Prof. Dr. med. Franz Porzsolt
Prof. Dr. sc. hum. habil. Christel Weiß
Prof. Dr. med. Manfred Weiß MBA
Dipl.-Kfm. Albert G. Müller
Dr. med. Susanne Isabel Becker MPH postgrad
Prof. Martin Eisemann PhD
Prof. Robert M. Kaplan PhD

Versorgungsforschung braucht dreidimensionale Standards zur Beschreibung von Gesundheitsleistungen – Teil 2

Zusammenfassung

Die drei Dimensionen einer standardisierten Gesundheitsleistung bestehen aus den Antworten auf die Fragen: Can it work? Does it Work? Is it worth it? Die erste Antwort beschreibt die **Efficacy**, d.h. den prinzipiellen Funktionsnachweis (Proof of Principle) unter idealen Studienbedingungen. Die zweite Frage beschreibt die **Effectiveness**, den spezifischen Funktionsnachweis, d.h. die Alltagstauglichkeit. Die dritte Frage beschreibt den **Value**, d.h. den Wert aus Sicht des individuellen Patienten und der Gesellschaft. Die Messung von Effectiveness und Value ist nur unter Alltagsbedingungen sinnvoll. Die einheitlich definierte Terminologie schließt die Bedeutung des Unterschiedes zwischen klinischer Relevanz und statistischer Signifikanz sowie die Definition des „Versorgungsalltags" ein. Die Funktionen, Stärken und Schwächen von Randomized Controlled Trials (RCTs) und von Pragmatic Controlled Trials (PCTs), die nicht miteinander konkurrieren, sondern sich ergänzen, werden diskutiert. Die Daten zeigen, dass RCTs und PCTs unterschiedliche Aussagen unterstützen, aber hinsichtlich der Validität ihrer Aussagen miteinander vergleichbar sind. Neu ist die explizite Unterscheidung der drei Dimensionen, die auch verdeutlicht, dass psychologische Effekte (einschließlich Placebo-Effekte) beim Nachweis des Proof of Principle unter Idealbedingungen weitgehend vermieden werden können. Nicht neu ist, dass die schwer quantifizierbaren psychologischen Effekte der Arzt-Patient-Beziehung die Versorgungsqualität wesentlich beeinflussen.

Schlüsselwörter

Dreidimensionale Bewertung von Gesundheitsleistungen, Efficacy, Effectiveness, Value, Real World Conditions, Pragmatic Controlled Trial, Observational study, Non-randomized trial

Crossref/doi

http://doi.org/10.24945/MVF.04.19.1866-0533.2163

Im ersten Teil des Aufsatzes haben wir unter dem Titel „Zusammenhang von Digitalisierung und Versorgungsforschung" die Voraussetzungen diskutiert, die bestehen müssen, um den Versorgungsalltag im Gesundheitssystem abbilden zu können (1). Bildlich gesprochen sind die Funktion einer digitalen Datenautobahn und die der Versorgungsforschung ähnlich: Funktionstüchtige Autobahnen sind wertlos, wenn die Abfahrten geographisch unglücklich angelegt und die verwendeten Fahrzeuge nur bedingt fahr- oder transporttauglich sind. Eine Datenautobahn ist eine notwendige, aber noch keine hinreichende Bedingung, um die Gesundheitsversorgung zu verbessern. Sie eignet sich zur Verbesserung der Gesundheitsversorgung, wenn sie dazu beiträgt, die Gesundheit der Bürger, anstatt nur einen Surrogatparameter, z.B. die Compliance mit den ärztlichen Verordnungen, zu verbessern. Und die Versorgungsforschung hat nur dann Zukunft, wenn sie es schafft, ihren Nutzen für die Gesellschaft zu beweisen. Dazu braucht sie Instrumente, die tatsächlich in der Lage sind, Effekte unter Alltagbedingungen zu beobachten und zu messen, d.h.: Sie braucht einen dreidimensionalen Standard zur Beschreibung komplexer Gesundheitsleistungen.

>> Bei jeder Aktion – so auch bei der Digitalisierung im Gesundheitssystem – sind Aufwand und Ertrag miteinander zu vergleichen. Das bedeutet, dass am Ende eine klassische ökonomische Analyse durchzuführen ist, die beschreibt, was aufzugeben ist und was man dafür als Gegenleistung erhält. Sinn macht diese Analyse nur, wenn mindestens zwei Handlungsmöglichkeiten bestehen (2, 3). Die Anwendung der ökonomischen Analyse auf die Bewertung der Erfolge des Gesundheitssystems beinhaltet drei Aspekte:

- Der Wert von Gesundheit kann aus verschiedenen, z.B. einer individuellen oder aus einer gesellschaftlichen Perspektive eingestuft werden. In beiden Fällen handelt es sich um eine subjektive Einschätzung von Effekten, die unter Alltagbedingungen zu beobachten und dort messbar sind.
- Anhand vergleichender Messungen unter Alltagsbedingungen kann geschätzt werden, ob der erbrachte Aufwand für Gesundheit zu mehr als einer zufällig bedingten Verbesserung führt.
- Ebenso lässt sich zeigen, ob der gleiche Erfolg auch mit einer anderen, weniger aufwendigen untersuchen Methode oder auch unter Inkaufnahme geringerer Belastungen erreicht werden kann.

Terminologie: Ökonomie, Kommerz, Gesundheitsökonomie und Klinische Ökonomik

Ökonomie und Kommerz

Wenn das Wort „ökonomisch" im Sinne monetärer Kosten verwendet wird, besteht das Risiko, die Bedeutung ökonomischer Analysen zu unterschätzen (4). Das Wort *OIKONOMIA* (Griechisch: die Haushaltung, Verwaltung) setzt sich zusammen aus *OIKOS* (das Haus, Hauswesen, Haushalt) und *NOMOS* (Gesetz, Brauch, Verfahren). Ökonomisieren bedeutet demnach nicht „weniger ausgeben", sondern „vergleichend abwägen". In der Literatur wird häufig das Ökonomisieren der Gesundheitsversorgung beklagt. Gemeint ist aber die Erzielung von Gewinn für das eigene Unternehmen oder das eigene Einkommen, d.h. die Kommerzialisierung der Gesundheitsversorgung.

Das Wort Kommerzialisierung leitet sich aus *COMMERCIUM* (Lateinisch: Handel) ab, das sich zusammensetzt aus CUM (mit, miteinander) und *MERX* (das Handelsgut). Wenn der Unterschied zwischen Ökonomie und Kommerz nicht wahrgenommen wird, wird auch die Bedeutung ökonomischer Abwägungen in der Gesundheitsversorgung als überflüssige Belästigung, aber nicht als Voraussetzung für Fortschritt und Gewinn wahrgenommen.

Fortschritt und Gewinn sind eng gekoppelt, weil jeder, der Fortschritt erzielt, den Lohn für seine Leistung erwartet. Gelegentlich entsteht der Eindruck eines zweigeteilten Systems: Die „Guten" erzielen Fortschritte, während die „Anderen" Gewinne einstreichen. Letzteres wird als „Interessenkonflikt" subsumiert. Jeder, der ein konkretes Ziel verfolgt, geht das Risiko eines Interessenkonflikts ein.

Der Unterschied besteht – wie meistens – in der Verhältnismäßigkeit. Niemand wird sich weigern, den Fortschritt angemessen zu vergüten, wenn dieser auch die richtige Zielgruppe erreicht.

