

**Vorsorgeuntersuchung Neugestaltung:
Screening auf Depression
(Version 1.1)**

Univ.-Prof. Dr. Johannes Wancata

6. Mai 2004
(Internetpublikation)

Inhaltsverzeichnis

	Seite
1. Ausgangslage	3
1.1. Epidemiologie depressiver Erkrankungen	3
1.2. Erkennen depressiver Erkrankungen durch Praktische Ärzte.....	3
1.3. Theoretische Rahmenbedingungen des Depressions-Screenings.....	4
2. Bearbeitung der einzelnen Fragestellungen.....	6
2.1. Kriteriums-Validität und zeitökonomische Anwendung von Depressions-Screening-Instrumenten	6
2.1.1. Kriteriums-Validität – Internationale Studien.....	6
2.1.2. Kriteriums-Validität – Deutschsprachige Studien	11
2.1.3. Zeitökonomische Anwendung von Depressions-Screening-Instrumenten	13
2.1.4. Kriteriums-Validität – Kombination aus Kurz-Screening gefolgt von längerem Screening-Instrument.....	14
2.1.5. Für welche Personengruppen liegen ausreichend Daten vor, die ein Screening auf Depression rechtfertigen?	17
2.2. Die potenziell schädlichen Folgen von Screening	19
2.2.1. Vermeidung der theoretisch möglichen Probleme bei Screening-Positiven	19
2.2.2. Hinweise auf potenziell negativen Folgen von Screening aus randomisiert-kontrollierten Studien	20
2.2.3. Screening-Negative	23
2.3. Überprüfung des Screening-Resultates.....	25
2.3.1. Vorgangsweise bei Screening-Positiven	25
2.3.2. Psychiatrische Interviews für Praktische Ärzte.....	27
2.3.3. Vorgangsweise bei Screening-Negativen	27
2.4. Stehen Möglichkeiten einer effektiven Behandlung mit geringen Nebenwirkungen zur Verfügung?	28
2.4.1. Behandlung mit Antidepressiva.....	28
2.4.2. Psychotherapie.....	28
2.4.3. Soziotherapie.....	28
2.5. Effektivität von Screening, gefolgt von diagnostischer Abklärung und Behandlung	29
2.6. In welchen Abständen sollte ein Screening auf Depressionen durchgeführt werden?	29
2.7. Welche Maßnahmen zur Qualitätssicherung könnten nützlich sein?	30
3. Schlussfolgerungen und Zusammenfassung.....	31
4. Verwendete Literatur.....	35
5. Anhang	
• Anhang A: Beschreibung der verwendeten Kennwerte der Kriteriums-Validität	40
• Anhang B: Depressions-Screening-Instrumente: Verwendete Abkürzungen	41

Auch wenn im Text überwiegend nur die männliche Form verwendet wird, sind immer Männer und Frauen gemeint.

1. AUSGANGSLAGE

1.1. Epidemiologie depressiver Erkrankungen

International ist davon auszugehen, dass ca. 20% der Bevölkerung im Verlauf des Lebens zumindest einmal unter einer klinisch-relevanten depressiven Erkrankung leidet. Die Häufigkeit einer aktuell behandlungsbedürftigen depressiven Störung wird mit einer Punktprävalenz von ca. 6% bezogen auf die letzten 4 Wochen angegeben (Härter et al 2003), wobei für Frauen häufig höhere Raten als für Männer berichtet werden (Kühner 2003).

Depressive Störungen sind ernste und häufig schwere Erkrankungen. Depressive Erkrankungen gehen mit einer erhöhten Mortalität, vor allem durch Suizide einher. Die Suizidrate bei Personen mit Depressionen ist etwa 30 Mal höher als in der Durchschnittsbevölkerung. Bei einer Reihe somatischer Erkrankungen, wie z.B. koronare Herzkrankheit und Herzinfarkt, Diabetes mellitus oder Rückenschmerzen ist das gleichzeitige Vorliegen einer depressiven Erkrankung ein entscheidender Prognosefaktor im Hinblick auf Chronifizierung bzw. erhöhte Mortalität.

Die durchschnittliche Dauer einer depressiven Episode liegt nach epidemiologischen Studien bei ca. 3, unbehandelt bei 4-6 Monaten. Nur bei ca. einem Drittel der Betroffenen kann mit einer vollständigen und anhaltenden Remission (ohne restliche Symptome und ohne wieder zu erkranken) gerechnet werden. Weitaus häufiger neigen Personen mit depressiven Erkrankungen zu Rezidiven. So geht man davon aus, dass etwa jeder zweite Patient im Laufe seines Lebens eine weitere depressive Episode erleidet, nach zweimaliger Erkrankung liegt diese Wahrscheinlich bei 70% und nach der dritten Episode bei 90% (Härter et al 2003).

Die Angaben zur Prävalenz depressiver Erkrankungen beim praktischen Arzt schwanken in den verschiedenen internationalen Studien. Die US Preventive Services Task Force (2002) berichteten in der Übersicht eine Punktprävalenz für depressive Episoden beim praktischen Arzt zwischen 6 und 8%. Eine vor einigen Jahren durchgeführte Studie der WHO (Üstün & Sartorius 1995) berichtet eine Prävalenz für depressive Episoden bei praktischen Ärzten in 15 verschiedenen Ländern von 10,4%. Zusätzlich wurden in dieser Studie 2,1% Patienten mit Dysthymie (eine chronisch verlaufende depressive Erkrankung) gefunden. An der genannten WHO-Studie nahmen zwei Zentren aus Deutschland teil, für die Prävalenzraten beim praktischen Arzt vorliegen: Mainz: depressive Episode: 11,2%, Dysthymie 0,9%; Berlin: depressive Episode 6,1%, Dysthymie 0,5%. Für Österreich liegen leider keine aktuellen Prävalenzstudien beim praktischen Arzt vor. International liegen jedoch die meisten Berichte für depressive Störungen zwischen 5 und 10%.

1.2. Erkennen depressiver Erkrankungen durch Praktische Ärzte

Nachdem der Großteil jener Personen, die unter Depressionen leiden, beim Praktischen Arzt und nicht in psychiatrischen Facheinrichtungen betreut wird, liegt es am Praktischen Arzt die psychiatrische Diagnose zu stellen. Nach Goldberg und Huxley (1980) hatten mehr als 90% der psychisch kranken Erwachsenenbevölkerung mit ihrem Hausarzt und nur etwa 7% mit einem Facharzt für Psychiatrie bzw. einer ambulanten oder stationären psychiatrischen Einrichtung Kontakt. Bei praktischen Ärzten fand die große epidemiologische Studie der WHO (Üstün & Sartorius 1995), dass beträchtliche Teile der psychisch Kranken von ihren Ärzten nicht als psychisch krank erkannt oder falsch diagnostiziert wurden. Auch aus österreichischen Allgemeinkrankenhäusern wird berichtet (Wancata et al. 2000), dass die nicht-psychiatrischen Ärzten der jeweiligen Fachgebiete etwa 45% aller psychischen Erkrankungen nicht erkannten. Sie berichteten weiters, dass die Rate der psychischen

Erkrankungen, die richtig erkannt werden, nicht nur von der Diagnose, sondern auch von den diagnostischen Subtypen abhängt. So wurden 61,5% der Patienten mit „Major Depression“ (entsprechend DSM-III-R), aber nur 41% jener mit einer „Minor Depression“ erkannt.

