

Evidence-Based Medicine: Interpretation von Studienergebnissen zu therapeutischen Verfahren

Franz Porzsolt und Jörg Sigle

AG Klinische Ökonomik, Klinikum der Universität Ulm, 89070 Ulm

Nach Abschluß der Diagnostik ist die Wahl der besten einzuleitenden Therapie ein Problem, das für jeden Patienten individuell zu lösen ist. Jeder erfahrene Praktiker weiß um die Schwierigkeit dieses Prozesses, weil für diese Entscheidung die Vorgeschichte des Patienten, seine Begleiterkrankungen und derzeitigen Therapien, seine persönlichen Präferenzen und sein soziales Umfeld, um nur einige dieser Aspekte zu nennen, ebenso zu berücksichtigen sind, wie neue medizinische Erkenntnisse. Seit einigen Jahren haben ökonomische Aspekte, d.h. der Vergleich von Kosten und Konsequenzen alternativer Handlungsmöglichkeiten (Drummond), diese Liste erweitert. Dazu sind bei jedem einzelnen Patienten die "Kosten" und Konsequenzen der Handlungsmöglichkeiten gegeneinander abzuwägen. Unter "Kosten" sind die vom Patient "in Kauf zu nehmenden" Risiken und Belastungen der Behandlung ebenso wie die finanzielle Belastung der Solidargemeinschaft zu verstehen. Die Konsequenzen beziehen sich auf den Nutzen des Patienten und die Vorteile der Solidargemeinschaft, die aus der durchgeführten Therapie resultieren.

Therapieentscheidungen sind in der täglichen Praxis aus drei Gründen schwierig:

Zunächst ist die klinische Relevanz wissenschaftlicher Behandlungsergebnisse (externe Evidenz) mit der Erfahrung des Arztes und der Kenntnis individueller Faktoren des Patienten (persönliche oder individuelle Evidenz) zu verbinden. Dazu ist es erforderlich, Kosten und Konsequenzen alternativer Handlungsmöglichkeiten möglichst exakt zu benennen. Ihre Bewertung und Integration werden durch Evidence-Based Medicine unterstützt (Por 1997). Hierzu dienen verschiedene Methoden, die anhand eines klinischen Szenarios dargestellt werden.

Die zweite Schwierigkeit entsteht aus dem Zeitdruck der täglichen Praxis. Wenn externe Evidenz neben der täglichen Routinearbeit gesucht, gefunden und bewertet werden soll, müssen diese Arbeitsschritte rasch zu erledigen sein; andernfalls wäre die Forderung nach der Integration externer Evidenz in die tägliche Routine unrealistisch.

Die dritte Schwierigkeit bei der Therapieentscheidung entsteht aus der notwendigen Unterscheidung, ob eine geplante

Therapie zur Linderung eines bestehenden oder zur Abwendung eines drohenden Problems durchgeführt werden soll: Bei der Linderung bestehender Probleme ist der Erfolg leicht zu beurteilen und der Zusammenhang zwischen Intervention und Eintritt des erwünschten Ergebnisses meist offensichtlich. Bei der Abwendung drohender Probleme ist die Beurteilung des Erfolgs erheblich aufwendiger, weil oft nicht bekannt ist, ob das drohende Problem ohne Intervention tatsächlich aufgetreten wäre.

Als Beispiel sollen Knochenmetastasen in der Wirbelsäule betrachtet werden: um das bestehende Problem „Schmerzen“ zu lindern, wird eine Chemotherapie oder eine Schmerztherapie verordnet; der Erfolg kann in der Praxis durch Befragung des Patienten leicht überprüft werden und der Zusammenhang mit der Therapie ist meist offensichtlich. In einem anderen Fall warnt der Radiologe vor dem drohenden Problem einer „Wirbelfraktur“. Nach einer Bestrahlung tritt keine Fraktur auf - wie aber wäre der Verlauf ohne diese Maßnahme gewesen?

Im folgenden Szenario soll beispielhaft dargestellt werden, wie die Qualität einer externen Evidenz beurteilt, und diese gegebenenfalls zu einer Therapieentscheidung herangezogen werden kann.

Das klinische Szenario

Der 58 jährige Pressesprecher eines Konzerns hat sich zur empfohlenen Krebs-Vorsorge-Untersuchung beim Hausarzt vorgestellt. Es bestanden angeblich keine Krankheitszeichen. Bei der Anamnese und klinischen Untersuchung wurden außer einem verdächtigen Palpationsbefund der Prostata keine pathologischen Befunde erhoben. Die Bestimmung des Prostata-spezifischen Antigens (PSA) ergab einen auf das dreifache der Norm erhöhten Wert; es wurde eine Vorstellung beim Urologen vereinbart. Die weiterführenden Untersuchungen ergaben die Diagnose eines Prostatakarzinoms im Stadium D2. Als Therapie wurde eine Orchiektomie vorgeschlagen. Wegen der möglichen Nebenwirkungen bat der Patient seinen Hausarzt, unabhängige Informationen zu Vor- und Nachteilen verschiedener Behandlungsmöglichkeiten einzuholen.