Gesundheitsökonomie und Klinische Ökonomik

Die ökonomische Bewertung von Gesundheitsleistungen wird aus zwei Perspektiven vorgenommen. Traditionell haben Wirtschaftswissenschaftler Ende der 90er Jahre begonnen, den Nutzen von Gesundheitsleistungen aus der Sichtweise sowie mit den Fähigkeiten und dem Wissen der Wirtschaftswissenschaften zu beschreiben (5). Parallel zu dieser Entwicklung in den Wirtschaftswissenschaften wurde in der Medizin die „Evidence-based Medicine (EBM)" verbreitet. Naturgemäß betraf der Fokus der wirtschaftswissenschaftlichen Bewertung die monetären Kosten, während der Fokus der medizinischen Bewertung auf den erzielten gesundheitlichen Mehrwert gerichtet war.

Die unvermeidbare monetäre Bewertung von Gesundheitsleistungen war für Mediziner absolut plausibel. Dennoch empfanden Kollegen, deren Interesse primär der Versorgung von Patienten galt, gesundheitsökonomische Analysen damals als nahezu unethische Verweigerung notwendiger ärztlicher Leistungen. Dieser Konflikt zwischen den Praktikern und Wissenschaftlern wurde in einem EBM-Kurs an der Universität Jena thematisiert (6):

„ … Würden wir eine Generation von Ärzten ausbilden, die alle wenig belegten Therapien bezweifelt, wäre die messbare Verschlechterung der medizinischen Versorgung programmiert…, weil die Qualität der Arzt-Patient-Beziehung mehr Einfluss auf das Ergebnis der Versorgung hat, als wir bisher angenommen haben. Würde man darauf verzichten, Ärzte auszubilden, die Zweifel am Nutzen unzureichend gesicherter Maßnahmen äußern, würde das Angebot an vielversprechenden Gesundheitsleistungen innerhalb kürzester Zeit unüberschaubar werden."

Dieser Kommentar bestätigt, dass die Diskussion der damaligen Medizinstudenten in Jena dazu beigetragen hat, unser erstes Buch „Klinische Ökonomik – Effektivität und Effizienz von Gesundheitsleistungen" (7) auf den Weg zu bringen. Den Begriff „Clinical Economics" hatten zwar John M. Eisenberg, Jeffrey Sachs und andere Wissenschaftler bereits geprägt, aber mit dem Ziel angewandt, ökonomische Überlegungen und Methoden auf die Kosten medizinischer Entscheidungen anzuwenden. Wir verstehen unter Klinischer Ökonomik / Clinical Economics / CLINECS die Analyse des gesundheitlichen Mehrwerts primär aus der Perspektive des individuellen Patienten und sekundär aus der gesellschaftlichen Perspektive. Das Problem, die Theorie und die Anwendung des Konzepts sind beschrieben (7, 8, 9).

Die Diskussion zur Hierarchie der Perspektiven war zwar kontrovers und heftig. Dennoch ist es aber unwahrscheinlich, den Wert von Gesundheitsleistungen für individuelle Patienten von deren gesellschaftlichem Mehrwert ableiten zu können. Die erste Frage wird immer sein, ob eine neue Intervention für konkrete Patienten nützlich ist; als zweite Frage ist immer auch der gesellschaftliche Nutzen zu diskutieren.

Der dreidimensionale Standard zu Beschreibung von Gesundheitsleistungen

Nihil novum sub sole (Nichts Neues unter der Sonne): Dass nicht alles neu erfunden werden muss, trifft auch auf die dreidimensionale Bewertung von Gesundheitsleistungen zu. Sir Archie Cochrane und Sir Austin Bradford Hill haben vor 80 Jahren gefordert, vor der Einführung von neuen Interventionen in den Versorgungsalltag drei Fragen zu beantworten (10):

- Can it work?
- Does it work?
- Is it worth it?

Umsetzbar war das Konzept bisher noch nicht, weil zusätzlich erforderliche Überlegungen und Methoden, besonders zum Nachweis der Alltagstauglichkeit nicht zur Verfügung standen. Im vorausgegangenen Teil unseres Beitrags haben wir das Theorem von Bayes (11) und anderer nicht-experimenteller Verfahren (12) angesprochen, deren Tauglichkeit für den Erkenntnisgewinn seit dem Zweiten Weltkrieg durch die Entschlüsselung des deutschen Geheimcodes ENIGMA bestätigt wurde.

Die Beiträge von Sir Ronald Aylmer Fisher und Sir Austin Bradford Hill zur Randomisierung haben die Medizin ohne Zweifel erheblich weiterentwickelt. Nachteilig war aber, dass nicht-randomisierte Konzepte generell abgelehnt wurden. Es bedurfte einer Menge an Überzeugungsarbeit, Daten zu publizieren, die darauf hinweisen, dass bedeutende Effekte in der Gesundheitsversorgung auch mit nicht-experimentellen Beobachtungsstudien ohne Verzerrung gemessen werden können (13-22). Solche Effekte sind bedeutend, wenn nicht der prinzipielle Wirkungsnachweis, sondern die Alltagstauglichkeit bedeutend ist. Beispiele dafür sind die Erkennung von Unter-, Über- und Fehlversorgung, die Erstellung von Versorgungsleitlinien und juristische Entscheidungen, die eine standardisierte Dokumentation der Versorgungsergebnisse unter Alltagsbedingungen voraussetzen.

Das Konzept der dreidimensionalen Bewertung ist in Abbildung 1 dargestellt und beruht auf den drei Cochrane-Hill Fragen. Um Fehler und Missverständnisse bei der Übersetzung von Begriffen zu vermeiden, beziehen wir uns auf die englische Terminologie, weil nahezu alle erklärenden Beiträge und Definitionen in englischer Sprache verfasst wurden. Demnach beschreiben die Antworten auf die drei genannten Fragen die drei Outcome-Dimensionen:

1.) Proof of Principle oder Efficacy,
2.) Effectiveness und
3.) Value.

Cochrane-Hill Fragen

Question	Answer	Condition	Perspective	Type of study	Tool
Can it work?	Proof of principle **Efficacy**	Ideal Study Conditions **(ISC)**	Clinical Research	Explanatory or interventional*	Randomized Controlled Trial **[RCT]**
Does it work?	Real world **Effectiveness**	Real World Conditions **(RWC)**	Health Services Research	Pragmatic or observational*	Pragmatic Controlled Trial **[PCT]**
Is it worth it?	Subjective **Value**		Economic Research	Complete Economic Analysis	Cost Effectiveness Analysis **[CEA]**

Abb. 1: Die Antworten auf die Cochrane-Hill Fragen erfordern eine dreidimensionale Beschreibung der Effekte von Gesundheitsleistungen: Den Nachweis des Wirkprinzips (Proof of Principle; Efficacy), den Nachweis der Alltagstauglichkeit (Real World-Effectiveness) und die Beschreibung des Wertes (Value). Diese drei Effekte können unter zwei Bedingungen, aus drei unterschiedlichen Perspektiven, mit drei verschiedenen Studientypen und Instrumenten gemessen werden. Um Fehler und Missverständnisse bei der Verbindung von Begriffen und Inhalten zu vermeiden, werden die englischen Bezeichnungen verwendet, deren Inhalte in der Literatur beschrieben sind. * Thiese MS. Observational and interventional study design types. Biochemia Medica 2014;24:199-210. http://dx.doi.org/10.11613/BM.2014.022

Terminologie und das Risiko ihrer Änderung

Die erste der genannten Dimensionen, der Proof of Principle oder die Efficacy, kann nur unter idealen Studienbedingungen (Ideal Study Conditions, ISC) erhoben werden, weil für den Nachweis des Proof of Principle oder der Efficacy alle Störfaktoren auszuschließen sind, welche den Effekt einer Intervention beeinflussen können. In unserem vorausgehenden Beitrag sind sieben Kriterien beschrieben, die erfüllt sein müssen, um einen Patienten in eine experimentelle Studie unter Idealbedingungen einschließen zu können. Neben diesen sieben Kriterien beeinflussen unbewusst angewandte Kriterien von Ärzten und Patienten die Rekrutierungsrate. Der Anteil der eligiblen Patienten variiert in experimentellen Studien erheblich und kann bis auf etwa 10% der versorgten Patienten absinken. Diese geringe Rekrutierungsrate erklärt die hochgradige Selektion von Patienten in experimentellen Studien. Der Vorteil dieser Selektion besteht in der erreichten Homogenität der untersuchten Population. Der Nachteil besteht in der Einschränkung der externen Validität. Mit anderen Worten: Eine experimentelle Studie ist sicher nicht geeignet, um die Alltagstauglichkeit einer Intervention nachzuweisen; experimentelle Studien sind aber sehr wohl geeignet, um den Proof of Principle bzw. die Efficacy unter ISC nachzuweisen.