Dies ist insofern von Bedeutung, da Depressionen nicht nur zu persönlichem Leid der Betroffenen führen, sondern auch das Risiko für langdauernde Einschränkungen und für eine erhöhte Sterblichkeit aufgrund von Suiziden erhöhen. Depressive Patienten nehmen Gesundheitseinrichtungen häufiger als andere in Anspruch und haben eine erhöhte Zahl von Krankenstandstagen, was zu beträchtlichen Kosten führt (Whooley et al. 1997). Dies führt dazu, dass Depressionen weltweit zu den häufigsten Ursachen für Behinderungen gezählt werden (Global Burden of Disease Study, zit. nach Aschauer et al. 2001)

Zahlreiche Studien wurden daher mit der Fragestellung durchgeführt, wie nun das Erkennen psychiatrischer Krankheitsbilder verbessert werden könne. Neben gezielten Fortbildungen wurde hier vor allem der Einsatz von Screening-Instrumenten vorgeschlagen. Als Vorteil von Screening-Instrumenten wurde immer wieder betont, dass sie auch im (oft hektischen) klinischen Alltag praktikabel und rasch auszufüllen seien.

1.3. Theoretische Rahmenbedingungen des Depressions-Screenings

Wie erwähnt war von verschiedenen Autoren vorgeschlagen worden, Praktische Ärzte sollten Screening-Instrumente für Depressionen routinemäßig verwenden, um Depressionen besser identifizieren zu können. Derartige Screening-Instrumente sind üblicherweise kurze Fragebögen, deren positive Antworten zu einem Summenscore zusammengezählt werden. Personen, die über einem bestimmten Grenzwert dieses Summenscores (= Cut-off) liegen, haben eine erhöhte Wahrscheinlichkeit unter einer Depression zu leiden (Screening-Positive). Dieses Vorgehen macht jedoch nur dann einen Sinn, wenn mehrere Voraussetzungen zutreffen. Diese Voraussetzungen wurden von der US Preventive Services Task Force (2002) in einer vereinfachten Form in einer Abbildung dargestellt (Abbildung 1).

Die folgenden Voraussetzungen sollten zutreffen:

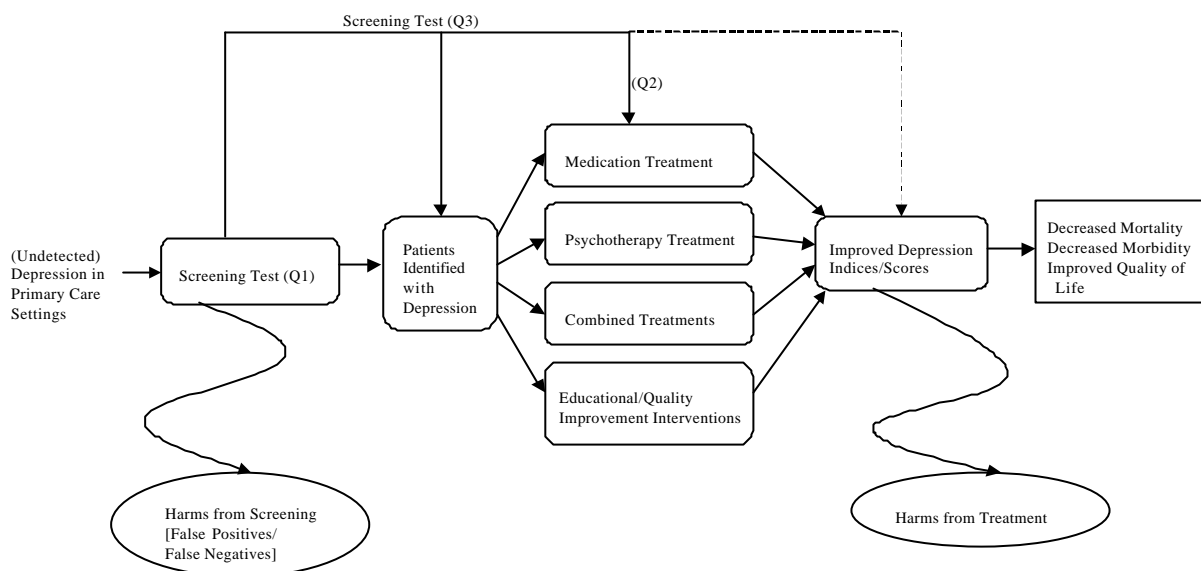
1. Screening-Instrumente sollen möglichst exakt vorhersagen, ob jemand an einer Depression leidet (= Kriteriums-Validität), und gleichzeitig zeitökonomisch in der Anwendung sein.
2. Die Verwendung von Screening-Instrumenten und die anschließende diagnostische Abklärung soll für die Betroffenen keine nachteiligen Folgen (potenzielle Schäden) haben.
3. Überprüfung des Screening-Resultates: Wenn das Screening-Resultat auf das Vorhandensein einer Depression hinweist, soll möglichst einfach geklärt werden können, ob wirklich eine Depression vorliegt. (Bei der Planung der diagnostischen Abklärung und Behandlung muss insbesondere die österreichische Situation berücksichtigt werden.)
4. Es sollen Möglichkeiten einer effektiven Behandlung zur Verfügung stehen, wobei die verfügbaren Behandlungstechniken im Vergleich zu ihrem Nutzen ein möglichst geringes Risiko für potenzielle Schäden (Nebenwirkungen) haben sollen.
5. Die gesamte Intervention (Screening, gefolgt von diagnostischer Abklärung und Behandlung) soll für die Betroffenen bessere Ergebnisse (Verringerung der Symptomatik bzw. der Krankheitsfolgen) bringen im Vergleich zu keiner Intervention.

Die folgenden Fragen scheinen im Zusammenhang mit einer etwaigen Implementierung in Österreich relevant:

6. Gibt es Hinweise darauf, in welchen Abständen eine derartige Screening-Prozedur sinnvoll ist?
7. Welche Maßnahmen zur Qualitätssicherung könnten nützlich sein?

Abbildung 1: Strukturelle Rahmenbedingungen des Screenings auf Depression
(entnommen aus: Research Triangle Institute 2002)

Figure 1: USPSTF Analytic Framework for Depression



Q1. What is the accuracy of screening instruments for depression in primary care populations?

Q2. Is treatment of depression in primary care patients (with pharmacologic therapy, psychotherapy, combinations of the 2, or educational interventions) effective in improving outcomes?

Q3. Is screening more effective than usual care in identifying patients with depression, facilitating treatment of patients with depression, and improving outcomes?

Im Falle einer routinemäßigen Implementierung im Rahmen der Vorsorgeuntersuchungen in Österreich muss die spezielle österreichische Situation Beachtung finden – dies wurde im Folgenden bei allen Teilen, die die praktische Planung und Umsetzung betreffen, berücksichtigt.

2. BEARBEITUNG DER EINZELNEN FRAGESTELLUNGEN

2.1. Kriteriums-Validität und zeitökonomische Anwendung von Depressions-Screening-Instrumenten

2.1.1. Kriteriums-Validität – Internationale Studien

Im Folgenden wird der Frage nachgegangen, inwieweit die neueren Kurz-Screening-Fragebögen bestehend aus 1 oder 2 Items mit den umfassenderen Screening-Fragebögen vergleichbar sind. Zu diesem Zweck wurde auf Basis der aus verschiedenen Studien verfügbaren Daten überprüft, wie genau die jeweiligen Screening-Instrumente das Vorhandensein einer Depression (PPV = Positive Predictive Value) bzw. die Abwesenheit einer Depression (NPV = Negative Predictive Value) vorhersagen können. Weiters wurde die Rate aller falsch zugeordneten Patienten (OMR = Overall Misclassification Rate) bestimmt (siehe Anhang A). Da diese Kennwerte von der Prävalenz, also der Häufigkeit des Krankheitsbildes, beeinflusst werden, wurden zu diesem Zweck die Werte für jene Prävalenz-Raten berechnet, die laut internationaler Literatur bei den Patienten von Praktischen Ärzten vorkommen (zumeist zwischen 5 und 10%).

