

I. Einleitung

Epitheliale Malignome sind für mehr als 80% der Krebstodesfälle verantwortlich; in den westlichen Industriestaaten fallen ihnen jährlich mindestens 1 Million, in der Bundesrepublik Deutschland allein über 200 000 Menschen zum Opfer. Zu den epithelialen Tumoren zählen unter anderem fast sämtliche Krebse der Lokalisationen Luftröhre, Bronchus, Lunge, Magen, Kolon/Rektum, Oesophagus, Brust, Blase, Ovar, Uterus, Pankreas, Kopf/Hals, und Leber¹.

Obwohl genaue Erhebungen zur Art der Behandlung von Patienten mit fortgeschrittenen Tumoren fehlen, muß man doch vermuten, daß die Mehrzahl von ihnen vor ihrem Tod irgendeine Form der systemischen zytotoxischen Therapie erhalten haben und daß eine solche Therapie in fast allen Fällen zumindest erwogen wird. So zählen DeVita et al. (1989) sämtliche rund 557 000 US-amerikanischen Patienten pro Jahr², bei denen ein disseminierter Tumor festgestellt wird oder eine Rekurrenz nach einer initialen Behandlung einer lokalisierten Erkrankung auftritt, zu den Kandidaten für Chemotherapie³. Für alle Patienten mit fortgeschrittenen Krebserkrankungen ist die Abwägung des Nutzens und Schadens einer möglichen zytostatischen Behandlung von Bedeutung. Zwei Fragen sind in diesem Zusammenhang entscheidend; erstens: Ist die Therapie imstande, die Überlebenszeit der Patienten zu verlängern? Und zweitens: Vermag sie zum Wohlbefinden (zur »Lebensqualität«) der Patienten beizutragen? In der vorliegenden Arbeit soll versucht werden, beide Fragen nach dem heutigen Stand des Wissens zu beantworten. Die Antworten sind nicht zuletzt auch von wirtschaftlichem Interesse, denn auch bei konservativer Schätzung liegen die jährlichen Kosten der palliativen Chemotherapie in der Bundesrepublik Deutschland im Milliardenbereich.

Der Glaube an die lebensverlängernde Wirkung

Der Laie, aber auch der mit der klinischen Onkologie nicht so vertraute Mediziner wird zumindest die Frage nach der lebensverlängernden Wirkung einer von erfahrenen Klinikern durchgeführten Chemotherapie für rein rhetorisch und überflüssig halten, denn selbstverständlich wird er voraussetzen, daß eine notorisch toxische und kostspielige Behandlung, wenn sie schon nicht heilen kann, doch wenigstens einen günstigen Einfluß auf die Prognose hat. Gestützt wird diese Auffassung durch die unbestreitbaren, weithin bekannten und z. T. dramatischen Erfolge bei manchen nichtepithelialen Malignomen (z. B. Keim-

¹ Nicht unter die epithelialen Malignome fallen vor allem die Lymphome und Leukämien, Sarkome sowie Keimzelltumoren. Sämtliche Ausführungen dieses Artikels beziehen sich, sofern nicht anders vermerkt, auf epitheliale Tumoren. Mit dem Attribut »fortgeschritten« werden nicht radikal resezierbare, rezidivierende oder disseminierte Erkrankungen bezeichnet.

² In derselben Publikation wird die Zahl der Krebsneuerkrankungen ohne nicht-melanomatöse Hautkrebs und ohne Zervixkarzinom in situ in den USA auf 598 000 beziffert.

³ Unter Chemotherapie wollen wir generell nur zytostatische Chemotherapie verstehen.

zelltumoren, akuten Leukämien, Morbus Hodgkin oder hochmalignen Non-Hodgkin-Lymphomen; s. z. B. Fritze, 1986), aber auch durch Erfolgsbehauptungen für epitheliale Tumoren, die sich in wissenschaftlichen Publikationen und onkologischen Standardwerken ebenso wie in den für die Öffentlichkeit bestimmten Broschüren und Stellungnahmen finden (z. B. DeVita et al., 1982, 1985; Berdel u. Fink, 1984; Chabner et al., 1984; Scheel u. Aumiller, 1984, S. 66; Frei 1985; Ross et al., 1985; Senn, 1985; Nagel u. Wander, 1986; Hossfeld, 1986; Greenfield et al., 1987; Buzdar, 1988; Ullmann, 1988; Schraub u. Berneheim, 1988; Hryniuk, 1988a⁴; Wilmanns, 1988⁵; Jacquillat u. Christen, 1989; Schoen u. Wendt, 1989; Weidmann u. Niederle, 1995). Häufig ist in Publikationen – gewollt oder ungewollt mißverständlich und vage – von »Erfolgen« die Rede, ohne daß spezifiziert würde, worin die Erfolge bestehen. So heißt es, um nur ein Beispiel anzuführen, in dem umfangreichen Werk »Current Cancer Therapeutics« von Kirkwood et al (1994; S. 19): »The antitumor antibiotics comprise a heterogeneous group of chemotherapy agents that have proven to be extremely successful in the treatment of a variety of tumors, both epithelial and mesenchymal.«