Dem Konzept der Evidenz-Based Medicine folgend, wurde zunächst folgende beantwortbare Frage formuliert:

- Wie sind, bei einem 58-jährigen, bisher symptomlosen Patienten mit einem Prostatakarzinom im Stadium D2 die Ergebnisse verschiedener Therapieformen in Bezug auf Verlängerung der Überlebenszeit, Verhinderung der Tumorprogression und Symptomkontrolle?

Anschließend wurde eine Literaturrecherche durchgeführt, um die besten Quellen von Evidenz für die Beantwortung der Frage zu finden:

- Eine Suche in „Best Evidence“ (CD Rom, British Medical Journal Publishing Group) einer Datenbank mit Abstracts von Artikeln, die bestimmte Qualitätskriterien erfüllen, ergab kein passendes Abstract.

In der MEDLINE-Datenbank wurde im Zeitraum von 1995-1997¹ mit verschiedenen Suchstrategien gesucht: Eine Recherche nach „Prostate Carcinoma and Castration“ ergab 81 Artikel, „Prostate Cancer and Castration“ ergab 620 Artikel, „Advanced Prostate Cancer and Castration“ ergab 17 Artikel, darunter randomisierte Studien und ein Overview über 22 Studien². Nach einer Durchsicht der Titel und Abstracts dieses Suchergebnisses erschien eine Arbeit von Chodak et al. als am besten geeignet, um Methoden der EBM am Beispiel des genannten Patienten zu demonstrieren. Wir haben bewusst eine Arbeit gewählt, die Beispiele für mögliche Schwächen in der Methodik enthält, welche ein Leser erkennen und kritisch bewerten können sollte.

In dieser Arbeit wurde die Therapie mit Bicalutamide (BIC), einem nicht-steroidalen Antiandrogen, mit einer operativen oder medikamentösen Kastration (KA) in Bezug auf Häufigkeit von Therapieversagen, Progression der Erkrankung, Überlebenszeit und Beeinflussung verschiedener Variablen aus dem Bereich der Lebensqualität verglichen³.

Für die Beratung des Patienten ist zu klären, ob die Ergebnisse dieser Studie hinreichend valide, bedeutend und für das spezifische Problem des Patienten zutreffend sind.

Konzept der Evidence-Based Medicine

Im englischen Sprachgebrauch werden mit dem Begriff „Evidence-Based Medicine“ die externen Informationen bezeichnet, die mit wissenschaftlichen Methoden erhoben wurden. Im deutschen Sprachgebrauch wird der Begriff „Evidenz“ wesentlich breiter gebraucht. Viele deutsche Kollegen verstehen auch ihre persönliche Erfahrung als „Evidenz“. Um

¹ Je nach Fragestellung und zur Verfügung stehender Zeit kann es sinnvoll sein, eine Suche zunächst auf neuere Veröffentlichungen zu beschränken und sie erst bei Unergiebigkeit auszudehnen.

² Die Suche nach „((PROSTATE CANCER) OR (PROSTATE CARCINOMA)) AND ((RANDOM* IN TG) OR (RANDOM* IN MESH) OR (RANDOM* IN PT)) and CASTRATION) and CASTRATION and CASTRATION“ ergibt ebenfalls 18 Treffer. Eingehende Recherchen erfordern, neben MEDLINE auch in anderen Datenbanken z.B. Embase nachzusehen.

³ Für unser Beispiel relevante Teilergebnisse sind weiter unten im Text angegeben.

die Verwendung identischer Begriffe mit unterschiedlichen Inhalten zu verhindern, sprechen wir in unserem Aufsatz von „externer Evidenz“ wenn wir den englischen Begriff „Evidence“ meinen und verwenden die Bezeichnung „interne Evidenz“ wenn die persönliche Erfahrung einer einzelnen Person gemeint ist.

Um die Abwägung von Kosten und Konsequenzen einer Therapie vorzunehmen und die externe Evidenz in die persönliche Erfahrung des Arztes zu integrieren, werden vier Schritte empfohlen:

- (1) die benötigte Information in beantwortbare Fragen umzusetzen,
- (2) die beste verfügbare Evidenz zu finden,
- (3) die Validität (Gültigkeit), Bedeutung (Effektgröße) und Brauchbarkeit (klinische Anwendbarkeit) der gefundenen Evidenz kritisch zu bewerten und
- (4) die bewertete Evidenz in die klinische Erfahrung zu integrieren und auf die aktuelle Fragestellung anzuwenden.

Die Schritte (1) und (2) wurden im Beispiel schon illustriert, auf (3) und (4) wird im folgenden eingegangen.