Die zweite Dimension, die Effectiveness ist per definitionem unter Alltagsbedingungen (real world conditions; RWC) nachzuweisen. Die aktuelle Diskussion bestätigt aber, dass das Consort-Statement aus dem Jahr 2008 (23) dem soeben vertretenen Standpunkt – eine experimentelle Studie sei sicher nicht geeignet, um die Alltagstauglichkeit einer Intervention nachzuweisen – widerspricht:

„The CONSORT statement is intended to improve reporting of randomised controlled trials and focuses on minimising the risk of bias (internal validity). The applicability of a trial's results (generalisability or external validity) is also important, particularly for pragmatic trials. A pragmatic trial (a term first used in 1967 by Schwartz and Lellouch) can be broadly defined as a randomised controlled trial whose purpose is to inform decisions about practice. This extension of the CONSORT statement is intended to improve the reporting of such trials and focuses on applicability."

Das Consort-Statement in der Fassung aus dem Jahr 2008 schlägt eine Umbenennung vor: von „pragmatic trials", welche nach Aussagen verschiedener Autoren (24-26) Effekte unter Alltagsbedingungen beschreiben sollen (24-26), in „pragmatic randomised trials". Diese Umbenennung würde die von Cochrane und Hill vorgeschlagene Differenzierung wieder aufheben und mehr Verwirrung verursachen als Klarheit schaffen, weil durch die Akzeptanz dieser Änderung beide Dimensionen, die Efficacy und die Effectiveness nur unter Idealbedingungen gemessen werden sollten.

In der exakt beschriebenen und häufig zitierten Arbeit von Schwartz & Lelloch (24) wird das Wort „Randomisation" nicht erwähnt. Deshalb sollte diese renommierte Publikation auch nicht verwendet werden, um eine Umbenennung der „pragmatic trials" in „pragmatic randomised trials" zu begründen. In der Fassung des Consort-Statements aus dem Jahr 2010 und in zahlreichen Arbeiten anderer Autoren ist das Ergebnis der Umbenennung bereits konsolidiert (27). Consort 2010 bietet keine neuen Erklärungen an, sondern verweist bezüglich pragmatischer Studien auf das Consort-Statement 2008. Hier besteht Diskussionsbedarf.

Die Dringlichkeit des Diskussionsbedarfs wird deutlich, wenn die Unterschiede zwischen Efficacy und Effectiveness bedacht werden: Die klinische Forschung ist unter idealen Studienbedingungen in einem explanatory bzw. interventional trial mit einem RCT durchzuführen, während die Versorgungsforschung ein pragmatic oder observational trial benötigt, um Effekte unter Alltagsbedingungen (real world conditions; RWC) nachzuweisen. Ein nicht-experimentelles Studiendesign verteilt die Probanden nicht auf eine begrenzte Zahl unterschiedlicher Interventionen, sondern überlässt die Wahl der Intervention dem Konsens zwischen Behandler und Patient.

Die Diskussion der dritten Dimension, des Wertes von Gesundheitsleistungen, wird in einem eigenständigen, nachfolgenden Beitrag geführt.

„Klinische Relevanz" und „statistische Signifikanz": Ethische und ökonomische Aspekte

Die wissenschaftliche Literatur ist sich zwar weitgehend darüber einig, dass es sinnvoll ist, neben der „(statistischen) Signifikanz" eines Unterschiedes auch dessen „(klinische) Relevanz" zu beschreiben. Zur Bedeutung der statistischen Signifikanz besteht ein Konsens aber nicht zur Definition der klinischen Relevanz. Die Statistik kann durch Berechnung einer Prüfgröße anhand der Stichprobendaten von zwei zu vergleichenden Therapiegruppen die Empfehlung abgeben, ob ein beobachteter Unterschied als Zufallsbefund oder als tatsächlicher Unterschied interpretiert werden sollte. Die Statistik kann aber kaum Unterstützung anbieten, wenn zu klären ist, ob ein beobachteter Unterschied vom individuellen Patienten selbst oder vom behandelnden Arzt als klinisch bedeutend oder unbedeutend eingestuft werden soll.

Dazu ein Beispiel: Wenn ein Schlafmittel den Schlaf in einer Studie durchschnittlich um fünf Minuten verlängert und die Studie sehr viele Patienten enthält, kann sich der beobachtete kleine Unterschied als signifikant darstellen. Hätte die Studie deutlich weniger Patienten eingeschlossen, könnte zwar ebenfalls ein durchschnittlicher Unterschied der Schlafdauer von fünf Minuten resultieren, der Standardfehler der Schätzung wäre aber sicher größer und der Unterschied könnte anhand der akzeptierten Definition in der Statistik nicht von einem Zufallsbefund unterschieden werden.

Wenn man jedoch die klinische Relevanz, die Minimal Important Difference (MID) oder die Minimal Clinically Important Difference (MCID) bewerten will, wird eine Verlängerung der Schlafdauer um fünf Minuten bei einer durchschnittlichen Schlafdauer von sechs Stunden (5 von 360 Minuten) klinisch nicht relevant sein, aber sehr wohl bei einer Schlafdauer von 30 Minuten (5 von 30 Minuten). Ob die Verhältnisse, die bei der Schlafdauer zutreffend sind, auch auf andere Szenarien zutreffen, kann vordergründig mit „nein" beantwortet werden. Wegen der Vielzahl und Komplexität der Einflussfaktoren ist das Problem noch als ungelöst zu interpretieren, was allerdings nicht ausschließt, dass sich hinter den Konzepten der MID oder der MCID möglicherweise dennoch ein Naturgesetz verbirgt, dessen Formel wir nur noch nicht verstanden haben.

Mit diesen Ausführungen wollen wir auf die bisher nicht immer beachtete ethische und ökonomische Bedeutung der klinischen Relevanz hinweisen. Im folgenden Abschnitt wird erklärt, weshalb PCTs, d.h. Studien unter Alltagsbedingungen, eine erheblich größere Zahl von Teilnehmern einschließen als RCTs. Wenn identische Informationen alleine durch logische Überlegungen auch mit weniger Probanden und damit schneller gewonnen werden können als ohne diese Überlegungen, entsteht ein ethisches und ökonomisches Problem. Die Überlegung, die wir zur Diskussion stellen wollen, betrifft die

Drei Entscheidungen bis zur Auswertung		
1. Der Patient (P) wählt den Zeitpunkt	**2. Der Arzt & P legen die Auswertung fest**	**3. Der Wissenschaftler wählt Ziel und geeigneten Test**
Der Patient nimmt ärztliche Hilfe wegen beunruhigender Zeichen, Symptome oder erhaltener Informationen in Anspruch. Der Prozess startet immer unter Alltagsbedingungen (RWC).	Der Arzt bietet die erbetene Hilfe an. Bei Interesse an einer systematischen Datenerfassung holt er dazu das notwendige Einverständnis des Patienten ein.	Wenn Arzt und Patient mit einer systematischen Auswertung einverstanden sind, entscheidet der Wissenschaftler, ob ein Proof of Principle (Efficacy) oder die Alltagstauglichkeit (Effectiveness) oder der Wert (Value) der Intervention geprüft wird.