Eine detaillierte Literatur-Suche und Bearbeitung war aufgrund der knappen zur Verfügung stehenden Zeit nicht möglich. Die internationale Studien zur Untersuchung dieser Fragestellung wurde daher einer der neueren, umfassenden Übersichten (Research Triangle Institute 2002) entnommen.

Für die folgenden Auswertungen wurden nur jene Studien verwendet, die

1. im allgemein-medizinischen Setting durchgeführt wurden (d.h. Pflegeheime, stationäre Patienten, Allgemeinbevölkerungsstudien ausgeschlossen) ,
2. übliche Screening-Instrumente und übliche Fall-Kriterien verwendeten (z.B. eigene Definitionen von Depression ausgeschlossen),
3. die Diagnosen von Fachleuten verwendet hatten,
4. die üblichen Versionen von Screening-Instrumenten verwendeten (z.B. Versionen mit speziellen Item-Kombinationen wurden ausgeschlossen),
5. Studien, die einen klaren Cut-off-Wert angaben.

Die folgenden Studien (Tabelle 1) wurden verwendet, um eigene Berechnungen bezüglich Positive Predictive Value, Negative Predictive Value und Overall Misclassification Rate durchzuführen. Mit Ausnahme einer einzigen Studie, die nur Frauen einschloss, haben alle Studien sowohl Männer als auch Frauen eingeschlossen, wobei der Anteil weiblicher Patientinnen meist etwas überwog. Die meisten Studien machten Angaben zu den untersuchten Altersgruppen. Aus der Alterszusammensetzung der Stichproben kann daraus geschlossen werden, dass für alle Personen über dem Alter von 18 Jahren ausreichende Informationen vorliegen.

Die meisten Studien wurden bei Praktischen Ärzten durchgeführt, einzelne Studien schlossen spezielle Gruppen von Patienten ein. Es kann vermutet werden, dass die Validitäts-Kennwerte der beiden Studien, die nicht bei Praktischen Ärzten durchgeführt wurden, für die hier vorliegende Fragestellung mit hinreichender Genauigkeit verwendbar sind.

Tabelle 1: Beschreibung der Studien aus der Übersicht des Research Triangle Institute (2002), die bezüglich der Validitätskennwerte weiter ausgewertet wurden

Autor, Jahr	Screening Setting	Alter (Jahre)	Anteil weiblicher Patientinnen	Diagnose	Größe der Stichprobe
Broadhead et al 1995	Praktischer Arzt	18-70	73%	Major Depression	388
Fechner-Bates et al 1994	Praktischer Arzt	17-80	76%	Major Depression	425
Hendrie et al 1995	Praktischer Arzt	>60	68%	Major Depression	125
Holcomb et al 1996	Gynäkologische Ambulanz	--	100%	Depression	105
Klinkman et al 1997	Praktischer Arzt	--	77%	Major Depression	425
Leon et al 1996	Praktischer Arzt	18-70	66%	Major Depression	501
Lustman et al 1997	Diabetes Ambulanz – Patienten aus Allg. Bevölkerung	--	48%	Major Depression	172
Lyness et al 1997	Praktischer Arzt	>60	58%	Major Depression	130
Myers & Weissman 1980	Praktischer Arzt	≥18	57%	Major Depression	515
Upadhyaya & Stanley 1997	Praktischer Arzt	65-90	49%	Depression	72
van Marwijk et al 1995	Praktischer Arzt	65-94	60%	Major Depression; Dysthymie	586
Zich et al 1990	Praktischer Arzt	≥18	<u>63%</u>	Major Depression	475

In Tabelle 2 sind die Werte für die Sensitivität und Spezifität dargestellt, die in den jeweiligen Studien berichtet wurden. (Wenn in einzelnen Studien mehrere Cut-off-Werte angegeben wurden, wurden für die folgende Darstellung entsprechend den internationalen Gepflogenheiten jene gewählt, wo Sensitivität und Spezifität am nächsten beisammen lagen.) Die Kennwerte Positive Predictive Value, Negative Predictive Value und Overall Misclassification Rate wurden für eine angenommene Prävalenz von 5% und 10% berechnet. Da dies die Ober- und Untergrenze der zumeist in der Literatur berichteten Prävalenz-Raten darstellt, kann angenommen werden, dass die real bei Praktischen Ärzten in Österreich zu beobachtenden Werte in diesem Bereich liegen. Bei den dargestellten Screening-Instrumenten handelt es sich solche, die in der internationalen Literatur sehr zu finden sind.

In den letzten Jahren waren von einigen Autoren auch Kurz-Screening-Fragebögen aus einem oder zwei Fragen bestehend für das Erfassen von Depressionen vorgeschlagen worden. Diese Kurz-Screening-Fragebögen waren entwickelt worden, da man davon ausging, dass kürzere Fragebogen im klinischen Alltag eher verwendet werden als längere Fragebögen.

Die beiden Fragen für den 2-Item Fragebogen waren von Whooley et al. (1997) aus dem PRIME-MD (Spitzer et al. 1994) entnommen worden. Spitzer et al. (1994) hatten für die Kombination dieser beiden Fragen eine Sensitivität von 86% und eine Spezifität von 75% berichtet. Dieser Fragebogen besteht aus den beiden folgenden Fragen (Übersetzung von Härter et al. 2003):

„Haben Sie sich im letzten Monat oft niedergeschlagen, depressiv oder hoffnungslos gefühlt?“

„Hatten Sie im letzten Monat oft wenig Interesse oder Freude an Dingen, die Ihnen früher Spaß gemacht haben?“

Whooley et al. (1997), die die Validität des 2-Item Fragebogens mit anderen Screening-Instrumenten verglichen hatten, berichteten eine Sensitivität von 96% und eine Spezifität von 57% (Tabelle 2). Diese Autoren hatten auch Kennwerte für jene Personen berichtet, die nicht unter Substanzmissbrauch leiden. Diese Kennwerte scheinen aber für die Aufgabenstellung eines Routine-Screenings nicht relevant, da ja alle Patienten von Praktischen Ärzten untersucht werden sollen, also auch jene, die unter Substanzmissbrauch leiden.

In einer kürzlich veröffentlichten Studie mit dem 2-Item Fragebogen aus Neuseeland berichteten Arroll et al. (2003) eine Sensitivität von 97% und eine Spezifität von 67%.

Williams et al. (1999) hatten die Frage untersucht, ob ein Screening auf Depression auch mittels einer einzigen Frage möglich sei. Sie verwendeten dafür die folgende Frage:

„Have you felt depressed or sad much of the time in the past year?“

Diese Frage bezieht sich auf das letzte Jahr – dies muss als Nachteil gewertet werden, da für die Behandlungsplanung der aktuelle Zustand bzw. die letzten Wochen relevant sind, nicht jedoch irgendwelche Episoden des letzten Jahres. Mit dieser Frage wird also auch erfasst, ob jemand im vergangenen Jahr eine depressive Episode hatte, die jedoch bereits abgeklungen ist.

Wenn man die Validitäts-Kennwerte der vier Studien zu den Kurz-Screening-Fragebögen mit jenen der umfassenderen Screening-Fragebögen vergleicht (Tabelle 2), ergibt sich folgendes Bild: der Median der Sensitivität der Kurz-Screening-Fragebögen liegt mit 91% um etwa 10% über den Medianen der umfassenderen Screening-Fragebögen. Die Spezifität (Median) der Kurz-Screening-Fragebögen ist aber etwa 15-20% niedriger als jene der umfassenderen Instrumente. Die Kurz-Screening-Fragebögen übersehen also nur einen sehr geringen Anteil der depressiven Patienten. Die Kurz-Screening-Fragebögen klassifizieren aber einen deutlich höheren Anteil der psychisch Gesunden fälschlich als depressiv.