Bewertung der Validität (Gültigkeit) der Aussagen einer Studie

Bevor die Ergebnisse einer Studie als valide (gültig) akzeptiert werden, sollte zunächst deren Methodik überprüft werden. Hierzu ist ein Vorgehen in zwei Schritten sinnvoll: Zunächst können ad hoc die beiden Fragen beantwortet werden, ob in der vorliegenden Studie die Zuordnung der Patienten in die Therapiegruppen zufällig, d.h. durch Randomisierung⁴, erfolgte, und ob die Daten aller Patienten, die in die Studie aufgenommen wurden, in den Schlußfolgerungen der Studie angemessen berücksichtigt wurden. Wenn beide Fragen zu verneinen sind, ist die Validität dieser verfügbaren Quelle von Evidenz sicher begrenzt. Bei positiver Antwort kann in einem zweiten Schritt die für die Bewertung zusätzlich notwendige Information eingeholt werden (Tab .I)

⁴ In einer randomisierten Studie werden die Patienten nach dem Zufallsprinzip (nicht nach Entscheidung durch den Arzt oder Patienten oder sonstigen Faktoren z.B. Geburtstag) den verschiedenen (Behandlungs-) Gruppen zugeordnet. Nur mit diesem Verfahren können bekannte und unbekannte Einflußfaktoren gleichmäßig auf beide Gruppen (wenn die Gruppen ausreichend groß sind) verteilt werden.

Erster Schritt: Therapiezuordnung und Vollständigkeit der Daten.

Der Wahrheitsgehalt einer Studie kann durch unbeabsichtigte Selektion von Patienten in die möglichen Behandlungsgruppen erheblich beeinflußt werden. Wenn die Zuordnung nicht zufällig erfolgt, besteht immer das Risiko, Patienten ungleich auf die Behandlungsgruppen zu verteilen. Demnach ist in einer randomisierten Studie vorab zu klären, ob bei einem Patienten jede der untersuchten Therapien durchgeführt werden könnte.

Die randomisierte Zuordnung zu den Behandlungsarmen ist bedeutend, weil es zahlreiche überraschende Beispiele gibt, in welchen Ergebnisse aus nicht-randomisierten Studien oder Erfahrung aus der täglichen Praxis in randomisierten Studien nicht bestätigt werden konnten (Tab II). Ein Vergleich der Ergebnisse einiger Studien, in welchen die Therapiezuteilung randomisiert bzw. nicht-randomisiert erfolgte, zeigte, daß die beobachteten Effekte in den nicht-randomisierten Studien größer als in den randomisierten Studien waren (Sacks HS et al, Chalmers TC et al, Colditz GA et al, Emerson JD et al.).

Selbst wenn für eine anstehende Entscheidung keine randomisierte Studie gefunden wird, ist eine Entscheidung oder Empfehlung abzugeben. Dabei sollte aber berücksichtigt werden, daß die Gültigkeit der verfügbaren Information einer nicht-randomisierten Studie geringer ist als die einer randomisierten Studie.

Ebenso bedeutend ist die Frage, ob alle Patienten, die in die Studie aufgenommen wurden, bei der Datenauswertung angemessen berücksichtigt wurden und zu den Schlußfolgerungen der Studie beigetragen haben. Herausgefallene Patienten führen zu fehlenden Informationen, was die Ergebnisse einer Studie allzu leicht verfälschen kann: Es ist denkbar, daß in der Kontrollgruppe die Informationen zu Patienten mit gutem Ergebnis und in der experimentellen Gruppe die Ergebnisse zu Patienten mit ungünstigem Ergebnis fehlen.

Bei randomisierten Studien ist außerdem darauf zu achten, daß alle Patienten in der Gruppe ausgewertet werden, in die sie randomisiert wurden, und zwar unabhängig davon, ob sie tatsächlich die in dieser Gruppe vorgesehene Behandlung erhalten haben ("intention-to-treat analysis"). Dieses Prinzip schützt vor einer Verfälschung des Ergebnisses der Randomisierung und nimmt dafür ein schwächer ausgeprägtes Studienergebnis in Kauf.

Als Beispiel könnte es vorkommen, daß ein Teil der Patienten, die in eine Gruppe A randomisiert wurden, die in dieser Gruppe vorgesehene Therapie im nachhinein ablehnt. Würden diese Patienten nun nachträglich als der Gruppe B zugehörig

ausgewertet (was der tatsächlich erhaltenen Therapie entspräche), dann käme dies einer Selektion von Patienten mit schlechter Compliance in die Gruppe A und von Patienten mit guter Compliance in die Gruppe B gleich⁵, wobei der Effekt der Randomisation verlorengelht und offensichtliche Ungleichheiten zwischen den Gruppen entstehen. Wertet man nach dem „intention-to-treat“-Prinzip aus, so tritt dieser Selektionseffekt nicht auf, allerdings mag der Behandlungseffekt nicht mehr so klar erkennbar sein, da ja der experimentell erzeugte Unterschied zwischen Gruppe A und Gruppe B in Bezug auf die Therapie wegen der teilweise abgelehnten Behandlungen geringer ausfällt.

