Johanniskraut: Versorgungsforschung folgt Doppelblindstudien

Reinhard P. T. Rychlik

Klinische Forschung und Versorgungsforschung

Evidenzbasierte Medizin (EbM) ist der gewissenhafte, ausdrückliche und vernünftige Gebrauch der gegenwärtig besten externen, wissenschaftlichen Evidenz für Entscheidungen in der medizinischen Versorgung individueller Patienten. Die Praxis der evidenzbasierten Medizin bedeutet die Integration individueller klinischer Expertise mit der bestmöglichen externen Evidenz aus systematischer Forschung (1).

Auf dieser Grundlage werden indikationsbezogene Leitlinien generiert, die in der individuellen Behandlung angemessene Entscheidungshilfen für den Arzt und den Patienten anbieten.

Das gesamte verfügbare Wissen aus Studien jeder Art wird themenorientiert und nach Evidenzklassen kategorisiert in systematischen Reviews und Metaanalysen zusammengefasst.

Die EbM bezieht sich also auf die Anwendung der besten verfügbaren Evidenz zur Beantwortung einer Forschungsfrage, die Informationen im Rahmen der Patientenversorgung liefern kann. Hier werden sowohl Informationen aus der klinischen Forschung als auch aus der Versorgungsforschung benötigt.

Therapeutische Wirksamkeit (efficacy) erfordert in der evidenzbasierten Medizin die klinische Studie (RCT) mit dem Anforderungsprofil:

- klare Definition des Hauptzielparameters
- strikte Auswahlkriterien des Patientenguts und der zugrunde gelegten Methoden
- Randomisierung
- Kontrollgruppen
- homogene Stichproben
- Verblindung
- Good Clinical Practice
- festgelegte Prüfzeit und -zeremonien.

RCTs stellen jedoch nur die eine Seite der evidenzbasierten Medizin dar, sie tragen wenig zur gesundheitspolitischen Entscheidungsfindung bei.

So werden die Termini «efficacy» und «effectiveness» zeitweilig immer noch synonym verwendet und lassen sich zum Teil auch mit RCT abbilden.

Grundsätzlich ist jedoch die klinische Wirksamkeit (efficacy) von der praktizierten Alltagswirksamkeit (effectiveness) zu unterscheiden.

Efficacy ist Gegenstand der biomedizinischen Forschung und vergleicht zwei unterschiedliche Behandlungsalternativen unter optimalen und sorgfältig kontrollierten, experimentellen Bedingungen, während effectiveness die Versorgungsrealität repräsentiert. Diese stellt jedoch oft nur eine Teilmenge der evidenzbasierten Medizin dar. Aus diesem Grund ist die Alltagswirksamkeit besser zur Stützung gesundheitspolitischer Entscheidungen geeignet: Sie bildet Über-, Unter- und Fehlversorgung ab, efficacy nur die wissenschaftliche, gewünschte (evidenzbasierte) Regelversorgung. Darüber hinaus gehört zur Alltagswirksamkeit der Vergleich mit einer Behandlungsalternative; die Behandlung mit Plazebo stellt keine Versorgungsalternative dar.

Die Frage stellt sich deshalb, ob klinische Studien (RCT) in diesem Kontext nicht grundsätzlich überbewertet werden und nicht auch andere Studientypen zum Erkenntnisgewinn verwendet werden können. Concato et al. (2) stellten in diesem Zusammenhang fest, dass Beobachtungsstudien (well designed!) im Sinne von Kohorten- oder Fallkontrollstudien keinesfalls die Grösse der Effekte überschätzen – wie lange Zeit angenommen wurde –,sondern durchaus mit RCT vergleichbar sind.

In der Entwicklung neuer Arzneimittel müssen Wirksamkeit, Sicherheit und Verträglichkeit nachgewiesen werden. Im Rahmen klinischer Prüfungen wird ein Studienprogramm (Abfolge mehrerer Studien nacheinander) mit dem Hinblick auf die spätere Zulassung aufgelegt. Um die Wirksamkeit von Arzneimitteln nachzuweisen, muss innerhalb der klinischen Prüfung ein möglichst störungsfreier Untersuchungsrahmen geschaffen werden. Mit Kontrolle, Randomisierung, strengen Ein- und Ausschlusskriterien wie beispielsweise dem Ausschluss multimorbider oder vorbehandelter Patienten sollen mögliche Störgrössen (Bias) ausgeschlossen werden. Diese Minimierung von Störgrössen wird in der klinischen Studie als interne Validität bezeichnet. RCT sind daher normalerweise durch eine hohe interne Validität geprägt zulasten der externen Validität, das heisst der Verallgemeinerung der Studienergebnisse. Nur unter diesen Idealbedingungen kann die statistische Assoziation zwischen der Anwendung einer bestimmten Therapie und dem Behandlungsergebnis als unverzerrter Kausalzusammenhang interpretiert werden.