Tab. 1: Die Bewertung von Alltagsergebnissen ist das Ergebnis dreier Entscheidungen.

Einbeziehung der klinischen Relevanz in das Studienprotokoll. Dort sollte bereits beschrieben sein, welcher Unterschied erwartet wird, um eine Differenz als bedeutend für den Patienten einzustufen. Falls die empirisch ermittelte Differenz der Mittelwerte den geforderten Unterschied unterschreitet, ist die Anwendung statistischer Methoden nicht erforderlich. Wenn die Hypothese einer Studie weder ohne noch mit statistischer Berechnung bestätigt werden kann, soll auf die statistische Berechnung verzichtet werden, weil statistische Power (die von der Anzahl der untersuchten Patienten abhängt) sinnvoller genutzt werden kann, um eine bedeutende Hypothese zu bestätigen, als einen klinisch ohnehin irrelevanten Unterschied abzusichern. Die Ergebnisse werden schneller und kostengünstiger mit als ohne diese Überlegungen zur Verfügung stehen.

Definition des „Versorgungsalltags", die Entscheidungen und die Messung von Effekten

Das Pragmatic Controlled Trial haben wir mit Unterstützung durch verschiedene Gruppen über einen Zeitraum von zehn Jahren entwickelt (13, 14), um Effekte messen zu können, die nicht unter idealisierten Bedingungen, sondern unter den Bedingungen des Versorgungsalltags zu beobachten sind.

Um die Entscheidungen und die Messung von Effekten im Versorgungsalltag nicht zu unterschätzen, sollte man sich unter Alltagsversorgung ein „natural chaos" vorstellen, das durch nicht-interventionelle Regeln in ein auswertbares System zu überführen ist. Unter Alltagsbedingungen ist generell jeder Patient (unabhängig vom Spektrum seines individuellen Leidens) zu versorgen. Es ist zweckmässig, den Schwerpunkt der Versorgung zu definieren (z.B. Abteilung für Unfallchirurgie oder Diabetes-Sprechstunde oder Praxis für Physiotherapie). Da aber alle Patienten zu versorgen sind, existieren unter Alltagsbedingungen im Gegensatz zur experimentellen Studienbedingungen keine Ausschlusskriterien.

Zweitens entscheidet jeder Therapeut (abhängig vom Spektrum seiner individuellen Präferenzen) zusammen mit dem Patienten über die Wahl der individuellen Intervention. Diese Definition wird von einigen Autoren sinngemäß bestätigt (24 - 26), während die Mehrzahl aller Autoren davon ausgehen, dass der Zufall auch unter Alltagsbedingungen über die Allokation entscheidet. Wir stimmen dieser Annahme nicht zu, weil die präferenzbasierte Allokation den Alltag von der zufallsbasierten Allokation im Experiment unterscheidet. Präferenz-basiert bedeutet, dass der Therapeut eine für seinen Patienten geeignete Therapie auswählt. Diese Auswahl mag einfach sein, wenn nur ein einziges Gesundheitsproblem besteht. Die Auswahl wird zur medizinischen Herausforderung, wenn der Therapeut bei seiner Entscheidung alle Gesundheitsprobleme dieses Patienten und alle Therapien berücksichtigt, die dieser Patient bereits von anderen Therapeuten verordnet bekommen hat. Auch eine elektronische Gesundheitskarte wird dem Therapeuten nicht die Abwägung zwischen dringend erforderlichen und optionalen Interventionen sowie zwischen unbedenklichen und gesundheitsgefährdenden Kombinationen verschiedener Interventionen ersparen.

Ebenso sollte man sich vor Augen halten, dass jede Form der Versorgung unter Alltagsbedingungen beginnt und dass nur in den wenigsten Fällen der Versorgung die Ergebnisse systematisch dokumentiert werden. Mit anderen Worten, es gibt bisher keine systematische Qualitätskontrolle, die bestätigt, dass im Alltag auch tatsächlich das erreicht wird, was die Forschung in Aussicht stellt. Die Dokumentation von Outcomes unter Alltagsbedingungen ist von drei voneinander unabhängigen Entscheidungen abhängig (Tab. 1):

- Jeder entscheidungsfähige Patient hat primär festzulegen, in welchem Gesundheits-Stadium er ärztliche Hilfe in Anspruch nimmt.
- Zweitens, wenn der Arzt plant, eine systematische (nicht nur zufällige) Dokumentation der Outcomes seiner Intervention vorzunehmen, muss er für diese zweite Entscheidung nicht nur hinreichend selbstkritisch sein, sondern zudem auch das Einverständnis des Patienten einholen.
- Falls die zweite Entscheidung zugunsten einer systematischen Dokumentation der Outcomes ausfällt, ist als dritte Entscheidung zu wählen, ob der Proof of Principle in einem experimentellen Design oder die Effectiveness bzw. der Value in einem nicht-experimentellen, beobachtenden Design nachgewiesen werden soll.

Möglicherweise erklären diese Hürden, weshalb wir den Eindruck haben, dass Reklamationen in der Automobilindustrie häufiger als im Gesundheitssystem vorkommen.

Man kann sich die Messung von Effekten unter Alltagsbedingungen auch einfach machen, indem man annimmt, dass kein Unterschied zwischen idealisierten Studienbedingungen und Alltagbedingungen besteht. Unter dieser Annahme würde sich ein RCT auch eignen, um Effekte im Versorgungsalltag abzubilden. Wir vertreten aber den Standpunkt, dass ein RCT eben nicht geeignet ist, den Versorgungsalltag abzubilden. Unser Standpunkt erschwert die Messung von Alltagseffekten, weil dazu zwei Probleme zu lösen sind: Es ist ein Algorithmus zu finden, mit dem vergleichbare Patientengruppen ohne Randomisierung generiert werden können und ein zweiter Algorithmus, um bei der großen Zahl unterschiedlicher Therapien jene zu identifizieren, die in derselben Gruppe zusammengefasst werden können.

Was RCTs und PCTs nicht voneinander unterscheidet

Bei jedem PCT – wie auch bei jeder anderen lege artis durchgeführten Studie – ist die Studienhypothese vor Projektbeginn zu definieren. Die Studienhypothese beschreibt die verglichenen Patientengruppen, die Interventionen, die erwartete Richtung (z.B. Überlegenheit) und die Größe der erwarteten Effekte. Die meisten Studien werden als Richtung der Hypothese die Überlegenheit (superiority)

