

**Stellungnahme zur Arbeit von L. Edler:
Chemotherapie mit komplementärer Misteltherapie
in:
Arzneimittel-,Therapie-Kritik 2003, 895-903**

zur Publikation: **„Chemotherapie mit komplementärer Misteltherapie.
Wie evident ist ihre Wirksamkeit wirklich?“**
Von Dr. Lutz Edler, Abt. Biostatistik, DKFZ, Heidelberg,

erschieden gleichzeitig als gleichlautende Publikationen in:

1. Internistische Praxis, 43, (203), 895-904
2. Chirurgische Praxis, 62 (2003), 167 ff.
3. Gynäkologische Praxis, 27 (2003), 765 ff.
4. Tägliche Praxis , 44 (2003), 895 ff.
5. Arzneimittel-, Therapie-Kritik, 35 (2003), 337 ff.

In dem Artikel beurteilt der Autor kritisch einige retrospektive (bzw. retrolektive) epidemiologische Studien zur Misteltherapie und Enzymtherapie bei Tumorerkrankungen und kommt zum Schluss, "dass auf der Grundlage der vorliegenden Studiendaten und einer kritischen Beurteilung der biometrischen Methodik auch die neuen Studien eine Wirksamkeit der Misteltherapie nicht objektiv belegen, die Methodik der epidemiologischen Kohortenstudie die grundsätzlichen Schwachstellen retrospektiver Studien nicht überwindet und in einzelnen Studien Selektions- und Informationsverzerrungen zugunsten der Mistelbehandlung und zuungunsten der Kontrollgruppe nicht ausgeschlossen werden können".

Wenn auch den Ausführungen des Autors zu den 'Schwächen' von epidemiologischen Kohortenstudien (EKS) zuzustimmen ist, so ist die Schlussfolgerung nachdrücklich zurückzuweisen, da sie im Widerspruch zu geltendem Recht steht und in letzter Konsequenz jegliche epidemiologische Forschung in Frage stellt.

Nach der seit 1999 gültigen gesetzlichen Regelung der EU (letzte Novellierung von 2001), ist für lang eingeführte Arzneimittel unter definierten Bedingungen („well established use“) der Nachweis der therapeutischen Wirksamkeit und Unbedenklichkeit unter anderem auch mit epidemiologischen Studien, bevorzugt mit vergleichenden epidemiologischen Kohortenstudien möglich, ohne dass weitere

prospektive klinische Studien oder präklinische Untersuchungen erforderlich wären¹
^{2 3}. Außerdem ist auch nach den, von Edler so häufig zitierten, EBM- Kriterien mit
den optimierten epidemiologischen Kohortenstudien (Evidenzklasse II,
Empfehlungsgrad B) der Nachweis der Wirksamkeit möglich.

Auch nach deutschem Recht stellen Anwendungsbeobachtungen (zu denen die
epidemiologischen Kohortenstudien zu zählen sind) ein zulässiges und gültiges
Erkenntnismaterial zur Nutzen/Risikobewertung von Arzneimitteln dar (wozu auch
der Nachweis der Wirksamkeit gehört). In der Bekanntmachung des BfArM vom 4.
Dezember 1998⁴ wird festgestellt: " Derartige Anwendungsbeobachtungen können,
soweit sie wissenschaftlich geplant und sorgfältig durchgeführt wurden, als anderes
wissenschaftliches Erkenntnismaterial im Sinne von §22 Absatz 3 AMG in die
Nutzen/Risiko-Bewertung bekannter Arzneimittel einbezogen werden. Die
Verordnung nach §26 AMG über die Arzneimittel-Prüfrichtlinien (BAnz. Nr. 96a vom
20. Mai 1995) führt im 5. Abschnitt Nr. 1 aus, daß neue Untersuchungen nicht zu
fordern sind, wenn sich die für die Beurteilung von Wirksamkeit und Unbedenklichkeit
notwendigen Angaben aus dem anderen wissenschaftlichen Erkenntnismaterial,
darunter auch Anwendungsbeobachtungen, entnehmen lassen."

Die ablehnende Schlussfolgerung des Autors ist auch aus seinen Erörterungen nicht
nachzuvollziehen, zumal er auf die 'Schwachstellen' kontrollierter klinischer Studien
nicht eingeht und ihm bei der Beurteilung der epidemiologischen Studien
Ungenauigkeiten und Fehler unterlaufen sind. Im Folgenden soll darauf näher
eingegangen werden.