Die strengen Einschlusskriterien verhindern, dass die in der RCT dokumentierten Auswirkungen für einen Grossteil einer Patientenpopulation nicht repräsentativ sind. So können RCT seltene oder unregelmässige Nebenwirkungen und/oder Langzeitwirkungen nicht adäquat darstellen. Die Studiendauer einer RCT umfasst möglicherweise interessante oder relevante Outcomes nicht, da diese erst in weiterer

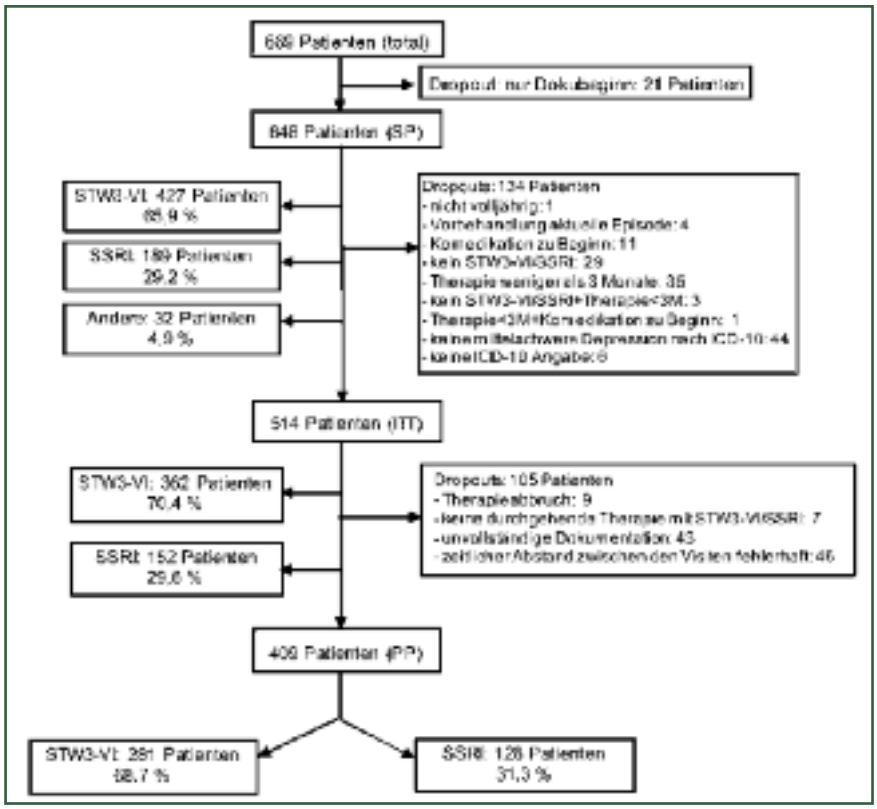

Abbildung 1: Darstellung der Studienpopulation

Zukunft abzulesen sind. Dieses Problem könnten retrospektive Beobachtungsstudien ergänzen (3).

Aus diesen und weiteren Gründen stellt die RCT nicht immer das einzige beziehungsweise das optimale Studiendesign für Fragestellungen zu den Auswirkungen auf die Gesundheitsversorgung dar (4). Selbst bei Fragestellungen nach der Wirksamkeit muss das optimale Studiendesign nicht zwangsläufig die RCT darstellen, sondern ist vielmehr abhängig von der jeweiligen Forschungsfrage und dem zu betrachtenden Outcome.

Die Versorgungsforschung hingegen beschreibt, analysiert und interpretiert die Versorgung der Bevölkerung unter Alltagsbedingungen.