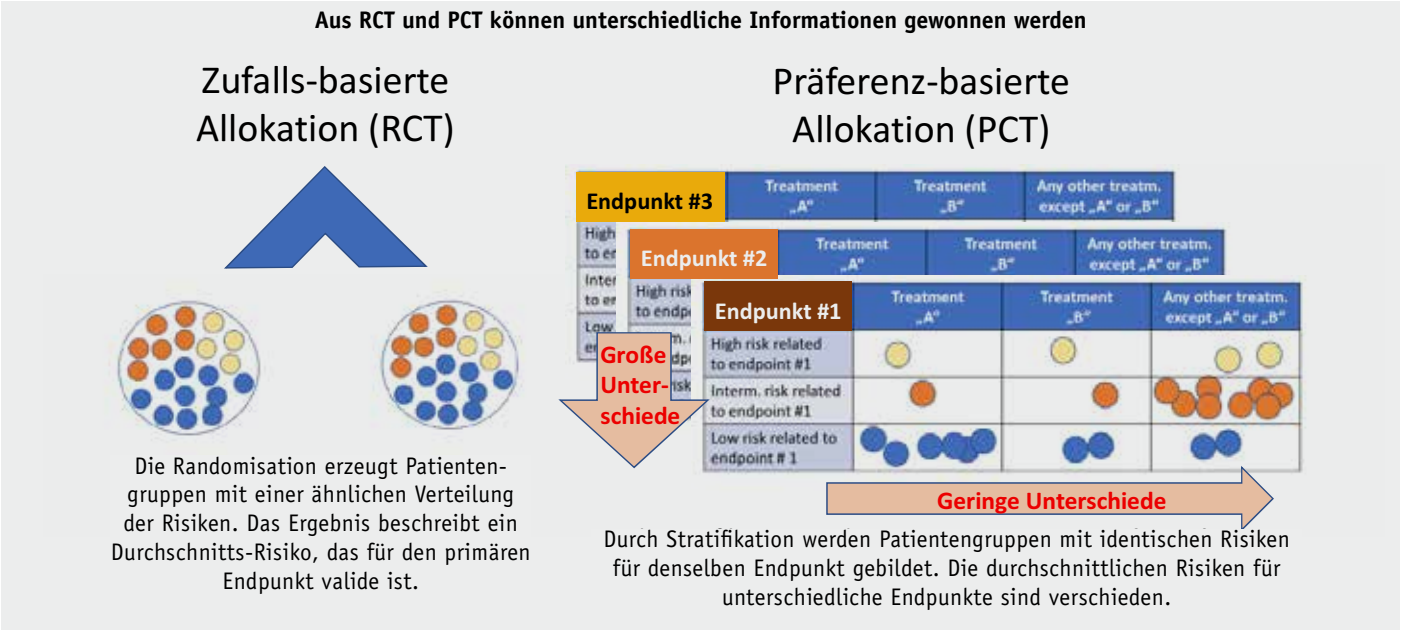


Abb. 2: Vergleich des Designs eines Randomized Controlled Trial (RCT) und eines Pragmatic Controlled Trial (PCT).

prüfen. Sofern bereits Daten vorliegen, die den Proof of Principle bestätigen (28), kann auch auf Gleichwertigkeit (equivalence) oder Nicht-Unterlegenheit (non-inferiority) geprüft werden. Als wesentlicher Bestandteil der Hypothese ist in allen Fällen auch die Größe des minimal erwarteten Effektes anzugeben, der noch als bedeutend angesehen wird.

Dieser Aspekt ist bei Versorgungsstudien (PCTs) bedeutender als bei Studien zum Nachweis des Proof of Principle (RCTs), u.a., weil PCTs ein Vielfaches der Patienten von RCTs einschließen und deshalb bereits bei wesentlich geringeren Effekten den statistischen Nachweis signifikanter Unterschiede erbringen können. Aus diesem Grund ist es sinnvoll, die klinische Relevanz nicht erst im Nachhinein anhand beobachteter Werte zu definieren, sondern sich bereits vor Erhebung der Daten darauf zu einigen, welche Unterschiede als klinisch bedeutungsvoll angesehen werden.

Was RCTs und PCTs voneinander unterscheidet

RCTs und PCTs können sich gegenseitig ergänzen, nicht ersetzen. RCTs können den prinzipiellen Nachweis des Effekts (Proof of Principle; Efficacy) erbringen. PCTs können die Effekte unter Alltagsbedingungen (Real World Effectiveness, RWE) bestätigen.

Bei den experimentellen Studien (RCTs) waren zwei Schwachpunkte aufgefallen, die bei der Entwicklung des Pragmatic Controlled Trials vermieden werden sollten. Diese beiden Aspekte der RCTs betreffen den Unterschied der beabsichtigten und unbeabsichtigten Interventionen und den Unterschied der erwünschten und tatsächlich erreichten Gleichverteilung der Risikofaktoren. Beide Schwachpunkte sind nicht unabhängig voneinander.

Beabsichtigte und unbeabsichtigte Interventionen in RCTs

Die in einem RCT verabreichten Interventionen werden einzeln durch Randomisation ausgewählt. In einem PCT wählt der Arzt zusammen mit dem Patienten die am besten geeignete Therapie. In beiden Studienarten findet Kommunikation zwischen Arzt und Patient statt. Jede Form der Kommunikation vermittelt Effekte, weil auch die beabsichtigte „Nicht-Kommunikation" eine Form der Kommunikation darstellt. Allerdings ist die Quantifizierung dieser Effekte schwierig, weil jede Form einer Therapie-Empfehlung die Erwartungshaltung eines Patienten beeinflusst.

Dazu ein Beispiel: Wenn in einer Studie zwei Therapien (A und B) miteinander verglichen werden, können in sehr seltenen Fällen beide Therapien von jeweils 50% der Patienten bevorzugt werden. Wesentlich häufiger wird aber zu beobachten sein, dass z.B. 60% oder 80% der Patienten Therapie A aber nur 40% oder 20% der Patienten Therapie B bevorzugen.

Werden diese Patienten randomisiert, werden in beiden Gruppen z.B. 80% der Patienten die Therapie A und 20% der Patienten die Therapie B bevorzugen. Wenn nun eine dieser Gruppen Therapie A und die andere Gruppe Therapie B erhält, wird – trotz Randomisation – unschwer nachzuweisen sein, dass bei gleicher Wirksamkeit der Therapien A und B die Ergebnisse in der Gruppe, in welcher die Mehrzahl der Patienten die präferierte Therapie erhalten haben, besser sein werden als in der anderen Gruppe (13).

Diese Überlegung deutet darauf hin, dass zum einen die Information, die ein Patient von seinem Arzt, von der Familie, Freunden, oder Bekannten oder von den Medien angeboten bekommt, seine Erwartungshaltung beeinflusst. Dieser Einfluss wird um so wirksamer sein, je stärker das Vertrauensverhältnis zwischen dem Patienten und dem Ratgeber ist. Zum anderen bestätigt das Beispiel, den Zusammenhang zwischen dem Einfluss einer Information auf die Prägung der Erwartungshaltung und den Effekt der Erwartungshaltung auf das beobachtete Ergebnis (29 – 31).

In experimentellen Studien kann der Effekt der Kommunikation auf die Erwartungshaltung/Präferenz der Patienten durch die Verblindung reduziert, aber nicht komplett vermieden werden. Die Bedeutung des Effekts der Kommunikation auf die Erwartungshaltung der Patienten, den Erwartungseffekt, haben wir im Jahr 2004 am Beispiel des Placebo-Effekts diskutiert (32). Unsere Diskussion bestätigt die Beobachtungen anderer Gruppen aus den 90er Jahren (29, 30). Die Macht dieses Erwartungseffektes lässt sich durch das menschliche Grundbedürfnis nach Sicherheit erklären. Am Beispiel der Sicherheitsschleife kann gezeigt werden, dass die Kommunikation sehr wahrscheinlich der entscheidende Faktor ist, der den Zusammenhang zwischen objektiven

Risiken und deren subjektiver Wahrnehmung steuert (33 – 36).

Eine kritische Analyse dieser Überlegungen könnte helfen, politische Fehlentscheidungen zu unterschiedlichen Therapieverfahren zu vermeiden. Maßgeblich sollten letztlich die Erfolge zur Linderung oder Verhinderung von Gesundheitsproblemen sein. Die Erklärung eines Wirkprinzips ist akademisch wertvoll. Praktisch bedeutend sind die Erfolge im Versorgungsalltag.