Nach einem Leitsatz des Bundesverwaltungsgerichts⁵ ist die therapeutische
Wirksamkeit eines Arzneimittels unzureichend begründet, "wenn sich aus dem
vorgelegten Material nach dem jeweils gesicherten Stand der wissenschaftlichen
Erkenntnisse nicht ergibt, daß die Anwendung des Arzneimittels zu einer größeren

¹ EC Commission Directive 1999/83/EC of 8. September 1999. EC Official Journal, L243, 9-11, 1999.

² Richtlinie 2001/83/EG des Europäischen Parlaments und Rates vom 06. November 2001, zur Schaffung eines
Gemeinschaftskodexes für Humanarzneimittel. Amtsblatt der Europäischen Gemeinschaften, L311/67-128,
2001.

³ Keller K., The European Union CD 1999/83/EEC on „well established use“ – application to herbal medicinal
products; Herbal Medicines Working Group, EMEA, London, Internet communication, 2001.

⁴ BAnz. Nr. 229 vom 4. Dezember 1998, Seite 16884

⁵ BVerwG 3. Senat, 14. Oktober 1993, AZ: 3 C 21/91

Zahl an therapeutischen Erfolgen führt als seine Nichtanwendung". Der Vergleich der mit dem zu prüfenden Arzneimittel (Test) erzielten Erfolge mit den bei anderen (oder keiner) Behandlung (Kontrolle) festgestellten Erfolgen und die Extrapolation dieser Beobachtungen auf die Gesamtheit der möglichen Anwendung mit statistischen Verfahren sind somit unerlässlich für jede Wirksamkeitsaussage. Der Vergleich kann aber nur dann als valide angesehen werden, wenn die Patienten der Test- und Kontrollgruppe vergleichbar und repräsentativ für die Gesamtheit der Anwender sind. Unbestritten wird dies am ehesten bei randomisierten, kontrollierten klinische Studien erreicht, die daher mit Recht das Standardverfahren zum Nachweis der Wirksamkeit darstellen. Durch die randomisierte Zuteilung wird zwar sichergestellt, dass die Patienten der Test- und Kontrollgruppe in gleicher Weise als Zufallsauswahl aus der Patientenpopulation angesehen werden können, die der Studie zugrunde liegt (entsprechend der getroffenen Patientenauswahl und Ein- und Ausschlusskriterien). Sie garantiert allerdings nicht, dass die konkret ausgewählten Patienten beider Gruppen sich in allen Ausgangsbedingungen nicht unterscheiden. Wenn man - wie im Artikel von Herrn Edler - sehr viele Ausgangsbedingungen als relevant für die Wirksamkeit annimmt, dann wird man häufig bei randomisierten Studien in einzelnen dieser Bedingungen erhebliche Unterschiede zwischen Prüf- und Kontrollgruppe feststellen. Um einen unverzerrten Vergleich zu erhalten, müssen dann auch die von Herrn Edler kritisierten statistischen Ausgleichsverfahren angewandt werden. Ein weiterer kritischer Punkt bei kontrollierten Studien betrifft die Selektion der Studienpatienten. Diese unterscheiden sich häufig von den späteren Anwendern, z. B., wenn die Studie in Kliniken durchgeführt wird, das Medikament aber auch oder sogar vornehmlich in der Allgemeinpraxis angewendet werden soll. Ein solche Selektion beeinträchtigt die Validität des Therapievergleichs nicht, wenn keine Wechselwirkung zwischen dem Medikamenteneinfluss und dem Einfluss der Selektionsbedingungen vorliegt; d.h. die Unterschiede im Therapieergebnis zwischen Prüf- und Kontrollbehandlung auch bei Anwendern mit anderen, von den Studienpatienten abweichenden Bedingungen ähnlich sind. Dies kann aber nicht a priori vorausgesetzt werden. Auch der Aussagekraft kontrollierter Studien sind somit Grenzen gesetzt.