Möglicherweise sind die zitierten Wissenschaftler nicht repräsentativ für die internistischen Onkologen. Denn soweit es den mir zugegangenen Äußerungen zu entnehmen war, stand bereits in den achtziger Jahren die Mehrzahl von ihnen in Wahrheit der Zytostatikatherapie durchaus kritisch gegenüber, und diese distanzierte Haltung hat in den letzten Jahren an Boden gewonnen, wie man an Äußerungen renommierter Onkologen ablesen kann (z. B. Hossfeld, 1990; Bonadonna, 1990; Mende, 1992; Possinger et al., 1993; Kaufmann, 1994; Kleeberg, 1994)⁶; doch haben die positiven Stellungnahmen das Bild, das man sich in der Öffentlichkeit von der Chemotherapie macht, zweifellos beeinflußt.

Besondere Erwähnung verdienen die Schätzungen von DeVita et al. (1985, 1989), weil sie – nicht zuletzt wegen des Renommées der Autoren – in mehreren nachfolgenden Publikationen übernommen wurden. DeVita et al. vertraten die Meinung, daß die relativen 5-Jahres-Überlebensraten⁷ ein gutes Maß zur Bestimmung der therapeutischen Erfolge und der Heilbarkeit von Krebs darstellen. Aufgrund von Vergleichen der Raten aus aufeinanderfolgenden Perioden kamen die Autoren zu dem Schluß, daß ein großer Teil des rund 30-prozentigen Anstiegs

⁴ Der Glaube dieses Autors an Therapieerfolge der Chemotherapie kommt bereits im Titel seiner Publikation (»More is better«) zum Ausdruck.

⁵ Wilmanns zitiert K. Brunner mit den Worten: »Bei 30% führt die palliative wirksame Therapie zu einer Lebensverlängerung.«

⁶ Bemerkenswert ist auch, daß die neueste Auflage des amerikanischen Standardwerks über Onkologie (DeVita et al, 1993) im Kapitel über zytostatische Therapie die früher aufgestellten Erfolgsbehauptungen nicht mehr enthält.

⁷ »Relativ« bedeutet hier: Nach Adjustierung für die in einer Normalbevölkerung mit gleicher Altersstruktur auftretende Mortalität. Relative Überlebensraten von Krebspatienten ab dem Zeitpunkt der Diagnosestellung sind für mehrere Staaten publiziert worden, so z. B. für die USA (Ries et al., 1983; Boring et al., 1992), Finnland (Hakulinen et al., 1981), Norwegen (The Cancer Registry of Norway, 1980) und Großbritannien (Cancer Statistics Group, 1982). Die finnische und die norwegische Übersicht enthalten zudem gesonderte Statistiken für disseminierte (»non-local«) Krebse.

der 5-Jahres-Überlebensraten seit den fünfziger Jahren auf den Einsatz von Chemotherapie zurückzuführen ist. Auch behaupteten sie, daß durch Chemotherapie ein Teil der Patienten mit disseminierten kleinzelligen Bronchialkarzinomen und Ovarialkarzinomen geheilt werde und daß eine relevante lebensverlängernde Wirkung der palliativen Chemotherapie im Falle des Mamma-, Endometrium- und Magenkarzinoms sowie der Plattenepithelkarzinome des Kopf/Halsbereichs nachweisbar sei.

Das in diesen Behauptungen ausgedrückte Vertrauen in die Analysen säkularer Trends ist aus zwei Gründen erstaunlich: Erstens ist die Gleichsetzung der Heilung mit dem 5-Jahres-Überleben irreführend. Sie wird auch nicht durch den Hinweis gerechtfertigt, daß die relativen Überlebensraten nach 5 Jahren verhältnismäßig stabil bleiben, was sich darin zeigt, daß die 20-Jahres-Überlebensraten rund 85% der 5-Jahres-Überlebensraten betragen, denn diesen hohen Prozentsatz erhält man nur bei ungerechtfertigter Zusammenfassung der primär disseminierten mit den lokalisierten Tumoren. Für die disseminierten Tumoren stellt er eine außerordentliche Beschönigung der Verhältnisse dar. Für sie ist einzig die differenzierte Betrachtungsweise von Interesse, die die vorhandene Information über die Ausbreitung des Tumorleidens berücksichtigt. Zweitens ist seit geraumer Zeit bekannt, daß historisch vergleichende Untersuchungen von Therapieergebnissen mit zahlreichen Fehlerquellen zu kämpfen haben und zur Quantifizierung und Beurteilung von Erfolgen fast immer inadäquat sind (vgl. Enstrom u. Austin, 1977; Doll u. Peto, 1980; Sacks et al., 1982; Pocock, 1983, S. 54–55; Smith, 1983; Cairns, 1985; Feinstein et al., 1985; Zelen, 1985; Tonkin u. Tannock, 1988). In *Tabelle 1* sind die hauptsächlichen Fehler und Verzerrungen aufgelistet. Sie alle führen tendenziell dazu, die in neuerer Zeit gefundenen Therapieergebnisse oder -erfolge zu überschätzen, und sie sind der Grund dafür, daß man heute der Auffassung ist, wirklich zuverlässige Informationen über den relativen Wert zweier Therapien seien nur in randomisierten Vergleichen zu gewinnen.