Zweiter Schritt: Verblindung, Gruppengleichheit, Begleittherapien.

Wenn die beiden Fragen im ersten Schritt positiv beantwortet sind, benötigt der Arzt, der eine bisher nicht auf Validität geprüfte Therapie auswählt, zusätzlich Informationen zur Vergleichbarkeit der Gruppen, zur "Verblindung" und zu den Begleittherapien.

Die Vergleichbarkeit der Gruppen kann durch die Randomisierung alleine nicht immer gewährleistet werden. Je kleiner die Gruppen sind, um so höher ist das Risiko, daß sich auch randomisierte Gruppen auf Grund der Zufallsschwankungen bereits zu Beginn der Therapie unterscheiden. Um dieses Risiko so gering wie möglich zu halten, sollte die Verteilung der bekannten prädiktiven Faktoren auf die randomisierten Gruppen geprüft und dokumentiert sein. Falls sich bei Auswertung der Studie zeigt, daß die Prädiktoren ungleich verteilt waren, können die Ergebnisse mit statistischen Methoden adjustiert werden. Das Vertrauen in den Wahrheitsgehalt der Studie wird erhalten bleiben, wenn die Schlußfolgerungen mit und ohne Adjustierung identisch sind.

Einfach zu messende und aussagekräftige Prädiktoren für die Überlebenszeit sind einzelne Dimensionen der Lebensqualität, deren einmalige Bestimmung zu Beginn einer Studie wertvolle Informationen liefert (Coates A).

Patienten, die wissen, daß sie eine neue experimentelle Therapie erhalten, werden ebenso wie ihre Ärzte in der Regel eine positive Erwartungshaltung einnehmen; diese Haltung kann die Wahrnehmung und die Berichterstattung von Ereignissen beeinflussen. Es ist bekannt, daß auch Testergebnisse unterschiedlich ausfallen, je nachdem, ob das Personal der Studie "verblindet" ist oder ob es die erwarteten Ergebnisse kennt (Guyatt GH et al 1984).

⁵ In verschiedenen Studien wurden bei Patienten mit guter Compliance zum Beispiel längeres Überleben beobachtet, als bei Patienten mit schlechter Compliance, selbst wenn die Behandlung lediglich aus einem Placebo bestand (Coronary Drug Project Research Group, Asher WL et al, Hogarty GE et al, Fuller R et al, Pizzo PA et al, Horwitz RI et al).

Letztlich sollte gewährleistet sein, daß die zusätzlich zur experimentellen Therapie durchgeführten Behandlungen, die sogenannten Begleittherapien, in den Behandlungsgruppen gleich sind. Leider werden Begleittherapien in vielen Studien nicht erfaßt oder nach Ermessen der behandelnden Ärzte angewandt; in diesen Fällen ist zu überlegen, ob die beschriebenen Ziele der Studientherapie auch durch andere, nicht dokumentierte Interventionen (z.B. Steroidtherapie) verursacht sein können. In einer sorgfältig geplanten Studie sind die zulässigen Begleittherapien im Abschnitt "Methoden" beschrieben.

Bewertung der Bedeutung (Effektgröße)

Für die Beschreibung der Effektgröße existieren verschiedene Darstellungsarten, zum Beispiel Relative Risiko-Reduktion oder Odds-Ratio (bei Metaanalysen) im Vergleich zwischen Experimental- und Kontrolltherapie („Senkung der Wahrscheinlichkeit eines Re-Infarkts um 14%" oder „die Odds-Ratio für einen Re-Infarkt unter Experimental vs. Kontrolltherapie ist 6:7). Die Bedeutung dieser Aussagen für die Praxis ist unter anderem noch abhängig von der Prävalenz der betrachteten Ereignisse.

Für die Praxis interessanter wäre zum Beispiel die Frage: wie viele Patienten muß ich therapieren (z.B. Beta-Blocker direkt nach dem Herzinfarkt geben), um schließlich bei einem Patienten das unerwünschte Ereignis (z.B. eine tödliche Arrhythmie) zu verhindern?

Die Antwort auf diese Frage läßt sich leicht errechnen (number needed to treat, NNT):

- Zunächst bestimmt man den Anteil der Patienten, bei welchem das unerwünschte Ereignis in der Kontrollgruppe auftritt (control event rate, CER = Anzahl Ereignisse / Anzahl der Patienten).
- Dann bestimmt man den Anteil der Patienten, bei welchem das unerwünschte Ereignis in der Experimentalgruppe auftritt (experimental event rate, EER = Anzahl Ereignisse / Anzahl der Patienten).
- Die Differenz zwischen diesen beiden Werten ergibt die absolute Risikoreduktion (absolute risk reduction, ARR = CER - EER)⁶.

