

Maßnahmen zur Verbesserung der Qualität von Anwendungsbeobachtungen

Tankred Wegener, Berthold Schneider und die Arbeitsgruppe
»Klinische Prüfung pflanzlicher Arzneimittel« der Gesellschaft für Phytotherapie

Die nach der Zulassung eines Arzneimittels (oder im Status der fiktiv zugelassenen bzw. registrierten Arzneimittel) fortdauernde und praxisnahe klinische Forschung wird gemäß internationalem Konsensus als Phase-IV-Forschung, im englischen als *post-marketing surveillance* (PMS), bezeichnet. Sie umfasst sowohl die Methoden der klinischen Prüfung der Phase IV als auch die der Anwendungsbeobachtung (AWB). Das englische Synonym einer AWB ist die Bezeichnung *observational cohort study*, evtl. auch *observational study* oder *drug monitoring study* (3, 24, 38).

Einen besonderen Stellenwert haben Anwendungsbeobachtungen (AWBen) für pflanzliche Arzneimittel im Zusammenhang mit Artikel 4 Nr. 8a) ii der EU-Richtlinie 75/318/EEC bzw. der Änderungsrichtlinie 1999/83/EC zum *well-established medicinal use*. Hier wird der Nachweis der allgemeinen medizinischen Verwendung mit bibliographischen Belegen (Publikationen) geführt

und umfasst mit den »postmarketing studies« im eigentlichen Sinne AWBen, epidemiologische und vergleichende epidemiologische Studien. Explizit muss erwähnt werden, dass mit der »rechten« klinischen Prüfung der Phasen II–III und auch IV nicht die reguläre medizinische Anwendung abgebildet wird: Dies ist bedingt durch die Patientenselektion über die Ein- und Ausschlusskriterien und durch die Vorgabe der diagnostisch-therapeutischen Maßnahmen einer klinischen Studie. Hier können nur Konzepte, die den Praxisalltag, das Verordnungsverhalten bzw. die Selbstmedikation mit erfassen, den *well-established medicinal use* dokumentieren. Insofern kommt der sorgfältig geplanten, durchgeführten, ausgewerteten und dokumentierten bzw. berichteten AWB eine entscheidende Funktion für den Nachweis der allgemeinen medizinischen Verwendung zu.

Als die wichtigsten Zielsetzungen der gesamten Phase-IV-Forschung wer-

den folgende Bereiche benannt (1, 5, 9, 18, 37):

- ▶ Gewinnung von neuen Erkenntnissen zur Wirksamkeit: insbesondere unter den Bedingungen der alltäglichen medizinischen Praxis bei regulären ambulanten Patienten, Erweiterung der untersuchten Patientenpopulationen, Erfassung von erwünschten Langzeiteffekten, Therapievergleich zu anderen pharmakologischen oder nichtpharmakologischen Maßnahmen;
- ▶ Gewinnung von neuen Erkenntnissen zur Verträglichkeit: Erfassung seltener Nebenwirkungen, Aufdecken von Wechselwirkungen, Erfassung von unerwünschten Langzeiteffekten;
- ▶ Auffindung/Hinweise von/auf neue/n Indikationen: z. B. die kardioprotektive, thrombozytenaggregationshemmende Wirkung der Acetylsalicylsäure, die blutdrucksenkende Wirkung der Thiaziddiuretika/Beta-blocker/Kalziumantagonisten, die lipidsenkende Wirkung von Artischockenextrakten;
- ▶ Erkenntnisse zur Praktikabilität, Akzeptanz, Compliance der Patienten;
- ▶ Erkenntnisse zum Verordnungs- und Verschreibungsverhalten unter alltäglichen Praxisbedingungen.

Bis auf eine gezielte Prüfung zum Nachweis neuer oder erweiterter Indikationen ist die Beantwortung all dieser Fragestellungen mittels der Methodik der AWB möglich.

AWB in der Kritik

Die wissenschaftliche Bedeutung der AWBen war in der Vergangenheit bei Ärzten und Behörden nur sehr gering und ist es heute häufig noch. Viele Negativbeispiele, vor allem aus dem Bereich des Produktmarketings, hatten die Methodik in Misskredit gebracht. Zu nennen sind vor allem die reinen Vermarktungsstudien zur Platzierung der betreffenden Produkte bei den Ärzten. Hier stand nicht die sorgfältige Erhebung wissenschaftlicher Daten im Vordergrund, sondern die Verbesserung des

Zusammenfassung

Die Methodik der Anwendungsbeobachtung stellt für pflanzliche Arzneimittel eine sinnvolle Ergänzung zu klinischen Prüfungen dar. Bedeutsam ist vor allem die Dokumentation der Wirkung unter alltäglichen Praxisbedingungen sowie die Erfassung der Verträglichkeit. Eine Arbeitsgruppe der Gesellschaft für Phytotherapie hat in Ergänzung zu einer früheren Aktivität erneut Empfehlungen erarbeitet, um die Akzeptanz dieser klinischen Methodik weiter zu erhöhen. Erarbeitet wurden Maßnahmen zur Verbesserung der organisatorischen Qualität, Empfehlungen zur Anwendung akzeptierter Befund- und Symptomenscores und zur Gesamtdarstellung der Ergebnisse. Anwendungsbeobachtungen können so geeignet sein, den *well-established medicinal use* im Sinne der EU Richtlinie 2001/83/EC zu bekräftigen.

Bekanntheitsgrades des Produktes. Für eine wissenschaftliche Bewertung, z. B. im Rahmen von klinischen Gutachten zur Wirksamkeit und Verträglichkeit als »sonstiges« Erkenntnismaterial, sind solche AWBen (sofern überhaupt publiziert) kaum zu verwenden. Hinzu kommt, dass in vielen früheren Publikationen von AWBen wenige Hinweise und Daten zur Planung und Durchführung angegeben sind, sodass eine inhaltliche Aussage oder Bewertung kaum möglich ist.