Tabelle 1 Verzerrungsquellen bei Therapievergleichen mit historischen Kontrollen und der Analyse von Trends in den Überlebensraten von Krebspatienten

- Verbesserungen in der Früherkennung und Metastasendiagnostik
 - Stadienwanderung
 - Intensivierung und Verbesserung der supportiven Maßnahmen
 - Unterschiede bzw. Änderungen in der Struktur des Patientenguts
 - Auswahlkriterien (z. B. konsequenter Ausschluß prognostisch ungünstig gelagerter Fälle)
 - Selbstselektion (z. B. Präferenzen gewisser Patientengruppen für bestimmte Therapieformen)
 - Herkunft und Zusammensetzung des Patientenguts
 - Unterschiede in der Qualität des Monitorings und der Datenerhebung
 - Publikationsbias
-

Nebenbei bemerkt können nicht nur historische Gegenüberstellungen, sondern auch Vergleiche von gleichzeitig behandelten Therapiekohorten irreführend sein, da bereits aufgrund der unterschiedlichen Patientenstruktur erhebliche Überlebenszeitunterschiede auftreten können. So haben mehrere in den USA durchgeführte Untersuchungen deutliche Zusammenhänge zwischen soziodemographischen Charakteristika einerseits und der Überlebenszeit andererseits festgestellt (z. B. Chirikos u. Horner, 1985; Hankey u. Myers, 1987; LeMarchand et al., 1987; Cella et al., 1991). Auch wurden aufgrund der unterschiedlichen Patientenstruktur erhebliche Überlebenszeitunterschiede zwischen verschiedenen Hospitälern festgestellt (Mettlin et al., 1987); große Differenzen in der Überlebenszeit bestanden z. B. sogar zwischen weißen und schwarzen Brustkrebspatientinnen gleichen Tumorstadiums, die in derselben Institution mit identischer Therapie behandelt wurden (Kimmick et al., 1991; Siegel et al., 1994)

Von besonderer Bedeutung für die Analyse säkularer Trends sind die beiden in der Tabelle zuerst genannten Effekte. Je besser und früher die Diagnose der Krankheit oder der Metastasierung gelingt, um so länger wird die Überlebenszeit ab dem Diagnosezeitpunkt sein, auch ohne daß irgendwelche Fortschritte in der Therapie erzielt worden wären. Auf das Phänomen der Stadienwanderung haben u. a. Feinstein et al. (1985) hingewiesen (vgl. auch Tonkin u. Tannock, 1988). Das Ansteigen der diagnostischen Sensitivität durch die Anwendung fortschrittlicher Diagnosetechniken führt dazu, daß die Stadienverteilung der Patienten sich zunehmend in Richtung der ungünstigen Stadien verschiebt. Dies hat die auf den ersten Blick paradoxe Folge, daß sich die Prognose sowohl in den frühen als auch in fortgeschrittenen Stadien verbessert: Die frühen Stadien profitieren davon, daß ihnen weniger prognostisch ungünstige Patienten fälschlich zugeordnet werden, die fortgeschrittenen Stadien davon, daß zusätzlich relativ günstig gelagerte Fälle (nämlich solche, deren ungünstiges Stadium früher unerkannt geblieben wäre) in sie aufgenommen werden. Insgesamt besteht das Dilemma, daß bei Zusammenfassung aller Stadien scheinbare Erfolgstrends in den Überlebenszeitstatistiken auf Verbesserungen der Frühdiagnostik und primären chirurgischen Therapie zurückgeführt werden können, während bei differenzierter Betrachtung nach Ausbreitungsstadien die Stadienwanderung den Schluß auf Fortschritte in der systemischen palliativen Therapie verbietet.

Auf den letzten in Tabelle 1 aufgeführten Punkt, den »Publikationsbias« werden wir weiter unten noch ausführlicher eingehen, denn er spielt nicht nur bei Vergleichen mit historischen Kontrollen eine Rolle, sondern ist quasi eine ubiquitäre Verzerrungsquelle in der Medizin (und nicht nur dort!).

Untersuchungen zur lebensverlängernden Wirkung

Ziel der vorliegenden Arbeit ist eine kritische Durchsicht der Literatur im Hinblick auf die Frage der lebensverlängernden Wirkung der Chemotherapie bei fortgeschrittenen epithelialen Malignomen. Die vorhandene Evidenz stammt aus Studien mit ganz unterschiedlicher Form und Aussagekraft. Um die Schlüssigkeit der von den Verfassern vorgetragenen Argumente und Interpretationen beurteilen zu können, bedarf es einiger analytischer Betrachtungen der »Logik«,

d. h. der erkenntnistheoretischen Aspekte, der klinisch-onkologischen Forschung. Diese Analysen bilden den zweiten Schwerpunkt der Darstellung.