Bei einer Prävalenz von 5% beträgt der Positive Predictive Value (Median) der Kurz-Screening-Fragebögen 12,5%. Dies heißt, dass bei Verwendung eines Kurz-Screening-Fragebogens nur etwa jeder achte Screening-Positive wirklich unter einer Depression leidet. Bei einer Prävalenz von 10% erhöht sich der Positive Predictive Value (Median) der Kurz-Screening-Fragebögen auf 23,1%. Die Positive Predictive Values der umfassenderen Screening-Fragebögen liegen zwischen 20% und 28% bei einer Prävalenz von 5% und zwischen 34% und 46% % bei einer Prävalenz von 10%. Der Anteil der screening-positiven Personen, die wirklich unter einer Depression leiden, ist also bei Verwendung der umfassenderen Screening-Fragebögen deutlich höher.

Die Negative Predictive Values liegen für alle Instrumente durchwegs über 97%, wobei sich die Kurz-Fragebögen nur minimal von den längeren Fragebögen unterscheiden. Die Kurz-Fragebögen führen insgesamt dazu, dass rund ein Drittel der Patienten falsch zugeordnet werden (Overall Misclassification Rate = OMR), während diese Fehlzuordnungen bei den anderen Fragebögen deutlich geringer sind.

Wenn man die Kennwerte für alle Screening-Instrumente betrachtet, zeigen sich für das Beck-Depression-Inventary (BDI) insgesamt die günstigsten Ergebnisse. Nur etwa 11%

werden insgesamt falsch zugeordnet. Die Kennwerte der Center for Epidemiological Studies Depression Scale (CES-D) zeigen insgesamt die zweitbesten Ergebnisse¹.

Tabelle 2: Gegenüberstellung der Validitäts-Kennwerte von international gebräuchlichen Screening-Instrumenten (Research Triangle Institute 2002) mit jenen der Kurz-Screening-Fragebögen (Abkürzungen der Screening-Instrumente siehe Anhang B)

Sens = Sensitivität, Spez = Spezifität, PPV = Positive Predictive Value, NPV = Negative Predictive Value, OMR = Overall Misclassification Rate

Autor, Jahr	Screening-Instrument	Sens	Spez	Prävalenz 5%			Prävalenz 10%		
				PPV	NPV	OMR	PPV	NPV	OMR
Gebräuchliche Screening-Instrumente									
Fechner-Bates et al 1994	CES-D >16	79,5	71,1	12,6	98,5	28,5	23,4	96,9	28,1
Hendrie et al 1995	CES-D >16	80,0	91,0	31,9	98,9	9,6	49,7	97,6	10,1
Zich et al 1990	CES-D >27	100,0	81,0	21,7	100,0	18,1	36,9	100,0	17,1
Klinkman et al 1997	CES-D >16	81,0	72,0	13,2	98,6	27,6	24,3	97,2	27,1
Lyness et al 1997	CES-D >21	92,0	87,0	27,1	99,5	12,8	44,0	99,0	12,5
Myers & Weissman 1980	CES-D >17	63,6	94,4	37,4	98,0	7,1	55,8	95,9	8,7
	Median	80,5	84,0	24,4	98,7	15,4	40,5	97,4	14,8
BDI - Beck Depression Inventory									
Holcomb et al 1996	BDI >16	83,0	89,0	28,4	99,0	11,3	45,6	97,9	11,6
Lustman et al 1997	BDI >13	85,0	88,0	27,2	99,1	12,2	44,0	98,1	12,3
Zich et al 1990	BDI >16	100,0	89,0	32,4	100,0	10,5	50,3	100,0	9,9
	Median	85,0	89,0	28,4	99,1	11,3	45,6	98,1	11,6
GDS-30 - Geriatric Depression Scale									
Lyness et al 1997	GDS-30: >10	100,0	84,0	24,8	100,0	15,2	41,0	100,0	14,4
van Marwijk et al 1995	GDS-30: >11	55,0	86,0	17,1	97,3	15,6	30,4	94,5	17,1
	Median	77,5	85,0	20,9	98,6	15,4	35,7	97,2	15,8
SDDS-PC - Short Depression Symptom Scale									
Leon et al 1996	SDDS-PC >2	71,0	87,0	22,3	98,3	13,8	37,8	96,4	14,6
Broadhead et al 1995	SDDS-PC >2	90,4	77,2	17,3	99,3	22,1	30,6	98,6	21,5
	Median	80,7	82,1	19,8	98,8	17,9	34,2	97,5	18,1
HAD-D - Hospital Anxiety and Depression Scale									
Upadhyaya & Stanley 1997	HAD-D: >9	70,0	87,0	22,1	98,2	13,9	37,4	96,3	14,7
Kurz-Screening-Fragebögen									
Spitzer et al. 1994	2 Fragen	86,0	75,0	15,3	99,0	24,5	27,7	98,0	23,9
Whooley et al. 1997	2 Fragen	96,0	57,0	10,5	99,6	41,1	19,9	99,2	39,1
Arroll et al. 2003	2 Fragen	97,0	67,0	13,4	99,8	31,5	24,6	99,5	30,0
Williams et al. 1999	1 Frage	85,0	66,0	11,6	98,8	33,1	21,7	97,5	32,1
	Median	91,0	66,5	12,5	99,3	32,3	23,1	98,6	31,0

Von praktischer Relevanz ist auch die Frage, wie hoch der Anteil Screening-Positiver von allen mit einem bestimmten Screening-Instrument Untersuchten ist. Diese Frage ist insofern wichtig, da alle Personen, die ein positives Screening-Ergebnis zeigen, dann in einem

¹ **BDI - Beck Depression Inventory:** Das Beck Depression Inventory wurde ursprünglich von Beck (1978) entwickelt. Die deutsche Übersetzung erfolgte durch Hautzinger et al. (1994). Das Beck Depression Inventory besteht aus 21 Items.

CES-D - Center for Epidemiological Studies Depression Scale: Die Center for Epidemiological Studies Depression Scale wurde ursprünglich von Radloff (1977) entwickelt. Die deutsche Übersetzung erfolgte durch die Arbeitsgruppe um Hautzinger. Die Center for Epidemiological Studies Depression Scale besteht aus 20 Items.

zweiten Schritt genauer untersucht werden müssen, um abzuklären, ob wirklich eine Depression vorliegt.

Bei Verwendung eines Kurz-Screening-Fragebogens (Tabelle 3) sind 36,4% (bei Prävalenz 5%) bzw. 39,2% (bei Prävalenz 10%) screening-positiv und müssen einer genaueren diagnostischen Abklärung unterzogen werden. Für die längeren Fragebögen sind diese Werte günstiger und liegen zwischen 15% und 21% (bei Prävalenz 5%), bzw. zwischen 19% und 24% (bei Prävalenz 10%).

Auch hier zeigen sich für das Beck-Depression-Inventory (BDI) insgesamt die günstigsten Ergebnisse, gefolgt von der Geriatric Depression Scale (GDS-30) und der Center for Epidemiological Studies Depression Scale (CES-D).