Zusammen mit offenen, unkontrollierten Studien aus früheren Jahren (von denen viele bei einer nachträglichen Bewertung des gewählten Studiendesigns unter heutigen Gesichtspunkten eher als Anwendungsbeobachtungen einzustufen wären) stellten jedoch bisher AWBen ein wichtiges und gängiges Verfahren zur Gewinnung von Anwendungsdaten für registrierte oder zugelassene Fertigarzneimittel dar. Die gewonnenen Ergebnisse trugen entscheidend zur Akzeptanz der betreffenden Arzneistoffe oder Fertigarzneimittel bei, z. B. in den Stoff- und Drogenmonographien der Aufbereitungskommissionen bis 1994.

Empfehlungen von Fachgesellschaften und dem BfArM

Aufgrund dieser unbefriedigenden Situation hatten sich in der Vergangenheit verschiedene Arbeitsgruppen mit der Methodik der AWB befasst und Empfehlungen zur Durchführung, der Auswertung und Ergebnisdarstellung erarbeitet (17, 21, 22, 31, 37). Eine Arbeitsgruppe der Gesellschaft für Phytotherapie hatte ebenfalls Empfehlungen erarbeitet, hier primär bezogen auf die Belange pflanzlicher Arzneimittel (1, 18).

In Zusammenarbeit mit der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie wurden daraufhin vom Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM) Kriterien zur Planung und Durchführung von AWBen erstellt und als Bekanntmachung im Bundesanzeiger veröffentlicht (5). Hierin wurde erstmalig seitens der Zulassungsbehörde

der methodische Stellenwert von Anwendungsbeobachtungen akzeptiert. Wie in dieser Bekanntmachung formuliert, können lege artis durchgeführte AWBen als anderes wissenschaftliches Erkenntnismaterial gemäß AMG § 22, 3 in die Nutzen-Risiko-Bewertung einbezogen werden. Dies wird in der Richtlinie 199/83/EG der EG-Kommission (Amtsblatt der Europäischen Gemeinschaft, L 243/9 vom 15.9.1999) ausdrücklich bestätigt, in der es heißt:

»Insbesondere ist zu klären, dass »bibliographischer Verweis« auf andere Informationsquellen (beispielsweise Untersuchungen nach dem Inverkehrbringen, epidemiologische Studien, mit ähnlichen Erzeugnissen durchgeführte Prüfungen) und nicht nur Versuche und Prüfungen als gültiger Nachweis für die Sicherheit und Wirksamkeit eines Erzeugnisses dienen können, wenn der Antragsteller hinreichend erläutert und begründet, warum er diese Informationsquellen anführt.«

Die Verwendung von AWBen zum Nachweis der Wirksamkeit und Sicherheit setzt allerdings voraus, dass Planung, Durchführung und Auswertung wissenschaftlich nachvollziehbaren Kriterien entsprechen.

Heutige Bedeutung von AWBen für Phytotherapeutika

Für pflanzliche Arzneimittel stellt die Methodik einer AWB aus verschiedenen Gründen eine sinnvolle Ergänzung zu klinischen Prüfungen gemäß § 40/41 AMG dar.

Zunächst ist festzustellen, dass ein Großteil phytotherapeutischer Fertigarzneimittel zur Therapie vorübergehender Gesundheitsstörungen, bei Befindlichkeitsstörungen mit Krankheitswert und bei Erkrankungen mit einem chronischen Verlauf eingesetzt wird. Dies ist in vielen Fällen auf nur »leicht« ausgeprägte Wirkungen, bei gleichzeitig guter Verträglichkeit und ausgezeichneter Akzeptanz durch die Patienten zurückzuführen.

Beispielhaft seien die für die Betroffenen imponierenden subjektiven

Symptome einer chronischen venösen Insuffizienz, einer Dyspepsie oder rheumatischer Beschwerden erwähnt. Objektive Messverfahren sind hier wegen großer Standardabweichungen und Messfehler ungenau und reflektieren kaum die Ergebnisse bez. der Verbesserung der subjektiven Empfindungen der Patienten. Der hohe Umsatz von oralen und topischen Venentherapeutika, der vielen pflanzlichen Antidyspeptika, der Tonika und Rheumaeinreibungen sind typische Beispiele einer Bevorzugung pflanzlicher Arzneimittel durch die Verbraucher.

Mit AWBen können Daten gewonnen werden, die das Anwenderspektrum und die Wirkeigenschaften der betreffenden Produkte im Praxisalltag bzw. der Routineanwendung exakt aufzeigen. Darüber hinaus können die Indikationsaussagen der früheren Aufbereitungsmonographien der Kommission E gestützt bzw. bestätigt werden. Ebenfalls können mit AWBen leicht Daten zur Abschätzung und Bewertung der Langzeitwirkung und -verträglichkeit gewonnen werden. In all diesen Fällen können AWBen einen wesentlichen Beitrag zum Erkenntnisstand leisten und bieten bei manchen Fragestellungen einen Vorteil gegenüber der klinischen Prüfung.

Kohortenstudien

Eine Bewertung der Wirksamkeit mittels Anwendungsbeobachtungen setzt voraus, dass nicht nur Beobachtungen an Patienten gemacht werden, die mit dem interessierenden Arzneimittel (der Prüftherapie) behandelt wurden, sondern auch an Patienten, die vergleichbare Mittel oder – bei einer Prüftherapie, die zusätzlich zu einer Standardtherapie angewandt wird – nur die Standardtherapie erhalten haben (Kontrolltherapie). Solche Studien, bei denen eine repräsentative Auswahl aus einer definierten Patientenpopulation ohne Einschränkung auf eine bestimmte Behandlung getroffen wird, werden Kohortenstudien genannt. Eine Kohortenstudie wird nach Feinstein (11) als retroektiv bezeichnet, wenn mit der Therapie bereits vor der Studie begonnen wurde, und als prolektiv, wenn die Zuteilung zur Prüf- und

Kontrolltherapie erst nach Studienbeginn erfolgt.