Der Unterschied der Versorgungsforschung zur klinischen Forschung besteht somit darin, dass sie sich mit der Effektivität diagnostischer und therapeutischer Verfahren in der Gesundheitsversorgung befasst sowie strukturelle Voraussetzungen für die Umsetzung neuer medizinischer Entwicklungen unter Alltagsbedingungen beschreibt (5). Um den Nutzen über die klinische Studie hinaus im breiten routinemässigen Anwendungsbereich durch den Arzt und Patienten zu bestimmen, ist die Durchführung von Studien unerlässlich, die die Wirksamkeit im Versorgungsalltag sowie Lebensqualität und weitere patientenrelevante Outcomes wie Anwendbarkeit (ease of use), Compliance und Zufriedenheit untersuchen.

Beispiel Depression

In der Versorgung von Patienten mit Depression bestehen erhebliche Optimierungspotenziale: Häufig liegt im hausärztlichen beziehungsweise allgemeinärztlichen Bereich, wo Patienten am häufigsten behandelt werden, eine schwierige diagnostische Situation vor, zumal die Patienten vorwiegend mit somatischen Symptomen vorsprechen und psychische Symptome oftmals nicht angegeben werden (6). Zudem bestehen seitens der Patienten und ebenso der behandelnden Ärzte Vorbehalte gegenüber medikamentösen oder psychotherapeutischen Therapien, wodurch eine erfolgreiche Behandlung erschwert wird. Patienten erhalten selbst bei korrekter Diagnosestellung häufig vom Arzt keine beziehungsweise eine zu kurze oder zu niedrig dosierte Verordnung von Antidepressiva. Eine medikamentöse Therapie wird von den Patienten aus Angst vor Abhängigkeit oder Nebenwirkungen oftmals abgelehnt, sodass letztlich eine Vielzahl der depressiven Patienten keine oder eine nur unzureichende Behandlung erhält.

Wenngleich randomisierte, klinische Studien die verlässlichsten Ergebnisse hinsichtlich der klinischen Wirksamkeit (Efficacy) von Pharmakotherapien liefern, sind diese aufgrund enger Vorgaben und Einschränkungen im Studienplan nicht uneingeschränkt auf die klinische Versorgung

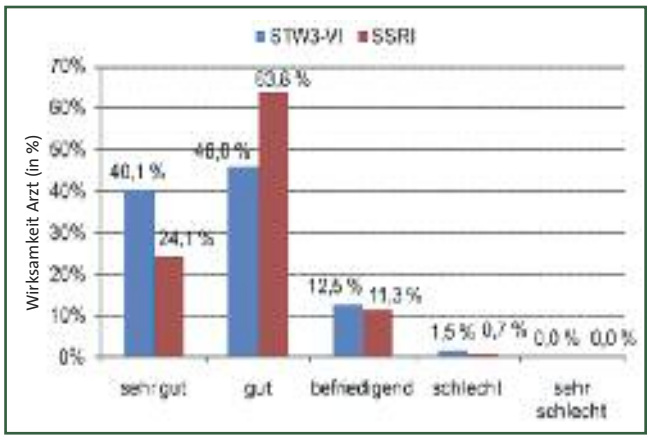

Abbildung 2: Beurteilung der Wirksamkeit durch den behandelnden Arzt, Visite 4 (fünfstufige Skala, in %)

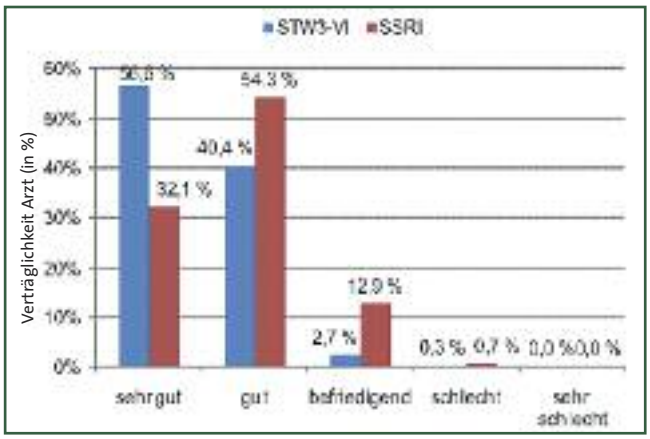

Abbildung 3: Beurteilung der Verträglichkeit durch den behandelnden Arzt zur Visite 4 (fünfstufige Skala, in %)