Tabelle 3: Anteil Screening-Positiver von allen Untersuchten: Gegenüberstellung von international gebräuchlichen Screening-Instrumenten (Research Triangle Institute 2002) und der Kurz-Screening-Fragebögen (Abkürzungen der Screening-Instrumente siehe Anhang B)

Autor, Jahr	Screening-Instrument	Anteil Screening-Positiver	
		Prävalenz 5%	Prävalenz 10%
Gebräuchliche Screening-Instrumente			
Fechner-Bates et al 1994	CES-D >16	31,4	34,0
Hendrie et al 1995	CES-D >16	12,6	16,1
Zich et al 1990	CES-D >27	23,1	27,1
Klinkman et al 1997	CES-D >16	30,7	33,3
Lyness et al 1997	CES-D >21	17,0	20,9
Myers & Weissman 1980	CES-D >17	8,5	11,4
	Median	20,1	24,0
Holcomb et al 1996	BDI >16	14,6	18,2
Lustman et al 1997	BDI >13	15,7	19,3
Zich et al 1990	BDI >16	15,5	19,9
	Median	15,5	19,3
Lyness et al 1997	GDS-30: >10	20,2	24,4
van Marwijk et al 1995	GDS-30: >11	16,1	18,1
	Median	18,1	21,3
Leon et al 1996	SDDS-PC >2	15,9	18,8
Broadhead et al 1995	SDDS-PC >2	26,2	29,6
	Median	21,1	24,2
Upadhyaya & Stanley 1997	HAD-D: >9	15,9	18,7
Kurz-Screening-Fragebögen			
Spitzer et al. 1994	2 Fragen	28,1	31,1
Whooley et al. 1997	2 Fragen	45,7	48,3
Arroll et al. 2003	2 Fragen	36,3	39,4
Williams et al. 1999	1 Frage	36,6	39,1
	Median	36,4	39,2

2.1.2. Kriteriums-Validität – Deutschsprachige Studien

Es wird im Folgenden die Frage untersucht, inwieweit die neueren Kurz-Screening-Fragebögen mit 1 oder 2 Items mit den auf deutsch übersetzten umfassenderen Screening-Fragebögen vergleichbar sind.

Leider erfolgte bislang weder für den 2-Item Fragebogen noch für die beschriebene einzelne Screening-Frage eine Untersuchung in einer deutschsprachigen Stichprobe. Aus diesem Grund werden die Werte aus den englischsprachigen Untersuchungen verwendet. Wir konnten auch keine Studien über diese Kurz-Screening-Fragebögen aus einem anderen europäischen Land finden.

Die Suche nach deutschsprachigen Versionen von Screening-Instrumenten erfolgte einerseits mittels internationaler elektronischer Suchmaschinen (MEDLINE, EMBASE – d.h. Bücher und deutschsprachige Zeitschriften nicht erfasst) und andererseits mittels Durchsicht deutschsprachiger Übersichtsarbeiten. Außerdem wurden einige deutschsprachige Experten persönlich kontaktiert und um Informationen gebeten.

Insgesamt fand sich nur eine sehr kleine Zahl von Studien aus dem deutschen Sprachraum, die in Tabelle 4 wiedergegeben sind. In vorliegende Darstellung wurden alle Studien eingeschlossen, die auch depressive Störungen einschlossen. Da manche Studien die Ergebnisse nicht spezifisch für die Diagnose „Depression“ angaben, wurden diese aus den folgenden Auswertungen ausgeschlossen.

Mit Ausnahme von zwei Studien, die nur Frauen einschlossen, haben alle Studien sowohl Männer als auch Frauen eingeschlossen. Die Mittelwerte des Alters der untersuchten Stichproben weisen darauf hin, dass für Erwachsene ausreichende Informationen vorliegen. Keine dieser Studien hatte Kinder oder Jugendliche eingeschlossen. Bei mehreren Studien wurden auch ältere Personen (über 65-Jährige) eingeschlossen.

Die meisten Studien wurden bei Praktischen Ärzten durchgeführt, einige Studien wurden aber auch bei stationären Krankenhauspatienten durchgeführt.

Tabelle 4: Validitätsstudien von Screening-Instrumenten in deutscher Sprache, die auch depressive Störungen einschlossen

Autor, Jahr	Stichprobe	Geschlecht	Alter (Mittelwert)	Größe der Stichprobe	Untersuchte psychiatrische Diagnosen
Arolt 1997	Stationäre Patienten (Interne & Chirurgie)	W,M		224	Alle (inklusive Depression)
Bergant et al. 1998	Stationäre Patientinnen postpartal	W	28,6	110	Depression
Goldberg et al. 1997, 1998	Praktischer Arzt	W,M		Je 400	Alle (inklusive Depression)
Katschnig et al. 1980	Praktischer Arzt	W,M		77	Depression
Muzik et al. 2000	Ambulante Patientinnen postpartal	W	28	50	Major depression
Schmitz et al. 1999 a	Praktischer Arzt	W,M	42,7	421	Alle (inklusive Depression)
Schmitz et al. 1999 b	Praktischer Arzt	W,M	42,7	408	Alle (inklusive Depression)
Wancata et al. 2001	Stationäre Pat. (Interne, Chirurgie, Gynäkologie, Rehabilitation)	W,M	46,3	511	Depression, Angst, psychosomatische Störungen

In Tabelle 5 sind die Werte für Sensitivität und Spezifität dargestellt, die in den jeweiligen Studien berichtet wurden. Die Kennwerte Positive Predictive Value, Negative Predictive Value und Overall Misclassification Rate wurden wiederum für eine angenommene Prävalenz von 5% und 10% berechnet. Da dies die Ober- und Untergrenze der zumeist in der Literatur berichteten Prävalenz-Raten darstellt, kann angenommen werden, dass die real bei Praktischen Ärzten in Österreich zu beobachtenden Werte in diesem Bereich liegen.

Wenn man die Validitäts-Kennwerte der Kurz-Screening-Fragebögen mit jenen der umfassenderen deutschsprachigen Screening-Fragebögen vergleicht (Tabelle 5), ergibt sich folgendes Bild: der Median der Sensitivität der Kurz-Screening-Fragebögen liegt mit 91% nur knapp (4%) über dem Median der umfassenderen Screening-Fragebögen. Die Spezifität (Median) der Kurz-Screening-Fragebögen ist aber cirka 20% niedriger als die Spezifität (Median) der umfassenderen Instrumente. Die Kurz-Screening-Fragebögen klassifizieren also einen deutlich höheren Anteil der psychisch Gesunden fälschlich als depressiv, während die Fähigkeit depressive Patienten richtig zu erfassen ähnlich hoch wie bei den längeren Fragebögen ist.

Bei einer Prävalenz von 5% beträgt der Positive Predictive Value (Median) der Kurz-Screening-Fragebögen 12,5%, während dieser Wert für die längeren Fragebögen (Median) etwa doppelt so hoch ist. Auch bei einer Prävalenz von 10% ist Positive Predictive Value (Median) der längeren Fragebögen nahezu doppelt so hoch wie bei den Kurz-Screening-Fragebögen. Der Anteil der screening-positiven Personen, die wirklich unter einer Depression leiden, ist also bei Verwendung der umfassenderen Screening-Fragebögen deutlich höher.

Die Negative Predictive Values liegen für alle Instrumente über 97%, wobei sich die Kurz-Fragebögen nicht von den längeren Fragebögen unterscheiden. Die Kurz-Fragebögen führen insgesamt dazu, dass rund ein Drittel der Patienten falsch zugeordnet werden (Overall Misclassification Rate = OMR), während diese Fehlzuordnungen bei den längeren Fragebögen nur cirka 14% betragen.