⁶ Bezieht man die absolute Risiko Reduktion (ARR) auf den Ausgangswert, d.h. auf die Ereignisrate in der Kontrollgruppe (CER), ergibt sich die relative Risikoreduktion (relative risk reduction, RRR = ARR/CER), die nur zusammen mit der ARR berichtet werden sollte, um Fehlschlüsse zu vermeiden:

Eine Absolute Risikoreduktion von CER = 0,001% auf EER = 0,0005%, also

- Die NNT ergibt sich aus dem Kehrwert der ARR ($NNT = 1/ARR$).

Bewertung der Brauchbarkeit (klinische Anwendbarkeit)

Für die Beurteilung der klinischen Anwendbarkeit von Studienergebnissen ist es erforderlich, daß konkrete Vorstellungen zu den Zielen und Risiken der Therapie bestehen. Diese Ziele und Risiken des Patienten sollten mit jenen der bewerteten Studie näherungsweise übereinstimmen.

Generell kann davon ausgegangen werden, daß eine Therapie aus der Sicht des Patienten als nützlich angesehen wird, wenn Quantität und/oder Qualität des Lebens günstig beeinflusst werden können. Der Nachweis der Wirksamkeit alleine, z.B. durch Verbesserung biochemischer oder physikalischer Parameter, wird aus der Sicht des Patienten nicht notwendigerweise als nützlich empfunden; entscheidend ist der Einfluß auf die beiden genannten Dimensionen, die wir zur Quantifizierung des Nutzens medizinischer Maßnahmen diskutiert haben (Por 1995, QOL).

Anwendung des Konzepts auf das Szenario

Validität der Studienergebnisse

Erster Schritt: Therapiezuordnung und Vollständigkeit der Daten.

Die Patienten wurden den Therapiegruppen durch Randomisierung zugeordnet. Patienten, welchen eine Orchiektomie wegen des körperlichen Gesundheitszustandes nicht zugemutet werden konnte, wurden vor der Randomisierung ausgeschlossen.

Es wird nicht berichtet, wieviele Patienten insgesamt in die Studie eingeschlossen wurden, somit ist nicht klar, welcher Anteil der Patienten letztlich in die Auswertung eingegangen ist.

$ARR = 0,001\% - 0,0005\% = 0,0005\%$, entspräche einer relativen Risikoreduktion von $RRR = ARR/CER = 0,0005\%/0,001\% = 50\%$ - also auf den ersten Blick bemerkenswert!

Da das zu verhindernde Ereignis aber schon von sich aus nur selten auftritt, wird das Risiko des einzelnen Patienten durch die untersuchte Intervention zwar um 50% gesenkt, aber eben nur von 1 in 100.000 auf 1 in 200.000. Man müßte somit $NNT = 1/ARR = 1/0,0005\% = 1/0,000005 = 200.000$ Patienten behandeln, um bei einem einzigen den erwünschten günstigen Effekt zu erzielen!

Deshalb sollte statt der oder zusätzlich zur RRR immer die ARR und/oder die NNT angegeben werden. Bitte beachten: 0,001% und 0,00001 sind zwei verschiedene Schreibweisen für exakt denselben Zahlenwert.

Alle 486 Patienten, die den Endpunkt "Therapieversagen" oder die "Kontrolle nach 3 Monaten" erreicht hatten, wurden ausgewertet.

Für die Analyse der Überlebenszeit wurden 516 Patienten berücksichtigt.

Subjektive Kriterien und die Variablen der Lebensqualität wurden nur bei Patienten erhoben, die zu Beginn der Studie symptomatisch waren.

Günstig wäre es, explizit über Ein- und Ausschlußkriterien, die Zahl der in die Studie eingeschlossenen Patienten, der randomisierten Patienten, der im Studienverlauf herausgefallenen Patienten sowie über die Begründungen dieser Ausfälle zu Berichten.

Möglicherweise wäre es sinnvoll gewesen, bei allen Patienten den Verlauf der Lebensqualität zu dokumentieren, oder die Begründung dafür zu berichten, weshalb dies nicht gemacht wurde.

Zweiter Schritt: Gruppengleichheit, Verblindung, Begleittherapien.

Zur Verblindung werden keine Angaben gemacht.

Die Gruppen waren zu Beginn der Behandlung vergleichbar, eine Tabelle mit diesbezüglich überprüften Parametern ist eingefügt.

Begleittherapien werden nicht erwähnt.

Günstig wäre es, anzugeben, ob Patienten und Personal (soweit möglich) bezüglich der Therapiezuordnung der Patienten verblindet waren, und wie die Geheimhaltung der Therapiezuordnung sichergestellt war.

Sofern keine Begleittherapien durchgeführt wurden, wäre eine entsprechende Aussage im Text ebenfalls informativ.

