

Wirksamkeit des Johanniskrautextrakts STW3-VI

Eine Reanalyse gepoolter Daten zweier Placebo-kontrollierter Studien

Hans-Peter Volz, Werneck, J. Müller und C. Kolb, Darmstadt

Um weitere Hinweise für die Wirksamkeit eines Johanniskrautextrakts (STW3-VI) zu erlangen, wurden die Daten zweier randomisierter, Placebo-kontrollierter Studien (gesamte Teilnehmerzahl: 401) unterschiedlichen weiteren Analyseschritten unterzogen: zum einen einer gepoolten Analyse des Verlaufs des Gesamtwerts der Hamilton-Depressions-Skala-17-Items (HAMD-17) sowie der Einzelitems, einer Faktorenanalyse zu Beginn und zu Ende der Studien sowie zu der Differenz (der HAMD-17); des Weiteren wurde der prädiktive Wert der initialen Besserung für die Remission (HAMD-17-Gesamtscore ≤ 8) untersucht. Es zeigte sich, dass der Gesamtwert des HAMD-17 in der STW3-VI-Gruppe deutlich stärker abfiel als in der Placebo-Gruppe. Dieser Unterschied war bereits in der dritten Woche statistisch signifikant und betrug zum Ende der Studie (Woche 6) 4,4 Punkte ($p < 0,01$). In sämtlichen HAMD-17-Einzelitems außer bei depressiver Hemmung, Gewichtsverlust und Krankheitseinsicht zeigte sich ebenfalls ein statistisch signifikanter Unterschied zwischen der Verum- und der Placebo-Gruppe. Die Faktorenanalyse ergab zum Zeitpunkt Tag 0 eine multifaktorielle Lösung, zum Zeitpunkt Woche 4 eine vierfaktorielle Lösung. Die initiale Besserung (bis zur Woche 2 um mindestens 20 % des HAMD-17-Gesamtwerts) besaß einen hohen prädiktiven Wert für die Remission am Ende der Behandlung. Diese Er-

gebnisse werden im Kontext anderer Untersuchungen mit ähnlicher Methodik diskutiert. Die Ergebnisse sind in guter Übereinstimmung zu Studien mit synthetischen Antidepressiva.

Schlüsselwörter: Johanniskrautextrakt, Depression, STW3-VI, Faktorenanalyse, Re-Analyse, Antidepressiva, Phytopharmakon

Psychopharmakotherapie
2018;25:170–6.

Johanniskrautextrakte sind die einzigen Phytopharmaka mit einem eindeutigen antidepressiven Wirksamkeitsnachweis in erster Linie für leichte bis mittelschwere Depressionen [9, 10]. Zahlreiche Autoren, so auch Linde et al. [9, 10] weisen darauf hin, dass es unterschiedliche Extrakte gibt und somit die positiven Ergebnisse nur für die Extrakte, die in den eingeschlossenen Studien verwendet wurden, gelten. Dies zeigt die Problematik von Metaanalysen oder überhaupt vergleichenden Analysen im Bereich der Phytopharmaka auf, nämlich dass aufgrund der Unterschiedlichkeit der einzelnen Anteile, insbesondere der wirksamkeitsbestimmenden Inhaltsstoffe, Wirksamkeitsunterschiede der verschiedenen Extrakte vorhanden sein können. Aus diesem Grunde kann es durchaus sinnvoll sei, die Wirksamkeit einzelner Extrakte individuell zu analysieren.

Mit dem Extrakt STW3-VI wurden mit einer Dosierung von 900 mg pro Tag zwei randomisierte, Placebo-kontrollierte Studien in der Indikation Depression durchgeführt [4, 18]. (Eine weitere randomisierte Placebo-kontrollierte Studie [3] wurde mit 600 mg pro Tag

des sehr ähnlichen Extrakts STW3 mit 50 mg Sertralin als interne Kontrolle über einen Beobachtungszeitraum von 12 Wochen durchgeführt. Auf diese Studie wird aufgrund des unterschiedlichen Designs und der vergleichsweise niedrigen Dosierung des Extrakts nicht weiter eingegangen.) Beide Studien zeigten im Vergleich zu Placebo eine signifikant bessere Wirksamkeit (u. a. im Haupteffektivitätskriterium, dem Gesamtwert der 17-Item-Hamilton Rating Scale for Depression, HAMD-17).

Um diese beiden Studien weiter zu analysieren, wurden verschiedene Ansätze gewählt:

- Gepoolte Analyse für den HAMD-17-Gesamtwert und die Einzelitems
- Faktorenanalyse über die Symptomstruktur (erfasst durch den HAMD-17) zu Beginn, zum Ende und über den Verlauf
- Analyse der initialen Besserung in Bezug auf deren prädiktiven Wert für die Gesamtbesserung

Der erste Ansatz diente dazu, die antidepressive Wirksamkeit des Extrakts STW3-VI besser zu charakterisieren, als dies in Einzelstudien möglich ist, wobei vor allen Dingen untersucht werden sollte, wann im Studienverlauf die Differenzierung zwischen Verum und Placebo einsetzt.

Mithilfe des zweiten Ansatzes sollte überprüft werden, ob die Faktorenlösungen der mit der HAMD-17 erfass-

Prof. Dr. Hans-Peter Volz, Krankenhaus für Psychiatrie, Psychotherapie und Psychosomatische Medizin Schloss Werneck, Balthasar-Neumann-Platz 1, 97740 Werneck, E-Mail: hans-peter.volz@kh-schloss-werneck.de
Dr. Jürgen Müller, Dr. Christiane Kolb, Steigerwald Arzneimittelwerk GmbH, Havelstraße 5, 64295 Darmstadt

ten depressiven Symptomatik der in die verwendeten beiden Studien eingeschlossenen Patientenpopulationen jenen Faktorenlösungen entspricht, die sich bei anderen, beispielsweise mit synthetischen Antidepressiva durchgeführten Studien ergaben. Falls dies der Fall ist, deutete dies auf den Einschluss einer vergleichbaren Patientenpopulation hin.