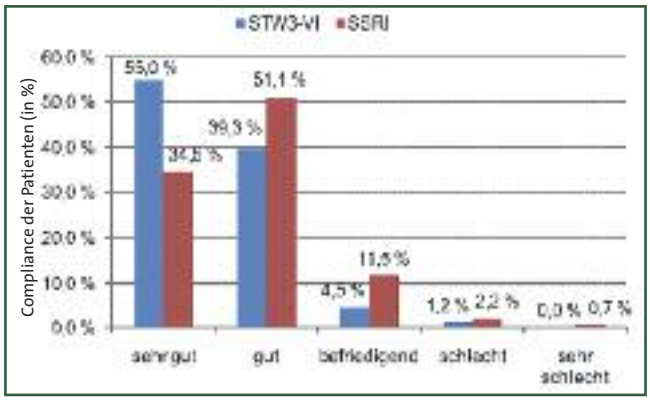

Abbildung 4: Beurteilung der Compliance durch den behandelnden Arzt zur Visite 4 (fünfstufige Skala, in %)

der Patienten im ambulanten Praxisalltag (effectiveness) übertragbar. Studien, die Daten zu Wirksamkeit und Unbedenklichkeit medikamentöser Therapiealternativen bei Patienten mit Depressionen im allgemeinärztlichen Versorgungskontext bei Berücksichtigung eines ausreichend langen Behandlungszeitraumes und damit den Nutzen unter Versorgungsbedingungen untersuchen, liegen zurzeit kaum vor (7). In der vorliegenden Versorgungsforschungsstudie sollten entsprechende Daten unter Berücksichtigung einer Therapie mit dem hoch dosierten Hypericum-Extrakt STW3-VI im Vergleich zu SSRI bei Patienten mit mittelschwerer Depression generiert werden.

Aus einer Metaanalyse 2008 (6) war ersichtlich, dass Johanniskraut mit synthetischen Standardantidepressiva vergleichbar ist.

Ziel

Primäres Ziel der Versorgungsforschungsstudie war die Untersuchung des Nutzens zweier Therapieansätze unter Alltagsbedingungen im niedergelassenen Bereich bei Patienten mit mittelschwerer Depression.

Der Nutzen sollte dabei anhand von Daten zu Alltagswirksamkeit (effectiveness) und Lebensqualität sowie Verträglichkeit der Therapien dargestellt werden. Daneben sollten Daten zu Compliance und Kosten der Patienten bezüglich der Behandlung mit dem gewählten Arzneimittel erhoben werden.

Patientenpopulation

In der Studie wurden Patienten mit einer gemäss ICD-10 mittelschwer eingestuften Depression betrachtet, bei denen entweder eine Phytotherapie mit einem Hypericumextrakt (STW3-VI) oder eine Behandlung mit einem beliebigen Wirkstoff aus der Gruppe der SSRI vorgesehen war (Gruppe STW3-VI beziehungsweise SSRI). Dabei sollte es sich um Patienten handeln, deren aktuelle Episode noch nicht mit einem anderen Antidepressivum antherapiert war und bei denen der jeweilige Wirkstoff als Monotherapie eingesetzt wurde. Weitere Kriterien für die Teilnahme an der Versorgungsforschungsstudie waren Volljährigkeit sowie eine schriftliche Einwilligung nach vorheriger Aufklärung des Patienten.

Tabelle 1:

Direkte und indirekte Kosten

Kostenart	Insgesamt (Euro)	STW3-VI (Euro)	SSRI (Euro)
Behandlungskosten	273,82	272,24	277,55
Arzneimittelkosten	160,56	157,72	167,28
nicht medikamentös	117,63	111,49	132,16
Klinikaufenthalt	33,38	0,00	112,28
direkte Kosten	585,38	541,44	689,27
Erwerbsminderung	114,62	121,52	98,31
Arbeitsunfähigkeit	506,69	427,56	693,77
indirekte Kosten	621,32	549,08	792,09
Gesamtkosten	1206,70	1090,53	1481,36

Design und Durchführung

Die Erhebung wurde als nicht interventionelle prospektive Beobachtungsstudie (Anwendungsbeobachtung gemäss § 67 Absatz 6, Arzneimittelgesetz/Deutschland) bei 77 niedergelassenen Ärzten für Allgemeinmedizin, praktischen Ärzten, Internisten sowie Neurologen/Psychiatern in Deutschland durchgeführt.

