeigt veranschaulicht werden.

Vor dem Hintergrund des Studiendesigns, aktueller Literatur und unter Berücksichtigung des Stands der Wissenschaft sind die Studienergebnisse und ihre Interpretation im Studienbericht und in einer Publikation ausführlich zu diskutieren und vorhandene methodische Limitationen sind darzulegen. Die kritische Diskussion trägt entscheidend zur klinischen Bewertung der Ergebnisse bei. In der Publikation der ALIFE-Studie wurden die gewonnenen Ergebnisse mit anderen Ergebnissen randomisierter kontrollierter Studien, in denen ebenfalls die Effekte von Heparin auf die Vermeidung von Fehlgeburten untersucht wurden, verglichen und die widersprüchliche Datenlage diskutiert. Schlussendlich konnte aufgrund der vorliegenden Studiendaten die Anwendung von Antikoagulanzen bei Frauen mit wiederkehrenden Fehlgeburten nicht empfohlen werden.

Obwohl RCTs als Goldstandard für die Evidenz gelten, wird die Generalisierbarkeit, also die Übertragbarkeit der Studienergebnisse auf die breite Patientenversorgung (externe Validität) oft infrage gestellt, da standardisierte und kontrollierte Studienbedingungen unzureichend die klinische Versorgungsrealität reflektieren. Hinzu kommt die Selektion der Studienpopulation, die oft nicht den Patienten in der täglichen Praxis entspricht, die viele Begleiterkrankungen und -medikationen haben. Um die Wirksamkeit und Sicherheit der Therapie an einer größeren und heterogenen Studienpopulation im klinischen Alltag zu überprüfen, werden nach der Zulassung Phase-IV-Studien durchgeführt, die in der Regel RCTs sind. Epidemiologische Studien wie zum Beispiel Kohortenstudien sind besonders geeignet, seltene Nebenwirkungen aufzudecken.

## Resümee

RCTs sind ein ideales Instrument für die Beantwortung der Frage nach einer kausalen Beziehung zwischen Therapie und Wirksamkeit (24). Die aktuellen Diskus-



Darstellung des Patientenflusses in einer randomisierten kontrollierten Studie (nach [23])

sionen sowie das neue Arzneimittelmarktneuordnungsgesetz (AMNOG, [BGBl. I S. 2262]) zur Nutzenbewertung von Arzneimitteln und Medizinprodukten machen deutlich, dass RCTs weiterhin als Standard für den Beleg von Wirksamkeit und Sicherheit bei der Zulassungsentscheidung gelten, dass jedoch darüber hinaus Studien zur Nutzenbewertung einer neuen Therapie im Vergleich zur etablierten als erforderlich angesehen werden.

Das IZKS Mainz wird unterstützt durch die Förderung „Klinische Studienzentren, Förderkennzeichen FK 01KN1103, IZKS Mainz“ des Bundesministeriums für Bildung und Forschung.

## Danksagung

Die Autoren danken Daniel Wachtlin, Interdisziplinäres Zentrum Klinische Studien (IZKS), für die hilfreiche Diskussion.

## Interessenkonflikt

Die Autoren erklären, dass kein Interessenkonflikt besteht.

## Manuskriptdaten

eingereicht: 23. 2. 2011, revidierte Fassung angenommen: 28. 6. 2011

## LITERATUR

1. Harboure R, Miller J: A new system for grading recommendations in evidence based guidelines. *BMJ* 2001; 323: 334–6.
2. Phillips B, Ball C, Sackett D, Badenoch D, Straus S, Haynes B, Dawes M (2001): Oxford Centre for Evidence-based Medicine Levels of Evidence. [www.cebm.net/levels\\_of\\_evidence.asp](http://www.cebm.net/levels_of_evidence.asp)
3. Meinert CL: Clinical Trials: Design, conduct, and analysis. Oxford University Press: New York 1986.
4. Shein-Chung C, Jen-Pei L: Design and analysis of clinical trials: concepts and methodologies. John Wiley & Sons: New Jersey 2004.

# KERNAUSSAGEN

- In der klinischen Forschung sind randomisierte kontrollierte Studien der Goldstandard für den Beleg von Wirksamkeit und Sicherheit einer neuen Therapie.
- Nur bei methodisch korrekter und der Fragestellung angemessener Planung, Durchführung und Auswertung führen randomisierte kontrollierte Studien zu belastbaren Ergebnissen.
- Geeignete Methoden zur Vermeidung von systematischen Fehlern, wie die Randomisierung und Verblindung, können der Verzerrung von Studienergebnissen entgegenwirken.
- Um die Robustheit der Studienergebnisse zu überprüfen, erfolgt die statistische Auswertung der Studiendaten basierend auf a priori definierten Patientenkollektiven.
- Für die Qualität einer randomisierten kontrollierten Studie ist neben den methodischen Standards die protokollgemäße, klinische Durchführung essenziell.