Zusammenfassende Bewertung der Validität

In der betrachteten Arbeit können mögliche methodische Schwächen insbesondere in Bezug auf Patientenselektion und ausreichende Erhebung der Lebensqualitäts-Parameter nicht ausgeschlossen werden. Deshalb können ihre Ergebnisse nicht ohne Vorbehalt als valide betrachtet werden.

Bedeutung der Studienergebnisse

In der Studie wird als „Wirksamkeit“ die Verhinderung der Tumorprogression definiert. Im Sinne dieser Definition ist die Wirksamkeit der Kastration größer (ARR = 11%; NNT = 9)⁷, als diejenige der BIC-Therapie. Allerdings wird dieser Effekt ausschließlich durch objektive Krankheitsprogressionen, nicht durch Todesfälle oder durch Ausschlüsse gestützt. In Bezug auf die Überlebenszeit wurde kein Unterschied zwischen den Therapiegruppen beobachtet.

Bei der Verbesserung der Symptome ergab sich ein Vorteil der BIC-Gruppe (ARR = 6,7%; NNT = 15)⁸ gegenüber der KA-Gruppe.

Unter den sonstigen Wirkungen war die Reduktion des Prostata-spezifischen Antigens und des Testosterons in der KA-Gruppe wesentlich deutlicher ausgeprägt als in der BIC-Gruppe.

Unter den Nebenwirkungen traten Schmerzen in der Brust und Gynäkomastie in der BIC-Gruppe häufiger auf, während urogenitale Symptome und Hitzewallungen in der KA-Gruppe häufiger beobachtet wurden.

Brauchbarkeit der Studienergebnisse

In der Studie wurden alle klinisch bedeutenden Ereignisse und Ergebnisse berücksichtigt; Begleiterkrankungen und deren Therapien sind im vorliegenden Fall nicht entscheidend. Die Patienten der Studie sind mit demjenigen des Szenarios vergleichbar.

Allerdings lassen die oben beschriebenen Schwächen es problematisch erscheinen, alleine auf diese Arbeit eine Empfehlung an einen Patienten zu stützen. Um sicherzustellen,

⁷ In der Studie wurden nicht Experimentaltherapie vs. keine Therapie verglichen, sondern Therapie BIC vs. Therapie KA. Deshalb bezieht sich die ARR auf die Differenz der Risiken einer Tumorprogression unter KA bzw. unter BIC. Die NNT = 9 bedeutet dann, daß 9 Patienten sich einer Kastration anstatt einer BIC-Therapie unterziehen müßten, um hierdurch bei 1 zusätzlichen Patienten eine Tumorprogression zu verhindern.

⁸ Es müßten also NNT = 15 Patienten mit BIC behandelt werden, anstatt sich einer Kastration zu unterziehen, um hierdurch die definierte Verbesserung der Symptome bei 1 zusätzlichen Patienten zu erreichen.

daß wirklich die beste externe Evidenz für die Therapieentscheidung beim individuellen Patienten berücksichtigt werden kann, sollten hier die Ergebnisse der MEDLINE-Recherche nach weiteren möglicherweise verwendbaren Arbeiten durchsucht werden, die dann genauso wie die hier behandelte zu prüfen sind.

Für die im Szenario vorgestellte Fragestellung kann jedoch wegen der uneinheitlichen Aussagen der Studien auch nach Durchsicht aller relevanten Arbeiten keine klare Empfehlung abgegeben werden, die ausschließlich auf externer Evidenz beruht. Unser Beispiel zeigt, daß in diesem Fall - wenn die externe Evidenz keine einheitliche Lösungen anbietet - die persönliche Erfahrung des Arztes, die „interne Evidenz“, für die Entscheidung und im vorliegenden Beispiel für die Beratung des Patienten erforderlich ist.

Zusammenfassung

Anhand eines Patientenbeispiels wurde gezeigt, wie Methoden der Evidence-Based Medicine in der Praxis angewandt werden können. Eine Arbeit aus der Literatur wurde anhand weniger Fragen auf ihre Qualität und auf ihre Anwendbarkeit auf den Patienten des Szenarios überprüft. Es wurde dargestellt, daß nach dem Erkennen von methodischen Schwächen einer Arbeit deren Ergebnisse nicht unkritisch verwendet werden sollten. Die vorhandene externe Evidenz kann in vielen Fällen unser klinisches Handeln leiten, wenn sie eindeutige Antworten gibt und ihre Qualität den dargestellten Anforderungen entspricht. Externe Evidenz ersetzt keine klinische Erfahrung aber klinische Erfahrung ohne Berücksichtigung der externen Evidenz wird dem zeitgemäßen Anspruch der Medizin auch nicht gerecht.