Gleichverteilung der Risikofaktoren in RCTs

Generell wird angenommen, dass durch die Randomisation alle denkbaren Risiken in den untersuchten Gruppen einer Studie ähnlich verteilt werden. Die Wahrscheinlichkeit, dass diese Annahme zutrifft, steigt u.a. mit der Größe der untersuchten Gruppen. Je kleiner aber die Zahl der eingeschlossenen Patienten ist, um so größer ist das Risiko, dass die Gruppen nicht wirklich vergleichbar sind. Die Gruppengröße alleine ist aber nicht der einzige Faktor, der die Gleichverteilung der Ausgangsrisiken beeinflussen kann. Neben anderen hier nicht angesprochenen Faktoren sind die oben erwähnten Kommunikationseffekte zu berücksichtigen, die je nach vorbestehenden Wertvorstellungen die Ergebnisse einer Intervention unterschiedlich beeinflussen. Die Risiken jeder anderen Gruppe außerhalb dieser Studie unterscheiden sich sehr wahrscheinlich anhand ihres Risikoprofils (geringe externe Validität), auch wenn die Ein- und Ausschlusskriterien der verglichenen Studien ähnlich sind.

Als Fazit lässt sich ableiten, dass das RCT ein theoretisches Konzept beschreibt, dessen reale Umsetzung in doppelt verblindeten Studien möglich ist, wenn diese Studien nach einer Modellberechnung, die fünf Annahmen voraussetzt, mehr als 1.000 Probanden enthält (Ch. Weiss pers. Mitteilung).

Lösung beider Probleme in Pragmatic Controlled Trials (PCTs)

In einem PCT können beide Probleme gelöst werden, die Vermeidung unbeabsichtigter Interventionen (Placebo-Effekte) und die ungleiche Verteilung der Risikofaktoren. Unbeabsichtigte Interventionen sind nur unter idealen Studienbedingungen ein Problem, nicht aber unter Alltagsbedingungen. Unter Alltagsbedingungen wird jeder Patient und Therapeut unabhängig von der Art der gewählten Therapie alles tun, um das bestmögliche Versorgungsergebnis zu erzielen. Dabei ist unbedeutend, ob es sich um eine spezifische oder unterstützende Intervention, z.B. eine vertrauensvolle Arzt-Patient-Beziehung, handelt. Diese unterstützenden Maßnahmen können bisher nicht quantifiziert werden. Deshalb ist davon auszugehen, dass die unter Alltagsbedingungen erzielten Versorgungsergebnisse immer durch Mischformen spezifischer und unterstützender Interventionen bedingt sind. Die unterstützende Intervention einer experimentellen Studie kommt durch die Kommunikation zustande, die erforderlich und obligat ist, um die Ziele des Patienten (gesund zu werden) mit den Zielen des Wissenschaftlers (Erkenntnisse zu gewinnen) in Übereinstimmung zu bringen. In einer pragmatischen Studie existiert diese Übereinstimmung der Ziele a priori. Die Kommunikation kann in einer pragmatischen Studie auf die Bedürfnisse des Patienten abgestimmt werden und wird dort mehr zur Stabilisierung der Arzt-Patient-Beziehung beitragen, als ein Gespräch, das die Rahmenbedingungen einer experimentellen Studie zu gewährleisten hat.

Die Randomisierung, die in experimentellen Studien eine ungleiche Verteilung der Risikofaktoren verhindern soll, wird in pragmatischen Studien durch eine Dreifach-Stratifikation ersetzt (Abb. 2).

Dazu sind die folgenden drei Schritte durchzuführen: Die Patienten sind in vergleichbaren Gruppen zusammenzufassen. Patienten sind miteinander vergleichbar, wenn:

1. bei diesen Patienten die gleichen Versorgungsziele angestrebt werden. Diese sind in der Regel. 1. Die Hauptzielkriterien (z.B. Überlebenszeit und spezifische Aspekte der gesundheitsbezogenen Lebensqualität). 2. Die unerwünschten Effekte (Nebenwirkungen oder Komplikationen der Therapie). 3. Kosten der Versorgung. Die Endpunkte und Kriterien zur Messung der Endpunkte werden von der Studienleitgruppe vorab im Studienprotokoll definiert.
2. jeder Patient bezüglich jedes Versorgungsziels einer Risikoklasse (hoch, intermediär, niedrig) zugeordnet ist, um nur Patienten mit gleichen Ausgangsrisiken miteinander zu vergleichen. Dazu ist bei Aufnahme jedes Patienten in ein PCT dessen individuelles Risikoprofil zu erheben. Zur Erhebung des Risikoprofils hat die Studienleitgruppe zunächst eine Risiko-Checkliste erstellt, die alle Risiken enthält, die das Erreichen eines der gewählten Endpunkte beeinflussen kann. Zudem hat die Studienleitgruppe einen zweiten Algorithmus definiert, der jeden Patienten anhand seines individuellen Risikoprofils für jeden Endpunkt gesondert einer Risikoklasse zuordnet.
3. die im IT-System des PCT hinterlegten Algorithmen jeden Patienten bezüglich jedes Endpunkts einer exakt definierten Risikoklasse zuordnen können und bei der Auswertung der Ergebnisse damit gewährleisten, dass nur Patienten aus identischen Risikoklassen miteinander verglichen werden.

Eine analoge Dokumentation und Hinterlegung ist für die Kategorisierung der Interventionen (Therapien und deren Änderungen) und der Endpunkte zu den vorab definierten Zeitpunkten erforderlich.

Vergleich der gemessenen Ergebnisse von RCTs und PCTs

Ungeachtet der unterschiedlichen Aussagen, die aus den Messdaten eines RCTs und PCTs abgeleitet werden können, lässt sich die erforderliche Dokumentation bei Durchführung eines PCT aus der klinischen Dokumentation der Routineversorgung entnehmen. Wenn die individuellen Risikofaktoren eines Patienten (einmalig), die Daten zur Therapie und zu den Änderungen der Therapien im Beobachtungszeitraum regelmäßig erfasst werden, und die Versorgungsergebnisse an den vorab definierten Messpunkten dokumentiert werden, entsteht kein zusätzlicher Dokumentationsaufwand. Der wesentliche Unterschied zwischen einem RCT und PCT besteht in der deutlich aufwendigeren Konzeption, aber wesentlich einfacheren Durchführung des PCT im Vergleich zum RCT. Die geschätzte Anzahl der erfassten Patienten beträgt in einem PCT rund das 20-fache der in einem RCT rekrutierten Patienten. In einem RCT erfolgt die Zuordnung vorselektierter Patienten nach dem Zufallsprinzip. In einem PCT werden die nicht vorselektierten Patienten dreifach stratifiziert, d.h. nach dem Endpunkt der Messung, dem am Endpunkt orientierten Ausgangsrisiko und anhand der vom Behandler gewählten Intervention.

Dieser Strategie liegen mehrere Überlegungen zugrunde. Wenn bei zwei Patienten unterschiedliche Effekte hinsichtlich zweier verschiedener Endpunkte beobachtet werden, wird vorab sichergestellt, dass die jeweiligen Ausgangsrisiken der verglichenen Patienten und Endpunkte übereinstimmen. Inzwischen wurde auch bestätigt, dass die Berechnung des Propensity Scores deutlich verbessert werden kann, wenn die Selektion der Faktoren zur Berechnung des Scores am Studienendpunkt orientiert wird, was wir mit dem Design eines PCT vorschlagen (33). Mit zunehmender Zahl der gemessenen Endpunkte sind entsprechende Korrekturen für multiples Testen, z.B. nach Bonferroni, zu berücksichtigen. Das Design des PCT sollte geeignet sein,

Ergebnisse zu generieren, die mit der Validität der Ergebnisse von RCTs vergleichbar sind.

