

Patientenrelevante Endpunkte und *patient-reported outcomes* in klinischer Forschung und medizinischer Praxis

Ingrid Mühlhauser, Hardy Müller

Prolog

Die Bemühungen des Gesundheitssystems und der gesetzlichen Krankenversicherung richten sich nach § 1 SGB V auf den Erhalt, die Wiederherstellung der Gesundheit und die Linderung von Krankheiten. Der Begriff der Gesundheit wird bewusst nicht näher definiert. Um Gesundheit aber erfassen und bemessen zu können, werden Indikatoren oder Substrate zur modellhaften Abbildung von Gesundheit notwendig.

Die Auswahl dieser Parameter, an denen der Erfolg von medizinischen Maßnahmen beurteilt wird, ist kultur- und kontextabhängig. Sie wird geprägt einerseits vom Selbstverständnis der Ärzte und von den Interessen anderer Akteure im Medizin- und Gesundheitswesen, andererseits von den Einstellungen, Wünschen und Erwartungen der Patienten und Bürger. In unserer Gesellschaft wirkt immer noch systembeherrschend die traditionelle Rollenverteilung mit dem Arzt als dem Wissenden und väterlich benevolenten Heiler und dem Patienten als dem Unwissenden, der sich mit Demut und Dankbarkeit dem Arzt anvertraut. Gestützt wird dieses Abhängigkeitsverhältnis durch die Ausblendung von Sterben und Tod und den Allmachtzu- und -anspruch der Medizin. Die Hoffnung auf Heilung soll bis zuletzt erhalten werden. Konsequenterweise ist das primäre medizinische Erfolgskriterium der Sieg über Krankheit und Tod, auch wenn dieser unter Vergrößerung und Verlängerung von Leid erkaufte werden muss. Es reichen der Verdacht und die Vermutung, dass durch ärztliches Handeln das genannte Ziel erreicht werden könnte. Am Umgang mit Krebserkrankungen lässt sich dies beispielhaft illustrieren. Auch noch nach aktuellen Behandlungsempfehlungen definiert vornehmlich der Arzt Ziele und Wege der Diagnostik und Behandlung. Die Aufgabe des Patienten reduziert sich auf das Befolgen der ärztlichen Anweisungen (1), das Gesundheitssystem hat Strukturen und finanzielle Mittel für den reibungslosen Ablauf zu garantieren.

Das Ideal des ewig jung bleibenden, funktionstüchtigen und lebensbejahenden vitalen Menschen hat zu einer idealtypischen Normierung der als gesund

geltenden Körperfunktionen und -zustände geführt. Abweichungen von diesen statisch definierten Normzuständen gelten als behandlungsbedürftig und legitimieren medizinische Maßnahmen, selbst bei Menschen, die sich gesund fühlen. Als Beispiele seien hier die über die letzten Jahrzehnte wiederholten und immer weiteren Absenkungen von Grenzwerten für normalen Blutzucker, Cholesterin oder Blutdruck genannt. So werden immer mehr Gesunde, oft von einem Tag zum anderen, zu Patienten mit Diabetes, Fettstoffwechselstörungen oder Bluthochdruck, ohne dass sich etwas an ihren Körperfunktionen oder Risiken verändert hätte. Der Vitalitätsstatus der Bevölkerung wird damit zu einem Produkt von Norm-Ausschüssen.

Auch Befindlichkeitsstörungen wie Schwitzen, Bauchgrimmen, Unterleibsziehen, Aufstoßen, Beinkribbeln oder Persönlichkeitsmerkmale wie sexuelle Unlust, Lebhaftigkeit, Melancholie, Lustlosigkeit, Schüchternheit oder Ängstlichkeit wurden zunehmend zu behandlungsbedürftigen Erkrankungen erklärt und erlangten somit den Rang patientenrelevanter Ergebnisparameter. Im Extremfall wird dies zum gesellschaftlichen Dogma wie der Anspruch auf ein Leben und Sterben ohne depressive Verstimmungen und mit völliger Schmerzfreiheit. Das Erfinden von Krankheiten hat inzwischen einen Namen erhalten, *disease mongering* (2). Die Industrie und medizinische Leistungsanbieter nutzen dies als Strategie zur Vermarktung von alten und neuen Therapieverfahren. Anwendungsbereiche werden ins öffentliche Bewusstsein manövriert und gleichzeitig die Produkte und Leistungen als Lösungen für diese Probleme angeboten.¹

Die Kulturabhängigkeit medizinischer Ergebnisparameter wird auch an den intensiven Bemühungen von Gesundheitspolitik und Ärzteschaft sichtbar, Krankheiten durch Vorsorge und Früherkennung verhindern zu wollen oder besser behandeln zu können. Hier gelten schon alleine die Teilnahme an Gesundenuntersuchungen und das Aufdecken von Risiken oder vermeintlichen Krankheitsvorstufen als Erfolgskriterien (4).

Der mit Anfang der 90er Jahre eingeleitete Paradigmenwechsel von der Erfahrungsmedizin zu einer wissenschaftsbasierten Medizin (evidenzbasierte Medizin, ebM) ist mit einer Änderung des Rollenverständnisses sowohl der Leistungsanbieter als auch der Patienten verbunden (5). Die ebM sieht ausdrücklich die Einbeziehung der Patienten in medizinische Entscheidungen vor. Vorstellungen und Werte der Patienten sollen berücksichtigt werden. Die Konzepte der partizipativen Entscheidungsfindung (*shared decision making*)

1 Das *disease mongering* läuft nach einem einfachen Schema ab. In einem Beitrag zum 1. April 2006 hatte Moynihan im renommierten British Medical Journal eine neue Krankheit MoDeD (*motivational deficiency disorder*) propagiert und vorgestellt, einschließlich Diagnose, medikamentöser Therapie und Kosten bei Nichtbehandlung (3). Die in kabarettistischer Manier erfundene Erkrankung wurde jedoch vom Publikum, vor allem den Ärzten und den Medien, dankbar und enthusiastisch aufgenommen. Die Aufdeckung des Ganzen als Aprilscherz wurde später nur schleppend und ungläubig akzeptiert.

(6) und der evidenzbasierten informierten Entscheidung (*evidence-based informed decision making*) (7-14) durch Patienten und Bürger lösen das Konzept der einseitigen alleinigen Entscheidungshoheit durch die Ärzte ab. Diese Entwicklungen werden auch den zunehmenden Ansprüchen der Patienten und Bürger gerecht, die bessere Informationen und Einbeziehung in medizinische Entscheidungen fordern (15-17).

Die Einsicht in die Notwendigkeit einer rationalen Medizin im Sinne der ebM wird befördert durch das Ende scheinbar uneingeschränkter finanzieller Ressourcen. Unter diesen Bedingungen verlieren die Normierung biologischer Parameter oder die Verlängerung der Lebenszeit um jeden Preis an Gewicht. Ersetzt werden diese durch Behandlungs- und Gesundheitsziele, die mit wissenschaftlich fundierten Verfahren angestrebt werden. Die Behandlungsverfahren wie die angestrebten Behandlungsziele sind unter Berücksichtigung der Patientenpräferenzen zu wählen. Die patientenbasierten Indikatoren (*patient-reported outcomes*) ersetzen bzw. ergänzen zunehmend die klinisch berichteten Ergebnisparameter (*clinic-reported outcomes*). Die Aktivitäten müssen sich einer Kosten-Nutzen-Bewertung stellen.

Die medizinische Forschung hat diesen Wandel aufgegriffen. Es gibt kaum noch ein Krankheitsbild, zu dem sich nicht interdisziplinäre Arbeitsgruppen formiert haben, die patientenrelevante Ergebnisparameter definieren und Methoden zu deren Erhebung entwickeln und überprüfen (18-25). *Outcome assessment, patient-reported outcomes, quality of life, patient participation, shared decision making, patient empowerment* und *patient information* sind mittlerweile etablierte Forschungsfelder. Diese Entwicklung wurde nicht zuletzt auch durch die Pharmaindustrie forciert, deren neue Produkte häufig übersteuert, aber nur selten echte Innovationen mit klar erkennbarem Nutzen sind (26-28). Wenn schon keine Verringerung der Morbidität oder Mortalität nachzuweisen ist, dann soll zumindest die Verbesserung von Lebensqualität oder Patientenzufriedenheit die Zusatzkosten rechtfertigen (29).

Rahmen und Schwerpunkte des vorliegenden Artikels

Mit dem vorliegenden Artikel werden beispielhaft aktuelle Entwicklungen und Kontroversen um patientenrelevante Endpunkte aufgezeigt. Spezifische Aspekte sind abhängig vom jeweiligen Krankheits- bzw. Beschwerdebild. Die Auswahl der Beispiele wird durch die praktischen und wissenschaftlichen Arbeitsfelder der Autoren geleitet, insbesondere der inneren Medizin mit Schwerpunkt Diabetologie und den Gesundheitswissenschaften sowie dem Versorgungsmanagement in der gesetzlichen Krankenversicherung. Letztlich wird deutlich werden, dass für jede einzelne Problemstellung, sei es im Rahmen eines Forschungsprojekts oder der Betreuung individueller Patienten in

der Praxis, die Frage nach relevanten Zielparametern neu gestellt und die jeweils angemessene Methode zur Überprüfung des Erfolgs definiert werden muss. Da patientenrelevante Ergebnisparameter für klinische Forschung und (Versorgungsmanagement-)Praxis eine große Überschneidung zeigen, werden die exemplarischen Darstellungen nicht generell nach diesen Einsatzbereichen getrennt. Die jeweiligen Besonderheiten werden jedoch deutlich gemacht.

Forschung und Praxis

Unser Gesundheitssystem ist nach wie vor dazu gedacht, Patienten mit Krankheiten zu behandeln und Leiden zu mildern. Die Wirtschaftsförderung der Gesundheitsanbieter oder der Pharmaindustrie gehört nicht zu den primären Zielen des Gesundheitssystems. Es geht letztlich um die Beschwerden, die die Patienten zum Arzt führen. Erfolg misst sich zwangsläufig auch an der Einschätzung der Patienten, in welchem Ausmaß ihnen geholfen wurde. Die Literatur ist reich an Analysen, die belegen, dass die Bedürfnisse der Patienten durch die Ärzte nicht ausreichend wahrgenommen oder Beschwerden nicht richtig gedeutet werden (20, 30, 31). Ähnlich wie beim Begriff Gesundheit wird auch zur Definition von Krankheit auf Labor- und Körperparameter zurückgegriffen, um damit die Beschwerden oder den Gesundheitszustand der Patienten zu erfassen. Die Erfahrung zeigt, dass diese Modellierungen oft fehlerhaft waren, daher ist eine Rückbesinnung auf die von den Patienten vorgebrachten Beschwerden und Symptome notwendig (32).

Wenn es um Tertiär-, Sekundär- oder Primärprävention geht, sind neben Beschwerden und Symptomen der Patienten Morbiditäts- und Mortalitätsergebnisse relevante Erfolgsparameter.

Die amerikanische (Federal Drug Administration, FDA) und die europäische Zulassungsbehörde für Medizinprodukte (European Medicines Agency, EMA) stellen für Produktanbieter Dokumente zur Durchführung von klinischen Studien zur Verfügung. Dazu finden sich auch, zum Teil krankheitsspezifisch, Anleitungen und Begründungen zur Auswahl von validen Ergebnisparametern (33–38). Auch das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG), das die wissenschaftliche Beweislage für spezielle Fragestellungen identifiziert und bewertet, definiert in seinem Methodenpapier, was relevante Endpunkte sind und wie diese erhoben werden sollen (39a). Im Folgenden soll auf einige Aspekte vertieft eingegangen werden.