Der Beobachtungszeitraum betrug 6 Monate je Patient. Studienrelevante Daten wurden zu Beginn der Studie (Einschluss des Patienten), nach ungefähr 3 bis 4 Wochen, 3 und 6 Monaten sowie zusätzlich zum Ende eines Quartals innerhalb des Beobachtungszeitraumes (aktuelle Therapien) dokumentiert. Neben soziodemografischen Angaben und anamnestischen Daten wurde die Symptomatik der Patienten anhand des Fragebogens Quick Inventory of Depressive Symptomatology (QIDS) erhoben. Der QIDS-Fragebogen dient der Beurteilung des Schweregrads der depressiven Symptomatik. Er wird als Fremdbeurteilungsinstrument durch den Arzt (QIDS-C) und als Selbstbeurteilungsinstrument für den Patienten (QIDS-SR) eingesetzt und widerspiegelt, neben dem medizinischen Ergebnis, auch die krankheitsbezogene Lebensqualität.

Weitere Inhalte der Dokumentation waren die aktuelle medikamentöse und die nicht medikamentöse Therapie des Patienten, die Lebensqualität (SF-36) sowie die Bewertung von Verträglichkeit, Wirksamkeit und Compliance anhand von fünfstufigen Skalen (1 = sehr gut, 2 = gut, 3 = befriedigend, 4 = schlecht, 5 = sehr schlecht) im Verlauf. Die Wirksamkeit im Therapieverlauf wurde auch anhand des QIDS zu Studienbeginn und zu den jeweiligen Beobachtungszeitpunkten sowie anhand des QIDS zu Studienbeginn und zu den jeweiligen Beobachtungszeitpunkten beurteilt.

Insgesamt war der Einschluss von 1000 Patienten geplant. Aufgrund des verzögerten Patienteneinschlusses wurde die Rekrutierung nach 5 Monaten beendet. Bei der Auswertung wurde zwischen der Population für die Verträglichkeitsanalyse und der für die Wirksamkeitsanalyse unterschieden: Die Safety-Population (SP, Verträglichkeitsanalyse) bildeten diejenigen Patienten, von denen neben der Einschlussdokumentation mindestens ein dokumentierter Folgebesuch vorlag. In die Wirksamkeitsanalyse (Intention-To-Treat-[ITT-]Population) wurden alle Patienten einbezogen, die die Zielpopu-

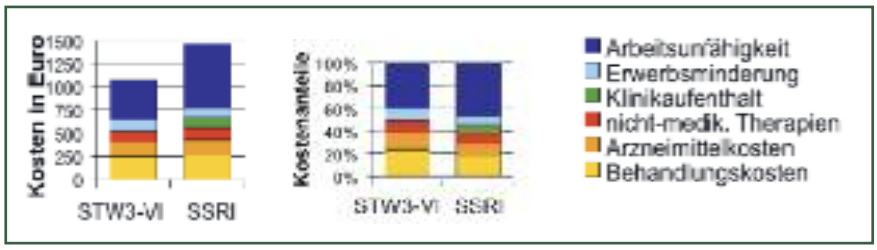

Abbildung 5: Durchschnittliche direkte und indirekte Kosten der STW3-VI/SSRI-Behandlung je Patient

Tabelle 2:

Kosten-Nutzwert-Analyse

Parameter	Insgesamt	STW3-VI	SSRI
Anzahl Patienten	226	154	72
MW Gesamtkosten	904,51 Euro	833,38 Euro	1056,65 Euro
QALY	0,3171	0,3220	0,3068
Kosten/Nutzwert	2852,12 Euro	2588,41 Euro	3444,03 Euro

lationskriterien erfüllten und eine Monotherapie mit Hypericumextrakt beziehungsweise SSRI von mindestens 3 Monaten erhielten.

Ergebnisse

Zum Ende der Feldphase lagen 669 Dokumentationen aus 77 beteiligten Arztpraxen (77,9% Allgemeinmediziner, praktische Ärzte und Internisten und 22,1% Neurologen/Psychiater) vor. Nach Eliminierung der Drop-out-Patienten standen für die Verträglichkeitsanalyse 648 Patienten und für die Auswertung der Wirksamkeit 514 Patienten zur Verfügung. Die Gründe für die Einstufung als Drop-out-Patient sowie die Verteilung der Studienpatienten auf die Therapiegruppen sind in *Abbildung 1* dargestellt.