Im dritten Ansatz sollte überprüft werden, ob auch unter einer Pharmakotherapie mit Johanniskrautextrakt die initiale Besserung den gesamten Therapieerfolg prädiziert.

Methode

In beide Studien wurden ambulante Patienten, die an einer mittelgradigen Depression litten (ICD-10: F32.1/F33.1), eingeschlossen. Der initiale HAMD-17-Gesamtwert musste zwischen 20 und 24 betragen. Die Patienten wurden über sechs Wochen mit 900 mg/Tag STW3-VI bzw. Placebo behandelt. Neben der HAMD-17 als primärem Wirksamkeitsparameter kamen andere Instrumente zum Einsatz, hierauf wird an dieser Stelle nicht eingegangen. In der Studie von Gastpar et al. [4] wurde auch eine dritte Gruppe, die mit 20 mg Citalopram behandelt wurde (aktive Referenz), eingeschlossen, auf diese Gruppe wird nicht weiter eingegangen. Die zwei kombinierten Placebo-kontrollierten Studien umfassen zusammen 398 Patienten. Da die kleinere der beiden Studien (140 Patienten; [18]) keine Daten zum Therapieerfolg nach sieben Tagen enthielt, wurden diese aus den Daten der im Verlauf sehr ähnlichen größeren Studie über eine Regressionsgleichung getrennt für Verum und Placebo geschätzt. Das durchschnittliche Alter war ähnlich ($46,4 \pm 12,5$ [STW3-VI] bzw. $43,3 \pm 12,6$ Jahre [Placebo] in der Studie von Uebelhack et al. und $50,8 \pm 12,1$ [STW3-VI] bzw. $49,4 \pm 12,7$ [Placebo] in der Studie von Gastpar et al. [4]), ebenso wie die initialen HAMD-17-Gesamtwerte ($22,8 \pm 1,1$ bzw. $22,6 \pm 1,2$ und $21,9 \pm 1,2$ bzw. $22,0 \pm 1,2$) wie auch der das Verhältnis männlicher zu weiblichen Patien-

ten ($21/49$ bzw. $25/45$ und $46/85$ bzw. $35/95$).

Folgende Methoden wurden zur weiteren Analyse dieser Daten angewendet:

1. In einer „pooled analysis“ wurden die von Ausgangswerten und Messzeitpunkten sehr gut vergleichbaren Daten der beiden Studien für die weitere Analyse kombiniert. (Dieses Vorgehen ist auch sehr empfehlenswert für die weiter unten beschriebene Faktorenanalyse.) Die Ergebnisse des Therapieerfolgs wurden hinsichtlich des HAMD-17-Gesamtwerts und der 17 Einzelitems mithilfe einer Kovarianzanalyse (ANCOVA) mit den Ausgangswerten als Kovariate untersucht. Die Kovarianzanalyse ist ein statistisches Verfahren, das Varianzanalyse und lineare Regressionsanalyse verbindet und so zu einer Erhöhung der statistischen Aussagekraft führt. Der Verlauf des HAMD-17-Gesamtscores über die Messzeitpunkte wurde mit einer Varianzanalyse mit Messwiederholung studiert.

2. Mithilfe einer Faktorenanalyse wurde die Struktur der Ergebnisse des HAMD-17 näher aufgeklärt. Auch hierzu war es sinnvoll, die Daten der beiden Studien zusammenzuführen, da beim Vorliegen von 17 Items (der HAMD-17) eine größere Patientenzahl zu einem stabileren Ergebnis der Faktorenanalyse führt, als dies bei Faktorenanalysen der Einzelstudien der Fall wäre. Die Faktorenanalyse ist ein gängiges multivariates statistische Verfahren, das dazu dient, aus vielen Variablen (hier die 17 Items des HAMD-17) auf wenige zugrunde liegende latente Variablen („Faktoren“) zu schließen. Somit kann auch von einem dimensionsreduzierenden Verfahren gesprochen werden. Hierzu wurden drei Faktorenanalysen gerechnet und die Ergebnisse miteinander verglichen: Zu Beginn der Therapie (symptomatischer Zustand), am Ende der Therapie (gebesserter Zustand) sowie die Differenzen der beiden Zeitpunkte. Das geschah unter der Vorstellung, dass gerade die Differenzen Hinweise auf die sehr interessanten änderungssensitiven Faktoren geben könnten.

3. Um die Besserung der Depression bereits nach sieben Tagen zur Besserung am Ende der Therapie nach 42 Tagen ins Verhältnis zu setzen, wurde eine in der neueren Literatur (z. B. Lemke et al., [8]) genannte Formel verwendet:

$$PPV = N_x / (N_x + N_y)$$

(PPV: positive predictive value; N: Anzahl Patienten; $N_x = N$ [frühe Besserung + Remission], $N_y = N$ [frühe Besserung, keine Remission])

Man erhält damit einen positiven Vorhersagewert, der angibt, wie groß die Chance auf eine Heilung am Therapieende ist, wenn bereits nach sieben Tagen entsprechende Anfangsverbesserungen eingetreten sind.