Nicht gegen-, sondern miteinander

Jedes Studiendesign hat seinen Wert, wenn man weiß, wozu, wann und wie man es einsetzen kann. Dazu eine einfache Analogie: Man kann mit einer Zange einen Nagel einschlagen, doch wird dazu ein Hammer besser geeignet sein. Ähnlich verhält es sich mit RCT und PCT. Während erstere zwar nach wie vor als Goldstandard bei klinischen Untersuchungen angesehen wird, interpretieren wir den Vorschlag von Cochrane und Hill als wegweisenden Fortschritt. Ein RCT kann den **prinzipiellen Nachweis** des Effekts einer Intervention (Proof of Principle; Efficacy) unter idealen Studienbedingungen erbringen. Mehr aber auch nicht. Basierend auf diesem prinzipiellen Funktionsnachweis kann mit einem PCT der **spezifische Funktionsnachweis**, d.h. der Effekt unter den jeweils vorliegenden Alltagsbedingungen (Real World-Effectiveness) geprüft werden. Diese klare Unterscheidung zwischen Eff & Eff (Abb. 1) entspricht einem neuen Vorschlag, der nur durch die künftige Diskussion verworfen, akzeptiert oder modifiziert werden kann. Die aktuelle „französische Diskussion" (Homöopathie) droht das Kind mit dem Bade auszuschütten, weil der bisher fehlende Nachweis der Efficacy einer breit angewandten Intervention zwar eine politische, aber keine wissenschaftliche Entscheidung unterstützen kann. Aus unserer Sicht bieten sich bei einem fehlenden Nachweis der Efficacy zwei Optionen an:

- Wenn eine neue Hypothese zur Efficacy formuliert wird, sollte diese geprüft werden (32); die nochmalige Wiederholung einer bereits mehrfach ohne überzeugendes Ergebnis geprüften Hypothese ist sicher keine vielversprechende Option.
- Wenn eine neue Hypothese noch nicht konsensfähig ist, bietet sich an, die Effectiveness der breit angewandten Intervention zu prüfen, deren Alltagstauglichkeit angezweifelt wird. Diese Prüfung wird aber nur möglich sein, wenn die Befürworter der miteinander konkurrierenden Heilmethoden bereit sind, an dieser Prüfung mitzuwirken. <<

Autorenerklärung

Die Autoren erklären, dass keine Interessenkonflikte vorliegen.

Zitationshinweis

Porzsolt et al.: „Versorgungsforschung braucht dreidimensionale Standards zur Beschreibung von Gesundheitsleistungen – Teil 2" in „Monitor Versorgungsforschung" (04/19), S. 53-60, doi: 10.24945/MVF.04.19.1866-0533.2163

Literatur

1. Porzsolt F, Becker IS, Weiss M, Wiedemann F, Weiss Ch. Der Zusammenhang von Digitalisierung und Versorgungsforschung. „Monitor Versorgungsforschung" 2019;3:49-53. http://doi.org/10.24945/MVF.03.19.1866-0533.2147
2. What is economic analysis. http://www.businessdictionary.com/definition/economic-analysis.html
3. Economic analysis definition. http://www.investorwords.com/16379/economic_analysis.html
4. Porzsolt F. Ökonomisierung oder Kommerzialisierung der Gesundheitsversorgung. Public Health Forum 2013;21:33e1-33e3. DOI:10.1016/j.phf.2013.09.018
5. Drummond M, O'Brien B, Stoddard GL, Torrance GW. Methods for the Economic Evaluation of Health Care Programmes. Oxford, U.K., Oxford University Press, 1997, pp 305. DOI: 10.1177/0272989x9801800417
6. Porzsolt F, Strauss B; Evidenzbasierte Medizin: Konflikt ist lösbar. Dtsch Ärzteblatt 2002;99: A-761
7. Porzsolt F, Williams AR, Kaplan RM (eds): Klinische Ökonomik. Effektivität und Effizienz von Gesundheitsleistungen. Ecomed Verlagsgesellschaft 2003, p1-372. ISBN 3-609-16153-1
8. Porzsolt F. Klinische Ökonomik: Die ökonomische Bewertung von Gesundheitsleistungen aus der Sicht des Patienten. In: Porzsolt F, Williams AR, Kaplan RM (eds): Klinische Ökonomik. Effektivität und Effizienz von Gesundheitsleistungen. Ecomed Verlagsgesellschaft 2003, p 17-40. ISBN 3-609-16153-1
9. Porzsolt F, Kaplan RM. Optimizing Health – Improving the value of healthcare delivery. 2006. Springer New York. ISBN 0-387-33920-5
10. Haynes B. Can it work? Does it work? Is it worth it? The testing of healthcare interventions is evolving. BMJ. 1999;319:652-653.
11. Grayne SB. The Theory That Would Not Die: How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian Submarines, and Emerged Triumphant from Two Centuries of Controversy. Yale University Press 2011. New Haven London.
12. Pearl J, Mackenzie D. The Book of Why: The New Science of Cause and Effect. Basic Books, New York. First edition. 2018
13. Porzsolt F, Eisemann M, Habs M, Wyer P. Form Follows Function: Pragmatic Controlled Trials (PCTs) have to answer different questions and require different designs than Randomized Controlled Trials (RCTs). J Publ Health 2013;21:307-313. DOI 10.1007/s10389-012-0544-5
14. Porzsolt F, Rocha NG, Toledo-Arruda AC, Thomaz TG, Moraes C, Bessa-Guerra TR, Leão M, Migowski A, Araujo de Silva AR, Weiss C. Efficacy and Effectiveness Trials Have Different Goals, Use Different Tools, and Generate Different Messages. Pragmatic and Observational Research 2015;6:47-54. DOI http://dx.doi.org/10.2147/POR.S89946
15. Porzsolt F. Clinical Economics and Nursing. Rev. Latino-Am. Enfermagem 2015;23: On-line version ISSN 1518-8345. http://dx.doi.org/10.1590/0104-1169.0000.2640
16. Porzsolt F. Clinical Economics – It is about Values not about Money. Brazil J Medicine Human Health 2016;4 (3). DOI: http://dx.doi.org/10.17267/2317-3386bjmhh.v4i3.1052
17. Porzsolt F, Correia LCL. Clinical Economics is based on Real World Conditions not only on Ideal Study Conditions. (Editorial). Arquivos Brasileiros de Cardiologia [Brazilian Archives of Cardiology] 2017. DOI: 10.5935/abc.20170084
18. Porzsolt F, Jauch KW. Real-World Usefulness Is Missing. Dtsch Arztebl Int. 2018;115:114-115. doi: 10.3238/arztebl.2018.0114c.
19. Porzsolt F, Becker S, Kamga Wambo O, Weiss M, Wiedemann F, Weiss Ch. Die dreidimensionale Bewertung des Nutzens von Gesundheitsleistungen. Ein Bei-trag zur Harmonisierung der wissenschaftlichen Diskussion. DFGMA 2018;5:1-5
20. Porzsolt F. The assessments of three different dimensions "Efficacy", "Effectiveness", and "Value" require three different tools: the Randomized Controlled Trial (RCT), the Pragmatic Controlled Trial (PCT), and the Complete Economic or Cost-Effectiveness Analysis (CEA). Surg Rehabil, 2018;2(4):1. doi: 10.15761/SRJ.1000145
21. Porzsolt F, Wiedemann F, Schmaling K, Kaplan RM. The Risk of Imprecise Terminology: Incongruent Results of Clinical Trials and Incongruent Recommendation in Clinical Guidelines. Abstract EBM Live 2019, Oxford/UK, July 15-17, 2019.
22. Porzsolt F, Becker SI, Kern AO, Langer D, Müller AG, Wiedemann H, Kaplan RM. Die Bewältigung des Pflegenotstands: Strategie und Maßnahmen. Monitor Pflege. 2019;2:24-31
23. Zwarenstein M, Treweek S, Gagnier JJ, Altman DG, Tunis S, Haynes B, Oxman AD, Moher D for the CONSORT and Pragmatic Trials in Healthcare (Practihc) groups. Improving the reporting of pragmatic trials: an extension of the CONSORT statement. BMJ. 2008; 337: a2390. doi: 10.1136/bmj.a2390: 10.1136/bmj.a2390 PMCID: PMC3266844 PMID: 19001484
24. Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutic trials. J. chron. Dis. 1967;20:637-648.
25. Grimes DA, Schulz KF. An overview of clinical research: the lay of the land. Lancet 2002;359:57-61
26. Thiese MS. Observational and interventional study design types. Biochemia Medica 2014;24(2):199-210. http://dx.doi.org/10.11613/BM.2014.022
27. Schulz KF, Altman DG, Moher D for the CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. BMC Medicine 2010, 8:18 http://www.biomedcentral.com/1741-7015/8/18
28. Schumi J, Wittes JT. Through the looking glass: understanding non-inferiority. Trials. 2011;12:106. doi: 10.1186/1745-6215-12-106.
29. Heckhausen, H. Achievement motivation and ist constructs: A cognitive model. Motivation and Emotion. 1977; 1,283-329.
30. Ross L, Lepper MR, Strack F, Steinmetz J. Social Explanation and Social Expectation: Effects of Real and Hypothetical Explanations on Subjective Likelihood. J Personality and Soc Psychol 1977; 35: 817-829. DOI: 10.1037/0022-3514.35.11.817
31. Rasmussen C, Johnson G. The Ripple Effect of Virginia Tech: Assessing the Nationwide Impact on Campus Safety and Security Policy and Practice. ERIC Number: ED502232. 2008. https://files.eric.ed.gov/fulltext/ED502232.pdf. Letzter Download July 7, 2019
32. Porzsolt F, Schlotz-Gorton N, Biller-Andorno N, Thim A, Meissner K, Roeckl-Wiedmann I, Herzberger B, Ziegler R, Gaus W, Pöppel E: Applying Evidence to Support Ethical Decisions: Is the Placebo Really Powerless? Science and Engineering Ethics 2004;10:119-132
33. Porzsolt F. Safety means perception of risk. J Med Safety 2016;Oct:18-24
34. Porzsolt F, Weiss Ch. Die Wahrnehmung lebensverändernder Diagnosen durch Patienten und deren Angehörige – Eine Online Befragung. MVF 2018;6:40-44
35. Porzsolt F. Syntopie zur Vermeidung persönlicher und konzeptioneller Konflikte: „Schwindelfrei Geld verdienen". Zur Publikation eingereicht
36. Porzsolt F, Matosevich R, Kaplan RM. Cancer Screening should focus reliable endpoints, specificity controls, and outcomes that are important to patients. Submitted for publication.
37. Ferdinand D, Otto M, Weiss Ch. Get the most from your data: a propensity score model comparison on real-life data. Internat. J Gen Medicine 2016:9 123–131