- Schumacher M, Schulgen G: Methodik klinischer Studien: Methodische Grundlagen der Planung, Durchführung und Auswertung. Berlin, Heidelberg: Springer-Verlag 2008.
- Röhrig B, du Prel, Blettner M: Study Design in medical research: Part 2 of a series on evaluation of scientific publications. Dtsch Arztebl Int 2009; 106(11): 184–9
- Röhrig B, du Prel du J, Wachtlin D, Blettner M: Types of study in medical research: part 3 of a series on evaluation of scientific publications. Dtsch Arztebl Int 2009; 106(15): 262–8.
- Spiestersbach A, Röhrig B, du Prel J, Gerhold-Ay A, Blettner M: Descriptive statistics: the specification of statistical measures and their presentation in tables and graphs: part 7 of a series on evaluation of scientific publications. Dtsch Arztebl Int 2009; 106(36): 578–83.
- Sauerbrei W, Blettner M: Interpreting results in 2x2 tables: part 9 of a series on evaluation of scientific publications. Dtsch Arztebl Int 2009; 106(48): 795–800.
- Victor A, Elsässer A, Hommel G, Blettner M: Judging a plethora of p-values: how to contend with the problem of multiple testing: part 10 of a series on evaluation of scientific publications. Dtsch Arztebl Int 2010; 107(4): 50–6.
- Zwiener, I, Blettner M, Hommel G: Survival analysis: part 15 of a series on evaluation of scientific publications. Dtsch Arztebl Int 2011; 108(10): 163–9.
- Kaandorp SP, et al.: Aspirin plus heparin or aspirin alone in women with recurrent miscarriage. NEJM 2010; 362: 1586–96.
- Heinzl S: Können ASS und Heparin wiederholte Fehlgeburten verhindern? Dtsch Arztebl 2010; 107(27): A-1355.
- ICH E6: Guideline for good clinical practice. London UK: International Conference on Harmonization 1996; CPMP/ICH/135/95.
- du Prel J, Röhrig B, Hommel G, Blettner M: Choosing statistical tests: part 12 of a series on evaluation of scientific publications. Dtsch Arztebl Int 2010; 107(19): 343–8.
- ICH E9: Statistical Principles for Clinical Trials. London UK: International Conference on Harmonization 1998; CPMP/ICH/363/96.
- Röhrig B, du Prel JB, Wachtlin D, Kwecien R, Blettner M: Sample size calculation in clinical trials: part 13 of a series on evaluation of scientific publications. Dtsch Arztebl Int 2010; 107(31–32): 552–6.
- ICH E10: Choice of Control Group and Related Issues in Clinical Trials. London UK: International Conference on Harmonization 2000; CPMP/ICH/364/96.

- Ellenberg JH: Intention-to Treat Analysis. In: Redmond C, Colton T (eds): Biostatistics in Clinical Trials. New Jersey: John Wiley & Sons: 2001.
- World Medical Association Declaration of Helsinki: Ethical principles for medical research involving human subjects. 2008. [www.wma.net/en/30publications/10policies/b3/17c.pdf](http://www.wma.net/en/30publications/10policies/b3/17c.pdf).
- Verordnung über die Anwendung der Guten Klinischen Praxis bei der Durchführung von klinischen Prüfungen mit Arzneimitteln zur Anwendung am Menschen (GCP-Verordnung – GCP-V). GCP-Verordnung vom 9. August 2004 (BGBl. I S. 2081), die zuletzt durch Artikel 4 der Verordnung vom 3. November 2006 (BGBl. I S. 2523) geändert worden ist.
- Clinical trial registration: a statement from the International Committee of Medical Journal Editors (ICMJE). Lancet 2004; 364: 911–2.
- Schulz KF, Altman DG, Moher D: CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. BMJ 2010; 340: c332.
- Windeler J: Bedeutung randomisierter klinischer Studien mit relevanten Endpunkten für die Nutzenbewertung. In: Diskussionsforum zur Nutzenbewertung im Gesundheitswesen: Begriffsdefinitionen und Einführung. 2007. [www.gesundheitsforschung-bmbf.de/\\_media/DLR\\_Nutzenbewert\\_07-11-22\\_Druckversion.pdf](http://www.gesundheitsforschung-bmbf.de/_media/DLR_Nutzenbewert_07-11-22_Druckversion.pdf).

## Anschrift für die Verfasser

Prof. Dr. rer. nat. Maria Blettner  
Institut für Medizinische Biometrie, Epidemiologie und Informatik (IMBEI)  
Universitätsmedizin der Johannes-Gutenberg-Universität Mainz  
Obere Zahlbacher Straße 69, 55131 Mainz  
[blettner-sekretariat@imbei.uni-mainz.de](mailto:blettner-sekretariat@imbei.uni-mainz.de)

## SUMMARY

### Randomized Controlled Trials: Part 17 of a Series on Evaluation of Scientific Publications

**Background:** In clinical research, randomized controlled trials (RCTs) are the best way to study the safety and efficacy of new treatments. RCTs are used to answer patient-related questions and are required by governmental regulatory bodies as the basis for approval decisions.

**Methods:** To help readers understand and evaluate RCTs, we discuss the methods and qualitative requirements of RCTs with reference to the literature and an illustrative case study. The discussion here corresponds to expositions of the subject that can be found in many textbooks but also reflects the authors' personal experience in planning, conducting and analyzing RCTs.

**Results:** The quality of an RCT depends on an appropriate study question and study design, the prevention of systematic errors, and the use of proper analytical techniques. All of these aspects must be attended to in the planning, conductance, analysis, and reporting of RCTs. RCTs must also meet ethical and legal requirements.

**Conclusion:** RCTs cannot yield reliable data unless they are planned, conducted, analyzed, and reported in ways that are methodologically sound and appropriate to the question being asked. The quality of any RCT must be critically evaluated before its relevance to patient care can be considered.

## Zitierweise

Kabisch M, Ruckes C, Seibert-Grafe M, Blettner M: Randomized controlled trials: part 17 of a series on evaluation of scientific publications. Dtsch Arztebl Int 2011; 108(39): 663–8. DOI: 10.3238/arztebl.2011.0663



The English version of this article is available online:  
[www.aerzteblatt-international.de](http://www.aerzteblatt-international.de)