Patienten-Bewertungen als Endpunkte

An der Notwendigkeit der Beurteilung von Behandlungsergebnissen durch die Patienten selbst gibt es heute keine ernsthaften Zweifel mehr. Die amerikanische Arzneimittelbehörde FDA hat 2006 *patient-reported outcomes* (PRO) als Kriterium zur Bewertung von Arzneimitteln zugelassen (38).

Auch der in Deutschland von Patienten und von politischer Seite geforderte Ausbau der Patientenorientierung des Gesundheitssystems zeigt Konsequenzen. Patientenvertreter sind mittlerweile im Gemeinsamen Bundesausschuss vertreten, dem Gremium, das über den Umfang der Leistungen in der gesetzlichen Krankenversicherung wacht. Zwar noch ohne formales Stimmrecht, hat die Stimme der Patientenvertreter jedoch Gewicht. Bei den Krankenkassen ist im Zuge der freien Kassenwahl durch die Versicherten der Kunde ins Zentrum der Aufmerksamkeit gerückt. Krankenkassen definieren sich als »Anwalt ihrer Versicherten«. Anregungen zur Bewertung des medizinischen *Outcomes* durch die Betroffenen als »Experten in eigener Sache« kamen gerade auch von Krankenkassen-Seite (39b). So ist es politisch gewollt und von den Krankenkassen erwartet, dass die Versicherten auch Aussagen über medizinische Abläufe und Ergebnisse treffen (39c). Letztlich ist es die Patientenzufriedenheit, die die öffentliche Qualitätsanmutung und -rezeption der Leistungsanbieter begründet. In den USA wurde dieser Entwicklung durch verstärkte wissenschaftliche Bemühungen zur methodischen Absicherung der PRO entsprochen. In den Jahren 2004–2008 wurden 37 Millionen Dollar für Forschungsvorhaben im Zusammenhang mit dem *Patient-Reported Outcomes Measurement Information System* (PROMIS) aufgewandt (<http://nihroadmap.nih.gov/>).

Die methodischen Probleme der Erhebung und Interpretation sind jedoch erheblich und Gegenstand weltweiter Forschung. In klinischen Studien, die unterschiedliche medizinische Verfahren vergleichen, haben *patient-reported outcomes* bereits eine große Bedeutung (40a). Hingegen ist die Interpretation der Beurteilung von Behandlungsergebnissen durch einzelne Patienten in der täglichen Praxis durch eine Vielzahl von Einflüssen erschwert. Das Ergebnis selbst und die Deutung des Ergebnisses durch Arzt und Patient sind kontextabhängig. Hier werden sämtliche Faktoren wirksam, die üblicherweise unter Placeboeffekten zusammengefasst werden (40b). Im Einzelfall ist es daher vielfach nicht möglich zu beurteilen, ob ein gewünschtes gesundheitliches Ergebnis durch oder trotz therapeutischer Maßnahmen erfolgte. Genesung, die auch ohne medizinische Intervention eingetreten wäre, kann auf diese Art fälschlicherweise der erfolgten Behandlung zugeschrieben werden. Tatsächlich handelt es sich in einem solchen Fall um Überbehandlung mit allen negativen Auswirkungen. »Wer heilt, hat Recht« ist nur bedingt richtig, nämlich dann, wenn Genesung ohne Therapie nicht hätte stattfinden können. Anderer-

seits könnte ein Behandlungsergebnis, mit dem der Patient nicht zufrieden ist, durchaus als Erfolg zu werten sein, wenn ohne Intervention die Krankheit noch schlechter verlaufen wäre.

Aus diesen Gründen ist auch »Zufriedenheit des Patienten mit dem Behandlungsergebnis« als Ergebnisparameter, vor allem in der Praxis, nur bedingt nutzbar (41). Patienten können zufrieden sein, obwohl sie Besseres hätten erwarten können, und umgekehrt unzufrieden sein, obwohl Bestmögliches erreicht wurde. Als Beispiele seien genannt die Patientin, die nach einer Schilddrüsenoperation mit heiserer Stimme die goldenen Hände des Chirurgen lobt und nicht weiß, dass ihre Heiserkeit Folge einer vermeidbaren Nervenverletzung durch den Eingriff ist, oder die unzähligen, nicht wissenschaftsbasierten Diätvorschriften bei Diabetes. Das Zuckerverbot für Zuckerkrankte hat sich dermaßen in das gesellschaftliche Bewusstsein eingepreßt, dass Betroffene sich dem vermeintlich Unvermeidbaren lange Zeit fügten und es außerhalb des Möglichen und Vorstellbaren sahen, Unzufriedenheit über noch so prokrustische Ernährungsmaßnahmen zu äußern (11). Hingegen werden Einschränkungen durch Diät vorrangig genannt, wenn Betroffene nach ihrer Lebensqualität befragt werden (42). Patientenzufriedenheit als aussagekräftiger und nutzbarer Ergebnisparameter setzt Informiertheit und angemessene Erwartungen der Patienten voraus (41). Die Betroffenen müssten im Voraus verstehen können und wollen, welche medizinischen Leistungen für ihr Problem angeboten werden, welche Erfolge sie von einer bestimmten Maßnahme erwarten können und mit welchen Misserfolgen und unerwünschten Ergebnissen sie zu rechnen haben. Voraussetzungen hierfür sind umfassende, ausgewogene und verständliche Informationen nach Kriterien für evidenzbasierte Patienteninformationen (14, 43–49) sowie die Sicherstellung, dass die Betroffenen die Informationen nutzen können. Damit Patienten über medizinische Behandlungen und Ergebnisse urteilen können, müssen sie über den Sachverhalt informiert sein. Die Gesundheitsbildung der Betroffenen ist eine notwendige Voraussetzung für ein Urteil (*voting*) durch die Patienten (41, 50–54). Diese Bedingungen sind im Augenblick nicht erfüllt.

Allgemeine (*generic*) und spezifische (*disease-specific*) Lebensqualität (*quality of life*) mit ihren Dimensionen zur physischen, mentalen und sozialen Funktionsfähigkeit sind heute in der klinischen Forschung anerkannte Ergebnisparameter. Für viele Einsatzbereiche fehlen jedoch ausreichend validierte Instrumente.

Die Mängel der Erhebungsmethoden zur Lebensqualität, Patientenzufriedenheit und den *patient-reported outcomes* bieten reichlich Spielraum für Missbrauch. Als Beispiel sei hier der Verkaufserfolg der Insulinanaloga genannt. Für diesen Insulinersatz ist ein klinischer Zusatznutzen in qualitativ hochwertigen Studien nicht oder kaum nachweisbar (55, 56). Die meisten Studien wurden jedoch im nicht verblindeten Studiendesign durchgeführt. Die

Vermarktung erfolgte mit dem Versprechen auf freie Diät und infolge davon besserer Lebensqualität (12). Freie Diät und flexibler Lebensstil waren aber auch schon zuvor mit Insulin möglich (11, 42, 57, 58). Effekte auf Patientenzufriedenheit und Lebensqualität in offenen Studien sind nicht unerwartet, aber dann nicht kausal auf neue Behandlungen zurückzuführen.

Mortalität

Krankheitsspezifische Mortalität wird häufig in klinischen Studien als primärer Endpunkt verwendet. Dies kann zu Fehlinterpretationen führen. Deshalb empfehlen auch FDA und EMEA heute, bereits für die meisten medizinischen Maßnahmen auch Daten zur Gesamtmortalität zu erheben (35, 37). Die Sterblichkeit an allen Todesursachen ist aussagekräftiger, da es zahlreiche Beispiele gibt, bei denen durch medizinische Maßnahmen zwar eine Abnahme krankheitsspezifischer Mortalität, aber auch eine Zunahme an anderen Todesursachen berichtet wurde. Diskrepanzen können real sein oder durch Verzerrungen der Datenerhebung zustande kommen. So werden manche Todesfälle fälschlicherweise nicht der Krebserkrankung zugeordnet, wenn sie im Zuge von Operationen von Tumoren auftreten, die durch Früherkennungsuntersuchungen festgestellt wurden (59). Fast ein Viertel aller Menschen stirbt an einer Krebserkrankung. Sehr viel mehr Menschen haben mindestens eine Krebsdiagnose im Laufe des Lebens erhalten, manche können auch mehrere Krebserkrankungen gleichzeitig haben, was ebenfalls die Zuordnung der Todesursache erschwert. Bedeutender ist das Phänomen, dass manche Behandlungen zwar die Sterblichkeit an der primären Krankheit vermindern, dass durch unerwünschte Wirkungen jedoch andere Todesursachen zunehmen.

Bei der Behandlung in der Praxis sind Morbidität und Mortalität zwar unmittelbar für die Betroffenen relevant, können jedoch nur mit Einschränkungen im Einzelfall kausal interpretiert werden. Für häufige Ereignisse, wie Herzinfarkt, kann ein eindeutiger Bezug zu einer bestimmten Behandlung in der Praxis nur schwer hergestellt werden.

Krankheitsspezifische und Gesamt-Sterblichkeit zählen zu den sogenannten harten Endpunkten, da sie relativ verlässlich, unabhängig von subjektiven Einschätzungen durch Patienten oder andere Befunder, erhoben werden können. Harte Endpunkte wären auch klinisch manifeste Knochenbrüche, Erblindung, Amputationen oder Organversagen. Sie werden den sogenannten weichen Endpunkten gegenübergestellt. Zu diesen zählen auch die sogenannten Surrogate.

Surrogate

Surrogate werden als Ersatzparameter für patientenrelevante Endpunkte wie Morbidität oder Mortalität sowohl in klinischen Studien als auch bei der Behandlung individueller Patienten benutzt. Nach dem Prinzip des *pars pro toto* erheben sie den Anspruch, einerseits einfach und forschungsökonomisch günstig generierbar zu sein und andererseits mit großer Sicherheit die sonst schwer erfassbaren relevanten Endpunkte zu repräsentieren. Gerade auch für leistungsorientierte Vergütung werden sie inzwischen eingesetzt (60–62).

Surrogate sind zum Beispiel Stoffwechselfparameter wie Blutzucker oder Cholesterinwerte, Befunde auf Röntgenbildern, Funktionswerte von Organen, Knochendichte, Ergebnisse von Gewebeuntersuchungen oder auch die Zeit bis zum Nachweis eines Tumorrezidivs. Da es oft Jahre dauert, bis Behandlungen Ergebnisse auf Morbidität oder Mortalität zeigen, wird versucht, stattdessen biologische Parameter zu nutzen, von denen man meint, dass sie die tatsächlich relevanten Langzeitergebnisse vorhersagen können. Blutdruck-, Cholesterin- oder Blutzuckerwerte werden zum Beispiel als Ersatz für Schlaganfall- oder Herzinfarktereignisse benutzt. Auch der Arzt in der Praxis geht davon aus, dass ein gut eingestellter Blutdruck letztlich Herz-Kreislauf-Komplikationen verhindern kann. Der Blutdruck selbst oder erhöhte Cholesterinwerte machen selten Beschwerden. Die Frage ist, ob Blutdruck, Blutzucker oder Verkalkungen auf Röntgenbildern tatsächlich für den Patienten relevante Endpunkte sind.