Die Datenerfassung wird retrospektiv genannt, wenn die Daten aus bereits bei Studienbeginn abgeschlossenen Krankenakten erhoben werden, und prospektiv, wenn die Daten nach Studienbeginn nach vorgegebenen Vorschriften erfasst werden. Während bei prolektiver Behandlungszuteilung im Allgemeinen auch die Daten prospektiv erfasst werden, ist bei einer retrolektiven Zuteilung auch eine prospektive Datenerfassung möglich, wobei nur die Behandlungsdaten, die ab Studienbeginn anfallen, erfasst werden (»prospektive Anwendungsbeobachtung mit zurückverlegtem Anfangspunkt«) (5).

Charakteristikum einer AWB

Wesentliches Charakteristikum einer Anwendungsbeobachtung (und somit auch einer Kohortenstudie) ist, dass im Rahmen der Studie auf die bei den einzelnen Patienten zur Anwendung kommenden Behandlungen kein Einfluss genommen wird. Die Behandlungen sind vielmehr vom Arzt und/oder Patienten frei wählbar. Vorlieben des behandelnden Arztes sowie Charakteristika der Behandlungseinrichtung und des Patienten werden im Einzelfall die Entscheidung für eine Behandlung beeinflussen. Man kann daher im Allgemeinen nicht davon ausgehen, dass die Patienten der Prüfgruppe mit denen der Kontrollgruppe in ihren Ausgangsbedingungen und in den zusätzlichen Behandlungsbedingungen vergleichbar sind. Da aber diese Bedingungen auch das Behandlungsergebnis beeinflussen können, ist ein unmittelbarer Vergleich der Ergebnisse zwischen Prüf- und Kontrollgruppe nicht frei von Verzerrungen (*unbiased*) und so für eine valide Wirksamkeitsausgabe nicht geeignet.

Verzerrungsfaktoren erkennen

Es ist eine der wichtigsten Aufgaben bei der Auswertung und Bewertung der Ergebnisse von AWBen, die möglichen Verzerrungsfaktoren zu erkennen und ihren Einfluss auszugleichen. Für diesen Ausgleich stehen zwei äquivalente Verfahren zur Verfügung: Stratifikation und

Regressionsanalyse. Bei der Stratifikation werden alle erfassten Fälle in Untergruppen (Strata) mit ähnlichen Ausgangs- und Behandlungsbedingungen unterteilt. Der Ergebnisvergleich zwischen Prüf- und Kontrollbehandlung erfolgt jeweils innerhalb der homogenen Untergruppen und ist daher unverzerrt. Anschließend werden die Vergleichsergebnisse über die Untergruppen in geeigneter Form zusammengefasst. Bei der Regressionsanalyse wird die Abhängigkeit des Therapieerfolgs von den Ausgangs- und Behandlungsbedingungen mit einer geeigneten Regressionsfunktion erfasst, die aus den Daten der Studie geschätzt wird. Anschließend werden die Ergebnisse in beiden Gruppen auf gleiche Bedingungen (meist mittlere Bedingungen der Studie) umgerechnet (adjustiert) und diese bereinigten Ergebnisse zwischen Prüf- und Kontrollgruppe verglichen.

Da im Allgemeinen sehr viele Bedingungen die Zuteilung und das Therapieergebnis beeinflussen, ist eine unmittelbare Stratifikation oder Regression nach allen diesen Bedingungen schwierig und kann oft nur unzureichend durchgeführt werden. Dieses Problem kann dadurch überwunden werden, dass man den Ausgleich in zwei Schritten durchführt. Im ersten Schritt wird der Einfluss der Bedingungen auf die Zuteilung der Prüf- oder Kontrollbehandlung ermittelt und durch eine geeignete Funktion ausgedrückt. Eine solche Funktion ist der von Rosenbaum und Rubin eingeführte *propensity score* (28). Das ist die Wahrscheinlichkeit, einem Patienten die Prüfbehandlung zu geben, als Funktion seiner individuellen Ausgangs- und Behandlungsbedingungen. Dieser *propensity score* kann aus den Daten der Studie geschätzt werden. Er fasst für jeden Patienten die Vielzahl der Bedingungen, die seine Behandlung aber auch das Behandlungsergebnis beeinflussen können, zu einer einzigen Kennzahl zusammen. Anschließend ist das erzielte Behandlungsergebnis nur noch über diese Kennzahl auszugleichen. Dadurch werden die Stratifikation und Regressionsanalyse der Behandlungsergebnisse wesentlich vereinfacht. Das Verfahren wur-

de inzwischen bei zahlreichen Beobachtungsstudien erfolgreich angewandt.

Traditionelle pflanzliche Arzneimittel

Ergänzend sei auf die zukünftige EU-Richtlinie zu den traditionellen pflanzlichen Arzneimitteln (10) hingewiesen: Neben dem Nachweis und Beleg einer langjährigen praktischen Anwendung der betreffenden Produkte können auch hier adäquat geplante und durchgeführte AWBen einen entscheidenden Vorteil für ein erfolgreiches Zulassungsverfahren bieten: Gemäß Artikel 3, Abs. 3d dieser Richtlinie werden für diese Produkte bibliographische Belege zusammen mit einem Sachverständigengutachten zu Beleg und zur Bewertung der Anwendungssicherheit gefordert.