Beurteilung der Wirksamkeit durch die Ärzte und Patienten

Zu den Visiten 2, 3 und 4 (ca. 3 bis 4 Wochen und 3 Monate nach Einschluss sowie zum Ende des Beobachtungszeitraumes nach ungefähr 6 Monaten) sollten sowohl Ärzte als auch die Patienten die Wirksamkeit von Hypericumextrakt STW3-VI beziehungsweise des eingesetzten SSRI anhand einer fünfstufigen Skala (1 = sehr gut, 2 = gut, 3 = befriedigend, 4 = schlecht, 5 = sehr schlecht) beurteilen. In der Gruppe STW3-VI attestierten die Ärzte zum Beobachtungsende – ebenso wie in der SSRI-Gruppe – bei der grossen Mehrheit der Patienten eine sehr gute oder gute Wirksamkeit (STW3-VI: 86,1%; SSRI: 87,9%; p = 0,580).

Bei der Bewertung der Wirksamkeit durch die Patienten gab es keine wesentlichen Unterschiede zu der ärztlichen Bewertung, wenngleich die Patienten beider Gruppen die Wirksamkeit ihrer Therapie zu allen Beobachtungszeitpunkten ein wenig schlechter bewerteten als die Ärzte (82,6 [STW3-VI] bzw. 83,7% [SSRI] sehr gute oder gute Bewertungen nach 6 Monaten).

Abbildung 2 zeigt die Wirksamkeitsbeurteilung der Ärzte für beide Gruppen im Verlauf des Beobachtungszeitraumes.

Beurteilung der Verträglichkeit der medikamentösen Therapien

6 Monate nach Beginn der medikamentösen Therapie beurteilten die Ärzte die Verträglichkeit der Hypericumtherapie für 97,0 Prozent der Patienten als sehr gut oder gut, während die Verträglichkeit der SSRI-Therapie für 86,4 Prozent der Patienten als sehr gut oder gut dokumentiert wurde ($p < 0,001$). Die Beurteilung durch die Patienten stellte sich im Vergleich zu der ärztlichen Bewertung sehr ähnlich dar: Mit einer sehr guten oder guten Bewertung durch 94,6 Prozent der Patienten der Hypericumgruppe und 82,1 Prozent der mit SSRI behandelten Patienten zeigt sich ebenfalls ein signifikanter Vorteil ($p < 0,001$) zugunsten der pflanzlichen Therapiealternative (*Abbildung 3*).

Die Auswertung der unerwünschten Ereignisse (UE) basiert auf der Safety-Population, die 648 Patienten umfasst (vgl. *Abbildung 1*).

Insgesamt waren im Beobachtungszeitraum 25 Patienten (STW3-VI: 17; SSRI: 8) von 28 unerwünschten Ereignissen betroffen. Damit lag die Inzidenz von unerwünschten Ereignissen in der SSRI-Gruppe geringfügig, jedoch nicht signifikant höher als in der STW3-VI-Gruppe (p = 0,884). Je 5 unerwünschte Ereignisse in beiden Gruppen wurden als schwerwiegend klassifiziert (SUE), wobei für keines der Ereignisse ein Kausalzusammenhang mit dem verabreichten Wirkstoff dokumentiert wurde. Bei 9 der 10 SUE führte ein Krankenhausaufenthalt zu der Bewertung «schwerwiegend». In 1 Fall war das Ereignis lebenbedrohlich (Bronchialkarzinom; kein Kausalzusammenhang zu den in der Studie betrachteten Therapien).

In der Hypericumgruppe traten 10 unerwünschte Ereignisse mit einem zumindest möglichen Kausalzusammenhang auf, betreffend 2,3 Prozent der Patienten. 8 davon waren dem Formenkreis der gelisteten Nebenwirkungen zuzuordnen, betreffend Haut/Unterhautzellgewebe, Gastrointestinaltrakt beziehungsweise Insomnie/Müdigkeit. Hinzu kamen je 1 Fall von Kopfschmerzen und Tinnitus. In der SSRI-Gruppe wurden 5 unerwünschte Ereignisse, betreffend 2,6 Prozent der Patienten, als Nebenwirkungen klassifiziert, darunter Hyperhydrose, vermehrte Depression, gastrointestinale Symptome, Insomnie und Kopfschmerzen.