Die Anwendung von Surrogaten in der bisherigen Form ist jedenfalls in den letzten Jahren zunehmend in die Kritik geraten (63a–65). In klinischen Studien ist ihre alleinige Anwendung höchst problematisch und wissenschaftlich kaum zu rechtfertigen. Anhand von Beispielen für sogenannte Surrogat-Trugschlüsse soll dies erläutert werden. Von Surrogat-Trugschlüssen spricht man, wenn bestimmte medizinische Interventionen zwar erwünschte Effekte auf das Surrogat haben, jedoch keine oder sogar gegenteilige auf die für Patienten relevanten Endpunkte (63a, 63b). Eine Übersicht über solche Beispiele gibt die Tabelle (S. 52 ff.). Eine aktuelle Analyse hat gezeigt, dass zur Bewertung der Relevanz von medizinischen Maßnahmen weltweit Surrogate in »*Health Technology Assessment (HTA)*«-Berichten so gut wie nicht benutzt werden und deren Anwendung auch nicht empfohlen wird (64). Die Frage ist trotzdem berechtigt, ob es nicht verlässliche Surrogate gibt, die längere und größere Studien eventuell überflüssig machen könnten. Was einen validen Surrogatparameter kennzeichnet, ist daher von Interesse. Die bereits erwähnte systematische Analyse der Bedeutung von Surrogaten kommt zu dem Schluss, dass Surrogate nur ausnahmsweise und nur unter definierten Bedingungen zur Bewertung von neuen Therapieverfahren herangezogen werden könnten (64).

Von FDA und EMEA waren bisher nur wenige Surrogate als Ergebnisparameter zur Bewertung von Arzneimitteln für metabolische bzw. kardiovaskuläre Einsatzgebiete anerkannt. Dazu zählten das glykosylierte Hämoglobin (HbA1c), LDL-Cholesterin und Blutdruck (33, 36). Der Druck auf die Arzneimittelbehörden durch die Industrie nimmt jedoch zu, da diese die Anerkennung weiterer Biomarker fordern. Unter den Schlagworten der individuellen Risikovorhersage mit Biomarkern und darauf abgestimmten Angeboten individuell angepasster Therapieformen sollen Medikamente schneller zugelassen und auf kontrollierte Studien mit patientenrelevanten Endpunkten verzichtet werden können (26). Ein solches Vorgehen würde jedoch die Fehler, die bereits bisher mit den biochemischen und klinischen Surrogaten gemacht wurden, wiederholen (63b).

Exkurs: HbA1c als Surrogat-Parameter

Am Beispiel des HbA1c-Wertes soll skizziert werden, dass selbst die bisher anerkannten Surrogate infrage gestellt werden müssen. Der HbA1c-Wert gilt als Gedächtnis der Blutzuckerwerte über die zurückliegenden drei Monate. Bei HbA1c-Werten unter etwa 8 Prozent bestehen typischerweise keine klinischen Beschwerden durch erhöhte Blutzuckerwerte, und nur bei über Jahre bestehenden hohen Blutzuckerwerten treten Folgeschäden auf, im Wesentlichen an Augen, Nieren und Nerven. Das deutlich erhöhte kardiovaskuläre Risiko, das die Prognose von Personen mit Diabetes Typ 2 wesentlich bestimmt, besteht bereits bei Diagnose oder nur kurzer Diabetesdauer. Bei Diagnose im höheren Lebensalter sind blutzuckerabhängige Spätkomplikationen kaum noch zu erwarten, und der Diabetes hat so gut wie keinen Einfluss mehr auf die Lebenserwartung (24). Neue Analysen zeigen zudem, dass im Alter auch bei normalen Blutzuckerwerten die HbA1c-Werte ohne Krankheitswert höher liegen (66). Trotzdem sind in den vergangenen Jahren die Zielwerte für Blutzucker und HbA1c unabhängig vom Alter immer weiter gesenkt worden. Aktuelle Leitlinien der Amerikanischen Diabetesgesellschaft definieren das Behandlungsziel für Personen mit Typ-2-Diabetes generell als »to achieve and maintain glycemic levels as close to the non-diabetic range as possible and to change interventions at as rapid a pace as titration of medications allows« (67). Als Folge werden heute auch alte Menschen intensiviert unter Umständen mit täglichen Blutzuckerselbstmessungen und mehrfach täglichen Injektionen behandelt.

In den letzten beiden Jahren sind mehrere randomisiert-kontrollierte Studien mit Beobachtungszeiten bis zu 6 Jahren abgeschlossen worden, die diese Zielvorgaben und Verfahren bei Patienten mit Typ-2-Diabetes untersucht haben. Keine dieser Studien konnte die gewünschten Effekte auf Morbidität oder

Mortalität nachweisen. Im Gegenteil: Die in den Leitlinien geforderte Polypragmasie mit dem Ziel der Normalisierung von HbA1c-Werten um jeden Preis für fast alle Patienten war in einer der Studien sogar mit einer Zunahme der Gesamtsterblichkeit assoziiert, was zum vorzeitigen Abbruch der Studie führte (68). Gänzlich unerwartet war in einer weiteren Langzeitstudie trotz erheblich verbesserter HbA1c-Werte weder ein Einfluss auf kardiovaskuläre Morbidität und Mortalität zu finden noch auf die blutzuckerabhängigen Schäden an Augen, Nieren und Nerven (69). Nur für wenige blutzuckersenkende Medikamente liegen aussagekräftige Nutznachweise auf patientenrelevante Endpunkte vor. Für den weltweit am häufigsten verordneten Wirkstoff Metformin aus der Gruppe der Biguanide beruht der Nachweis einer Reduktion der Sterblichkeit lediglich auf einer einzigen randomisierten Langzeitstudie, publiziert im Jahr 1998, an einer Subpopulation von etwa 300 Patienten (70). Diese Ergebnisse stehen jedoch im Gegensatz zu einer früheren randomisierten Studie, die vorzeitig abgebrochen werden musste, weil mit einem verwandten Biguanid, dem Phenformin, die kardiovaskuläre Mortalität zunahm (71, 72). Für die in Deutschland am häufigsten verordnete Kombination von Metformin mit Sulfonylharnstoffen liegt lediglich eine Substudie einer randomisierten Studie vor, in der diese Medikamentenkombination mit einer erhöhten Mortalität assoziiert war (73). Für eine Vielzahl neuer Wirkstoffe und Wirkstoffgruppen gibt es entweder gar keine Ergebnisse zu patientenrelevanten Endpunkten oder, sofern solche Untersuchungen durchgeführt wurden, sind die Ergebnisse beunruhigend (74–78). So zeigte sich für die Wirkstoffgruppe der Glitazone, dass Rosiglitazon statt der gewünschten Abnahme zu einer erheblichen Zunahme kardiovaskulärer Komplikationen führt (76), ein verwandter Wirkstoff, Pioglitazon, diese negativen Effekte jedoch nur teilweise zeigt (77). Auch für die Gruppe der weit verbreiteten Sulfonylharnstoffe gibt es keinen sogenannten Klasseneffekt (79). Das bedeutet, dass HbA1c kein valider Surrogat-Endpunkt sein kann. Jeder Wirkstoff muss für das jeweilige Indikationsgebiet anhand patientenrelevanter Endpunkte in aussagekräftigen Langzeitstudien geprüft werden.

HbA1c-Werte werden zum Beispiel im britischen Gesundheitssystem als Parameter zur leistungsorientierten Vergütung benutzt (60, 61). Auch in anderen Bereichen des Versorgungsmanagements wird der HbA1c-Wert als Erfolgskriterium eingesetzt (18, 80). Ein möglichst großer Anteil von Patienten mit Diabetes soll HbA1c-Werte im oder nahe dem Normbereich erreichen. Vor dem Hintergrund der negativen Evidenz für eine generelle Normalisierung von HbA1c-Werten erscheint dieses Vorgehen problematisch. Zudem ist eine isolierte Bewertung ausschließlich eines Laborparameters ohne gleichzeitige Berücksichtigung anderer Komponenten der Behandlung nicht aussagekräftig. Eine Senkung der HbA1c-Werte bei gleichzeitiger Zunahme des Risikos an schweren Unterzuckerungen kann nicht als Erfolg gewertet wer-

den. Auch ist es für das Leben der Patienten nicht gleichgültig, ob die Absenkungen biochemischer Parameter ohne Medikamente, mit einem Wirkstoff oder einer Vielzahl an Substanzen erfolgt (68). Gerade für ältere Menschen oder solche, die an mehreren Krankheiten leiden, ist bei einem nicht diskriminierenden Vorgehen mit nicht vorhersagbaren Medikamenteninteraktionen oder anderen Auswirkungen auf das Befinden der Betroffenen zu rechnen. Dies gilt auch für Surrogate wie Blutdruck oder Blutfettwerte.

Eine isolierte Nutzung biochemischer oder anderer Parameter ist letztlich weder in der Praxis noch in Forschungsprojekten sinnvoll. Eine Blutdrucksenkung kann sowohl bei verminderter oder aber auch bei vermehrter Verschreibung von blutdrucksenkenden Medikamenten erfolgen. Eine HbA1c-Senkung kann mit erhöhtem oder auch vermindertem Risiko für Unterzuckerungen einhergehen (11, 12, 81). Gewichtszunahme kann auch Zeichen von Gesundung sein, zum Beispiel als Folge einer Stoffwechselrekompensation infolge Verbesserung einer sehr schlechten Blutzuckereinstellung oder Behandlung einer Hyperthyreose. Behandlungen können mit Schwächung oder Stärkung der Autonomie und Identität von Patienten einhergehen (82). Neue Bewertungsverfahren von Evidenz zu klinischen Fragestellungen wie GRADE² versuchen dieser Komplexität der Ergebnisse von medizinischen Interventionen gerecht zu werden (83a).

Patientendefinierte Behandlungsziele als Ergebnisparameter

Die Konzentrierung auf biologische Parameter missachtet die persönlichen Präferenzen der einzelnen Patienten (82). Eine Normalisierung von HbA1c-Werten ohne Rücksicht auf Alter und Begleiterkrankungen ist selbst aus Leistungsanbieterperspektive nicht zu rechtfertigen (24). Anstelle von Therapiezielen, die durch Ärzte definiert und anhand von Surrogaten erhoben werden, sollten Behandlungsziele treten, die mit bzw. durch Patienten definiert werden und letztlich auch von den Patienten berichtet werden. Die Ziele sind zwischen Ärzten und Patienten zu verhandeln. Dazu müssen den Betroffenen ausreichende, evidenzbasierte und verständliche Informationen verfügbar gemacht werden (8, 11, 83b). Konsequenterweise ist das Erreichen oder Nicht-Erreichen des gemeinsamen oder auch durch den Patienten letztlich alleine definierten Zieles der primär entscheidende Ergebnisparameter (83b). An diesem sollten Erfolge der medizinischen Maßnahmen abgelesen werden. Weiterer Qualitätsparameter ist der Nachweis darüber, dass der Patient durch entsprechende Aufklärung die Möglichkeit hatte zu verstehen, worum es geht

2 in: GRADE, The Grading of Recommendations Assessment, Development and Evaluation, siehe: <http://www.gradeworkinggroup.org>.

(83b). Vorausgesetzt wird hierbei, dass die Therapieziel festlegung auf Grundlage der aktuell am besten verfügbaren wissenschaftlichen Beweislage im Sinne der ebM erfolgt (84).

Als Positivbeispiel sind die inhaltlichen Regelungen in den deutschen *Disease-Management-Programmen* (»Chroniker-Programme«) für Diabetes hervorzuheben, die durch ihre ebM-basierten Aussagen und Empfehlungen auch immer eine an Symptomen, Alter und Begleiterkrankungen orientierte Festlegung wünschenswerter HbA1c-Werte verfolgen (80, 85). Obwohl die Stoffwechselziele zwischen Arzt und Patient verhandelt und festgelegt werden sollen, erfolgt die Evaluation der Programme allerdings nicht nach diesen individuell definierten Zielen, sondern weiterhin nach erreichten HbA1c-Werten in Patientengruppen.