Im Unterschied zu den Anforderungen der ICH- oder EU-Guidelines zur Good Clinical Practice (GCP) existierte bisher kein vergleichbares Regelwerk für die AWB. Trotz der publizierten Empfehlungen des BfArM und vieler Fachgesellschaften werden nach wie vor AWBen durchgeführt, deren Bedeutung für wissenschaftliche Zwecke eher bescheiden ist. Mitglieder der Arbeitsgruppe »Klinische Prüfung pflanzlicher Arzneimittel« der Gesellschaft für Phytotherapie haben sich daher erneut mit der Qualität von AWBen befasst und weitere Empfehlungen erarbeitet, damit die Aussagekraft und Bedeutung dieser klinischen Forschungsmethode weiter verbessert werden kann.

Vorschläge zur weiteren Verbesserung der Qualität von AWBen

Organisatorische Qualität

Zunächst ist darauf hinzuweisen, dass, um Verwechslungen zu vermeiden, die im Rahmen einer AWB zu verwenden Termini einheitlich verwendet werden sollten. So gelten bei einer AWB die Bezeichnungen Beobachtungsplan und Beobachtungsbogen (im Unterschied zu Prüf- oder Studienplan und Prüfbogen oder CRF = Case Report Form bei der klinischen Prüfung). Das zugelassene oder

registrierte (bzw. fiktiv zugelassene) Fertigarzneimittel, für das die Anwendung dokumentiert wird, sollte als Studienmedikation bezeichnet werden. Da auch die AWB als wissenschaftliche Studie definiert ist, kann der Terminus Studienmedikation wie bei der klinischen Prüfung der Phasen I–IV eingesetzt werden.

In vielen AWBen besteht ein Mangel an Qualität der erhaltenen bzw. erstellten Unterlagen. Oftmals fehlen Informationen zur Art der Rekrutierung der Ärzte und der Verteilung der Beobachtungsbögen, der Anzahl der verteilten/ausgegebenen Beobachtungsbögen, der Rücklaufquote und der Art der Betreuung der Ärzte (inkl. Fachgebiete) während der Laufzeit der AWB. Diese Daten sind sowohl zur Bewertung der Qualität einer AWB als auch für einen Vergleich mit anderen Studien oder AWBen von grundlegender Bedeutung und sollten daher immer mit erfasst werden.

Generell wird zur Behebung dieser und anderer Mängel und zur Sicherung der Qualität empfohlen, Standard-Arbeits-Vorschriften (SOP) zur Planung, Durchführung und Auswertung von AWBen aufzustellen. Sie sollten den kompletten Ablauf einer AWB mit den zu beachtenden Aspekten beschreiben – ohne jedoch den Umfang der SOP für eine klinische Prüfung abzubilden.

Auf die Begleitung und Lenkung einer AWB während der Praxisphase mit einem regelmäßigen Monitoring (telefonisch, brieflich, persönlich) zur Gewährleistung der Qualität wurde wiederholt hingewiesen. Dies wird idealerweise durch einen Monitor, ggf. auch durch einen sorgfältig ausgebildeten und eingewiesenen Außendienst sichergestellt.

Einigen Berichten zu AWBen (internen Abschlussberichten und Publikationen) können nicht alle Details der Planung, Durchführung und Auswertung entnommen werden. Eine Bewertung der methodischen und organisatorischen Qualitäten und damit die Beurteilung, inwiefern die weiter oben genannten Empfehlungen des BfArM bzw. der Fachgesellschaften berücksichtigt wurden, ist in diesen Fällen kaum möglich.

Sowohl in die Abschlussberichte als auch in Publikationen sollten daher ausreichende Detailinformationen aufgenommen werden.

Beobachtungsparameter

Der Methodik der unbeeinflussten ärztlichen Routineanwendung gemäß sind apparative oder andere objektive Messverfahren eher selten im Rahmen von AWBen anwendbar. Andererseits ist die Bedeutung der inhaltlichen Aussagen von AWBen gering, wenn individuell erstellte oder konstruierte Befund- oder Symptomenscores zur Anwendung kommen. Ein Vergleich der Ergebnisse zur Wirksamkeit und Verträglichkeit mit den Ergebnissen aus Studien mit anderen Pharmaka bei der gleichen Indikation, explizit als eines der Ziele von AWBen genannt, ist so kaum möglich. In vielen Fällen erlauben die gewählten »Scores« oder »Skalen« zumeist nur eine Aussage zur »allgemeinen« oder »globalen« Wirksamkeit und Verträglichkeit (von »sehr gut« bis »sehr schlecht«), die für eine eingehende Bewertung und Diskussion der Ergebnisse recht unbefriedigend sind. Gezielte Aussagen zu den therapeutischen Wirkungen, die sich bei typischen Einzelsymptomen unterschiedlich zeigen, sind schwer abzuleiten.

Werden dagegen in AWBen die gleichen methodischen Ansätze wie in klinischen Studien verwendet, wie z.B. bei Johanniskraut-Extrakten der HAMD-Score, ermöglicht sich hierdurch ein Vergleich mit den Ergebnissen anderer Studien (Vergleichs-, Referenz-, Plazebotherapie; auch mit den Erkenntnissen zu chemisch definierten Arzneistoffen).

Lebensqualität als Parameter

Einen bedeutenden Ansatz kann die Erfassung der therapeutischen Ergebnisse, des so genannten »Outcomes«, bei AWBen darstellen. Gesundheit ist nicht nur definiert als die reine Abwesenheit von Krankheiten, sondern auch als das Wohlbefinden auf körperlicher, seelischer und sozialer Ebene. Der Bereich der Lebensqualität bzw. des patientenzentrierten Gesundheitsstatus hat daher

eine zunehmende Bedeutung für jede therapeutische Intervention, besonders bei chronischen Erkrankungen. Die Entwicklung eines qualitativ hochstehenden Gesundheitswesens bedingt eine evidenz- und patientenorientierte und zugleich effiziente und damit kostengünstige Medizin. Dabei genügt es nicht mehr, in Studien lediglich die Wirksamkeit und Effizienz von präventiven, therapeutischen oder rehabilitativen Ansätzen zu untersuchen. In Anlehnung an die Qualitätsforschung ist auch eine kritische Überprüfung der in der Wirksamkeitsforschung gewonnenen Erkenntnisse in der Praxis wichtig. Erfasst wird auf diesem Wege das gesamttherapeutische Ergebnis, das Outcome.