Das Beweismaterial für die Wirkungen einer Therapie auf die Prognose wurde zunächst in *direkte* und *indirekte* Evidenz eingeteilt (Tabelle 2). Direkte Evidenz erlaubt ohne zusätzliche Annahmen den Schluß auf das Vorhandensein oder die Abwesenheit der behaupteten Therapiewirkungen; sie stammt immer aus einer randomisierten Studie. Es gibt drei Arten solcher Studien, die sich in ihrem Stellenwert und der Aussagekraft positiver und negativer Befunde hinsichtlich der uns interessierenden Frage deutlich voneinander unterscheiden:

1. Randomisierte Vergleiche von Patienten, die mit der fraglichen Therapie behandelt werden, mit Kontrollen, die diese Therapie nicht erhalten, aber ansonsten in gleicher Weise behandelt werden. Positive Ergebnisse in diesen Untersuchungen sind im Rahmen der durch den Studienplan vorgegebenen Irrtumswahrscheinlichkeiten schlüssig⁸, negative Ergebnisse hingegen strenggenommen nur dann, wenn der Vergleich mit *unbehandelten* Kontrollen vorgenommen wurde. Sofern es sich allerdings um Vergleiche von Chemotherapie plus Strahlentherapie mit Strahlentherapie allein handelt, sind auch positive Ergebnisse mit Vorbehalt zu versehen, da sie möglicherweise nicht auf unmittelbare zytostatische Effekte des Chemotherapeutikums, sondern nur auf eine Verstärkung der Strahlentherapie hindeuten könnten. Mit anderen Worten, positive Ergebnisse aus solchen Studien können nicht zur Rechtfertigung der Therapie mit dem fraglichen Zytostatikum unter Verzicht auf Radiatio dienen.

Tabelle 2 Formen von Untersuchungen, die Evidenz zur Frage der lebensverlängernden Wirkung zytostatischer Therapie liefern können.

Direkte Evidenz

1. Randomisierte Vergleiche von Patienten, die die fragliche Chemotherapie erhalten, mit solchen, die sie nicht erhalten, aber ansonsten in gleicher Weise behandelt werden (speziell: Vergleiche mit unbehandelten Kontrollen)
2. Randomisierte Vergleiche »Sofortige Chemotherapie versus verzögerte Chemotherapie«
3. Randomisierte Vergleiche unterschiedlich aggressiver Regimes (»Dosis-Wirkungs-Beziehungen«)

Indirekte Evidenz

1. Randomisierte Vergleiche unterschiedlicher Regimes
 2. Nichtrandomisierte Vergleiche unterschiedlicher Therapiekohorten, z. B.
 - Inter-Hospital-Vergleiche
 - Vergleiche aufeinanderfolgender Kohorten desselben Zentrums
 - Vergleiche mit Literaturkontrollen
-

⁸ Daß sie »schlüssig« sind, bedeutet nicht, daß sie frühere Paradigmen umstoßen können. Wie Edler (1990) richtig bemerkt und mit Beispielen belegt, entscheidet in der Praxis niemals eine einzige Studie über die Wirksamkeit oder Unwirksamkeit einer Therapie. »The process of scientific discovery is continuous, and conclusions always remain tentative.« (Greer, 1987)

2. Randomisierte Vergleiche des Typs »sofortige Chemotherapie versus verzögerte (z. B. bei starken Symptomen verabreichte) Chemotherapie«. Wenn sich in einer derartigen Studie die sofortige Chemotherapie als überlegen erweist, so wird man ihr einen positiven Effekt nicht absprechen können, wohingegen negative Ergebnisse weniger Gewicht besitzen: Sie zeigen vielmehr lediglich, daß es beim Einsatz der Chemotherapie nicht auf den frühen Zeitpunkt ankommt.

3. Vergleiche von Zytostatika-Regimes, die sich nur durch Weglassen zytostatischer Agenzien im einen Arm unterscheiden, insbesondere »Dosis-Wirkungs«-Studien. Solche Vergleiche unterschiedlich aggressiver Regimes können i. a. nur dann etwas über positive Effekte der Therapie auf die Lebenserwartung aussagen, wenn sie ein positives Resultat zeitigen, d. h. wenn die aggressivere Therapie im randomisierten Vergleich mit einer längeren Überlebenszeit einhergeht. Aus negativen Resultaten wird man nur dann auf die Wirkungslosigkeit der Therapie schließen können, wenn die niedrigste Dosis einer unbehandelten Kontrolle entspricht.