Von Compliance³ und Adhärenz⁴ zur informierten Patientenentscheidung

Compliance oder Adhärenz der Patienten zu ärztlichen Empfehlungen oder Verschreibungen werden sowohl in klinischen Studien als auch in der täglichen Praxis und in der Versorgungsforschung immer noch als Ergebnisparameter benutzt (86–89). Dazu zählen zum Beispiel das Einhalten von Arztterminen, die regelmäßige Einnahme von Medikamenten oder das Befolgen von Lebensstiländerungen. Obwohl solche Parameter zur Erklärung von Effekten nützlich sein können, sind sie keine für Patienten relevante Endpunkte.

Screening und Präventionsmaßnahmen richten sich an Gesunde bzw. sich gesund fühlende Menschen, die durch Aufdeckung und Behandlung von Risikofaktoren eine bessere Lebensprognose erhalten sollen. Besonders in der Primärprävention erleiden durch Screeningmaßnahmen meist mehr Menschen Schaden⁵ als Nutzen (7, 9, 10, 89, 90). Die Ermöglichung einer informierten Entscheidung der Zielgruppe, vor allem bei Krebsfrüherkennung, ist daher als ethisch notwendig definiert worden (53, 54, 90). Trotzdem werden immer noch Ergebnisparameter benutzt, die primär auf hohe Teilnahmequoten und eine hohe Rate an Krebsdiagnosen zielen. Dies sind keine für die Zielgruppe

3 Das Wort *Compliance* (englisch für Befolgung) steht für die Einhaltung von Anordnungen des Arztes durch den Patienten.

4 Adhärenz (englisch: *adherence*) steht ursprünglich für die Einhaltung von Vereinbarungen, die gemeinsam zwischen Arzt und Patient festgelegt wurden, tatsächlich wird damit aber oft *Compliance* gemeint.

5 Als Schaden bezeichnen wir eine falsche bzw. nicht gerechtfertigte Diagnose mit anschließender Behandlung (Überdiagnose und Übertherapie) und einen auch nur zeitweise im Raume stehenden falschen Verdacht auf eine Erkrankung. Die z. B. von den Brustkrebs-Screening-Programmen alleine als Schädigung konstatierte Strahlenbelastung ist dagegen u. E. vergleichsweise gering.

relevanten Ergebnisparameter. Krebsfrüherkennungsmaßnahmen führen zu Überdiagnosen und Übertherapien, das heißt zu einer Zunahme von Krebsvorstufen und Krebsdiagnosen, die sich ohne Screening für den Patienten nicht bemerkbar gemacht hätten. Die Diagnose eines Krebses im Frühstadium oder einer vermeintlichen Krebsvorstufe kann in solchen Fällen einen gesunden Menschen erst krank machen. Aus diesen Gründen ist die sehr beliebte Angabe von 5-Jahres-Überlebensraten nach Krebsdiagnosen kein valider Ergebnisparameter (91). Je mehr Überdiagnosen eingehen, desto besser kann fälschlicherweise die Prognose erscheinen.

Wenn eine informierte Entscheidung möglich sein soll, dann kann die alleinige Teilnahme am Screening oder an anderen Vorsorgeuntersuchungen nicht mehr als patientenrelevanter Ergebnisparameter angesehen werden. Die Nichtinanspruchnahme der Untersuchung ist ausdrücklich vorgesehen (53, 54, 90). Vielmehr wäre ein Indikator für relevante Qualität die Dokumentation, dass die Betroffenen über Vor- und Nachteile des Screenings informiert wurden, dass die Informationen verstanden wurden und dass die Betroffenen eine informierte Entscheidung entsprechend ihren Wertevorstellungen treffen konnten. Ethisch vertretbar ist nicht die möglichst hohe Teilnahmequote, sondern die Angemessenheit des Informations- und Beratungsangebots.⁶

Nicht nur bei Früherkennung und Vorsorge gewinnen informierte Entscheidungen als relevante Endpunkte an Bedeutung, sondern auch bei chronischen Erkrankungen wie Multipler Sklerose (13, 93, 94), psychiatrischen Leiden oder Krebserkrankungen (15, 17). Patientenorganisationen fordern die Bereitstellung von Studienergebnissen in verständlicher Art und Weise als Grundlage für Entscheidungen einschließlich der Option der Nichtinanspruchnahme einer medizinischen Maßnahme (15).

Beispiel Krebserkrankungen: In Studien zur Überprüfung von neuen Krebstherapien werden häufig »rezidivfreies Überleben« oder die Zeit bis zum Tumorrezidiv als primäre Endpunkte definiert. Doch die Behandlungsergebnisse sind komplex. Unerwünschte Nebenwirkungen und die Lebensqualität der Patienten müssten gleichwertig mitberücksichtigt werden. Diese Forderung wird exemplarisch illustriert durch die persönliche Geschichte einer Ärztin, die sich in England an die Spitze der Patientenbewegung zur Finanzierung des Brustkrebsmedikaments Herceptin® gesetzt hatte (92). Als sie später als Betroffene jedoch selbst vor der Entscheidung stand, verstand sie erstmals, was tatsächlich unter der propagierten Verminderung der Rezidivrate mit Herceptin® von 50 Prozent zu verstehen war, nämlich eine Reduzierung der Tumorrezidive bei 5 von 100 Frauen, die die Therapie anwendeten. Bei einer

6 Diese Einschätzung schlägt sich auch im neuen Nationalen Krebsplan des Bundesgesundheitsministeriums nieder. Die ursprünglichen Ziele, bei den Screening-Programmen möglichst hohe Teilnahmequoten zu erreichen, wurden in der Fachdiskussion präzisiert. Mittlerweile wird eine möglichst informierte Entscheidung der Teilnehmer angestrebt.

ebenso großen Anzahl von Frauen kam es jedoch durch die Behandlung zu klinisch relevanter Herzschwäche. Nachdem sie dies verstanden hatte, entschied sie sich gegen die Behandlung. Ihrer Forderung, dass es allen Patienten ermöglicht werden sollte, die Vor- und Nachteile von Therapieangeboten zu verstehen, sollte Gehör geschenkt werden.

In Zukunft werden Parameter, die den Grad von Patienten-*Empowerment*, vor allem bei chronischen Krankheiten, und *health literacy*, einschließlich *critical health literacy*⁷ (50, 51), abbilden, zunehmend an Bedeutung gewinnen. Das Management von Unsicherheit, nicht die Vermeidung von Unsicherheit, wird als Qualitätsparameter verwendet werden (13, 93–95).

Von klinisch relevanten Effektgrößen zu patientenrelevanten Effektgrößen

Üblicherweise werden in klinischen Studien die Stichprobengrößen anhand untersucherdefinierter Ergebnisparameter bestimmt. Die Experten definieren klinisch relevante Unterschiede, die in der Studie nicht übersehen werden sollen. Auch hier sollten die Betroffenen stärker mit einbezogen werden. Klinisch relevante Unterschiede könnten durch patientenrelevante Unterschiede ergänzt oder ersetzt werden. Auch dies setzt ein wissenschaftliches Grundverständnis der beteiligten Patienten voraus (96–100). Dort, wo Unwissenheit herrscht, besteht großes Potenzial für Manipulation. Und ein solches gibt es bei der Planung von klinischen Studien, insbesondere, wenn diese unter Federführung oder Einfluss der Produkthersteller erfolgen, in reichlichem Maße (27, 28, 101).

Skalierungen von Endpunkten

Häufig werden in klinischen Studien zur Ergebnisbewertung Skalen benutzt. Beispiele sind die Hamilton-Depressionsskala oder Skalen zur Bewertung des Demenzgrades. Bei ausreichend großen Stichproben führen bereits geringe Verschiebungen auf den Skalen zu statistisch hoch signifikanten Ergebnissen. Die Relevanz der Effekte ist jedoch trotzdem oft vernachlässigbar. So werden z. B. selbst in Studien bei Patienten mit schwerer Depression, die positive Ergebnisse berichten, im Schnitt nur etwa Unterschiede zwischen Medikamenten und Placebo von 3 bis 5 Punkten auf der 20 Punkte umfassenden Hamilton-Skala erzielt (102). In etwa der Hälfte der Studien gibt es gar keine Unterschiede. Da negative Studien mit Antidepressiva nicht immer publiziert

7 Kritische Gesundheitsbildung verknüpft die Konzepte Gesundheitsbildung und ebM, siehe auch Steckelberg (50, 51).

wurden, konnte sich die Botschaft durchsetzen, dass Antidepressiva bei Depression grundsätzlich wirksam sind (102–104). Millionen von Menschen auch mit nur leichten oder mittelschweren depressiven Verstimmungen werden mit nebenwirkungsreichen Medikamenten behandelt, die offenbar nicht mehr als Placebowirkungen haben (102).

Die Einbeziehung von wissenschaftsgebildeten Patienten in die Definition von relevanten Effekten könnte hier Abhilfe schaffen (105). Ein solches Vorgehen würde möglicherweise auch öffentliche Patiententrüstungen vermeiden, die durch solche zum Teil ja bewusst verbreiteten Missverständnisse über Behandlungseffekte geschürt werden. Ein Beispiel hierfür ist die Empörung der Betroffenen über die Beschränkung der Erstattung von Medikamenten für an Demenz erkrankte Personen durch das britische NICE (106) oder des Deutschen Diabetikerbundes über die Negativergebnisse bei der wissenschaftlichen Bewertung von Insulinanaloga durch das IQWiG bzw. den Gemeinsamen Bundesausschuss (107). Tatsächlich sind die Effekte dieser Medikamente klinisch gar nicht oder kaum relevant. Bei Aufklärung würde dies vermutlich auch durch die Betroffenen erkennbar werden.

Kategorisierung von Ergebnisparametern: ein Klasse(n)-Effekt

Durch Kategorisierung von kontinuierlichen Ergebnisvariablen können Effekte erheblich manipuliert werden. Dies wurde zum Beispiel für den Bereich der Depression oder für die Prävention von Diabetes nachgewiesen. So haben wir in einer eigenen Untersuchung mit Diabetesexperten gezeigt, dass die weit verbreitete Aussage, dass durch Lebensstiländerungen wie gesunde Ernährung und vermehrte körperliche Aktivität bei Hochrisikogruppen das Auftreten von Diabetes mellitus um bis zu 60 Prozent verhindert werden kann, in höchstem Maße irreführend ist. Die tatsächlichen Veränderungen des Blutzuckers von durchschnittlich 5 mg% oder des HbA1c-Wertes von 0,1 Prozent werden bei angemessener Datenpräsentation durch dieselben Diabetesexperten überwiegend als klinisch gar nicht oder wenig bedeutsam eingestuft (108, 109). Täuschungsphänomene durch verzerrte Darstellung wissenschaftlicher Studien sind weit verbreitet, wären aber mit wenig Aufwand und guter Absicht zu verhindern (52, 100).

Sicherheit und unerwünschte Wirkungen

Die meisten klinischen Studien sind primär auf den Nachweis von Wirksamkeit und Effektivität, aber nicht auf Sicherheit ausgerichtet. Die Stichprobengrößen sind daher fast immer zu klein. Es braucht oft viele Jahre, bis rele-

vante unerwünschte Wirkungen nach Zulassung der Medikamente erkannt werden (110). Zudem werden unerwünschte Wirkungen in Publikationen nicht ausreichend berichtet (27, 28, 102). Manche möglichen negativen Effekte werden erst gar nicht in Erwägung gezogen und daher nicht im Studiendesign berücksichtigt. Als Beispiel sind hier Interaktionen mit anderen Medikamenten zu nennen, was vor allem bei alten oder sehr kranken Menschen Bedeutung hat.