Was sind geeignete Parameter?

Es ist inzwischen unumstritten, dass Patienten durchaus in der Lage sind, ihre Symptome, Alltags- und Funktionseinschränkungen zuverlässig zu beurteilen. Vorausgesetzt, ihnen werden relevante Fragen in einer standardisierten Form gestellt. Auch von ärztlichen Fachgesellschaften und Verbänden wurde in den letzten Jahren gefordert, die Aussagekraft der klinischen Effektivität von Pharmaka vermehrt an den Effekt einer indikationsgemäßen Langzeittherapie, dem Outcome, zu koppeln und weniger an kurzfristig erzielbare Effekten auf der reinen Befundebene. Eine umfangreiche Auflistung findet sich z.B. in den Therapieleitlinien der Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF). Zulassungsbehörden haben solche Leitlinien aufgegriffen und empfehlen eine Berücksichtigung dieser Erkenntnisse für den Nachweis der therapeutischen Wirksamkeit; siehe hierzu z.B. die Guidelines der FDA (14). Verstärkt gilt dies für die Anwendung bei chronischen Krankheitsprozessen, die in den letzten Jahren in den Mittelpunkt der pharmakologischen Forschung gerückt sind und im besonderen ein Anwendungsgebiet der Phytopharmaka darstellen; z.B. in der Rheumatologie und Gastroenterologie. Wichtig ist hier die sorgfältige Auswahl für die jeweilige Fragestellung geeigneter Parameter und, sofern neuent-

wickelt, eine (Pilot-)Testung der Outcome-Instrumente, um deren Validität (Gültigkeit), Reliabilität (Zuverlässigkeit), und Sensitivität (Verlaufsempfindlichkeit) zu prüfen. An vier Beispielen soll dies im Folgenden näher erläutert werden.

Psychiatrische und psychologische Erkrankungen

Recht umfangreich sind mittlerweile die Erfahrungen, die aus dem Bereich der Neuropsychopharmakologie und Pharmakopsychiatrie vorliegen. Hier befassen sich Arbeitsgruppen als »Vorreiter« bereits seit langer Zeit mit der Quantifizierung therapeutischer Resultate mittels geeigneter standardisierter Bewertungs- oder Messskalen. Indikationen, für die validierte, reliable und sensitive Skalen vorliegen, sind z. B.:

- ▶ Depression, erfasst mit der Hamilton Depression Scale (HAMD),
- ▶ Angst, erfasst mit dem Anxiety Status Inventory (ASI) oder dem State-Trait-Anxiety Inventory (STAI),
- ▶ Demenz, erfasst mit der Beurteilungsskala für geriatrische Patienten (BGP) und
- ▶ Schlafstörungen, erfasst mit dem Schlaffragebogen A und B (SF-A/B).

Eine ausgezeichnete Sammlung und umfangreiche Dokumentation der verfügbaren Skalen ist vom Collegium Internationale Psychiatrica Scalarum CIPS herausgegeben worden (8).

Rheumatische Erkrankungen

Vorrangige Therapieziele einer Rheumabehandlung sind die Verbesserung der Lebensqualität der Patienten und die Erleichterung des normalen Alltags. Erreicht werden kann dies durch eine Linderung oder Befreiung von Schmerzen, eine Verbesserung der Bewegungsfunktionen bzw. eine Verhinderung bleibender Funktionseinschränkungen und ein Aufhalten der Erkrankungsprogredienz. Bei der Erfassung des Gesundheitsstatus bzw. der Lebensqualität lassen sich prinzipiell drei verschiedene Verfahren unterscheiden:

- ▶ die Beurteilung durch einen Mediziner,
- ▶ die Durchführung standardisierter Aktivitäten durch den Patienten und deren Beurteilung (z. B. Keitel Index) und
- ▶ die Beantwortung standardisierter Fragebögen seitens der Patienten (33).

Fragebögen zur Erfassung des krankheitsspezifischen Gesundheitsstatus eignen sich vor allem zur Erfassung von klinisch relevanten Veränderungen als Konsequenz der Behandlung. Da solche Fragebögen gewöhnlich für eine definierte Erkrankung bedeutsame Elemente beinhalten, sind sie verlaufsempfindlicher als allgemeinere Fragebögen (wie z. B. die in abgewandter oder vereinfachter Form häufig eingesetzte Clinical-Global-Impression [CGI]-Skala). Krankheitspezifische Fragebögen sind dabei genauso zuverlässig wie traditionelle klinische Messparameter (z. B. gemessene Gehstrecke) oder Laborwerte (30).

Es existieren bereits verschiedene validierte »Standardinstrumente« (Selbstbeurteilungsfragebögen) zur Erfassung des Gesundheitszustandes bei Patienten mit rheumatischen Erkrankungen. Sie alle erfassen Schmerzen und die körperliche Funktionsfähigkeit. So wurden z. B. für die Osteoarthritis der Lequesne-Index (20) und der Western-Ontario-McMaster-Universities-Osteoarthritis-Index (WOMAC) (32), für die rheumatoide Arthritis der Health Assessment Questionnaire (HAQ) (4, 15), die American College of Rheumatology (ACR) Kriterien (12, 13) und die Paulus-Kriterien (27) und für Rückenschmerzen der Arhus-Rückenschmerzindex (22) entwickelt und für die Anwendung empfohlen. Alle diese Selbstbeurteilungsfragebögen wurden und werden bereits in klinischen Studien eingesetzt und sind daher prinzipiell auch für eine Bewertung der Wirksamkeit in AWBen geeignet.