Die indirekte Evidenz kann aus zwei verschiedenen Untersuchungsformen stammen:

1. Randomisierte Vergleiche unterschiedlicher Therapieschemata (die nicht des oben angeführten Typs 3 sind).

Eine Überlebenszeit-verlängernde Wirkung eines Chemotherapie-Regimes würde man vermuten, wenn sich in einem solchen Vergleich ein bedeutender Unterschied zu seinen Gunsten herausstellt. Hierbei ist allerdings Vorsicht am Platze: Wegen der großen Zahl unterschiedlicher Zytostatikaschemata und der großen Zahl der durchgeführten Untersuchungen ist es nicht ausreichend, daß sich lediglich ein erkennbarer oder (schwach) signifikanter Unterschied ergibt (Multiplizität der statistischen Prüfungen!). Vielmehr muß gewährleistet sein, daß es sich um einen hochsignifikanten und möglichst einen reproduzierbaren Unterschied handelt. Von Bedeutung ist die Reproduzierbarkeitsforderung nicht zuletzt auch wegen der Gefahr eines Publikationsbias, der dadurch zustande kommt, daß positive Ergebnisse (Studien mit auffälligen, statistisch signifikanten Resultaten) eher veröffentlicht werden als sogenannte Nullergebnisse (s. u.). Die aus Vergleichen verschiedener Regimes stammende Evidenz ist deshalb nicht direkt, weil man nicht ausschließen kann, daß das schlechtere Abschneiden der unterlegenen Therapie nur darauf zurückzuführen ist, daß sie das Überleben im Vergleich zu unbehandelten Kontrollen verkürzt hat. In diesem Zusammenhang ist zu bedenken, daß Zytostatika bei Patienten, deren Tumor unter der Therapie weiter wächst, allenfalls Schaden anrichtet; zu beachten ist auch, daß es gelegentlich sogar zu therapiebedingten Todesfällen kommt, ein Problem, das namentlich bei Hochdosistherapien bedeutsam ist. Der Einwand entfällt übrigens bei den Dosis-Wirkungs-Studien, denn in ihnen ist der mögliche therapiebedingte Schaden in der Niedrigdosisgruppe niemals größer als in der Hochdosisgruppe. Als hinreichend für die Reproduzierbarkeit wird man eine Überlegenheit in der Form akzeptieren, daß sich der Überlebenszeitvorteil für die Therapie in unterschiedlichen randomisierten Studien im Vergleich zu verschiedenen Standardschemata ergibt.

Nullergebnisse aus Vergleichen unterschiedlicher Chemotherapieschemata lassen im Einzelfall keine sicheren Schlüsse zu. Allerdings besitzen sie in ihrer Gesamtheit ein beträchtliches Gewicht: Wenn nämlich zumindest einige der Chemotherapie-Regimes imstande wären, das Leben der Patienten zu verlängern, so wäre es geradezu ein ungeheurer Zufall, wenn sämtliche, ganz unterschiedlichen Schemata in paarweisen Vergleichen *identische* Effekte oder nichtsignifikante Unterschiede hervorbrächten. Dabei ist zu beachten, daß man seit den achtziger Jahren darauf vertrauen kann, daß die Studienprotokolle eine Fallzahlplanung enthalten. Diese Planung stellt sicher, daß wahre, relevante Unterschiede zwischen den Therapiegruppen in der Studie auch mit einer gewissen Mindestwahrscheinlichkeit (meist 80% oder mehr) entdeckt würden.

Dieses Argument wird auch nicht durch den Hinweis (s. Ciampi u. Till, 1980) entkräftet, daß a priori Therapieunterschiede in randomisierten Studien schon deshalb selten zu erwarten seien, weil sich Ärzte nur dann bereit finden, solche Studien durchzuführen, wenn nach ihrer doch meist zutreffenden Voreinschätzung die zu bewertenden Therapien gleichwertig sind. Dieser Einwand kann nur in Ausnahmefällen den Mangel an sichtbaren Erfolgen in randomisierten Studien erklären (erst recht ist er kein Beleg für die Existenz bisher unbeobachteter Effekte), denn tatsächlich vorhandene Unterschiede müßten auch und gerade bei umsichtiger Studienplanung noch erkennbar bleiben. Ganz abgesehen davon kommen fundierte Bedenken wegen einer mutmaßlich unterschiedlichen Wirksamkeit der Prüftherapien überhaupt nur dann ins Spiel, wenn vor der randomisierten Studie bereits auffällige und vielversprechende Ergebnisse aus Phase-II-Studien vorliegen.

2. Nichtrandomisierte Vergleiche verschiedener Therapiekohorten.

Wie wir gesehen haben, sind solche Vergleiche mit Vorsicht zu interpretieren. Damit sie aussagekräftig sind, dürfen sich die Kohorten nicht in der Diagnostik oder im Staging unterscheiden, eine Forderung, die vermutlich allenfalls bei historischen Vergleichen aufeinanderfolgender, im selben Zentrum behandelter Kohorten von Patienten erfüllt ist. Aber selbst dann werden nur wirklich bedeutende Unterschiede in der Überlebenszeit der Patienten überzeugen können. Einen Spezialfall bilden Analysen der säkularen Trends in publizierten Überlebensraten. Dies ist die schwächste Form der indirekten Evidenz. Von Ausnahmen mit dramatischen Fortschritten (vgl. die oben angeführten Beispiele nichtepithelialer Malignome) abgesehen, können derartige Untersuchungen niemals positive, sondern allenfalls negative Evidenz liefern. Negative Befunde, die also das Fehlen eines erkennbaren Fortschritts anzeigen, können hingegen wegen der erwähnten Tendenz solcher Analysen, künstliche Erfolge vorzuspiegeln, einen beträchtlichen Stellenwert haben. Auch dies gilt freilich nur, wenn sie durch Untersuchungen an wohldefinierten Patientengruppen (hier: Patienten mit fortgeschrittenen Malignomen) gewonnen werden⁹. Analy-