Unter Praxisbedingungen kann das Nutzen-Risiko-Verhältnis anders sein als unter Studienbedingungen. Ein Beispiel sind die deutlich höheren Blutungskomplikationen bei Behandlung mit Gerinnungshemmern zur Prävention von Schlaganfall unter Alltagsbedingungen. Diese können jedoch durch Patientenschulung vermindert werden (111). Ein für Patienten relevanter Endpunkt wäre daher die Fähigkeit der Betroffenen, unerwünschte Wirkungen von Behandlungen so weit als möglich zu vermeiden. Hier wird *functional literacy*⁸ bedeutsam, z. B. beim richtigen Dosieren von Medikamenten oder der Selbstanpassung von Insulin zur Vermeidung von Unterzuckerungen.

Tertiäreffekte

Für Patienten bedeutsame Effekte können bestehen, ohne dass diese in klinischen Studien unter experimentellen Bedingungen sichtbar wurden. Sie können auch auftreten, falls sich das Verhältnis von erwünschten und unerwünschten Wirkungen verschiebt. Diese Effekte können positiv oder negativ sein. Bei Fokussierung auf eine neue Medikamentenbehandlung könnten eventuell nichtmedikamentöse Maßnahmen vernachlässigt werden. Dies ist immer wieder in der klinischen Diabetologie zu beobachten. Studien zur Vermarktung von neuen Medikamenten versuchen irrelevante Unterschiede in metabolischen Ergebnissen oder im Risiko für Unterzuckerungen nachzuweisen. Dabei wird systematisch ausgeblendet, dass die Effektgrößen durch Patientenschulung oder etablierte andere Standardverfahren sehr viel größer sein können (12, 42, 58, 112).

Das Hinzufügen einer neuen Behandlung, insbesondere wenn sie aus mehreren Wirkstoffen besteht, kann vor allem bei Personen, die auch noch andere Medikamente einnehmen, zu nicht vorhersehbaren Auswirkungen führen. Wirksame Behandlungen könnten abgesetzt und gegen wenig wirksame ausgetauscht werden. Medikamenteninteraktionen mit unbekanntem Folgen sind möglich. Diese Auswirkungen werden so gut wie nie in klinischen Studien berücksichtigt. Daher muss auch der behandelnde Arzt im Einzelfall ohne ausreichende Evidenz entscheiden.

8 Functional literacy steht für Kompetenzen wie Lesen, Schreiben, Rechnen, die für das Funktionieren im täglichen Leben notwendig sind.

Kombinierte Endpunkte

Kombinierte Endpunkte sind beliebt in klinischer Forschung zur Reduzierung der Studienpopulation und Beobachtungszeit. Die Probleme, die damit verbunden sind, können jedoch erheblich sein. Die EMA rät davon ab, Parameter, die vom Untersucher bzw. Behandler abhängen, wie z. B. Krankenhauseinweisungen, mit harten Endpunkten wie Herzinfarkt oder Mortalität zu kombinieren (35).

Darstellung der Ergebnisse

Häufig werden in klinischen Studien die Ergebnisse als Ereignisse pro Patient-jahre dargestellt. Dies hat den Vorteil, dass unterschiedlich lange Beobachtungszeiten ausgewertet werden können. Nachteil ist die Nichtberücksichtigung der Tatsache, dass Teilnehmer in unterschiedlichem Ausmaß Ereignisse erleiden können. Ein Beispiel ist die Häufigkeit von schweren Unterzuckerungen bei Insulinbehandlung. Selbst bei den besonders gefährdeten jungen Patienten, die gar kein eigenes Insulin mehr produzieren, sind es meist weniger als 20 Prozent, die schwere Unterzuckerungen erleiden. Darunter gibt es aber solche, die nicht nur ein Ereignis pro Jahr haben, sondern mehrere (112). Die Summierung auf einen einzigen Wert von Ereignissen pro Anzahl Patienten pro Jahr wird somit der Bedeutung dieser Nebenwirkung für individuelle Patienten nicht gerecht. Andererseits ist die alleinige Darstellung als Anteil Personen mit Ereignissen nicht ausreichend, da die Gesamtzahl der Ereignisse nicht erkenntlich ist. Es müssen also beide Darstellungen genutzt werden.

Wenn die Studiendauer nur kurz ist, dürfen die Ergebnisse nicht auf längere Zeiträume hochgerechnet werden. Dies ist besonders problematisch, wenn Behandlungen ein Leben lang erfolgen sollen, die Studien aber längstens über einen Zeitraum von 5 bis 6 Jahren erfolgten. Beispiel hierfür ist die Behandlung mit Statinen, einer Substanzgruppe zur Senkung des Cholesterinspiegels.

Auch ist es für die Patienten nicht gleichgültig, wann Ereignisse auftreten. Bei der Behandlung mit Sexualhormonen in der Menopause treten zum Beispiel Thrombosen schon innerhalb der ersten Monate nach Beginn einer Hormonbehandlung vermehrt auf. Bei älteren Frauen gibt es auch schon im ersten Behandlungsjahr häufiger Herz-Kreislauf-Komplikationen. Wenn bei Ende der Studie nach 4 Jahren trotzdem nicht mehr Herzinfarkte als unter Placebo gezählt wurden, bedeutet das jedoch nicht, dass die Behandlung unschädlich war (113). Die von den negativen Effekten betroffenen Frauen sind dadurch schon früher verstorben oder hatten einen Herzinfarkt, den sie ohne Hormone erst später bekommen hätten.

Kosten

Auch die Kosten für die Gesundheitsversorgung erlangen zunehmend den Status patientenrelevanter Endpunkte. Auf Praxisebene in Deutschland sind aktuell bedeutsam die »individuellen Gesundheitsleistungen«, die sog. IGeL-Angebote (114). Auch hier bedarf es einer Grundausstattung an *critical health literacy* der Bürger, um diese bewerten zu können. Durch Beteiligung von Patientenvertretern an der Erstellung von wissenschaftlichen Gutachten im IQWiG und an Entscheidungen des Gemeinsamen Bundesausschusses definieren auch Betroffene die Bedeutung von patientenrelevanten Endpunkten im Verhältnis zu den Kosten. Eine durch wissenschaftliche Basiskompetenzen gestärkte Beteiligung von Patientenvertretern in ethischen Komitees für klinische Studien wäre ebenfalls höchst wünschenswert (96, 105).

Kosten für Patienten entstehen auch durch deren Teilnahme an den zahlreichen primär marketingorientierten Studien. Patienten können den fehlenden Nutzen dieser Studien nicht erkennen (105). Die finanziellen Mittel für die Weiterbildung von Patientenvertretern in Ethikkommissionen oder der Bürger insgesamt stehen jedoch nicht zur Verfügung.

Fazit

Erhalt, Wiederherstellung oder Besserung der Gesundheit zählen zu den primären Zielen des Gesundheitswesens. Der Begriff der Gesundheit bleibt abstrakt. Zur Erfassung von Gesundheit und der Gestaltung der Rahmenbedingungen sind Indikatoren und Ergebnisparameter notwendig.

Die Auswahl von Ergebnisparametern muss zukünftig besser begründet werden, sowohl für die Anwendung in klinischer Forschung als auch in der Praxis. Patienten sollten verstärkt in die Auswahl und Definition von Endpunkten einbezogen werden. Methoden zur Nutzung von Patientenbewertungen als Ergebnisparameter müssen verbessert werden. Bewertungen von Ergebnissen durch Patienten müssen Informations- und Verständnisstand der Betroffenen berücksichtigen. Surrogate alleine sind keine aussagekräftigen Parameter in klinischen Studien. Tatsächliche Messwerte sind Kategorisierungen vorzuziehen. Die Relevanz von Veränderungen klinischer Parameter, insbesondere wenn diese mit Skalen gemessen werden, muss vor deren Einsatz bestimmt werden (115). In klinischen Studien ist die Gesamtmortalität krankheitsspezifischer Mortalität überlegen und sollte zumindest immer mitberücksichtigt werden. Lebensqualität als alleiniger Erfolgsparameter erfordert eine Verblindung der Interventionen. Die Zeithorizonte sind zu berücksichtigen.

Tabelle Beispiele für Surrogat-Trugschlüsse (aktualisiert nach Tabelle 1 aus 63; Seite 1 von 5)

Präventionsziel (Endpunkt)	Surrogat-Parameter	Studien-population	Studien-dauer	Intervention	Surrogat-Ergebnis	Endpunkt-Ergebnis	Referenzen
Menopausale Hormontherapie							
Prävention von Herzinfarkt	Cholesterin-/Blutzuckerspiegel	HERS (1663 Frauen, 55 bis 80 Jahre, mit koronarer Herzkrankheit), WHI (16.608 Frauen)	HERS 4 Jahre; WHI vorzeitiger Abbruch nach ca. 5 Jahren	Östrogene, Gestagene	Senkung von Cholesterin- und Blutzuckerspiegel	Zunahme von Herzinfarkt, Schlaganfall, Lungenembolie	113, 116
Prävention von Demenz	Neuronale Funktion	WHI – wie oben	Wie oben	Östrogene, Gestagene	Neuroprotektiv	Zunahme von Demenz, Verschlechterung der Kognition	117, 118, 119
Prävention von Harninkontinenz	Trophik der Urogenital-Schleimhaut	WHI – wie oben	Wie oben	Östrogene, Gestagene	Verbesserung der Trophik der Schleimhaut	Zunahme und Verschlechterung von Inkontinenz	120a
Osteoporose							
Verminderung osteoporotischer Frakturen	Knochendichte (Densitometrie)	202 Frauen, postmenopausal mit Osteoporose	4 Jahre	75 mg NaF plus 1500 mg Calcium pro Tag vs. Placebo und 1500 mg Calcium pro Tag	Zunahme der Knochendichte	Zunahme der Extravertebralfrakturen um (72 vs. 24, p < 0,01) in der NaF-Gruppe	120b

Tabelle Beispiele für Surrogat-Tragschlüsse (aktualisiert nach Tabelle 1 aus 63; Seite 2 von 5)

Präventionsziel (Endpunkt)	Surrogat-Parameter	Studienpopulation	Studiendauer	Intervention	Surrogat-Ergebnis	Endpunkt-Ergebnis	Referenzen
Anämie							
Verbesserung von Leistungsfähigkeit bei Krebskranken	Hämoglobinwerte	Meta-Analyse		Erythropoietin	Zunahme der Hämoglobinwerte	Zunahme der Mortalität	121
Verbesserung von Leistungsfähigkeit bei Niereninsuffizienz	Hämoglobinwerte	Meta-Analyse		Erythropoietin	Zunahme der Hämoglobinwerte	Zunahme der Mortalität	122
Vitamine							
Verhinderung von Lungenkrebs	Vitamin-Serumspiegel	29.133 Männer, Raucher	5-8 Jahre	Vitamin E, 50 mg pro Tag oder Beta-Caroten, 20 mg pro Tag	Anstieg der Serumvitaminspiegel	Unter Beta-Caroten Anstieg von Lungenkrebs um 18 %, Gesamtmortalität um 8 %	123
Verhinderung von Lungenkrebs	Vitamin-Serumspiegel	18.314 Personen, Raucher oder Asbestarbeiter	Nach 4 Jahren Abbruch	Beta-Caroten, 30 mg plus 25 000 IU Vitamin A pro Tag	Anstieg des Serumvitaminspiegels	Zunahme von Lungenkrebs um 28 %, der Gesamtmortalität um 17 %	124
Prävention von Krebs	Vitamin-Serumspiegel	Meta-Analyse		Hochdosiert Vitamin E	Anstieg des Serumvitaminspiegels	Zunahme der Gesamtmortalität	125
Kardiovaskuläre Risikofaktoren							
Prävention von koronarer Herzkrankheit und vorzeitigem Tod bei Diabetes	Blutzucker	UGDP	Studienabbruch	Tolbutamid	Senkung des Blutzuckerspiegels	Zunahme der kardialen Mortalität	71, 71
Prävention von koronarer Herzkrankheit und vorzeitigem Tod bei Diabetes	Blutzucker	UGDP	Studienabbruch	Phenformin	Senkung des Blutzuckerspiegels	Zunahme der kardialen Mortalität	71, 126