Gastrointestinale Erkrankungen

Die Anwendung bei dyspeptischen Beschwerden ist eine Indikation vieler Phytopharmaka. Zu bemängeln ist, dass

der Terminus einer »antidyspeptischen« Wirkung verschiedene Wirkprofile umfasst, die zunächst bezogen auf die therapeutische Wirkung unter Berücksichtigung pharmakologischer Kenntnisse zu präzisieren sind.

Auf Basis eines international akzeptierten Klassifizierungssystems der Dyspepsie und dem heutigen Verständnis der Pathophysiologie wurden mittlerweile relevante und typische Beschwerdep Profile definiert (7, 16, 25, 26, 35). In jüngerer Zeit wurden daraufhin verschiedene standardisierte Messverfahren entwickelt, wobei die klinisch-diagnostische Forschung und die Entwicklung von Selbstbeurteilungsskalen noch nicht abgeschlossen sind. Einige der vorgeschlagenen Skalen sollten neben dem Einsatz in klinischen Studien auch für den Einsatz im Rahmen von AWBen geeignet sein (6, 19, 23, 29, 34, 36, 39, 40). Ein Mangel besteht noch an aussagekräftigen Studien, die mit einem entsprechenden Design durchgeführt und publiziert worden sind und so einen Vergleich mit den Ergebnissen einer AWB ermöglichen könnten.

Abschließende Bewertung von Wirksamkeit und Verträglichkeit

Die Formen der bisher erfassten bzw. abgefragten abschließenden Bewertung von »allgemeiner« oder »globaler« Wirksamkeit und Verträglichkeit von Patienten und Therapeuten ist sehr heterogen. Dadurch wird diese Bewertung als Vergleich zu anderen AWBen oder auch klinischen Studien unnötig erschwert oder verhindert.

Für eine standardisierte abschließende Bewertung der Wirksamkeit, einer Zustandsänderung (Gesamtzustand des Patienten) und der Verträglichkeit (unerwünschten Wirkungen bzw. Therapierisiken) empfiehlt sich die Anwendung des CGI (Clinical-Global-Impression-Skala). Diese Skala, wenn auch nicht unumstritten (2), wurde mittlerweile in vielen Studien als auch in AWBen eingesetzt. Ein Vergleich mit anderen Studienergebnissen ist damit in vielen Fällen bereits möglich.

Zur praktischen Anwendung von Selbstbeurteilungsskalen

Ergebnisse aus AWBen mit pflanzlichen Arzneimitteln, in denen Selbstbeurteilungsskalen eingesetzt worden sind, könnten bei vergleichbaren Strukturmerkmalen der Kollektive einen Vergleich mit anderen Pharmakotherapien bzw. Therapieverfahren erlauben. Die Bedeutung solchermaßen durchgeführter AWBen würde damit gewinnen.

Es ist allerdings zu beachten, dass der Einsatz von Selbstbeurteilungsfragebögen bzw. Outcome-Messinstrumenten eine ausreichende Erfahrung oder ein Training der Ärzte und auch motivierte Patienten voraussetzt. Vorteilhaft ist die Zusammenarbeit mit Ärzten, die bereits aus anderen klinischen Studien mit solchen Methoden vertraut sind. Sofern dies nicht zutrifft, müssen die Ärzte intensiv mit den Besonderheiten einer solchen Studie vertraut gemacht werden. Dies setzt eine individuelle Schulung bzw. ausführliche Einweisung vor Beginn der Therapiephase voraus.

Ergebnisdarstellung und Publikation

Auf die oftmals fehlenden Informationen zur Qualität in Planung und Durchführung, die eine abschließende Wertung ohne weitere Kenntnisse unmöglich machen, wurde bereits hingewiesen. Im Gegenteil wird oft die Beschreibung der Ergebnisse einer AWB durch eine umfangreiche Darstellung des sonstigen wissenschaftlichen Erkenntnisstandes zur Droge, zum Arzneistoff und zum Produkt begleitet und entspricht so eher einer Übersichtsarbeit. Die Ergebnisse der eigentlichen AWB sind auf diese Weise schwer analysierbar. Sofern als Original- oder Hauptpublikation vorgesehen, sollte sich die Darstellung der Ergebnisse einer AWB primär auf die eigentlichen Studienziele und Ergebnisse beschränken. Umfangreiche Gesamtdarstellungen einer Thematik oder eines Arzneistoffes sollten dann eher in separat verfassten Reviews zusammengestellt werden.

Bei vielen Patienten, für die die Therapie in AWBen mit pflanzlichen Arznei-

mitteln dokumentiert wird, ist die initiale Symptomausprägung eher gering. Zu bemängeln ist aber eine Darstellung der Ergebnisse, die z.B. zugunsten der oftmals beachtlichen Rückbildungsraten die Angabe der tatsächlichen initialen Ausprägung oder den Schweregrad eines Einzelsymptomes vermeidet. Für eine Wertung der therapeutischen Ergebnisse ist die Angabe der initialen Symptomausprägung jedoch von maßgeblicher Bedeutung.

Die Diskussion der Ergebnisse einer AWB erfordert eine sorgfältige Vorgehensweise. Die mit dieser Methodik gewonnenen Daten dürfen nicht überbewertet werden. Kritisch sind hier insbesondere spekulative Aussagen bezüglich neuer Indikationen zu sehen. Im unkontrollierten Vergleich sind »Nachweise« für neue Indikationen nicht möglich, allenfalls Hinweise können aufgedeckt werden.