Bei epidemiologischen Kohortenstudien bemüht man sich, ein möglichst repräsentative Auswahl der Anwender des Testmedikaments und der

Kontrollbehandlungen zu erreichen. In den von uns publizierten und von Herrn Edler zitierten Studien^{6 7} erfolgte die Auswahl in 2 Schritten. Im ersten Schritt wird eine repräsentative Auswahl von Zentren getroffen, in denen die zu vergleichenden Therapien angewendet werden. Im 2. Schritt werden in den Zentren alle Patienten ausgewählt, die in einem vorgegebenen Zeitraum wegen der indizierten Erkrankung (Nachsorge des Mammakarzinoms) behandelt wurden. Die Erfassung aller Patienten und die Vollständigkeit und Richtigkeit der erfassten Daten wurde durch unabhängige Monitore überprüft. Durch die Auswahl aller Patienten innerhalb des Zentrums wird eine zentrumsinterne Selektion vermieden. Bei der Auswahl der Zentren wurden nicht nur Zentren berücksichtigt, bei denen das Prüfmedikament (Enzymtherapie, Misteltherapie) angewandt wurde, sondern es wurde anhand von aktualisierten Adresslisten eine zufällige Auswahl anderer Zentren vorgenommen. Selbstverständlich kann die Studie (wie auch eine kontrollierte Studie) nur bei den Zentren durchgeführt werden, die sich zur Teilnahme bereit erklären. Dadurch kann eine Selektion erfolgen, die aber - wie oben ausgeführt - den Therapievergleich nicht tangiert, wenn keine Wechselwirkung besteht.

Bei einer retrospektiven Datenerfassung aus Krankenakten können nur die Daten, die in den Akten aufgezeichnet sind, erfasst werden. Man kann aber davon ausgehen, dass die für die Studie wesentlichen Angaben zur Therapie (z.B. Strahlentherapie, Chemotherapie, Hormontherapie), zu Zusatzerkrankungen, schweren Nebenwirkungen und Ereignissen (Metastasen, Rezidiv, Tod) vollständig und korrekt aufgezeichnet wurden.

Wie in der Arbeit von Herrn Edler richtig dargestellt wurde, besteht der kritische Punkt bei Kohortenstudien in der möglichen Verzerrung durch Heterogenität von Prüf- und Kontrollgruppe. Die Heterogenität muss ermittelt und nach Möglichkeit ausgeglichen werden. Dass dafür entsprechende statistische Methoden (Regression, Stratifizierung) zu Verfügung stehen, mit denen der "Einfluss prognostischer Faktoren .. korrigiert" werden kann, wird von Herrn Edler selbst festgestellt. Falsch ist aber seine Feststellung, dass diese Verfahren in der Praxis auf "unüberwindliche Probleme" stoßen. Wenn dies so wäre, dann wäre jegliche komparative epidemiologische Forschung praktisch unmöglich und auch viele kontrollierte

⁶ Beuth J, et al., Cancer Chemother Pharmacol (2001) 47 (Suppl): 45-54

⁷ Schneider B, Bock PR, Schweiz Z GanzheitsMed (2002); 13: 400-403

Studien, bei denen Ausgleichsverfahren (z.B. die Cox-Regression für Überlebenszeiten) angewandt werden, wären praktisch bedeutungslos. Selbstverständlich wird man nicht alle Faktoren, die das Therapieergebnis beeinflussen können, erfassen. Aber die meisten dieser Faktoren sind eng miteinander korreliert, so dass durch den Ausgleich mit den erfassbaren Faktoren meist auch der Einfluss von nicht erfassbaren Faktoren ausgeglichen wird. Dies gilt auch für Faktoren wie z.B. Verhaltensmuster und psychische Aspekte, die durch 'Surrogatparameter' wie Alter, Herkunft, Vorerkrankungen, Art der Praxis gut repräsentiert werden können. Außerdem hat Rosenbaum gezeigt, dass viele der statistischen Testverfahren (insbesondere die Verfahren zum Vergleich von Ereignisdauern) in einem weiten Rahmen robust gegen einen 'hidden bias' sind⁸.

Der Propensity Score, der in den Studien zum Ausgleich von Heterogenitäten benutzt wurde, ist selbst kein neues Ausgleichsverfahren, sondern ein Hilfsmittel, das einen Ausgleich von sehr vielen Einflussfaktoren erleichtert, indem es den komplexen Einfluss dieser Faktoren durch eine einzige Kenngröße ausdrückt, die den Einfluss aller in ihm enthaltenen Faktoren vollständig erfasst. Er ermöglicht so, eine große Zahl von dokumentierten Einflussgrößen zum Ausgleich der Heterogenitäten praktisch einzuschließen und den 'hidden bias' zu reduzieren. Wir haben uns in den oben zitierten Arbeiten nicht darauf beschränkt, den Ausgleich nur nach einem Modell durchzuführen, sondern verschiedene Modelle gewählt und die Homogenität zwischen den Behandlungsgruppen in Strata mit ähnlichem Propensity Score überprüft. Der von Herrn Edler kritisierte 'Fallstrick' (S.899) wurde also berücksichtigt und liegt bei den Studien nicht vor. Auch ein Ausgleich mit anderen wichtigen Faktoren zusätzlich zum Propensity Score wurde von uns durchgeführt. Die Feststellung von Herrn Edler zu diesem Punkt ist also falsch.