Ergebnisse

Gepoolte Analyse

In der gepoolten Analyse zeigte sich in der Placebo-Gruppe ein Abfall des HAMD-17-Gesamtwerts von 22,2 (Standardabweichung [SD] 1,2) auf 15,2 (6,6), in der Verum-Gruppe von 22,2 (1,2) auf 10,8 (5,8), der Unterschied war hochsignifikant ($p < 0,01$; Kovarianzanalyse) und war auch bereits am dritten Messzeitpunkt der doppelblinden Vergleichsperiode (Tag 21) hochsignifikant ($p < 0,01$; Kovarianzanalyse). Ebenfalls hochsignifikante Ergebnisse sind in der Varianzanalyse mit Messwiederholung zu sehen, die eine statistische Betrachtung des Gesamtverlaufs ermöglicht. Dieses Ergebnis ist auch in **Abbildung 1** dargestellt. Beachtenswert ist auch der Endwertunterschied zwischen den Gruppen in Höhe von 4,4 HAMD-17-Punkten, was auf die hohe klinische Relevanz dieses Unterschieds hinweist.

Sämtliche HAMD-17-Einzelitems wurden einer entsprechenden Analyse unterzogen: Mit Ausnahme der Items 8 (depressive Hemmung), 16 (Gewichtsverlust) und 17 (Krankheitseinsicht) zeigte sich jeweils ein signifikanter Wirkvorteil von STW3-VI (**Tab. 1**) im Vergleich zu Placebo. Die drei Schlafitems der HAMD-17 (Items 4 bis 6) können zu einem Summenitem zusam-

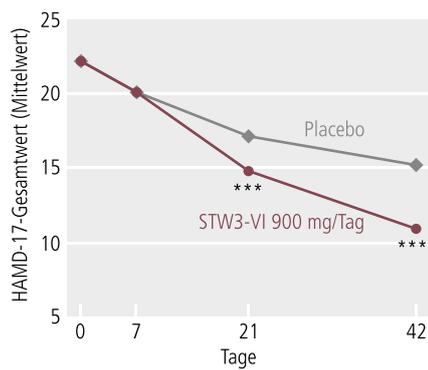


Abb. 1. Verlauf der Mittelwerte des gepoolten HAMD-17-Gesamtwerts über die Behandlungsdauer von 42 Tagen mit Verum oder Placebo. Am Tag 21 und am Tag 42 zeigt sich ein statistisch signifikanter Unterschied zugunsten des Verums ($p < 0,01$; Kovarianzanalyse bzw. Varianzanalyse mit Messwiederholung mit dem Ausgangswert als Kovariate).

mengefasst werden, der Verlauf dieses Summenitems über die Zeit ist in **Abbildung 2** für den Vergleich zu Placebo dargestellt.

Faktorenanalyse

In den **Abbildungen 3** und **4** sowie den **Tabellen 2** und **3** sind die Ergebnisse der Faktorenanalyse, die auf der Basis der HAMD-17-Einzelitem-Werte zum Beginn der Studie (Tag 0), zum Ende der Studie (Tag 42) sowie für die Differenzwerte (Tag 0 bis Tag 42) gerechnet

wurden, zusammenfassend dargestellt. Zur Interpretation wurden die Faktoren mit Eigenwert > 1 bzw. der „Knick“ im Verlauf der Eigenwertkurve herangezogen.

Zum Zeitpunkt Visite 1, Tag 0 (Tab. 2, Abb. 3)

Es ergab sich eine multifaktorielle (Eigenwertkriterium) bzw. 5-faktorielle („Knick“ des Eigenwertverlaufs) Lösung mit einem Hauptfaktor, der nur unwesentlich mehr Varianz aufklärt als die beiden nachfolgenden Faktoren. Festzuhalten bleibt, dass die Schlafitems (Faktor 2), Depression und Erregung (Faktor 5) sowie Suizidalität und Einsicht (Faktor 4) jeweils eigene Faktoren darstellen.

Zum Zeitpunkt Visite 4, Tag 42 (Tab. 3, Abb. 4)

Es ergab sich eine 4-faktorielle (Eigenwertkriterium) bzw. 2-faktorielle („Knick“ des Eigenwertverlaufs) Lösung mit einem Hauptfaktor, der bei weitem die größte Varianzaufklärung zeigt und hauptsächlich aus depressions-, schlaf- und angstbezogenen Items besteht. Aus der **Abbildung 4** ist zu erkennen, dass eine (statistische) Bündelung (angezeigt durch die relativ hohen

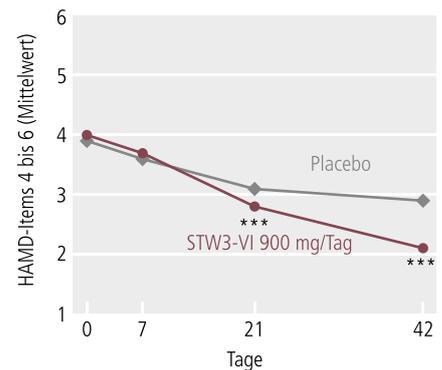


Abb. 2. Verlauf der Mittelwerte des gepoolten HAMD-Items 4 bis 6 (Schlafitems) über die Behandlungsdauer von 42 Tagen mit Verum oder Placebo. Am Tag 21 und am Tag 42 zeigt sich ein statistisch signifikanter Unterschied zugunsten des Verums ($p < 0,01$; Kovarianzanalyse bzw. Varianzanalyse mit Messwiederholung mit dem Ausgangswert als Kovariate). ns: nicht signifikant

Werte des Faktors 1 in Relation zu den Faktoren 2 bis 4) von acht Items auf einen sogenannten Hauptfaktor erfolgt ist, der damit auch die wesentlichen Aspekte der Depression auf sich vereint. Dieser Hauptfaktor zeigt einen hohen Eigenwert, der dann steil abfällt und somit die wesentlich geringere statistische Bedeutung der anderen Komponenten anzeigt.