⁹ Ganz unbrauchbar sind freilich auch die üblichen Gesamtüberlebenskurven (für alle Patienten ab Diagnosestellung) nicht, denn es gibt die Möglichkeit, aus ihnen unter schwachen mathematischen Annahmen die Überlebenskurven für inkurable Patienten herauszurechnen. Diese Konstrukte haben sogar den Vorteil, daß für sie die Studienwanderung keine Rolle spielt.

sen, in denen die Sterberaten in Populationen erkrankter *und* gesunder Individuen betrachtet werden (vgl. z. B. die viel diskutierten Untersuchungen von Bailer u. Smith (1986) oder auch die neueren Untersuchungen von Becker et al. (1989) für die Bundesrepublik Deutschland), eignen sich kaum dazu, einen Mangel an therapeutischem Fortschritt zu belegen, da in ihnen die Effekte von Risikofaktoren (die sich in Inzidenzraten ausdrücken), von Früherkennungsmaßnahmen und Therapie (Letalität der Erkrankten) vermengt werden. Hinzu kommt, daß man in solchen Untersuchungen ein eindeutig positives Ergebnis auch dann nicht erhalten wird, wenn (bei unveränderten Inzidenzraten) die Therapie zwar Fortschritte macht, aber wenn sich diese auf Verlängerungen der Überlebenszeit beschränken, ohne daß die Heilungsraten erhöht werden. Dies liegt daran, daß die Betrachtung altersstandardisierter oder altersspezifischer Mortalitätsraten im Vergleich zu Follow-up-Studien an Erkrankten ein zu unsensitives methodisches Instrument ist, um bloße (moderate) Verlängerungen der Lebensdauer bei Erkrankten anzuzeigen.

Einige zusätzliche Bemerkungen zur Interpretation von Dosis-Wirkungs-Studien sind am Platze. Hier ist allergrößte Vorsicht angezeigt, denn Fehlschlüsse sind unbedingt zu vermeiden, weil sie angesichts der Toxizität hochdosierter Chemotherapie schwere Folgen für die Patienten haben können.

Zunächst ist festzustellen, daß eine Beziehung zwischen der Dosis eines Zytostatikums und der Prognose (Überlebenszeit) höchstens in randomisierten Studien überzeugend nachgewiesen werden kann. In einer ganzen Reihe von Studien (z. B. Levin u. Hryniuk, 1987; Hryniuk, 1988b; s. auch De Vita et al. 1993, S. 283ff.) wurden Zusammenhänge zwischen der sogenannten »Dosis-Intensität« – definiert als Dosis pro Zeiteinheit und berechnet als Quotient aus Gesamtdosis (pro Quadratmeter Körperoberfläche) und Expositionsdauer – einerseits und der Wirkung (z. B. der Responserate oder der Überlebensdauer) andererseits analysiert. Dabei wurden allerdings entweder die Resultate unterschiedlicher Studien zusammengeworfen, oder aber man versuchte die prognostische Relevanz der Dosis-Intensität in einarmigen (unkontrollierten) Studien zu ermitteln (z. B. Bonomi et al., 1989). Beide Verfahren sind nicht schlüssig und tendieren dazu, artifizielle Zusammenhänge zu produzieren. Solche Untersuchungen sind nämlich mit mindestens drei methodischen Problemen belastet (vgl. z. B. Redmond et al., 1983):

1. Es ist gut möglich, daß nur deshalb eine Dosis-Wirkungs-Beziehung gefunden wird, weil eine intensive Therapie nur Patienten mit vergleichsweise gutem Allgemeinzustand verabreicht wird. Auch auf Patientenseite ist ein Zusammenhang zwischen der Zustimmung zu einer intensiven Therapie und der Prognose a priori durchaus nicht auszuschließen. Dieser Einwand trifft gleichermaßen im Falle der Analyse mehrerer Studien wie im Falle der retrospektiven Auswertung unterschiedlich behandelter Patienten eines einzigen Behandlungszentrums zu.

2. Mitunter geht die höhere Dosierung mit längerer Applikation einher. Dann kann es unter Patienten, die die höhere Dosis erhalten, per definitionem keine frühen Todesfälle geben, so daß die Überlebenszeitanalysen ab Therapiebeginn die höher dosierte Therapie fälschlich favorisieren. Diese Verzerrung entfällt in

randomisierten Studien, wo jeder Patient in dem vorgesehenen Therapiearm berücksichtigt wird.