Tabelle 5: Gegenüberstellung der Validitäts-Kennwerte von Screening-Instrumenten in deutscher Sprache, die ausschließlich depressive Störungen untersuchten, mit jenen der Kurz-Screening-Fragebögen (Abkürzungen der Screening-Instrumente siehe Anhang B) Sens = Sensitivität, Spez = Spezifität, PPV = Positive Predictive Value, NPV = Negative Predictive Value, OMR = Overall Misclassification Rate

Autor, Jahr	Screening-Instrument	Sens	Spez	Prävalenz 5%			Prävalenz 10%		
				PPV	NPV	OMR	PPV	NPV	OMR
Längere Screening-Instrumente									
Bergant et al. 1998	EPDS \geq 10	96,0	100,0	100,0	99,8	0,2	100,0	99,6	0,4
Katschnig et al. 1980	GHQ-30-B \geq 4	83,3	77,1	16,1	98,9	22,6	28,8	97,6	22,3
Muzik et al. 2000	EPDS \geq 11	87,0	87,0	26,0	99,2	13,0	42,6	98,4	13,0
	SDS \geq 50	89,0	77,0	16,9	99,3	22,4	30,1	98,4	21,8
	SCL-90-R \geq 63	78,0	87,0	24,0	98,7	13,5	40,0	97,3	13,9
	Median	87,0	87,0	24,0	99,2	13,5	40,0	98,4	13,9
Kurz-Screening-Fragebögen									
Spitzer et al. 1994	2 Fragen	86,0	75,0	15,3	99,0	24,5	27,7	98,0	23,9
Whooley et al. 1997	2 Fragen	96,0	57,0	10,5	99,6	41,1	19,9	99,2	39,1
Arroll et al. 2003	2 Fragen	97,0	67,0	13,4	99,8	31,5	24,6	99,5	30,0
Williams et al. 1999	1 Frage	85,0	66,0	11,6	98,8	33,1	21,7	97,5	32,1
	Median	91,0	66,5	12,5	99,3	32,3	23,1	98,6	31,0

Wie bereits erwähnt ist auch die Frage, wie hoch der Anteil Screening-Positiver von allen mit einem bestimmten Screening-Instrument Untersuchten ist, von praktischer Bedeutung.

Bei Verwendung eines Kurz-Screening-Fragebogens sind zwischen 36,4% und 39,2% (Median, in Abhängigkeit von der Prävalenz) screening-positiv und müssen einer genaueren diagnostischen Abklärung unterzogen werden (Tabelle 6). Für die längeren Fragebögen sind dieser Werte günstiger und liegen bei 17% (bei Prävalenz 5%), bzw. bei 20% (bei Prävalenz 10%).

Tabelle 6: Anteil Screening-Positiver von allen Untersuchten: Gegenüberstellung von Screening-Instrumenten in deutscher Sprache, die ausschließlich depressive Störungen untersuchten, und der Kurz-Screening-Fragebögen (Abkürzungen der Screening-Instrumente siehe Anhang B)

Autor, Jahr	Screening-Instrument	Anteil Screening-Positiver	
		Prävalenz 5%	Prävalenz 10%
Längere Screening-Instrumente			
Bergant et al. 1998	EPDS	4,8	9,6
Katschnig et al. 1980	GHQ-30-B	25,9	28,9
Muzik et al. 2000	EPDS	16,7	20,4
	SDS	26,3	29,6
	SCL-90-R	16,3	19,5
	Median	16,7	20,4
Kurz-Screening-Fragebögen			
Spitzer et al. 1994	2 Fragen	28,1	31,1
Whooley et al. 1997	2 Fragen	45,7	48,3
Arroll et al. 2003	2 Fragen	36,3	39,4
Williams et al. 1999	1 Frage	36,6	39,1
	Median	36,4	39,2

2.1.3. Zeitökonomische Anwendung von Depressions-Screening-Instrumenten

Um Screening-Instrumente im Rahmen der Vorsorgeuntersuchung einsetzen zu können, ist es erforderlich, dass sie in möglichst kurzer Zeit angewandt werden können. Die in Tabelle 7 beschriebenen Screening-Instrumente sind Fragebögen, die der Patient selbst ausfüllen kann.

Wie zu erwarten benötigen die Kurz-Fragebögen mit einem oder zwei Items die geringste Zeit. Aber auch alle anderen Fragebögen sind in wenigen Minuten auszufüllen.

Tabelle 7: Screening-Instrumente für Depression – zum Ausfüllen benötigte Zeit: die Angaben basierend überwiegend auf Mulrow et al. 1995, teilweise aber auch aus anderen Publikationen und eigenen Erfahrungen (Abkürzungen siehe Anhang B)

Instrument	Items	Benötigte Zeit (Minuten)
BDI	21	2-4
CES-D	20	2-4
SDDS-PC	5	<2
Zung SDS	20	2-4
GDS-30	30	3-5
GHQ-30	30	3-5
HAD-D	7	ca. 2
EPDS	10	ca. 2
SCL-90-R	13	2-4
Kurz-Fragebögen	1-2	<1

Bei der EPDS muss berücksichtigt werden, dass sie zur Erfassung von Depressionen speziell bei Frauen in der Zeit nach der Geburt entwickelt wurde. Der GHQ inkludiert zwar die Erfassung von Depressionen, ist aber nicht speziell dafür entwickelt worden. Alle anderen angeführten Screening-Instrumente (bzw. deren Subskalen) sollen ausschließlich Depressionen erfassen.

Auf Basis dieser Ergebnisse wäre zu überlegen, dass jeder Patient im Rahmen der Vorsorgeuntersuchung den Kurz-Screening-Fragebogen mit 2 Items ausfüllt. Diejenigen, die im Kurz-Screening positiv sind, werden gebeten einen der üblichen längeren Screening-Fragebögen auszufüllen. Nur diejenigen, die auch im längeren Screening positiv sind, werden vom Praktischen Arzt auf das Vorhandensein einer Depression untersucht.

Ob die Kombination eines Kurz-Screening-Fragebogens (Vorteil: wenig Zeit zum Ausfüllen) mit den Vorteilen eines gebräuchlichen längeren Screening-Instruments (Vorteil: bessere Validitäts-Kennwerte) insgesamt vorteilhaft ist, muss separat untersucht werden.

2.1.4. Kriteriums-Validität – Kombination aus Kurz-Screening gefolgt von längerem Screening-Instrument (basierend auf den angeführten internationalen Studien)

Wie erwähnt stellt sich die Frage, ob die Vorteile eines aus ein oder zwei Fragen bestehenden Kurz-Screening-Fragebogens (wenig Zeit zum Ausfüllen) mit den Vorteilen eines gebräuchlichen Screening-Instruments (bessere Validitäts-Kennwerte) kombiniert werden können.

Im Folgenden wird von dem Modell ausgegangen, dass jeder Patient im Rahmen der Vorsorgeuntersuchung den Kurz-Screening-Fragebogen mit 2 Items ausfüllt. Diejenigen, die im Kurz-Screening positiv sind, werden gebeten einen der üblichen längeren Screening-Fragebögen (BDI oder CES-D) auszufüllen. Nur diejenigen, die auch im längeren Screening positiv sind, werden vom Praktischen Arzt auf das Vorhandensein einer Depression untersucht.

Um die Vorteile und Nachteile eines solchen Vorgehens zu überprüfen, wurden die neuen resultierenden Validitäts-Kennwerte und die Rate jener, die in beiden Screening-Verfahren positiv sind (d.h. jene, die vom Praktischen Arzt genauer untersucht werden müssen) berechnet. In der Darstellung der Validitäts-Kennwerte der Kombination der beiden Screening-Verfahren wurden 2 Modelle berücksichtigt, da nicht mit Sicherheit ausgeschlossen werden kann, dass sich durch die Kombination der beiden Screening-Verfahren die Sensitivität oder die Spezifität etwas verschlechtern. (Für die folgenden Berechnungen wurden nur die oben errechneten Mediane verwendet.)