Faktorenanalyse der Differenzen Visite 1 – Visite 4 (Tab. 4)

Es ergab sich eine 5-faktorielle (Eigenwertkriterium) bzw. 2-faktorielle („Knick“ des Eigenwertverlaufs) Lösung mit einem Hauptfaktor, der einen deutlich größeren Anteil der Varianz aufklärte als die anderen Faktoren. Der Hauptfaktor besteht aus den Items, die sich auf Depression, Arbeit, Angst und allgemeine körperliche Symptome beziehen, der zweite Faktor repräsentiert die Schlafitems. Zu vermuten wäre, dass es sich hier um die Faktoren handelt, die wesentliche Aspekte der Verlaufsänderung (Besserung der Depression) repräsentieren. Diese Vermutung wird bestätigt, wenn man Faktoren als Subskalen betrachtet und den Verlauf berechnet.

Remissionsprädiktion

Hier wurde eine mindestens 20%ige Verbesserung nach einer Woche Behand-

Tab. 1. Prä-Post-Werte der Einzelitems der HAMD-17 einschließlich der entsprechenden Test-Ergebnisse auf Signifikanz (ANCOVA mit Baseline-Werten als Kovariate)

Item	Placebo		STW3-VI		Signifikanz
	Ausgangswert	Endwert	Ausgangswert	Endwert	
1	2,42	1,66	2,47	1,16	<0,01
2	1,49	1	1,58	0,57	<0,01
3	0,42	0,19	0,47	0,13	<0,01
4	1,33	1,02	1,29	0,73	<0,01
5	1,29	0,95	1,41	0,69	<0,01
6	1,27	0,93	1,31	0,62	<0,01
7	1,79	1,15	1,79	0,72	<0,01
8	0,95	0,55	0,93	0,45	n. s.
9	1,54	1,14	1,38	0,91	<0,01
10	2,21	1,4	2,22	1,04	<0,01
11	2	1,4	2,05	0,87	<0,01
12	0,67	0,39	0,64	0,24	<0,01
13	1,37	1,11	1,4	0,76	<0,01
14	0,77	0,52	0,78	0,4	<0,05
15	1,3	1	1,3	0,85	<0,05
16	0,46	0,2	0,35	0,15	n. s.
17	0,92	0,61	0,89	0,64	n. s.

Tab. 2. Ergebnis der Faktorenanalyse (Darstellung der Ladungen) zum Zeitpunkt Visite 1: Es ergab sich eine multifaktorielle (Eigenwertkriterium) bzw. 5-faktorielle („Knick“ des Eigenwertverlaufs) Lösung mit einem Hauptfaktor, der nur unwesentlich mehr Varianz aufklärt als die beiden nachfolgenden Faktoren. Es werden nur Ladungen $\geq 0,4$ oder $\leq -0,4$ dargestellt (siehe auch Abb. 3).

HAMD-17-Items	Faktoren					Bedeutung
	1	2	3	4	5	
1					0,508	Depressive Stimmung (5)
2						
3				0,727		Suizidalität (4)
4		0,509				Einschlafstörungen (2)
5		0,760				Durchschlafstörungen (2)
6		0,721				Schlafstörungen am Morgen (2)
7						
8	-0,563					Hemmung (1)
9					-0,874	Erregung (5)
10	0,696					Angst – psychisch (1)
11	0,493					Angst – somatisch (1)
12			0,785			Som. Symptome, gastrointestinal (3)
13	0,576					Som. Symptome – allgemeine (1)
14						
15	-0,733					Hypochondrie (1)
16			0,746			Gewichtsverlust (2)
17				-0,738		Einsicht (4)

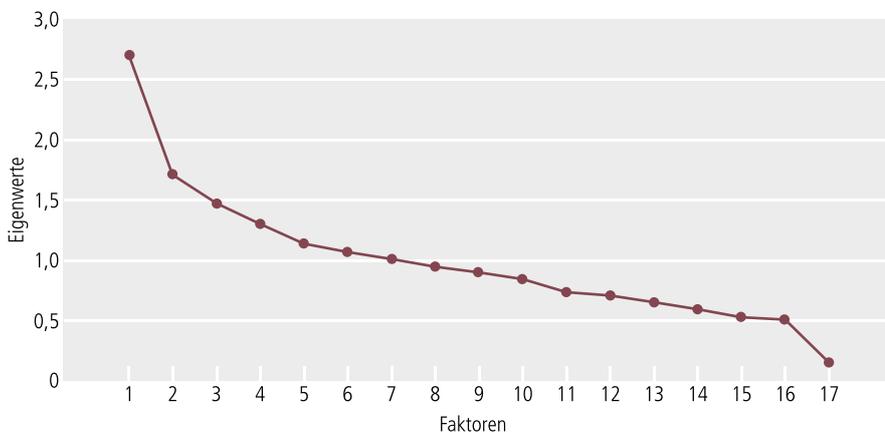


Abb. 3. Faktorenanalyse zum Zeitpunkt Visite 1: Darstellung der Eigenwerte der einzelnen Faktoren. Es zeigt sich deutlich, dass kein ausgeprägter Hauptfaktor zur Varianzaufklärung beiträgt.

lung in Beziehung gesetzt zur Remission (HAMD-17-Gesamtwert ≤ 8) nach sechswöchiger Behandlung. Es ergab sich ein positiver Vorhersagewert nach einer 7-tägigen Behandlung für Placebo zu 54,5% und für Verum zu 74,5%. Er zeigt in dieser Berechnung an, dass Patienten, die nach sieben Tagen Behandlung bereits eine Verbesserung des Gesamtscores von 20% oder mehr aufweisen, eine wesentlich höhere Chance auf eine deutliche Besserung ihrer Depression am Ende der Behandlung haben

(HAMD-17-Gesamtwert ≤ 8), wenn sie in der Verum-Gruppe sind.