3. Geht in die Auswertung jeder Patient nur mit der tatsächlich verabreichten Dosis ein oder werden gar Patienten, die die geplante volle Dosis nicht erhalten haben, von der Analyse ausgeschlossen (dies war z. B. in der Analyse von Arriagada et al., 1988, der Fall) so verbleiben in der höheren Dosisgruppe nur besonders robuste Patienten mit günstigem Immunausgangsstatus, die Toxizität der höheren Dosierung wird unterschätzt und Frühodesfälle (ob krankheits- oder therapiebedingt) werden in der Hochdosisgruppe fehlen oder unterrepräsentiert sein.

All dies führt zu einer artifiziellen Dosis-Wirkungs-Beziehung. Um so verständlicher ist es, daß eine aggressive Therapie mitunter mit Ergebnissen von Dosis-Intensitäts-Analysen begründet wird (vgl. z. B. Chokski et al., 1988a), wenn zugleich randomisierte Studien jegliche Evidenz für die Verbesserung der Lebenserwartung oder -qualität unter einer aggressiveren Therapie vermissen lassen (vgl. Henderson et al., 1988). In einer kritischen Analyse der Literatur kommen Dodwell et al. (1990) zu dem auch heute noch gültigen (vgl. Green, 1994) Schluß: »In summary therefore the hypotheses generated by retrospective analyses regarding dose intensity and total dose have yet to be proven«.

Vorsicht ist aber auch bei randomisierten Studien am Platze. Zwar wären sie grundsätzlich geeignet, Dosis-Wirkungs-Beziehungen zu etablieren, doch muß sichergestellt sein, daß die unter Punkt 2 und 3 genannten Verzerrungsquellen ausgeschaltet sind. Dies geschieht dadurch, daß in jedem Therapiearm alle zugewiesenen Patienten, unabhängig von der tatsächlichen Dauer oder Menge der applizierten Therapie, in der Auswertung berücksichtigt werden (zur eminenten Bedeutung dieses intention-to-treat-Prinzips bei der Auswertung klinischer Studien s. Windeler, 1993). Wenn eine der beiden Dosen mit längerer Applikationsdauer einhergeht, so wird man auch bei korrekter Auswertung den Dosis-Effekt nicht vom Therapiedauer-Effekt trennen können. Dies ist besonders kritisch, wenn die niedrige Dosierung über längere Zeit verteilt ist. In diesem Fall kann ein scheinbarer Überlebensvorteil zugunsten des höher dosierten Schemas in Wirklichkeit den Nachteil reflektieren, den eine langanhaltende Chemotherapie mit sich bringt.

Es ist möglich, daß günstige Wirkungen der Chemotherapie nur in gewissen Untergruppen von Patienten bestehen. In einem solchen Falle ist nicht unbedingt zu erwarten, daß in randomisierten Studien ein Vorteil für die Chemotherapie auch in der Gesamtkohorte gefunden werden muß, denn erstens können die Gruppen zu klein oder der Effekt zu gering sein, um sich statistisch bemerkbar zu machen, und zweitens ist es möglich, daß die positiven Wirkungen durch eine negative Wirkung (Toxizität) bei anderen Patienten kompensiert werden. Hinzu kommt, daß Wissenschaftler den Ergebnissen von Untergruppenanalysen – insbesondere dann, wenn sie statistisch auffällig sind – zu Recht mit erheblichem Mißtrauen begegnen. Denn diese Analysen werden in aller Regel post hoc durchgeführt, d. h. sie sind nicht bereits im Studienprotokoll vorgeplant gewesen. Und da die Zahl der möglichen Untergruppenanalysen an vorliegendem Datenmaterial einer Studie praktisch unbegrenzt ist, leuchtet es ein, daß immer in

irgendeiner Untergruppe ein signifikantes, positives Ergebnis für eine Therapie aufgefunden werden kann. Skepsis ist um so wichtiger, je mehr man annehmen muß, daß das Ergebnis der Untergruppenanalyse erwünscht war. Nelson (1984) bemerkt hierzu spöttisch: »With a small digital computer and a fertile imagination an investigator can test several scenarios in a few minutes. The abstract then can read ›nebulomycin improves disease-free survival ($p = 0,05$) in pre-menopausal women with breast cancer who have grey hair and six or seven nodes positive«.

Je intensiver die Suche nach derartigen Untergruppen vorgenommen wird, um so wichtiger ist die Forderung nach der Reproduzierbarkeit des Ergebnisses.

Im übrigen sei angemerkt, daß die Tatsache, daß einige Patienten unter Chemotherapie sehr lange überleben, kein Indiz für die Wirksamkeit der Therapie ist, denn auch bei unbehandelten Patienten gleicher Tumorlokalisation und gleichen Stadiums finden sich große Unterschiede in den Überlebenszeiten (s. z. B. Rosenthal u. Doermer (1989), mit weiteren Nachweisen) und gelegentlich schwer erklärliche Extremwerte (Bloom et al., 1962). Ganz entscheidend für die Prognose von Patienten mit fortgeschrittenen Krebsen ist die Wachstumsgeschwindigkeit, die sehr heterogen ist. Mehrere Autoren (z. B. Pearlman, 1983; von Fournier et al., 1994) kommen zu dem Schluß, daß die Tumorverdopplungszeiten in größeren Patientengruppen approximativ einer log-normalen Verteilung folgen.¹⁰ In der Untersuchung von von Fournier et al. (1994) an Brustkrebspatientinnen ergab sich, daß mehr als 5 Prozent von ihnen Tumorverdopplungszeiten von über 1000 Tagen aufwiesen.