2.1.4.1. Verwendung der Originaldaten der beiden Screening-Verfahren:

In Tabelle 8 zeigt sich, dass bei Verwendung der Originaldaten der beiden Screening-Verfahren die Kombination aus Kurz-Screening gefolgt von einem längeren Screening-Verfahren zu deutlich besseren Werten für die Positive Predictive Values (PPV) führt. Diese Werte sind sowohl bei einer angenommenen Prävalenz von 5% als auch von 10% deutlich besser als bei Verwendung einzelner Screening-Verfahren (nur Kurz-Screening oder nur gebräuchliches Screening-Instrument). Auch die Zahl jener, die insgesamt falsch zugeordnet werden (OMR) ist deutlich niedriger als bei Verwendung einzelner Screening-Verfahren. Die Negative Predictive Values sind minimal geringer. Allerdings sinkt die „effektive Sensitivität“ (also die aus der Kombination der beiden Screening-Verfahren resultierende Sensitivität) um etwa 7-8%.

2.1.4.2. Hypothetische Annahme, dass sich die Sensitivität und die Spezifität des längeren Screening-Instrumentes etwas verschlechtert:

Da das längere Screening-Instrument nur bei jenen verwendet werden soll, die im kürzeren Screening-Verfahren positiv sind, kann nicht ausgeschlossen werden, dass sich die Sensitivität oder die Spezifität etwas verändern. Im Falle einer hypothetischen Verbesserung einer der beiden Werte (Sensitivität, Spezifität) würden sich die Ergebnisse insgesamt verbessern. Im Falle einer hypothetischen Verschlechterung einer der beiden Werte wären die Ergebnisse schlechter als bei Verwendung der Originaldaten.

Im Folgenden (Tabelle 8) wird hypothetisch angenommen, dass sich sowohl die Sensitivität als auch die Spezifität des längeren Screening-Instrumentes um 5% verschlechtern. Für die Kombination aus Kurz-Screening gefolgt von einem längeren Screening-Verfahren sind die Positive Predictive Values (PPV) noch immer deutlich besser als bei Verwendung einzelner Screening-Verfahren (nur Kurz-Screening oder nur gebräuchliches Screening-Instrument). Auch die Zahl jener, die insgesamt falsch zugeordnet werden (OMR) ist deutlich niedriger als bei Verwendung einzelner Screening-Verfahren. Die Negative Predictive Values würden gering abnehmen. Allerdings sinkt die „effektive Sensitivität“ um etwa 12% im Vergleich zur ausschließlichen Verwendung eines üblichen längeren Screening-Instrumentes.

Tabelle 8: Gegenüberstellung der Validitäts-Kennwerte (Median) der international gebräuchlichen Screening-Instrumente und der Kurz-Screening-Fragebögen mit jenen eines kombinierten Screenings (Kurz-Screening-Fragebogen gefolgt von gebräuchlichen Screening-Instrumenten; Abkürzungen der Screening-Instrumente siehe Anhang B) Sens = Sensitivität, Spez = Spezifität, PPV = Positive Predictive Value, NPV = Negative Predictive Value, OMR = Overall Misclassification Rate, Eff-Sens = Effektive Sensitivität (aus der Kombination der beiden Screening-Verfahren resultierende Sensitivität), Effektive Spezifität (aus der Kombination der beiden Screening-Verfahren resultierende Spezifität)

	Sens	Spez	Prävalenz 5%			Prävalenz 10%		
			PPV	NPV	OMR	PPV	NPV	OMR
Kurz-Screening-Fragebögen								
Median	91,0	66,5	12,5	99,3	32,3	23,1	98,6	31,0
Gebräuchliche Screening-Instrumente								
CES-D Median	80,5	84,0	24,4	98,7	15,4	40,5	97,4	14,8
BDI Median	85,0	89,0	28,4	99,1	11,3	45,6	98,1	11,6
	Eff-Sens	Eff-Spez						
Kombination von Kurz-Screening mit längeren Fragebögen: Originaldaten								
Kurz-Screening plus CES-D (Median)	73,3	94,6	41,8	98,5	6,4	60,3	97,0	7,5
Kurz-Screening plus BDI (Median)	77,4	96,3	52,5	98,8	4,6	70,0	97,5	5,6
Kombination von Kurz-Screening mit längeren Fragebögen: Hypothetische Verminderung der Sensitivität und Spezifität um jeweils 5%								
Kurz-Screening plus CES-D (Median)	68,7	93,0	34,0	98,3	8,2	52,0	96,4	9,5
Kurz-Screening plus BDI (Median)	72,8	94,6	41,7	98,5	6,5	60,1	96,9	7,5

Diese Frage, wie hoch der Anteil von Personen ist, die vom Praktischen Arzt untersucht werden müssen, um abzuklären, ob wirklich eine Depression vorliegt, ist von großer praktischer Relevanz.

Bei Verwendung der Kombination aus Kurz-Screening gefolgt von einem längeren Screening-Verfahren wären deutlich weniger Personen vom Praktischen Arzt zu untersuchen als bei ausschließlicher Verwendung eines Kurz-Screening-Fragebogens oder ausschließlicher Verwendung eines gebräuchlichen Screening-Instrumentes (Tabelle 9). In Abhängigkeit von der Prävalenz und vom ausgewählten zweiten Screening-Instrument wären nur zwischen 7 und 12% vom Praktischen Arzt zu untersuchen.

Tabelle 9: Anteil jener, die bei der Kombination von Kurz-Screening mit einem längeren Fragebogen zu untersuchen ist, im Vergleich zur alleinigen Verwendung international gebräuchlicher Screening-Instrumente oder der Kurz-Screening-Fragebögen (Abkürzungen der Screening-Instrumente siehe Anhang B)

Autor, Jahr	Anteil Screening-Positiver	
	Prävalenz 5%	Prävalenz 10%
Kurz-Screening-Fragebögen		
Median	36,4	39,2
Gebräuchliche Screening-Instrumente		
CES-D Median	20,1	24,0
BDI Median	15,5	19,3
Kombination von Kurz-Screening mit längeren Fragebögen (Originaldaten)		
Kurz-Screening plus CES-D (Median)	8,8	12,2
Kurz-Screening plus BDI (Median)	7,4	11,1

Insgesamt bringt die Verwendung eines Kombinationsmodells (jeder Patient, der beim Kurz-Screening-Fragebogen positiv ist, wird mit einem längeren Screening-Fragebögen untersucht) einige Vorteile. Die Rate jener, die insgesamt falsch zugeordnet werden, und die Zahl jener, die vom Praktischen Arzt genauer untersucht werden müssen, sind deutlich geringer. (Leider liegen für die Kombination der genannten Screening-Verfahren bislang keine empirischen Daten vor, sodass nicht mit Sicherheit abgeschätzt werden kann, ob und wie sich die Validitäts-Kennwerte des längeren Screening-Fragebogens durch dieses Vorgehen verändern.)

Man darf auch nicht übersehen, dass durch die Kombination von zwei unterschiedlichen Screening-Verfahren die „effektive Sensitivität“ abnimmt. Die „effektive Sensitivität“ könnte verbessert werden, wenn beim längeren Screening-Fragebogen der Cut-off so verändert wird, dass die Sensitivität zunimmt und die Spezifität abnimmt (was voraussichtlich zu einer gering höheren Zahl jener führen würde, die der Praktische Arzt untersuchen muss). Dieses Vorgehen würde aber die Durchführung einer Studie zu speziell dieser Fragestellung erfordern.