Diskussion

In der vorliegenden Arbeit wurden zwei große, randomisierte, Placebo-kontrollierte Studien bei Patienten mit mittelgradiger Depression, die mit dem Johanniskrautextrakt STW3-VI in einer Dosierung von 900 mg/Tag durchgeführt wurden, einer gepoolten Analyse sowie einer Faktorenanalyse unterzogen.

In Bezug auf den HAMD-17-Gesamtwert zeigte sich ein deutlicher, hochsignifikanter Vorteil von STW3-VI über Placebo. Der Endwertunterschied im Gesamtwert der HAMD-17 betrug 4,4 Punkte. In der kleineren der beiden eingeschlossenen Untersuchungen [18] betrug der Endwertunterschied 7,4 Punkte, in der größeren Untersuchung [4] 2,7 Punkte, was zeigt, dass der große Endwertunterschied der gepoolten Analyse im Wesentlichen durch die kleinere Studie getrieben wird, aber auch in der größeren Studie vorhanden war. Ein wesentlicher Unterschied der beiden Studien bestand darin, dass in der Studie von Gastpar et al. [4] neben STW3-VI noch Citalopram 20 mg/Tag als aktiver Arm eingeschlossen war, ein den Untersuchern wohlbekannter selektiver Serotonin-Wiederaufnahmehemmer (SSRI), sodass die Erwartungshaltung in den Effekt der Verum-Therapie zumindest anders war als in der 2-Arm-Studie von Uebelhack et al. [18]. Zudem erhielt in der Studie von Uebelhack et al. jeder zweite Patient Placebo, in der Studie von Gastpar et al. nur jeder dritte, was die Erwartungshaltung sowohl der Untersucher wie auch der Patienten beeinflussen haben könnte. Der deutliche antidepressive Effekt von STW3-VI ließ sich auch auf HAMD-17-Einzelitem-Niveau sehr gut zeigen. Lediglich bei den Items 8 (depressive Hemmung), 16 (Gewichtsverlust) und 17 (Krankheitseinsicht) zeigten sich keine statistisch signifikanten Unterschiede, was mit der in diese beiden Studien eingeschlossenen Patientenpopulationen (mittelgradige depressive Episode, HAMD-17-Gesamtwert initial zwischen 20 und 24 Punkten) eine hinlängliche Erklärung insofern findet, da bei nicht sehr ausgeprägten Depressionen die Hemmung und der Gewichtsverlust gering ausgeprägt sind, die Krankheitseinsicht hingegen deutlich vorhanden ist, somit die Möglichkeiten der Veränderung (z. B. durch ein Antidepressivum) von vorneherein begrenzt sind. Die Schlafitems besserten sich parallel zur Reduktion des HAMD-17-Gesamtwerts. Es zeigte sich hier keine zeitliche

Tab. 3. Ergebnis der Faktorenanalyse (Darstellung der Ladungen der rotierenden Faktoren der Eigenwerte) zum Zeitpunkt Visite 4: Es ergab sich eine 4-faktorielle (Eigenwertkriterium) bzw. 2-faktorielle („Knick“ des Eigenwertverlaufs) Lösung mit einem Hauptfaktor, der aus den links dargestellten Einzelitems der HAMD-17 bestand und einen deutlich größeren Anteil der Varianz aufklärte als die anderen Faktoren. Es wurden nur Ladungen $\geq 0,5$ oder $\leq -0,5$ (siehe auch Abb. 4).

HAMD-17-Items	Faktoren				Bedeutung
	1	2	3	4	
1	0,721				Depressive Stimmung (1)
2					
3			0,679		Suizidalität (3)
4	0,721				Einschlafstörungen (1)
5	0,695				Durchschlafstörungen (1)
6	0,650				Schlafstörungen am Morgen (1)
7	0,624				Arbeit und sonstige Tätigkeiten (1)
8		0,700			Hemmung (2)
9					
10	0,565				Angst – psychisch (1)
11	0,504				Angst – somatisch (1)
12		0,664			Som. Symptome, gastrointestinal (2)
13	0,524				Som. Symptome – allgemeine (1)
14			0,806		Genitale Symptome (3)
15		0,626			Hypochondrie (1)
16		0,579			Gewichtsverlust (2)
17				-0,795	Einsicht (4)

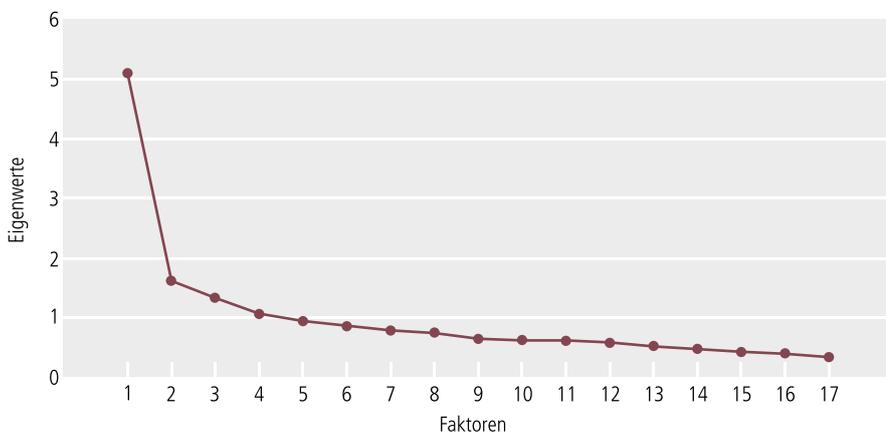


Abb. 4. Faktorenanalyse zum Zeitpunkt Visite 4: Darstellung des Eigenwerte der einzelnen Faktoren. Es zeigt sich deutlich, dass der erste Faktor (Hauptfaktor) mit Abstand den höchsten Eigenwert besitzt und somit auch am stärksten zur Varianzaufklärung beiträgt.