Tab I: Leitfaden für die Bewertung eines Artikels zur Therapie
(**modifiziert** nach Guyatt GH et al 1993)

Validität der Studienergebnisse

Erster Schritt:

Wurden die Patienten zufällig, d.h. durch Randomisation den Therapiegruppen zugeordnet?

Wurden die Daten aller Patienten, die in die Studie aufgenommen wurden, in den Schlußfolgerungen der Studie angemessen berücksichtigt?

Wie wird über fehlende Daten berichtet, wie werden sie erklärt?

Zweiter Schritt:

Waren die Patienten und das an der Studie beteiligte Personal "verblindet" (s. Anmerkung)

Waren die Gruppen zu Beginn der Behandlung vergleichbar? Wie und anhand welcher Parameter wurde dies überprüft?

Waren die Begleittherapien in den Behandlungsgruppen vergleichbar?

Bedeutung der Studienergebnisse

Wie groß war der absolute Behandlungseffekt?

Wie genau war der Schätzwert des Behandlungseffekts (Streuung)?

Brauchbarkeit der Studienergebnisse

Können die Ergebnisse auf meinen Patienten übertragen

werden?

Wurden alle klinisch bedeutenden Ereignisse und Ergebnisse berücksichtigt?

Überwiegt der zu erwartende Nutzen der Therapie die möglichen Nachteile und Belastungen?

Anmerkung: In einer Doppel-blind-(Medikamenten-)Studie wissen weder der Patient noch der Arzt oder das Studienpersonal, ob ein bestimmter Patient die experimentelle Behandlung oder die Kontrollbehandlung (z.B. Placebo) erhält. Bei manchen Therapien (z.B. Operationen) können die Beteiligten nur teilweise verblindet werden; bei diesen Studien ist es um so wichtiger, daß die Aus- und Bewertung der Ergebnisse soweit wie möglich "verblindet" durchgeführt wird.

Tab II: Unerwartete Befunde aus randomisierten Studien (nach Guyatt GH et al 1993)

Befund	Referenz
Keine wirksame Prävention apoplektischer Insulte durch extrakraniellen-intrakraniellen Bypass	Haynes RB et al 1987
Bei Patienten mit Sepsis wird die Mortalität durch Steroide eher erhöht als gesenkt	Bone RC et al 1987
Steroid-Injektionen in Gelenke lindern keine Rückenschmerzen	Carette S et al 1991
Plasmapherese ist bei Patienten mit Polymyositis nicht nützlich	Miller FW et al 1992

Literatur

Asher WL and Harper HW: Effect of human chorionic gonadotropin on weight loss, hunger, and feeling of well-being. Am J Clin Nutr 1973;26:211-218

Bone RC, Fisher CJ, Clemmer TP et al: Controlled trial of high-dose methylprednisolone in the treatment of severe sepsis and septic shock. N Engl J Med 1987;317:653-658

Carette S, Marcoux S, Truchon R et al: A controlled trial of corticosteroid injections into facet joints for chronic low back pain. N Engl J Med 1991;325:1002-1007

Chalmers TC, Celano P, Sacks HS, Smith H Jr: Bias in treatment assignment in controlled clinical trials. N Engl J Med 1983;309 1358-1361

Chodak G, Sharifi R, Kasimis B, Block NL, Macramalla E, Kennealey GT: Single-agent therapy with bicalutamide: A comparison with medical or surgical castration in the treatment of advanced prostate carcinoma. Urology 1995;46:849-855

Coates A, Porzsolt F, Osoba D: Quality of life in oncology

practice: Prognostic value of EORTC QLQ-C30 scores in patients with advanced malignancy. Eur J Cancer 1997;33:1025-1030

Colditz GA, Miller JN, Mosteller F: How study design affects outcomes in comparisons of therapy , I: Medical. Stat med 1989; 8:441-454

Coronary Drug Project Research Group. Influence on adherence treatment and response of cholesterol on mortality in the Coronary Drug Project. N Engl J Med 1980;303:1038-1041

Drummond MF, Stoddart GL, Torrance GW: Methods for the economic evaluations of health care programmes. Oxford 1987, Oxford Medical Publications 1987, pp 5 - 17

Emerson JD, Burdick E, Hoaglin DC et al: An empirical study of the possible relation of treatment differences to quality scores in controlled randomized clinical trials. Controlled Clinical Trials 1990;11:339-352

Fuller R, Roth H, Long S: Compliance with disulfiram treatment of alcoholism. J Chronic Dis 1983;36:161-170

Glossar: Siehe jeweils Anhang zur englischen oder deutschen Ausgabe der Zeitschrift Evidence-Based Medicine

Guyatt GH, Pugsley SO, Sullivan MJ et al: Effect of encouragement on walking test performance. Thorax 1984;39:818-822

Guyatt GH, Sackett DL, Cook DJ: Users' guide to medical literature. II. How to use an article about therapy or prevention. JAMA 1993;270:2598-2601