Herr Edler hat in 3 Schritten mögliche Gründe für die Verzerrung bei epidemiologischen Kohortenstudien aufgeführt. Im Schritte 1 führt er aus, dass keine Patientenaufklärung erfolgt. Dies ist bei retrospektiver Datenerfassung kaum möglich und auch nicht notwendig, da zumindest an forschenden Einrichtungen die Patienten im Behandlungsvertrag der anonymisierten Verwendung ihrer Daten zu Forschungszwecken zugestimmt haben. Auf eine strenge Anonymisierung der Daten

⁸ Rosenbaum PR: Observational Studies. Springer Series in Statistics, Springer-Verlag New York, Inc. 1995

wurde bei den Studien großer Wert gelegt. Selbstverständlich stellen die Patienten, die sich nach Aufklärung zur Studienteilnahme bereit erklären eine Selektion dar und die Patientenpopulation von epidemiologischen Studien unterscheidet sich von der kontrollierter Studien. Genau das ist auch beabsichtigt, wobei man bei epidemiologischen Studien möglichst gut die tatsächliche Anwenderpopulation erreichen will und auch kann. Es ist unverständlich wie nach Meinung von Herrn Edler dadurch eine Verzerrung der Wirksamkeitsaussage zustande kommen soll.

Im 2. Schritt wird die Auswahl der Studienzentren und des Zeitfensters kritisiert. Auf die Auswahl der Zentren wurde bereits oben eingegangen und erklärt, dass eine möglichst repräsentative Auswahl angestrebt wurde. Ein Zeitfenster für die Behandlung muss schon deshalb gewählt werden, weil das Testmittel erst ab einem bestimmten Zeitpunkt auf dem Markt ist und bei der Auswahl von früheren Behandlungsfällen gar nicht gegeben werden konnte. Außerdem ändern sich die Behandlungsbedingungen mit der Zeit, so dass auch kein allzu großen Zeitfenster gewählt werden kann. Da es sich bei den Studien um Nachsorgepatientinnen handelte, kann eine Überlappung von Behandlungsperioden nicht vorkommen, wohl aber unterschiedlich lange Behandlungsdauern. Dies wurde bei der statistischen Analyse berücksichtigt. Auch dieser Einwand im Schritt 2 trifft somit nicht zu.

Auf die Einwände im Schritt 3, nämlich die korrekte und vollständige Dokumentation, wurde bereits eingegangen. Selbstverständlich können nur die Daten für die Studie erfasst werden, die in den Krankenakten dokumentiert sind. Dies sind oft 'Surrogatparameter' für das vollständige Krankheitsgeschehen (z. B. Verordnungsdaten), aus denen sich aber das tatsächliche Geschehen meist gut extrahieren lässt. Es soll noch erwähnt werden, dass in unseren Studien die Kontrollgruppe klar definiert ist. Es sind die Patienten, die keine Testtherapie (Enzyme bzw. Misteltherapie) erhalten haben. Die Testtherapie wurde also nicht als Primärbehandlung sondern als Zusatzbehandlung zu einer Basistherapie angesehen. Aus den Studienergebnissen können daher sehr gut ärztliche Entscheidungen über den Wert diese Therapie in Abhängigkeit von der Basistherapie und den sonstigen Behandlungsbedingungen hergeleitet werden.

Abschließend soll noch einmal betont werden, dass epidemiologische Kohortenstudien weder als Ersatz noch als Alternative zu kontrollierten klinischen Studien anzusehen sind und von uns auch nicht angesehen wurden⁹. Sie stellen vielmehr eigene wissenschaftliche Methoden der pharmakoepidemiologischen Forschung dar, die valide und gültige Aussagen über den Einsatz, die Wirksamkeit und Unbedenklichkeit von auf dem Markt befindlichen Arzneimitteln unter den Bedingungen der medizinischen Praxis gestatten.

Verfasser:

Berthold Schneider

Prof. (em.) Dr. phil. nat., Medizinische Hochschule Hannover,

Institut für Biometrie, Carl-Neuberg-Strasse 1, D-30625 Hannover

⁹ Schneider B; Cancer Chemother Pharmacol (2001) 47 (Suppl.): 35-37