Tabelle Beispiele für Surrogat-Trugschlüsse (aktualisiert nach Tabelle 1 aus 63; Seite 3 von 5)

Präventionsziel (Endpunkt)	Surrogat-Parameter	Studienpopulation	Studien-dauer	Intervention	Surrogat-Ergebnis	Endpunkt-Ergebnis	Referenzen
Prävention von koronarer Herzkrankheit und vorzeitigem Tod bei Diabetes	Blutzucker	Meta-Analysen aus RCTs		Rosiglitazone	Senkung des Blutzuckerspiegels	Zunahme von Herzinfarkten und Herzinsuffizienz	76
Prävention von koronarer Herzkrankheit, diabetesbedingten Spätschäden und vorzeitigem Tod bei Diabetes	Blutzucker	UKPDS	10 Jahre	Intensivierte Diabetestherapie mit einer Kombination von Metformin und Sulfonylharnstoffen	Senkung des Blutzuckerspiegels	Zunahme der Gesamtmortalität	73
Prävention von koronarer Herzkrankheit, diabetesbedingten Spätschäden und vorzeitigem Tod bei Diabetes	Blutzucker	VADT	6 Jahre	Intensivierte Insulintherapie	Senkung des Blutzuckerspiegels	Keine Effekte auf Spätschäden, Trend zu höherer kardiovaskulärer Mortalität	69
Primärprävention der koronaren Herzkrankheit	Serumcholesterin	15.745 Männer, 30-59 Jahre alt, mit Cholesterinwerten im oberen Normbereich	5,3 Jahre	1,6 g Clofibrat pro Tag vs. Placebo	Senkung des Serumcholesterinspiegels um 9 %	Inzidenz von ischämischer Herzkrankheit um 20 % (p < 0,05) niedriger, Gesamtmortalität jedoch um 47 % höher	127, 128
Primärprävention der koronaren Herzkrankheit	Koronare Risikofaktoren	1.222 gesunde mit kardiovaskulären Risikofaktoren	5 Jahre, 15 Jahre	Multiple Risikofaktorenintervention, Clofibrat, Propranolol, Pindolol, Amilorid, Hydrochlorothiazid	Risikofaktorensenkung um 46 %	Zunahme der Gesamtmortalität um 45 %, der kardialen Mortalität um 143 % in der Interventionsgruppe	129, 130

Tabelle Beispiele für Surrogat-Trugschlüsse (aktualisiert nach Tabelle 1 aus 63; Seite 4 von 5)

Präventionsziel (Endpunkt)	Surrogat-Parameter	Studienpopulation	Studiendauer	Intervention	Surrogat-Ergebnis	Endpunkt-Ergebnis	Referenzen
Prävention von Herz-Kreislauf-Komplikationen bei Hypertonie	Blutdruckwerte	24.000 Personen mit Hypertonie und einem weiteren Risikofaktor	Abbruch nach 3,3 Jahren	Doxazosin im Vergleich zu Chlortalidon	Blutdrucksenkung	Zunahme von Herzinsuffizienz unter Doxazosin	131
Verminderung des plötzlichen Todes nach Herzinfarkt	Asymptomatische frühzeitige ventrikuläre Extrasystolen	3549 Patienten nach Herzinfarkt mit ventrikulärer Dysfunktion	1 Jahr	Ecaimid, Flecaimid, Moricizim* vs. Placebo	Vollständige oder teilweise Supprimierung der ventrikulären Extrasystolen	Überlebensrate höher in Placebo-Gruppe, häufiger kein Herzstillstand oder plötzlicher Herztod in Placebo-Gruppe	132, 133
Verbesserung von Herzinsuffizienz	Körperliche Belastbarkeit	Patienten mit Herzinsuffizienz	6 Monate	Mirinone Epoprostanol	Verbesserte Hämodynamik	Zunahme der Mortalität	134, 135
Verbesserung von Herzinsuffizienz	Herzleistung	Patienten mit Herzinsuffizienz	Abbruch	Ibopamine	Verbesserte Herzleistung	Zunahme der Mortalität	136
Prävention von Herzinfarkt	LDL-Cholesterin	15.067 Patienten mit hohem kardiovaskulärem Risiko	Abbruch nach 12 Monaten	Torcetrapib	Senkung von LDL-Cholesterin um 25 % Erhöhung von HDL-Cholesterin um 70 %	Zunahme von Herz-Kreislauf-Komplikationen um 25 %, der Mortalität um 58 %	137
Psychiatrische und psychologische Interventionen							
Prävention von Suizid bei Depression	Stimmungslage, Antrieb nach Depressionsskala	Meta-Analyse		Antidepressiva, SSRI	Verbesserung von Stimmung und Antrieb nach Depressionsskala	Hinweise auf Zunahme von Suizidalität	103
Prävention einer (chronischen) post-traumatischen Belastungsstörung	Akzeptanz, Zufriedenheit	Meta-Analysen		Debriefing	Zufriedenheit mit der Intervention, als hilfreich empfunden	Verzögerung der natürlichen Erholung nach psychologischem Trauma	138, 139, 140

Tabelle Beispiele für Surrogat-Trugschlüsse (aktualisiert nach Tabelle 1 aus 63; Seite 5 von 5)

Präventionsziel (Endpunkt)	Surrogat-Parameter	Studienpopulation	Studiendauer	Intervention	Surrogat-Ergebnis	Endpunkt-Ergebnis	Referenzen
Unfall-/Verletzungsprävention							
Unfallverhütung im Haushalt bei Kindern	Sicherheitsvorkehrungen im Haushalt	3428 sozial schwache Familien mit Kindern < 5 J., Nottingham, GB	2 Jahre	Kostenlose Ausstattung und Beratung	Verbesserte Sicherheitsmaßnahmen	Zunahme von Arztbesuchen wegen Unfallverletzungen	141
Prävention von Straftaten							
Prävention von Jugendkriminalität bei gefährdeten oder straffälligen Kindern	Akzeptanz	Meta-Analyse		Abschreckungsprogramme mit organisierten Besuchen in Gefängnissen	Hohe Akzeptanz bei den Teilnehmern, Lehrern und Eltern	Zunahme der Straftaten	142

Literaturverzeichnis

1. Beckmann MW, Adler G, Albers P et al. Dreistufenmodell optimiert Behandlung unter Kostendeckung. Dtsch Ärztebl 2007; 104:A3004-3009.
2. Moynihan R, Heath I, Henry D. Selling sickness: the pharmaceutical industry and disease mongering. BMJ 2002; 324:886-891.
3. Moynihan R. Scientists find new disease: motivational deficiency disorder. BMJ 2006; 332:745 (1 April).
4. Gemeinsamer Bundesausschuss. Richtlinie zur Umsetzung der Regelungen in § 62 für schwerwiegend chronisch Erkrankte (»Chroniker-Richtlinie«). Bundesanzeiger, 2007; 198: 7821; http://www.g-ba.de/downloads/62-492-140/RL_Chroniker_2007-07-19.pdf.
5. Sackett DL, Sharon ES, Richardson WS, Rosenberg W, Haynes R. How to practice and teach evidence-based medicine. Churchill Livingstone, Oxford 2000.
6. Charles C, Gafni A, Whelan T. Shared decision-making in the medical encounter: what does it mean? (or it takes at least two to tango). Soc Sci Med 1997; 44:681-692.
7. Mühlhauser I, Höldke B. Information zum Mammographiescreening – vom Trugschluss zur Ent-Täuschung. Radiologe 2002; 42:299-304.
8. Mühlhauser I, Berger M. Evidence-based patient information in diabetes. Diab Medicine 2000; 17:823-829.
9. Mühlhauser I, Filz M. Screening auf Zervixkarzinom. Informationen zur Beratung von Frauen. arznei-telegramm 2008; 39:29-38.
10. Mühlhauser I, Steckelberg A. Aufklärung über Krebsfrüherkennung am Beispiel Mammographie- und Darmkrebs-Screening. Z Allg Med 2005; 81:523-527.
11. Mühlhauser I. Evidence-based patient education in diabetes and beyond – application to other chronic diseases. From obedience training to informed decision making. In: Porta M (ed) Embedding education in diabetes practice. Frontiers in Diabetes. Karger, Basel. 2005; 132-146.
12. Mühlhauser I, Lenz M. Verbesserung der Therapieergebnisse durch Patientenwissen? Z Evid Fortbild Qual Gesundh wesen (ZEFQ) 2008; 102:223-230.
13. Kasper J, Heesen C, Mühlhauser I: Evidenzbasierte Patienteninformation – dargestellt am Beispiel der Immuntherapie bei Patienten mit Multipler Sklerose. Bundesgesundheitsblatt 2008 (im Druck).
14. Steckelberg A, Berger B, Köpke S et al. Kriterien für evidenzbasierte Patienteninformationen. Z Ärztl Fortbild Qualitätssich 2005; 99:343-351.
15. Tenter B. Gesundheitswesen: Behandlungsleitlinien und wissenschaftliche Studien. Deutscher Bundestag, Petitionsausschuss. 2007; <http://itc.napier.ac.uk/e-Petition/bundestag/>.
16. Müller H. Können wir uns Lebensqualität leisten? Forum DKG 2007; Heft 6:24-25.
17. Hamann J, Neuner B, Kasper J et al. Participation preferences of patients with acute and chronic conditions. Health Expect 2007; 10:358-363.
18. B.I.R.O. Best information through rational outcomes. Clinical review indicator development. June 2006. <http://biro-project.eu/documents/downloads>.
19. Caron-Flinterman JF, Broerse JEW, Bunders JFG. The experiential knowledge of patients: a new resource for biomedical research? Soc Sci Med 2005; 60:2575-2584.

20. Farin E. Patientenorientierung und ICF-Bezug als Herausforderungen für die Ergebnismessung in der Rehabilitation. *Rehabilitation* 2008; 47:67-76.
21. Mittelman MS. Psychosocial intervention research: challenges, strategies and measurement issues. *Aging & Mental Health* 2008; 12:1-4.
22. Garcia SF, Cella D, Clauser SB et al. Standardizing patient-reported outcomes assessment in cancer clinical trials: a patient-reported outcomes measurement information system initiative. *J Clin Oncol* 2007; 25:5106-5112.
23. Huang ES, Brown SES, Ewigman BG et al. Patient perceptions of quality of life with diabetes-related complications and treatments. *Diabetes Care* 2007; 30:2478-2483.
24. Huang ES, Zhang Qi, Gandra N et al. The effect of comorbid illness and functional status on the expected benefits of intensive glucose control in older patients with type 2 diabetes: a decision analysis. *Ann Intern Med* 2008; 149:11-19.
25. Pronovost PJ, Miller M, Wachter RM. The GAAP in quality measurement and reporting. *JAMA* 2007; 298:1800-1802.
26. Müller H. Mythos Innovation? Anmerkungen zu medizin-technischen Entwicklungen aus Sicht der Gesetzlichen Krankenversicherung. In: Groß, D, Jakobs EM (Hrsg.): *E-Health und technisierte Medizin. Neue Herausforderungen im Gesundheitswesen*. Anthropina. Aachener Beiträge zur Geschichte, Theorie und Ethik der Medizin, Bd. 2. Berlin 2007; 101-111.
27. Angell M. *The truth about the drug companies*. Random House 2004. New York, Toronto.
28. De Angelis DC. Impugning the integrity of medical science. The adverse effects of industry influence. *JAMA* 2008; 299:1833-1835.
29. Glasgow RE, Peeples M, Skovlund SE. Where is the patient in diabetes performance measures? *Diabetes Care* 2008; 31:1046-1050.
30. Berkanovic E, Hurwicz ML, Lachenbruch PA. Concordant and discrepant views of patients' physical functioning. *Arthritis Care & Research* 1995; 8:94-101.
31. Rockwood K, Fay S, Song X et al. on behalf of the video-imaging synthesis of treating Alzheimer's disease (VISTA) investigators. Attainment of treatment goals by people with Alzheimer's disease receiving galantamine: a randomized controlled trial. *CMAJ* 2006; 174:1099-1005.
32. Starfield B. Quality and outcomes framework: patient-centred? *The Lancet* 2008; 372:692-694.
33. European Medicines Agency. Concept paper on the need for revision of the note for guidance on clinical investigation of medical products in the treatment of diabetes mellitus. 2008; Doc. Ref. EMEA/CHMP/EWP/176348/2008.
34. European Medicines Agency. Guideline on the evaluation of anticancer products in man. 2005; CPMP/EWP/205/95/Rev.3/Corr.2 and Appendix 2008; Doc. Ref. EMEA/CHMP/EWP/27994/2008.
35. European Medicines Agency. Guideline on the evaluation of medicinal products for cardiovascular disease prevention. 2008; Doc. Ref. EMEA/CHMP/EWP/311890/2007.
36. European Medicines Agency. Note for guidance on clinical investigation of medicinal products in the treatment of diabetes mellitus. 2002; CPMP/EWP/1080/00.
37. US Department of Health and Human Services. Food and Drug Administration. Guidance for industry. Clinical trial endpoints for the approval of cancer drugs and biologics. 2007.