Hogarty GE and Goldberg SC: Drug and sociotherapy in the aftercare of schizophrenic patients: one-year relapse rates. Arch Gen Psychiatry 1973;28:54-64

Horwitz RI, Visoli CM, Berman L et al: Treatment adherence and risk of death after myocardial infarction. Lancet 1990;336:542-545

Miller FW, Leitman SF, Cronin ME et al: Controlled trial of plasma exchange and leukapheresis in polymyositis and dermatomyositis. N Engl J Med 1992;326:1380-1384

Pizzo PA, Robichaud KJ, Edwards BK et al: Oral antibiotic pro-

phylaxis in patients with cancer: A double-blind randomized placebo-controlled trial. J Pediatr 1983;102:125-133

Porzsolt F, Wirth A, Mayer-Steinacker R, Voigt W, Heymer B, Link K-H, Schulte M, Röttinger E: Qualitätssicherung in der Onkologie: Zieldefinition und Zielerreichung am Beispiel von Therapieempfehlungen des Interdisziplinären Onkologischen Konsils des Tumorzentrums Ulm. Münch. med. Wschr. 1994;136:371-376

Porzsolt F, Druckrey, E.: Zur Bewertung medizinischer Maßnahmen. Strahlenther. Onkol. 1995; 171:371-378

Porzsolt F, Wirth A, Mayer-Steinacker R, Schulte M, Negri G, Suhr P, Link K-H, Gaus W, Röttinger EM: Quality assurance by specification and achievement of goals in palliative cancer treatment. Cancer Treatment Reviews 1996;22:41-50

Porzsolt F: Klinische Ökonomik - ein Leitfaden für ökonomisches Arbeiten in der Klinik und Praxis. Med Klin 1996;91(8): 531-535

Porzsolt F, Kunz R: Evidence-Based Medicine: Medizinischer Fortschritt nicht nur an einer Front. (Editorial). Evidence Based Medicine (Deutsche Ausgabe) 1997;1(2):1

Sacks HS, Chalmers TC, Smith H Jr: Randomized versus historical assignment in controlled clinical trials. Am J Med 1983;309: 1353-1361

Bales GT, Chodak GW: A controlled trial of bicalutamide versus castration in patients with advanced prostate cancer. Urology. 1996; 47(1A Suppl): 38-43; discussion 48-53

AB: OBJECTIVES. Bicalutamide is a new, potent antiandrogen with potential efficacy in the treatment of men with advanced prostate cancer. Although no pure antiandrogen has been well studied versus castration, potentially fewer adverse effects could occur, making such an agent a potentially useful alternative therapy. To date, three randomized controlled trials have been performed comparing these two treatments. In preliminary studies, a dose of 50 mg bicalutamide per day was selected for these trials. In two of the studies (0302, 0303), this drug was compared to either medical or surgical castration, the latter choice being made by the patient. In the third study (0301), bicalutamide was compared to bilateral orchiectomy. METHODS. Using an intention-to-treat format, the outcomes assessed were time-to-treatment failure, time-to-objective disease progression, subjective response, and survival time in men with previously untreated metastatic disease. The incidence of breast tenderness, gynecomastia, and hot flushes was also determined in both treatment arms. A quality-of-life questionnaire was administered on multiple occasions after initiation of therapy. RESULTS. Based on an analysis of > 1000 patients, the objective and subjective results favored castration over bicalutamide (50 mg/day). The hazard ratios for time-to-treatment failure (1.59), time-to-disease progression (1.62), and median survival (1.44) were all significantly greater in the castration group ($P > 0.001$). Another difference noted at 3 months was a significantly lower median fall in prostate-specific antigen values in the bicalutamide group (86-88% versus 96-97%). Symptomatic patients receiving bicalutamide were only 0.43 times as likely to have subjective improvement as the patients treated by castration. A comparison of pharmacologic effects showed that only the incidence of hot flushes was lower in the bicalutamide group, whereas breast tenderness and gynecomastia were more common. This difference in hot flushes, however, translated into better quality of life during the first several months with regard to sexual relations and sexual functioning. CONCLUSIONS. Bicalutamide monotherapy at 50 mg/day appears inferior to castration in overall objective and subjective response rates. Whether higher doses of bicalutamide can compete more favorably will need to be tested in further clinical trials.

MESH: Aged-; Aged,-80-and-over; Middle-Age; Neoplasm-Staging; Prostatic-Neoplasms-mortality; Prostatic-Neoplasms-pathology; Survival-Rate

MESH: *Androgen-Antagonists-therapeutic-use; *Anilides-therapeutic-use; *Orchiectomy-; *Prostatic-Neoplasms-therapy

TG: Comparative-Study; Human; Male; Support,-Non-U.S.-Gov't

PT: CLINICAL-TRIAL; JOURNAL-ARTICLE; RANDOMIZED-CONTROLLED-TRIAL