Verzögerung, was gegen mitunter geäußerte Interpretationen spricht, Schlaf bessere sich unter der Gabe von Johanniskrautextrakten erst sekundär, nachdem sich die depressive Kernsymptomatik gebessert habe.

Faktorenanalyse

Eine klassische Methode zur Strukturauflösung von Daten ist die sogenannte Faktorenanalyse, die seit Jahrzehnten

beispielsweise in der psychologischen und medizinischen Forschung Verwendung findet. Die Faktorenanalyse kommt immer dann zum Einsatz, wenn die Bündelung von Variablen (in unserem Fall die Items des HAMD-17) von methodischem Interesse ist. Hier stellt sich die Frage, ob es möglich ist, die Vielzahl von Items (= 17) auf einige wenige zentrale Faktoren zu reduzieren. Das Ziel ist also die Verdichtung

von Informationen bzw. die Identifikation erklärungsrelevanter übergeordneter Zusammenhänge.

In der vorliegenden Arbeit wurde mithilfe einer Faktorenanalyse für den Zeitpunkt Tag 0 eine multifaktorielle Lösung gefunden. Die ersten fünf Faktoren umfassten folgende Items:

- Faktor 1: Hemmung, Angst psychisch und somatisch, körperliche Symptome allgemein sowie Hypochondrie
- Faktor 2: Schlaf
- Faktor 3: Somatische Symptome gastrointestinal und Gewichtsverlust
- Faktor 4: Suizidalität und Einsicht
- Faktor 5: Depressive Stimmung und Erregung

Für den Zeitpunkt Tag 42 wurden vier Faktoren gefunden, die einen Eigenwert > 1 aufwiesen:

- Faktor 1: Depressive Stimmung, alle drei Schlaf-Items, somatische und psychische Angst, körperliche Symptome – allgemein
- Faktor 2: Hemmung, somatische Symptome gastrointestinal, Hypochondrie, Gewichtsverlust
- Faktor 3: Suizidalität, genitale Symptome
- Faktor 4: Einsicht

Wenngleich massive Kritik an der Zuverlässigkeit der Faktorenstruktur des HAMD seit vielen Jahren geübt wird (z. B. [1, 11]) und die Faktorenanalysen eine unterschiedliche Anzahl relevanter Faktoren trotz ähnlicher Stichprobencharakteristika erbrachten (zwischen ein und sieben Faktoren; Bagby et al. [1] fanden in ihrer Übersicht zwei bis acht identifizierte Faktoren), gibt es doch eine gewisse Übereinstimmung in Bezug auf die Faktorenladungen. So berichteten Bagby et al. [1], dass in 15 von ihnen erfassten Studien in 13 der Faktor Schlaf identifiziert wurde, ein genereller Depressionsfaktor (depressive Stimmung, Schuld, Suizidalität) sei auch häufig beschrieben worden, nämlich in sechs dieser 15 Studien. Ebenfalls in sechs Studien zeigte sich ein Faktor, der aus Agitation und psychischer sowie somatischer Angst bestand. Somit sind die von uns hauptsächlich zum Tag 0 identifizierten Faktoren 1 und 2 in gu-

Tab. 4. Faktorenanalyse der Differenzen Visite 1 – Visite 4 (Darstellung der Ladungen): Es ergab sich eine 5-faktorielle (Eigenwertkriterium) bzw. 2-faktorielle („Knick“ des Eigenwertverlaufs) Lösung mit einem Hauptfaktor, der einen deutlich größeren Anteil der Varianz aufklärte als die anderen Faktoren. Es wurden nur Ladungen $\geq 0,5$ dargestellt.

HAMD-17-Items	Faktoren					Bedeutung
	1	2	3	4	5	
1	0,587					Depressive Stimmung (1)
2						
3					0,900	Suizidalität (5)
4		0,620				Einschlafstörungen (2)
5		0,777				Durchschlafstörungen (2)
6		0,696				Schlafstörungen am Morgen (2)
7	0,633					Arbeit und sonstige Tätigkeit (1)
8						
9			0,546			Erregung (3)
10	0,625					Angst – psychisch (1)
11	0,583					Angst – somatisch (1)
12				0,665		Som. Symptome, gastrointestinal (4)
13	0,659					Som. Symptome – allgemeine (1)
14						
15						
16				0,817		Gewichtsverlust (4)
17			0,739			Einsicht (3)

ter Übereinstimmung mit diesen älteren Faktorenanalysen.

Auch neuere Arbeiten fanden immer wieder, dass Schlaf eine wichtige Rolle spielt [2, 5, 12, 13, 15, 16, 19] ebenso wie Angst und körperliche Beschwerden/Hypochondrie [2, 5, 6, 12, 13, 15, 16, 19].