Ausgangspunkt der vorliegenden Recherche waren neuere Übersichtsarbeiten über den Stand der internistischen Onkologie, vor allem die von EORTC herausgegebenen kritischen Würdigungen aller randomisierten Studien für die wichtigsten Krebslokalisationen. Der Versuch, aufbauend auf diesem Material einen möglichst umfassenden Überblick zumindest über die direkte Evidenz zur Frage der lebensverlängernden Wirksamkeit der Chemotherapie zu erhalten, beinhaltete neben Datenbankrecherchen und der Durchsicht jüngster, auf Kongressen vorgetragener Resultate auch eine persönliche Anfrage mittels eines Rundbriefes an über 350 Onkologen und onkologisch tätige Zentren in der Welt. Ziel dieser Anfrage war es, unpublizierte oder dem Verfasser unbekannte Referenzen zu erhalten, aber auch Rechtfertigungen für den Einsatz von Chemotherapie seitens solcher Onkologen kennenzulernen, die nicht über direkte Evidenz für Überlebenszeitverlängerung durch die Therapie verfügen.

Ein Aspekt, den man bei der Beurteilung von Therapien aufgrund publizierten Materials stets im Auge behalten sollte, ist die Asymmetrie der methodischen Fehler. Aus vielerlei Gründen (für eine genauere Analyse und für Beispiele siehe Abel u. Windeler, 1995a) haben diese Fehler eher die Tendenz, in Richtung falsch-positiver als in Richtung falsch-negativer Ergebnisse zu wirken. Die vielleicht wichtigste Fehlerquelle, die auch in randomisierten Studien nicht immer ausgemerzt wird, ist der Ausschluß von »nichtauswertbaren« Patienten

¹⁰ Interessant ist übrigens auch die Feststellung von Pearlman, daß die Anzahl der Tumorverdopplungen bis zum Tod der Patienten stets gleich und unabhängig von der Therapie ist.

von der Analyse der Resultate. Dieser Ausschluß ermöglicht nicht nur eine Manipulation, sondern führt fast unweigerlich zu einer Beschönigung der Ergebnisse.

Hinzu kommt, daß die publizierten Studien nicht immer den wahren Stand der Forschung wiedergeben. Besondere Bedeutung kommt der Auswahlverzerrung durch den oben erwähnten Publikationsbias zu. Die Existenz dieses Phänomens wurde in mehreren Untersuchungen belegt (z. B. Dickersin et al., 1987; Berlin et al., 1989; Easterbrook et al., 1991; Boissel u. Haugh, 1993). Danach macht sich der Selektionseffekt bei Beobachtungsstudien stärker bemerkbar als bei randomisierten Studien und bei kleineren Studien stärker als bei größeren.

Für den Publikations-Bias gibt es mehrere Ursachen. Studienleiter und ihre Geldgeber sind möglicherweise weniger daran interessiert, Studien mit Nullergebnissen aufzuschreiben und einzureichen, sei es aus Desinteresse, aus Enttäuschung darüber, daß die Resultate die ursprünglich gehegten Hoffnungen nicht erfüllt haben, oder aus dem Kalkül heraus, daß die Publikation der Kommerzialisierung oder Anwendung der Therapie schaden könnte. Nach Dickersin et al. scheinen dies die Hauptgründe für Publikations-Bias in klinischen Studien zu sein. Denkbar ist auch, daß Studienleiter, die fest von der Wirksamkeit bzw. Überlegenheit einer neuen, in der Studie geprüften Therapie überzeugt sind, das Nullergebnis nicht wahrhaben wollen und auf ein positives Resultat einer neuen Studie »warten«. Dieses selektive Einreichen von Arbeiten dürfte fast ausnahmslos in Richtung falsch-positiver Einschätzungen über die geprüfte Therapie wirken.

Herausgeber und Reviewer wählen ihrerseits aus den Einreichungen eher die Studien mit auffälligen oder gar spektakulären Resultaten aus. Diese Selektion entspringt nicht nur dem – legitimen – Bedürfnis, dem Leser etwas Spannendes zu bieten und die Auflage der Zeitschrift zu fördern, sondern sie ist auch rational begründbar, und zwar durch die Asymmetrie der Relevanz positiver und negativer Resultate in der Medizin. Auffällige Ergebnisse im Sinne einer positiven Evidenz für die Wirksamkeit einer Therapie sind in der Regel nämlich per se, indem sie einen Therapiefortschritt implizieren, für den Patienten relevant, wohingegen die Bedeutung eines (wahren) Nullergebnisses an der jeweiligen Fragestellung zu bemessen ist. Hier ist aber, da das Feld der irrelevanten Fragestellungen unendlich groß ist, auf Editorensseite eine strenge Auswahl zu treffen.