In der biometrischen Auswertung ist neben einer Beschreibung der Daten vor allem eine sorgfältige Analyse der Bedingungen vorzunehmen, die der Zuteilung der Patienten zur Prüf- oder Kontrolltherapie zugrunde liegen. Der Einfluss dieser Bedingungen auf das Therapieergebnis ist zu ermitteln und vor einem Vergleich der Therapieergebnisse zwischen Prüf- und Kontrollgruppe auszugleichen. Geeignete Verfahren wurden oben kurz skizziert. Es sei hier insbesondere auf die Verwendung des *propensity scores* zur Erfassung der Einflussfaktoren für die Therapiezuweisung und zum Ausgleich der durch diese Faktoren induzierten Verzerrungen der Therapieergebnisse hingewiesen. Eine statistische Bewertung der auf gleiche Ausgangs- und Behandlungsbedingungen adjustierten Unterschiede im Wirksamkeitsparameter zwischen Prüf- und Kontrollgruppe durch Konfidenzintervalle oder Signifikanzwahrscheinlichkeiten ist bei AWBen genau so möglich und notwendig wie bei kontrollierten Studien. Durch die Adjustierung werden die Inhomogenitäten der Behandlungsgruppen ausgeglichen. Die Validität des Vergleichs hängt von der Güte der Adjustierung ab. Diese sollte überprüft werden. Auf die Verwendung Bayesiani-

scher Verfahren, bei denen für das Zusammenwirken der verschiedenen Einflussfaktoren auf das Therapieergebnis ein detailliertes Modell aufgestellt und den im Modell auftretenden Parametern a priori Wahrscheinlichkeiten zugeordnet werden, sei nachdrücklich hingewiesen.

Schlussfolgerungen

Die wissenschaftliche Methodik der AWB ist ein wichtiger Bestandteil der Forschung nach Zulassung eines Arzneimittels. Dies gilt insbesondere für Phytopharmaka, für die die Erfahrung aus dem Praxisalltag deren Bedeutung als eine sichere und wirksame Routine-therapie zutreffend beschreiben kann. Die genannten Vorschläge dienen insgesamt einer weiteren Verbesserung der methodischen Qualität in Planung, Durchführung und Auswertung sowie der Berichterstattung von AWBen. Mit dieser Methodik kann ein wichtiger Beitrag für den therapeutischen Stellenwert von Phytopharmaka geleistet werden – und sicherlich werden auch zukünftig gut geplante, organisierte und ausgewertete AWB zum Erkenntnisgewinn bei pflanzlichen Arzneimitteln beitragen.

Tankred Wegener
Consulting Herbal Medicinal Products
Braunholzstr. 17
33378 Rheda-Wiedenbrück
E-Mail: t.wegener@consulting-hmp.de

Prof. Dr. Berthold Schneider
Institut für Biometrie
Medizinische Hochschule Hannover
Carl-Neuberg-Str. 1
30625 Hannover
E-Mail: schneider.berthold@mh-hannover.de

Literatur

- AG »Anwendungsbeobachtungen« der GPHY: Empfehlungen der Gesellschaft für Phytotherapie (GPHY) zur Planung, Durchführung und Auswertung von Anwendungsbeobachtungen. *Klin Pharmakol akt* 1996; 7: 76–83.
- Beneke M, Rasmus W: »Clinical global impressions« (EDGEU): Some critical comments. *Pharmacopsychiat* 1992; 25: 171–176.