Allerdings erstaunt, dass wir keinen eigenständigen Faktor Depression mit einer starken Ladung des ersten HAMD-17-Items zum Untersuchungszeitraum 0 fanden. Dies ist aber, betrachtet man die vorliegende Literatur, auch bei zahlreichen anderen Faktorenanalysen der HAMD-17 der Fall [1].

Zudem war in unseren Studien der initiale HAMD-17-Gesamtwert streng reglementiert: Nur Patienten mit einem initialen HAMD-17-Gesamtwert zwischen 20 und 24 wurden in die beiden untersuchten Studien aufgenommen. Dies führt zu einer deutlichen Verminderung der initialen Varianz, was in der Folge dazu führen kann, dass das Item 1 (depressive Stimmung) in der Faktorenanalyse kein großes Gewicht erhält, was sich zum Untersuchungszeitpunkt 2 (Tag 42) ändert, wohl wegen der dann (wieder) gegebenen

größeren Varianz der HAMD-17-Gesamtwerte.

Interessant ist, dass sich im Verlauf der Studie in der Faktorenanalyse ein Übergang von einer multifaktoriellen Lösung zu einer vierfaktoriellen Lösung ergibt, wobei ein Hauptfaktor mit Abstand den höchsten Anteil der Varianz erklärt. Das zeigt, dass bei nur noch bestehender Restsymptomatik zumindest in der hier untersuchten Stichprobe wenige Faktoren ausreichen (im Vergleich zur vollsymptomatischen Stichprobe zu Tag 0), um die Symptomatik statistisch ausreichend zu erklären. Es hat ein Übergang in eine „Einfachstruktur“ stattgefunden.

Responseprädiktion

In der vorliegenden Arbeit konnte gezeigt werden, dass eine Besserung um mindestens 20% im HAMD-17-Gesamtwert zu Ende der ersten Behandlungswoche für Verum zu 74,5% eine Remission (HAMD-17-Gesamtwert ≤ 8) prädizierte. Hiermit reihen sich die Ergebnisse dieser Untersuchung in eine lange Reihe ähnlicher Befunde ein (für einen Überblick über die ältere Literatur siehe [20], ansonsten [17]). Hierbei

kann festgehalten werden, dass meist die initiale Besserung nach zweiwöchiger Behandlung als Prädiktionsmaß verwendet wurde; dass in der vorliegenden Studie bereits nach einer Woche eine Prädiktion für eine Remission am Ende der Studie (nach sechs Wochen) möglich war, kann im Sinne eines unterstützenden Arguments für die gute Wirksamkeit des Johanniskrautextrakts gewertet werden. Auf der anderen Seite zeigt das Ergebnis aber des Weiteren, dass auch für die Behandlung mit Johanniskraut dieser für zahlreiche Antidepressiva (wohl am besten durch die Metaanalyse von Szegeedi et al. [17] für Mirtazapin nachgewiesen) gezeigte Zusammenhang gilt.

Limitationen

Die Arbeit schloss zwei Studien ein, die zwar nach einem ähnlichen, aber nicht nach einem exakt identischen Design durchgeführt wurden; zudem musste für die Studie von Uebelhack et al. [18] der HAMD-17-Gesamtwert für den Tag 7 über eine Regressionsgleichung geschätzt werden. Auch könnte durch die eingeschlossene Referenz (Citalopram, 20 mg/Tag) in der Studie von Gaspar et al. [4], deren Patienten nicht in die Auswertungen mit eingeschlossen wurden, eine Verzerrung der Erwartungshaltungen der eingeschlossenen Patienten im Vergleich zu einer Zwei-Arm-Studie eingetreten sein. Insgesamt ist auch die deutlich unterschiedliche Studiengröße zu bemängeln.

Allgemeine Aspekte

Als generelles Fazit kann festgehalten werden: Die Re-Analyse zweier Placebo-kontrollierter, randomisierter Studien mit dem Johanniskrautextrakt STW3-VI zeigte neben interessanten Einzelbefunden insgesamt, dass sich eine Reihe aus der Analyse von Studiendaten synthetischer Antidepressiva gewonnener Erkenntnisse replizieren ließ, was als indirekter Hinweis für die Qualität der eingeschlossenen beiden Studien wie auch für den robusten Behandlungseffekt (im Vergleich zu Placebo) von STW3-VI interpretiert werden kann.

Allerdings wäre es im Feld der Metaanalysen unbedingt wünschenswert, beispielsweise mithilfe von Netzwerk-Metaanalysen die vergleichende Wirksamkeit unterschiedlicher Johanniskraut-Extrakte zu untersuchen. Zudem fehlen nach wie vor Daten zur Wirksamkeit in der Langzeittherapie.

Interessenkonflikterklärung

HPV war in den letzten zwei Jahren als Berater für folgende Firmen tätig: Lundbeck, Pfizer Pharma, Dr. Willmar-Schwabe, Bayer Vital, Janssen-Cilag, neuraxpharm. Von folgenden Firmen hat HPV in den letzten zwei Jahren Zuwendungen für Vorträge erhalten: Lundbeck, Astra Zeneca, Pfizer Pharma, Dr. Willmar-Schwabe, Bayer Vital, Janssen-Cilag, Servier und neuraxpharm. Außerdem besitzt HPV Aktien von Novartis.

JM und CK sind Mitarbeiter der Steigerwald Arzneimittelwerk GmbH.