Compliance der Patienten

Die Ergebnisse der zu allen Visiten vorgenommenen Bewertung der Compliance korrelieren mit den guten Ergebnissen hinsichtlich Wirksamkeit sowie Verträglichkeit der betrachteten Therapiealternativen: So dokumentierten die Ärzte bei der Mehrheit der Patienten am Beobachtungsende eine sehr gute oder gute Compliance (STW3-VI: 94,3%; SSRI: 85,6%; p = 0,002). Bei Betrachtung der Mittelwerte zeigt sich zu allen Beobachtungszeitpunkten ebenfalls eine bessere Compliance bei den mit Hypericumextrakt behandelten Patienten. Die Patienten selbst – hauptsächlich die mit einem SSRI behandelten – stuften ihre Compliance insgesamt besser ein als die Ärzte, jedoch ohne signifikante Unterschiede (sehr gute oder gute Bewertung V2/V3/V4 STW3-VI in Prozent: 89,5/92,6/94,3; SSRI: 85,4/90,3/87,1).

Die prozentuale Bewertung der Compliance seitens der Ärzte ist zum Zeitpunkt der Visite 4 (nach ca. 6 Monaten) in *Abbildung 4* enthalten.

Kosten

Bei der Betrachtung der Kosten konnte sowohl für die direkten als auch die indirekten Kosten Vorteile für Johanniskraut dargestellt werden (*Tabelle 1*), die im Wesentlichen auf eine Reduktion der nicht medikamentösen Therapien und der Klinikaufenthalte zurückzuführen sind.

Abbildung 5 zeigt die erhöhte Arbeitsunfähigkeit unter SSRI sowie notwendige Klinikaufenthalte.

Zur Kosten-Effektivitäts-Analyse konnten die Daten von 422 Patienten herangezogen werden.

Trotz marginal geringer Effekte für STW3-VI zeigte sich eine kostengünstigere Verteilung im Vergleich zu SSRI (125,03 Euro STW3-VI vs. 132,26 Euro SSRI).

Die Therapie mit STW3-VI generierte geringere Kosten, erzielte einen höheren QALY-Score und insgesamt ein besseres Kosten-Nutzwert-Verhältnis (gemessen auf der Basis der QALY und der Gesamtkosten) im Vergleich zur Behandlung mit SSRI.

Zusammenfassend zeigte sich eine vergleichbare Alltagswirksamkeit bei besserer Verträglichkeit, Compliance und Kostengünstigkeit. ◆

Anschrift des Referenten
Prof. Dr. Dr. med. Reinhard P. T. Rychlik
Institut für Empirische Gesundheitsökonomie
Am Ziegelfeld 28
D-51399 Burscheid
info@ifeg.de

Literaturverzeichnis:

1. Vgl. Sackett DL, Rosenberg WMC, Gray JAM et al.: Evidence based medicine: what it is and what it isn't. BMJ 1996; 312: S. 71–72.

2. Concato J, Shah N, Horwitz RI. Randomized, controlles trials, observational studies, and the hierarchy of research designs. N Engl J Med 2000; 342: 1887–92.

3. Black N.: Education and debate – Why we need observational studies to evaluate the effectiveness of health care. BMJ 1996; 312: 1215–1218.

4. Bekkering T, Kliejnen J. Verfahrensweisen und Methoden zur Nutzenbewertung von Arzneimitteln in Deutschland. Gutachten 8.4.2008. URL: http://www.vfa.de/de/presse/studienliste.html/methoden-zur-nutzenbewertung.pdf.

5. Ständige Kongresskommission Deutscher Kongress für Versorgungsforschung. Memorandum zur Versorgungsforschung in Deutschland. Situation-Handlungsbedarf, Strategien. Hamburg 28.3.2003. URL:http://www.gmds.de/pdf/publikationen/memoranden/memo_versorgungsforschung.pdf.

6. Spiessl H, Hübner-Liebermann B, Hajak G: Volkskrankheit Depression. Epidemiologie, Versorgungssituation, Diagnostik, Therapie und Prävention. Dtsch Med Wochenschr 2006; 131:35.40.

7. S3-Leitlinie/Nationale VersorgungsLeitlinie Unipolare Depression, Version 1.1, Dezember 2009.

8. Linde K, Berner MM, Kriston L. St. John's wort for major depression (review). Cochrane Database of Syst Rev 2008, 4: Art. No. CD000448, DOI: 10.1002/14651858. CD000448. Pub3. 2008.

Anmerkung der Redaktion:
Der Hypericumextrakt STW3-VI ist in der Schweiz unter der Marke Deprivita® im Handel.