38. US Department of Health and Human Services. Food and Drug Administration. Guidance for industry. Patient-reported outcome measures: use in medical product development to support labelling claims. 2006.
- 39a. Bitzer EM, Dörning H, Schwartz FW et al. Der Erfolg von Operationen aus Sicht der Patienten. Eine retrospektive indikationsbezogene Patientenbefragung zum Outcome elektiver chirurgischer Leistungen in der stationären Versorgung. St. Augustin 1999.
- 39b. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG). Methoden 3.0. 2008; www.iqwig.de.
- 39c. Straub C, Nebling T, Müller H. Translating research into practice: A German sickness fund supporting patient participation. *Patient Education and Counseling* 2008; 73:544-550.
- 40a. Patrick DL, Guyatt GH, Acquadro C on behalf of the Cochrane Patient Reported Outcomes Methods Group. *Cochrane Handbook*, Chapter 17. 2008.
- 40b. Windeler J. Placebo Effekte. *Z ärztl Fortbild Qual Gesundh wes (ZaeFQ)* 2007; 101:307-312.
41. Berger B, Lenz M, Mühlhauser I. Patient zufrieden, Arzt gut? Inwiefern ist Patientenzufriedenheit ein Indikator für die Qualität der hausärztlichen Versorgung? Eine systematische Übersichtsarbeit. *Z Evid Fortbild Qual Gesundh wes (ZEFQ)* 2008; doi:10.1016.
42. DAFNE Study Group. Training in flexible, intensive insulin management to enable dietary freedom in people with type 1 diabetes: dose adjustment for normal eating (DAFNE) randomised controlled trial. *BMJ* 2002; 325:746 online.
43. Trevena LJ, Davey HM, Barratt A et al. A systematic review on communicating with patients about evidence. *J Eval Clin Pract* 2006; 12:13-23.
44. Coulter A, Ellins J, Swain D et al. Assessing the quality of information to support people in making decisions about their health and healthcare. Picker Institute Europe 2006. Oxford.
45. Coulter A, Ellins J. Effectiveness of strategies for informing, educating, and involving patients. *BMJ* 2007; 335:24-27.
46. Harnischmacher U, Ihle P, Goebel J et al. TMF Band 3: Checkliste und Leitfaden zur Patienteneinwilligung – Grundlagen und Anleitung für die klinische Forschung. Medizinisch Wissenschaftliche Fachgesellschaft; Berlin 2006.
47. O'Connor A, Légaré F, Stacey D. Risk communication in practice: the contribution of decision aids. *BMJ* 2003; 327:736-740.
48. Domenighetti F, D'Avanzo B, Egger M et al. Women's perception of the benefits of mammography screening: population-based survey in four countries. *Internat J Epidemiol* 2003; 32:816-821.
49. Feldman-Stewart D, Brennenstuhl S, McIssac K et al. A systematic review of information in decision aids. *Health Expectations* 2006; 10:46-61.
50. Steckelberg A, Hülfenhaus C, Kasper J, Mühlhauser I. Ebm@school – a curriculum of critical health literacy for secondary school students: results of a pilot study. *Internat J Public Health* 2008 (im Druck).
51. Steckelberg A, Hülfenhaus C, Kasper J et al. How to measure critical health competences: development and validation of the Critical Health Competence Test (CHC Test). *Adv in Health Sci Educ* 2007; DOI 10.1007/s10459-007-9083-1.
52. Gigerenzer G, Gaissmaier W, Kurz-Milcke E et al. Helping doctors and patients make sense of health statistics. *Psychological Science in the Public Interest* 2008; 8:53-96.

53. General Medical Council. Protecting patients, guiding doctors. Seeking patients' consent: the ethical considerations. London 1998; www.gmc-uk.org.
54. General Medical Council. Consent: patients and doctors making decisions together. London 2008; www.gmc-uk.org.
55. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG). Nutzenbewertung kurzwirksamer Insulinanaloge zur Behandlung des Diabetes mellitus Typ 2. 2005; www.iqwig.de.
56. Holleman F, Gale EA. Nice insulins, pity about the evidence. *Diabetologia* 2007; 50:1783-1790.
57. Berger M, Mühlhauser I. Diabetes care and patient-oriented outcomes. *JAMA* 1999; 281:1676-1678.
58. Amiel SA. Type 1 diabetes: treatment without tears? *Diabetologia* 2005; 48:1963-1964.
59. Welch HG, Black WC. Are deaths within 1 month of cancer-directed surgery attributed to cancer? *J Natl Cancer Inst* 2002; 94:1066-1077.
60. Doran T, Fullwood C, Kontopantelis E, Reeves D. Effect of financial incentives on inequalities in the delivery of primary clinical care in England: analysis of clinical activity indicators for the quality and outcomes framework. *Lancet* 2008; 372:728-736.
61. Hughes E. Payment by results – a model for other diabetes healthcare systems? *Primary Care Diabetes* 2007; 1:111-113.
62. Müller H, Nebling T. Die Bedeutung von Pay-for-Performance-Programmen für die Gestaltung der Medizinischen Versorgung. In: Verein für Kommunalwissenschaften e.V. Berlin (Hrsg.): *Mythos wirkungsorientierte Steuerung (Aktuelle Beiträge zur Kinder- und Jugendhilfe)* Bd. 64. Berlin 2008.
63. Mühlhauser I, Meyer G. Evidence Based Medicine: Widersprüche zwischen Surrogatergebnissen und klinischen Endpunkten. *Psychother Psych Med* 2006; 56:193-201.
64. DIMDI. Surrogatendpunkte als Parameter der Nutzenbewertung. HTA 2009 (im Druck) www.dimdi.de.
65. Fleming T, MeMets DL. Surrogate end points in clinical trials: are we being misled? *Ann Intern Med* 1996; 125:605-613.
66. Pani LN, Korenda L, Meigs JB et al. Effect of aging on A1c levels in individuals without diabetes. *Diabetes Care* 2008; 31:1991-1996.
67. American Diabetes Association (ADA). Standards of Medical Care in Diabetes. *Diabetes Care* 2008; 31(Suppl 1):12-54.
68. Action to Control Cardiovascular Risk in Diabetes (ACCORD) Study Group. Effects of intensive glucose lowering in type 2 diabetes. *N Engl J Med* 2008; 358:2545-2559.
69. Duckworth WC, Abairra C. Veterans Administration Diabetes Trial (VADT). Vortrag auf dem Europäischen Diabeteskongress, Rom, 2008; www.easd.org; http://conference2web.com/easd2008_webcast/Sessions/UKPDS30yeardata.aspx.
70. UK Prospective Diabetes Study (UKPDS) Group. Effect of intensive blood-glucose control with metformin on complications in overweight patients with type 2 diabetes (UKPDS 34). *Lancet* 1998; 351:854-865.
71. University Diabetes Program. A study of the effects of hypoglycemic agents on vascular complications in patients with adult-onset diabetes. V. Evaluation of phenformin therapy. *Diabetes* 1975; 24 (Suppl 1):65-184.

72. University Group Diabetes Program. A study of the effects of hypoglycemic agents on vascular complications in patients with adult-onset diabetes mellitus. II. Mortality results. 1970; 10(Suppl 2):785-830.
73. UK Prospective Diabetes Study (UKPDS) Group. Intensive blood glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). *Lancet* 1998; 352:837-853.
74. Gale EA. Troglitazone: the lesson that nobody learned? *Diabetologia* 2006; 49:1-6.
75. Hauber A, Gale EA. The market in diabetes. *Diabetologia* 2006; 49:247-252.
76. Richter B, Bandeira-Echtler E, Bergerhoff K et al. Rosiglitazone for type 2 diabetes mellitus. *Cochrane Database Syst Rev* 2007; 3:CD006063.
77. Richter B, Bandeira-Echtler E, Bergerhoff K et al. Pioglitazone for type 2 diabetes mellitus. *Cochrane Database Syst Rev* 2006; 4:CD006060.
78. Richter B, Bandeira-Echtler E, Bergerhoff K, Lerch CL. Dipeptidyl peptidase-4 (DPP-4) inhibitors for type 2 diabetes mellitus. *Cochrane Database Syst Rev* 2008; 2:CD006739.
79. University Group Diabetes Program. A study of the effects of hypoglycemic agents on vascular complications in patients with adult-onset diabetes. VI. Supplementary report on non-fatal events treated with tolbutamide. *Diabetes* 1976; 25:1129-1152.
80. Risikostruktur-Ausgleichsverordnung (17. RSA-ÄndV). Das Bundesgesetzblatt im Internet: www.bundesgesetzblatt.de 2008; Teil 1 Nr. 11: 468-475.
81. Lenz M, Steckelberg A, Richter B, Mühlhauser I. Meta-analysis does not allow appraisal of complex interventions in diabetes and hypertension self-management. A methodological review. *Diabetologia* 2007; 50:1375-1383.
82. Abholz HH. Komplexität bei der Behandlung in der Allgemeinmedizin – Schwierigkeiten der Bestimmung von Outcome-Parametern. *Z Evid Fortbild Qual Gesundh.wesen (ZEFQ)* 2008; 102:351-356.
- 83a. Guyatt GH, Oxman AD, Vist GE et al for the GRADE Working Group. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008; 336:924-926.
- 83b. Lenz M, Mühlhauser I: Decision aids in diabetes. In: Edwards A, Elwyn G (eds): Shared decision making in health care: achieving evidence based patient choice, 2e. Oxford University Press, Oxford (in press).
84. Richter B, Berger M. Randomized controlled trials remain fundamental to clinical decision making in Type 2 diabetes mellitus: a comment to the debate on randomized controlled trials. *Diabetologia* 2000; 43:254-258.
85. Szecsenyi J et al. German diabetes disease management programs are appropriate for restructuring care. *Diabetes Care* 2008; 31:1150-1154.
86. European Network of Health Promoting Schools. The European Network of Health Promoting Schools – the alliance of education and health. 2004; www.euro.who.int/document/e62361.pdf.
87. Jorgensen KJ, Gotzsche PC. Presentation on websites of possible benefits and harms from screening for breast cancer: cross sectional study. *BMJ* 2004; 328:148 (online).
88. Meyer G, Steckelberg A, Mühlhauser I. Analysis of consumer information brochures on osteoporosis prevention and treatment. *GMS Ger Med Sci* 2007; 5:Doc01.
89. Mühlhauser I. Ist Vorbeugen besser als Heilen? *Z. ärztl. Fortbild. Qual. Gesundh.wes. (ZEFQ)* 2007; 101:293-299.