- 3 Benson K, Hartz AJ: A comparison of observational studies and randomized, controlled trials. *N Engl J Med* 2000; 342: 1878–1886.
- 4 Brühlmann P, Stucki G, Michel BA: Evaluation of a German version of the physical dimensions of Health Assessment Questionnaire in patients with rheumatoid arthritis. *J Rheumatol* 1994; 21: 1245–1249.
- 5 Bundesinstitut für Arzneimittel und Medizinprodukte. Bekanntmachung über die Zulassung und Registrierung von Arzneimitteln. Empfehlungen zur Planung und Durchführung von Anwendungsbeobachtungen. BAnz Nr. 229 vom 4.12.1998.
- 6 Chassany O, Marquis P, Scherrer B, et al.: Validation of a specific quality of life questionnaire for functional digestive disorders. *Gut* 1999; 44: 527–533.
- 7 Colin-Jones DG: Management of dyspepsia: Report of a working party. *Lancet* 1988; II: 576–579.
- 8 Collegium Internationale Psychiatriae Scalorum. Internationale Skalen für Psychiatrie, 4. Aufl. Göttingen: Beltz-Test; 1996.
- 9 Denes G, Gorbach T: Arzneimittelforschung nach der Zulassung. *Munch Med Wochenschr.* 1990, 132: 373–375.
- 10 EC Proposal for a Directive of the European Parliament and of the Council amending the Directive 2001/83/EC as regards traditional herbal medicinal products (2002). COM (2002) I final. 2002/0008 (COD).
- 11 Feinstein AR. *Clinical Epidemiology. The Architecture of Clinical Research.* Philadelphia: W.B. Saunders Company; 1985.
- 12 Felson DT, Anderson JJ, Boers M, et al.: American College of Rheumatology preliminary definition of improvement in rheumatoid arthritis. *Arthritis Rheum* 1995; 38: 1–9.
- 13 Felson DT, Anderson JJ, Boers M, et al.: The American College of Rheumatology preliminary core set of disease activity measures for rheumatoid arthritis clinical trials. *Arthritis Rheum* 1993; 36: 729–740.
- 14 Food and Drug Administration, Leitseite der Guidance Documents des Center for Drug Evaluation and Research (CDER): <http://www.fda.gov/cder/guidance/index.htm>.
- 15 Fries JF, Spitz PW, Kraines RG, et al.: Measurement of patient outcome in arthritis. *Arthritis Rheum* 1980; 32: 137–145.
- 16 Hotz J, Enck P, Goebell H, et al.: Konsensusbericht: Reizdarmsyndrom – Definition, Diagnosesicherung, Pathophysiologie und Therapiemöglichkeiten. *Z Gastroenterol* 1999; 37: 685–700.
- 17 Hönig R, Eberhardt R, Kori-Lindner C, et al. (Hrsg.): Anwendungsbeobachtung. Qualitätsstandards, praktische Durchführung, Beitrag zur Arzneimittelsicherheit und Nachzulassung. Berlin: E. Habrich, 1998.
- 18 Kraft K, Loew D, Schneider B, Kemper FH: Planung, Durchführung und Auswertung von Anwendungsbeobachtungen. Empfehlungen der Gesellschaft für Phytotherapie (GPHY). *Pharm Ind* 1997; 59: 755–759.
- 19 Leidy NK, Farup C, Rentz AM, et al.: Patient-based assessment in Dyspepsia. Development and Validation of Dyspepsia Symptom Severity Index (DSSI). *Digestive Dis Sci* 2000; 45: 1172–1179.
- 20 Lequesne MG, Mery C, Samson M, et al.: Indexes of severity for OA of the hip and knee. *Scand J Rheumatol* 1987; 65 (Suppl.): 85–89.
- 21 Linden M, Baier D, Beiting H, et al.: Leitlinien zur Durchführung von Anwendungsbeobachtungen (AWB) in der Psychopharmakotherapie. *Nervenarzt* 1994; 65: 638–644.
- 22 Manniche C, Asmussen K, Lauritsen B, et al.: Low back rating scale: Validation of a tool for assessment of low back pain. *Pain* 1994; 57: 317–326.
- 23 Mearin F, Balboa A, Zarate N, et al.: Placebo in functional dyspepsia: Symptomatic, gastrointestinal motor, and gastric somatosensory responses. *Am J Gastroenterol* 1999; 94: 116–125.
- 24 Medicines Control Agency, Committee on Safety of Medicines, Royal College of General Practitioners, British Medical Association, Association of British Pharmaceutical Industry. Guidelines for company-sponsored safety assessment of marketed medicines (SAMM). *Pharmacoepidemiology & Drug Safety* 1994; 3: 1–6.
- 25 Müller-Lissner S, Koelz HR: Dyspepsie: Definition, Ursachen und Vorgehen. *Dt Ärztebl* 1992; 89: B1420–B1425.
- 26 Müller-Lissner SA, Klausner AG: Funktionelle abdominale Beschwerden. Funktionelle Dyspepsie und irritables Kolon. *Internist* 1999; 40: 543–554.
- 27 Paulus HE, Egger MJ, Ward JR, et al.: Analysis of improvement in individual rheumatoid arthritis patients treated with disease-modifying antirheumatic drugs based on findings in patients treated with placebo. *Arthritis Rheum* 1990; 33: 477–484.
- 28 Rosenbaum PR, Rubin DB: The central role of propensity score in observational studies for causal effects. *Biometrika* 1983; 70: 41–55.
- 29 Sandha GS, Hunt RH, Veldhuyzen van Zanten SJ, et al.: A systematic overview of the use of diary cards, quality-of-life questionnaires, and psychometric tests in treatment of *Helicobacter pylori*-positive and negative non-ulcer dyspepsia. *Scand J Gastroenterol* 1999; 34: 244–249.
- 30 Sangha O, Stucki G: Patienten-zentrierte Evaluation der Krankheitsauswirkungen bei muskuloskelettalen Erkrankungen: Übersicht über die wichtigsten Outcome-Instrumente. *Z. Rheumatol.* 1997; 56: 322–333.
- 31 Selbmann HK: Checkliste für Publikationen von Anwendungsbeobachtungen. *Fortschr. Med.* 1996, 114: 123–124.
- 32 Stucki G, Meier D, Stucki S, et al.: Evaluation einer deutschen Version des WOMAC (Western-Ontario-and-McMaster Universities)-Arthrose-Index. *Z. Rheumatol.* 1996; 55: 40–49.
- 33 Stucki G: Outcome-Messinstrumente zur Beurteilung der klinischen Ergebnisqualität. *Akt. Rheumatol.* 1997; 22 (Suppl. 1): S3–S5.
- 34 Talley NJ, Haque M, Wyeth JW, et al.: Development of a new dyspepsia impact scale: The Nepean Dyspepsia Index. *Aliment Pharmacol Ther* 1999, 13: 225–235.
- 35 Talley NJ, Stanghellini V, Heading RC, et al.: Functional gastroduodenal disorders. *Gut* 1999; 45 (Suppl. 1): II37–II42.
- 36 Talley NJ, Verlinden M, Jones M: Validity of a new quality of life scale for functional dyspepsia: A United States multicenter trial of the Nepean Dyspepsia Index. *Am J Gastroenterol* 1999; 94: 2390–2397.
- 37 Victor N, Windeler J, Hasford J, et al.: Empfehlungen der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie GMDS. Empfehlungen zur Durchführung von Anwendungsbeobachtungen. *Inf Biom Epidem Med Biol* 1997; 28: 247–252.
- 38 Waller PC, Wood SM, Langman MJS, et al.: Review of company postmarketing surveillance studies. *BMJ* 1992; 304: 1470–1472.
- 39 Wiklund IK, Junghard O, Grace E, et al.: Quality of life in reflux and dyspepsia patients. Psychometric documentation of a new disease-specific questionnaire (QOLRAD). *Eur J Surg* 1998; Suppl. 583: 41–49.
- 40 Yacavone RF, Locke GR 3rd, Provenzale DT, et al.: Quality of life measurement in gastroenterology. What is available? *Am J Gastroenterol* 2001; 96: 285–297.