Health services research needs three-dimensional standards for description of health services – Part 2

The three dimensions of a standardised health service consist of the answers to the questions: Can it work? Does it work? Is it worth it? The first answer describes the **efficacy**, i.e. the proof of principle under ideal study conditions. The second question describes **effectiveness**, the specific proof of function, i.e. suitability for everyday use. The third question describes the **value**, i.e. the value from the point of view of the individual patient and society. The measurement of effectiveness and value only makes sense under everyday conditions. The uniformly defined terminology includes the meaning of the difference between clinical relevance and statistical significance, as well as the definition of „everyday supply". The functions, strengths and weaknesses of Randomized Controlled Trials (RCTs) and Pragmatic Controlled Trials (PCTs), which do not compete with each other but complement each other, are discussed. The data show that RCTs and PCTs support different statements but are comparable in terms of the validity of their statements. What is new is the explicit distinction between the three dimensions, which also makes it clear that psychologic effects (including placebo effects) can be avoided in the detection of the proof of principle under ideal conditions. It is not new that the psychological effects of the doctor-patient relationship, which are difficult to quantify, have a considerable influence on the quality of care.

Keywords

Three-dimensional assessment of health services, Efficacy, Effectiveness, Value, Real World Conditions, Pragmatic Controlled Trial, Observational study, Non-randomized Trial, Saving statistical power

Prof. Dr. med. Franz Porzsolt



wurde am Ontario Cancer Institute in Toronto/ONT, an der Philipps-Universität Marburg und an der Universität Ulm ausgebildet. Er hat für die Entwicklung des Fachs „Klinische Ökonomik" (Nutzen von Gesundheitsleistungen aus Sicht des Patienten und der Solidargemeinschaft) im Jahr 2012 den Deutschen IQ-Preis erhalten. 2013 hat er den wissenschaftlichen Verein „Institute of Clinical Economics (ICE) e.V. gegründet. Kontakt: mindset@clinical-economics.com

Prof. Dr. sc.hum. habil. Christel Weiß



ist Leiterin der Abteilung für Medizinische Statistik, Biomathematik und Informationsverarbeitung der Medizinischen Fakultät Mannheim. Studium der Mathematik und Physik an der Johannes-Gutenberg-Universität in Mainz. 1986-1992 wissenschaftliche Mitarbeiterin am Institut für Experimentelle Chirurgie der Karl-Ruprechts-Universität Heidelberg. 1991 Promotion. 2011 Erlangung der Venia Legendi für die Fächer Biomathematik und Epidemiologie. Kontakt: christel.weiss@medma.uni-heidelberg.de

Prof. Dr. med. Manfred Weiß MBA



ist Oberarzt am Universitätsklinikum Ulm, Klinik für Anästhesiologie, Ulm und Mitglied des Institute of Clinical Economics (ICE e.V.), Ulm. Er hat einen MBA in Betriebswirtschaft für Ärzte; und ist Mitglied im Wissenschaftlichen Arbeitskreis Intensivmedizin im Forum Epidemiologie/Ethik sowie im Translationalen Intensivmedizinischen Forschungsnetzwerk (TIFOnet) in der Forschungsgruppe Immunologie. Kontakt: manfred.weiss@uniklinik-ulm.de

Dipl.-Kfm. Albert G. Müller



ab 1986 mehr als 30 Jahre tätig beim Universitätsklinikum Ulm als stellvertr. Vorstand und Leiter Finanzen, in diesen Funktionen Mitglied bei der baden-württembergischen Schiedsstelle Krankenhaus, bei der bw. Krankenhausgesellschaft und im Verband der Universitätsklinika Deutschlands

Kontakt: a.g.mueller.bc@t-online.de

Dr. med. Susanne Isabel Becker MPH postgrad



baut derzeit als Chefärztin die Klinik für Seelische Gesundheit am MZG Bad Lippspringe auf und ist Mitglied des Institute of Clinical Economics (ICE) e.V.. Die Ausbildungen zur Epidemiologin und Fachärztin für Psychiatrie und Psychotherapie hat sie an der Johann Wolfgang Goethe-Universität in Frankfurt am Main, der Ludwig-Maximilians-Universität in München und dem IAK-KMO absolviert. Kontakt: s.becker@medizinisches-zentrum.de

Prof. Martin Eisemann PhD



ist Inhaber des Lehrstuhls für Psychologie an der UiT Arctic University of Norway in Tromsø. Er ist Gutachter u.a. bei der Europäischen Kommission, Gastprofessor an zahlreichen Universitäten und Verfasser zahlreicher wissenschaftlicher Publikationen.

Kontakt: martin.eisemann@uit.no

Prof. Dr. Robert M. Kaplan PhD



war Chief Science Officer der US Agency for Health Care Research and Quality (AHRQ) und stellvertretender Direktor für Verhaltens- und Sozialforschung am National Institute of Health. Er ist Distinguished Professor an der UCLA, Apl. Professor of Medicine in Stanford, war Past-President fünf professioneller Organisationen, u.a. der ISOQOL.
Kontakt: bob.kaplan@stanford.edu