90. Koch K, Mühlhauser I für den Fachbereich Patienteninformation des Deutschen Netzwerkes evidenzbasierte Medizin. Kriterien zur Erstellung von Patienteninformationen zu Krebsfrüherkennungsuntersuchungen. 2008; www.ebm-netzwerk.de.
91. Welch HG, Schwartz LM, Woloshin S. Are increasing 5-year survival rates evidence of success against cancer? *JAMA* 2000; 22:2975-2978.
92. Keidan J. Sucked into the Herceptin maelstrom. *BMJ* 2007; 334:18.
93. Kasper J, Köpke S, Mühlhauser I et al. Informed shared decision making about immunotherapy for patients with multiple sclerosis (ISDIMS): a randomised controlled trial. *European Journal of Neurology* 2008 (im Druck).
94. Kasper J, Kuch C, Heesen C. Shared decision-making als Interaktionsstil: eine konstruktivistische Perspektive. In: Scheibler F, Pfaff H (Hrsg.) *Shared Decision Making. Der Patient als Partner im medizinischen Entscheidungsprozess*. Weinheim, München: Juventa, 2003; 34-45.
95. Köpke S, Kasper J, Mühlhauser I et al. Patient education programme to enhance decision autonomy in multiple sclerosis relapse management: a randomised controlled trial. *Multiple Sclerosis* 2008 (im Druck).
96. Berger B, Steckelberg A, Meyer G et al. Empowerment of patient and consumer representatives through training in basics of evidence-based medicine: a pilot study. Unpublished Manuscript.
97. Dickersin K, Braun L, Mead M et al. Development and implementation of a science training course for breast cancer activists: Project LEAD (leadership, education and advocacy development). *Health Expect* 2001; 4:213-220.
98. Milne R, Oliver S. Evidence based consumer health information: developing teaching in critical appraisal skills. *Internat J Quality in Health Care* 1996; 8:439-445.
99. McIntyre P. Europe must tackle health illiteracy to avoid a health underclass. *Cancer World* 2008; 52-56.
100. Woloshin S, Schwartz LM, Welch G. The effectiveness of a primer to help people understand risk. *Ann Intern Med* 2007; 146:256-265.
101. Groves T. Mandatory disclosure of trial results for drugs and devices. *BMJ* 2008; 336:170.
102. Ioannidis JPA. Effectiveness of antidepressants: an evidence myth constructed from a thousand randomized trials? *Philosophy, Ethics, and Humanities in Medicine* 2008; 3:14 doi:10.1186/1747-5341-3-14.
103. Anonym. Antidepressiva: Lebensgefährliche Placebos? *arznei-telegramm* 2005; 36:45-47.
104. Turner EH, Matthews AM, Linardatos E et al. Selective publication of antidepressant trials and its influence on apparent efficacy. *NEJM* 2008; 358:252-260.
105. Berger B, Studt K, Krause K et al. Schutz von Studienteilnehmern durch standardisierte Informationen. Vorschlag für einen einheitlichen Kriterienkatalog für die Arbeit der Ethikkommissionen – Ergebnisse einer Expertenbefragung. *Z Arztl Fortbild Qualitätssich (ZaeFQ)* 2005; 99:389-395.
106. Alzheimer's Society. People with dementia and their carers set NICE a challenge. Press release vom 31.3.2005; <http://www.alzheimers.org.uk/>.
107. Deutscher Diabetikerbund. Über 100.000 Unterschriften in Berlin übergeben – gegen die Restriktion der Insulin-Analoga. Pressemitteilung 10.5.2006; www.diabetikerbund.de/.

108. Mühlhauser I, Kasper J, Meyer G, FEND. Understanding of diabetes prevention studies: questionnaire survey of professionals in diabetes care. *Diabetologia* 2006; 49:1742-1746.
109. Mühlhauser I. Diabetes experts' reasoning about diabetes prevention. *BMC Research Notes* 2008; 1:90 [Epub ahead of print].
110. Giezen TJ, Mantel-Teeuwisse AK, Straus SMJM et al. Safety-related regulatory actions for biologicals approved in the United States and the European Union. *JAMA* 2008; 300:1887-1896.
111. Sawicki PT for the Working Group for the Study of Patient Self-Management of Oral Anticoagulation. A structured teaching and self-management program for patients receiving oral anticoagulation. A randomized controlled trial. *JAMA* 1999; 281:145-150.
112. Mühlhauser I, Berger M. Patient education – evaluation of a complex intervention. *Diabetologia* 2002; 45:1723-1733.
113. Hulley S, Grady D, Bush T, et al. for the Heart and Estrogen/progestin Replacement Study (HERS) Research Group. Randomized trial of estrogen plus progestin for secondary prevention of coronary heart disease in postmenopausal women. *JAMA* 1998; 280:605-613.
114. Pressemeldung des MDK. Patientenbroschüre der Bundesärztekammer zu IGeL-Leistungen greift zu kurz. Essen, 14. März 2008.
115. Strech D. Vier Ebenen von Werturteilen in der medizinischen Nutzevaluation. Eine Systematik zur impliziten Normativität in der Evidenz-basierten Medizin. *Z ärztl Fortbild Qual Gesundh wes (ZaeFQ)* 2007; 101:473-480.
116. Writing Group for the Women's Health Initiative Investigators. Risks and benefits of estrogen plus progestin in healthy postmenopausal women. Principal results from the Women's Health Initiative randomized controlled trial. *JAMA* 2002; 288:321-333.
117. Shumaker SA, Legault C, Kuller L, et al. Women's Health Initiative Memory Study. Conjugated equine estrogens and incidence of probable dementia and mild cognitive impairment in postmenopausal women: Women's Health Initiative Memory Study. *JAMA* 2004; 291:2947-2958.
118. Shumaker SA, Legault C, Thal L, et al. for the WHIMS Investigators. Estrogen plus progestin and the incidence of dementia and mild cognitive impairment in postmenopausal women: the Women's Health Initiative Memory Study: a randomized controlled trial. *JAMA* 2003; 289:2651-2662.
119. Espeland MA, Rapp SR, Shumaker SA, et al. Women's Health Initiative Memory Study. Conjugated equine estrogens and global cognitive function in postmenopausal women: Women's Health Initiative Memory Study. *JAMA* 2004; 291:2959-2968.
- 120a. Hendrix SL, Cochrane BB, Nygaard IE, et al. Effects of estrogen with and without progestin on urinary incontinence. *JAMA* 2005; 293: 935-948
- 120b. Riggs BL, Hodgson SF, O'Fallon WM, et al. Effect of fluoride treatment on the fracture rate in postmenopausal women with osteoporosis. *N Engl J Med* 1990; 322:802-809.
121. Bennett CL, Silver SM, Djulbegovic B et al. Venous thromboembolism and mortality associated with recombinant erythropoietin and Darbepoetin administration for treatment of cancer-associated anemia. *JAMA* 2008; 299:914-924.

122. Promminticul A, Haas SJ, Elsik M, Krum H. Mortality and target haemoglobin concentrations in anaemic patients with chronic kidney disease treated with erythropoietin: a meta-analysis. *Lancet* 2007; 369:381-388.
123. The Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study Group: The effect of vitamin E and beta carotene on the incidence of lung cancer and other cancers in male smokers. *N Engl J Med* 1994; 330:1029-1035.
124. Omenn GS, Goodman GE, Thornquist MD, et al. Effects of a combination of beta-carotene and vitamin A on lung cancer and cardiovascular disease. *N Engl J Med* 1996; 334:1150-1155.
125. Miller ER, Pastor-Barriuso R, Dalal D, et al. Meta-analysis: high dosage vitamin E supplementation may increase all-cause mortality. *Ann Intern Med* 2005; 142:37-46.
126. University Group Diabetes Program (UGDP). Effects of hypoglycemic agents on vascular complications in patients with adult-onset diabetes. VIII. Evaluation of insulin therapy: final report. *Diabetes* 1982; 31(Suppl 5):1-81.
127. Committee of Principal Investigators. WHO cooperative trial on primary prevention of ischaemic heart disease with clofibrate to lower serum cholesterol: final mortality follow-up. *Lancet* 1984; 2:600-604.
128. Committee of Principal Investigators. A co-operative trial in the primary prevention of ischaemic heart disease using clofibrate. *Br Heart J* 1978; 40:1069-1118.
129. Strandberg TE, Salomaa VV, Naukkarinen VA, et al. Cardiovascular morbidity and multifactorial primary prevention: fifteen-year follow-up of the Helsinki businessmen study. *Nutr Metab Cardiovasc Dis* 1995; 5: 7-15.
130. Strandberg TE, Salomaa VV, Naukkarinen VA, et al. Long-term mortality after 5-year multifactorial primary prevention of cardiovascular diseases in middle-aged men. *JAMA* 1991; 266:1225-1229.
131. The ALLHAT Officers and Coordinators for the ALLHAT Collaborative Research Group. Major cardiovascular events in hypertensive patients randomized to Doxazosin vs Chlorthalidone. *JAMA* 2000; 283:1967-1975.
132. The Cardiac Arrhythmia Suppression Trial (CAST) Investigators: Preliminary report: Effect of encainide and flecainide on mortality in a randomized trial of arrhythmia suppression after myocardial infarction. *N Engl J Med* 1989; 321:406-412.
133. Epstein AE, Hallstrom AP, Rogers WJ, et al. Mortality following ventricular arrhythmia suppression by encainide, flecainide, and moricizine after myocardial infarction. The original design concept of the cardiac arrhythmia suppression trial (CAST). *JAMA* 1993; 270:2451-2455.
134. Califf RM, Adams FK, McKenna WJ, et al. A randomized controlled trial of epo-prostenol therapy for severe congestive heart failure. *Am Heart J* 1997; 134:44-54.
135. Packer M, Carver JR, Rodeheffer RJ, et al. Effect of oral milrinone on mortality in severe chronic heart failure. The PROMISE Study Research Group. *N Engl J Med* 1991; 325:1468-1475.
136. Hampton JR, van Veldhuisen DJ, Kleber FX, et al. Randomised study of effect of ibopamine on survival in patients with advanced severe heart failure. *Lancet* 1997; 349:971-977.
137. Barter PJ, Caulfield M, Eriksson M et al. for the ILLUMINATE Investigators. Effects of torcetrapib in patients at high risk for coronary events. *N Engl J Med* 2007; 357:2109-2122.

138. Emmerik AAP, Kamphuis JH, Hulsbosch AM, et al. Single session debriefing after psychological trauma: a meta-analysis. *Lancet* 2002; 360:766-771.
139. Rose S, Bisson J, Wessely S. A systematic review of single-session psychological interventions (>debriefing<) following trauma. *Psychother Psychosom* 2003; 72:176-184.
140. Michael T, Lajtman M, Margaf J. Frühzeitige psychologische Interventionen nach Traumatisierung. *Dtsch Ärztebl* 2005; 102:A 2240-2243.
141. Watson M, Kendrick D, Coupland C et al. Providing child safety equipment to prevent injuries: randomised controlled trial. *BMJ* 2005; 330:178-181.
142. Petrosino A, Turpin-Petrosino C, Buehler J. »Scared Straight« and other juvenile awareness programs for preventing juvenile delinquency. *The Cochrane Database of Systematic Reviews* 2002; Issue 2: CD002796.