Efficacy of the hypericum-extract STW3-VI – a reanalysis of pooled data of two placebo-controlled trials

To sample further information regarding the efficacy of a St. John's wort extract (STW3-VI), the data of two randomized, placebo-controlled trials (total number of participants: 401) were analyzed using different approaches: the course of the pooled data of the Hamilton Depression Scale (HAMD-17) total score and also of the single items of the HAMD-17 was calculated; a factor analysis of the HAMD-17 at the beginning, at the end, and also using the difference was performed; and the predictive power of the initial decrease in the HAMD-17 total score for the remission (HAMD-17 total score ≤ 8) at the end of the trial was analyzed. The HAMD-17-total score decreased in the STW3-VI-group considerably more than in the placebo group, the difference was already statistically significant at week 3 and reached difference of 4.4 points at week 6 ($p < 0.01$). In all HAMD-17 single items (beside retardation, loss of weight, and insight) a statistically significant difference was found, too. A factor analysis resulted in a multi-factorial result for week 0 and a four-factorial solution for week 6. The initial reduction of the HAMD-17 total score of at least 20% at week 2 predicted the remission at week 6. These results are discussed in the light of other investigations using similar methods. A

good concordance with studies using synthetic antidepressants was found.

Key words: Hypericum extract, depression, STW3-VI, factor analysis, re-analysis, antidepressants, phytopharmakon

Literatur

1. Bagby RM, Ryder AG, Schuller DR, Marshall MB. The Hamilton Depression Rating Scale: has the gold standard become a lead weight? *Am J Psychiatry* 2004;161:2163–77.
2. Broen MP, Moonen AJ, Kuijff ML, Dujardin K, et al. Factor analysis of the Hamilton Depression Rating Scale in Parkinson's disease. *Parkinsonism Relat Disord* 2015;21:142–6.
3. Gastpar M, Singer A, Zeller K. Efficacy and tolerability of hypericum extract STW3 in long-term treatment with a once-daily dosage in comparison with sertraline. *Pharmacopsychiatry* 2005;38:78–86.
4. Gastpar M, Singer A, Zeller K. Comparative efficacy and safety of a once-daily dosage of hypericum extract STW3-VI and citalopram in patients with moderate depression: a double-blind, randomised, multicentre, placebo-controlled study. *Pharmacopsychiatry* 2006;39:66–75.
5. Grover S, Sahoo S, Dua D, Chakrabarti S, et al. Scales for assessment of depression in schizophrenia: Factor analysis of Calgary depression rating scale and Hamilton Depression Rating Scale. *Psychiatry Res* 2017;252:333–9.
6. Guillén V, Santos B, Yllá L, Bulbena A, et al. Depressive dimensions and item response analysis of the Hamilton Depression Rating Scale-17 in eating disorders. *Compr Psychiatry* 2012;53:396–402.
7. Hammond MF. Rating depression severity in the elderly physically ill patient: reliability and factor structure of the Hamilton and the Montgomery-Asberg Depression Rating Scales. *Int J Geriatr Psychiatry* 1998;13:257–61.
8. Lemke M, Barthel B, Hajak G. Antidepressive Therapieeffekte unter Agomelatin über sechs Monate und prädiktiver Wert einer frühen Verbesserung. Ergebnisse der VITAL-Studie. *Psychopharmakotherapie* 2016;23:151–61.
9. Linde K, Berner M, Egger M, Mulrow C. St. John's wort for depression: meta-analysis of randomised controlled trials. *Br J Psychiatry* 2005;186:99–107.
10. Linde K, Mulrow CD, Berner M, Egger M. St. John's wort for depression. *Cochrane Database Syst Rev*. 2005;(2):CD000448.
11. Maier W, Philipp M, Gerken A. Dimensionen der Hamilton-Depressionsskala (HAMD). Faktorenanalytische Untersuchungen. *Eur Arch Psychiatry Neurol Sci* 1985;234:417–22.
12. Mitchell PB, Hadzi-Pavlovic D, Evoniuk G, Calabrese JR, et al. A factor analytic study in bipolar depression, and response to lamotrigine. *CNS Spectr* 2013;18:214–24.
13. Olden M, Rosenfeld B, Pessin H, Breitbart W. Measuring depression at the end of life: is the Hamilton Depression Rating Scale a valid instrument? *Assessment* 2009;16:43–54.
14. Onega LL, Abraham IL. Factor structure of the Hamilton Rating Scale for Depression in a cohort of community-dwelling elderly. *Int J Geriatr Psychiatry* 1997;12:760–4.
15. Shafer AB. Meta-analysis of the factor structures of four depression questionnaires: Beck, CES-D, Hamilton, and Zung. *J Clin Psychol* 2006;62:123–46.
16. Schennach R, Obermeier M, Seemüller F, Jäger M, et al. Evaluating depressive symptoms in schizophrenia: a psychometric comparison of the Calgary Depression Scale for Schizophrenia and the Hamilton Depression Rating Scale. *Psychopathology* 2012;45:276–85.
17. Szegei A, Jansen WT, van Willigenburg AP, van der Meulen E, et al. Early improvement in the first 2 weeks as a predictor of treatment outcome in patients with major depressive disorder: a meta-analysis including 6,562 patients. *J Clin Psychiatry* 2009;70:344–53.
18. Uebelhack R, Gruenwald J, Graubaum HJ, Busch R. Efficacy and tolerability of Hypericum extract STW3-VI in patients with moderate depression: a double-blind, randomized, placebo-controlled clinical trial. *Adv Ther* 2004;21:265–75.
19. Uher R, Farmer A, Maier W, Rietschel M, et al. Measuring depression: comparison and integration of three scales in the GENDEP study. *Psychol Med* 2008;38:289–300.
20. Volz HP, Müller H, Sturm Y, Preussler B, et al. Effect of initial treatment with antidepressants as a predictor of outcome after 8 weeks. *Psychiatry Res* 1995;58:107–15.