In Summe kann man auf Basis der verfügbaren Daten annehmen, dass die Kombination von zwei unterschiedlichen Screening-Verfahren Vorteile im Vergleich zur Verwendung einzelner Screening-Fragebögen hat.

2.1.5. Für welche Personengruppen liegen ausreichend Daten vor, die ein Screening auf Depression rechtfertigen?

Aus der Literatur gibt es Hinweise darauf, dass die Ergebnisse über die Effektivität von Screening, die in einer bestimmten Patientengruppe gewonnen wurden, nicht automatisch auf andere Patientengruppen übertragbar sind. So können beispielsweise die Ergebnisse von Screening, die Studien an ambulanten Patienten gewonnen wurden, nicht automatisch

auf stationär aufgenommene Krankenhauspatienten übertragen werden. Es gibt auch Hinweise, dass derartige Ergebnisse nicht problemlos auf andere Altersgruppen bzw. das andere Geschlecht übertragen werden können.

Aus diesem Grund wurden alle vorliegenden randomisierten, kontrollierten Studien bezüglich der Stichproben-Zusammensetzung ausgewertet (Tabelle 10).

Tabelle 10: Welche Personengruppen wurden in Studien eingeschlossen, die den Effekt von Screening bei Depressionen im Primary Care Setting untersuchten (-- meint: keine eindeutigen Angaben)

Autor, Jahr	Fachrichtung ¹	Alter (in Jahren)	Geschlecht	Ausgeschlossene Personen ²
Callahan et al. 1994	PA	60+	Männer + Frauen	<ul style="list-style-type: none"> • Sprachschwierigkeiten • Schwerhörigkeit • Suizidalität
Callahan et al. 1996	PA	60+	Männer + Frauen	--
German et al. 1987	PA	18 +	Männer + Frauen	--
Hoeper et al. 1984	PA, Internisten	--	Männer + Frauen	--
Johnstone & Goldberg 1976	PA	16 +	--	<ul style="list-style-type: none"> • Schwerhörigkeit • Sehbehinderung
Lewis et al. 1996	PA	18-70	--	<ul style="list-style-type: none"> • Sprachschwierigkeiten • zu schwer krank
Rand et al. 1988	PA	18 +	Männer + Frauen	<ul style="list-style-type: none"> • Sprachschwierigkeiten • zu schwer krank • geistig behindert
Zung & King 1983	PA	20 +	--	<ul style="list-style-type: none"> • Geistig behindert
Reifler et al. 1996	Internisten	--	Männer + Frauen	<ul style="list-style-type: none"> • Sprachschwierigkeiten • Sehbehinderung • geistig behindert • Leseschwäche
Williams et al. 1999	PA, Internisten	--	--	<ul style="list-style-type: none"> • Obdachlose
Linn & Yager 1980	Internisten	--	Männer + Frauen	<ul style="list-style-type: none"> • Sprachschwierigkeiten
Moore et al. 1978	PA	20-60	--	--
Dowrick 1995 b	PA	16-64	Männer + Frauen	<ul style="list-style-type: none"> • Suizidalität
Dowrick 1995 a	PA	16-64	Männer + Frauen	<ul style="list-style-type: none"> • Suizidalität
Mazonson et al. 1996	PA, Internisten	21-65	Männer + Frauen	
Magruder-Habib et al. 1990	PA	-90	Männer	<ul style="list-style-type: none"> • Demenz • Schwerhörigkeit • Sehbehinderung
Wells et al. 2000	PA	18+	Männer + Frauen	<ul style="list-style-type: none"> • Sprachschwierigkeiten • medizinische Notfälle

¹ PA = Praktische Ärzte

² Sprachschwierigkeiten: meint Patienten, die jene Sprache, in der die jeweilige Studie durchgeführt wurde, nicht ausreichend beherrschten

Mit Ausnahme von 2 Studien fanden alle Studien bei Praktischen Ärzten statt, wobei einige Studien auch Internisten eingeschlossen hatten, die in der medizinischen Primärversorgung tätig waren.

Die meisten Studien hatten sowohl Männer als auch Frauen eingeschlossen, eine einzige Studie hatte ausschließlich Männer eingeschlossen. Das in die Studien eingeschlossene Spektrum der Altersgruppen zeigt, dass für Personen über 20 Jahren ausreichend Daten vorliegen. Zahlreiche Studien hatte auch ältere Patienten (über 65-Jährige) eingeschlossen sechs Studien hatte unter anderem ältere Personen eingeschlossen, zwei Studien hatten gezielt Personen über 60 Jahren untersucht. Eine Studie hatte Höchstaltrige (über 90-Jährige) ausgeschlossen.

Die folgenden Personengruppen waren aus einzelnen oder mehreren Studien ausgeschlossen worden: Personen, die die jeweilige Sprache nicht ausreichend beherrschten, Personen mit ausgeprägten Sehbehinderungen oder Schwerhörigkeit, Personen mit ausgeprägten kognitiven Defiziten (durch eine geistige Behinderung oder eine Demenz verursacht), Personen mit Schreib- bzw. Leseschwächen, obdachlose Personen. Weiters waren medizinische Notfälle und körperlich schwer Kranke ausgeschlossen worden. Auch Personen, die Hinweise auf Suizidalität zeigten, waren nicht untersucht worden.

Zusammenfassend kann gesagt werden, dass ausreichend Daten für die Patienten von Praktischen Ärzten beiderlei Geschlechts und aller Altersgruppen über 20 Jahren vorliegen. Dies wird auch durch die Zusammensetzung der Stichproben der bereits beschriebenen Validitätsstudien (Tabellen 1 und 4) bestätigt. Aus praktischen Gründen scheint es nicht sinnvoll, Patienten mit ausgeprägten kognitiven Defiziten und Personen, die aufgrund von Sprachschwierigkeiten Probleme haben, den Fragebogen zu verstehen, einzuschließen. Auch Personen mit Schreib- bzw. Leseschwächen werden Probleme haben, einen Fragebogen auszufüllen. Für Personen mit ausgeprägten Sehbehinderungen wird es notwendig sein, spezielle Fragebögen (große Schrift) zu entwickeln oder die einzelnen Fragen dem Patienten vorzulesen. Es ist verständlich, dass bei prospektiven Studien (aufgrund von forschungs-organisatorischen Überlegungen) obdachlose Personen ausgeschlossen wurden, es gibt aber aus unserer Sicht keinen Grund, diese Personen von Vorsorgeuntersuchungen auszuschließen. Bei schwerst körperlich Kranken und medizinischen Notfällen stehen andere Probleme im Vordergrund, sodass es sinnvoll scheint, die Vorsorgeuntersuchung aufzuschieben. Patienten, die Hinweise auf Suizidalität zeigen, sind als Notfälle zu betrachten, die einer Weiterleitung zu einer Fachkraft erfordern.

2.2. Die potenziell schädlichen Folgen von Screening

Es stellt sich die Frage, ob ein Screening auf depressive Erkrankungen potenziell negative Folgen haben oder Schäden verursachen kann. Zu diesem Zweck wurde die wissenschaftliche Literatur umfassend bearbeitet.

2.2.1. Vermeidung der theoretisch möglichen Probleme bei Screening-Positiven

Wenn der praktische Arzt aufgrund zu geringen eigenen Wissens über psychische Erkrankungen dem Screening-Test blind vertraut und die Screening-Ergebnisse mit einer psychiatrischen Diagnose gleichsetzt, könnte dies dazu führen, dass alle Screening-Positiven einer Behandlung mit Psychopharmaka und/oder Psychotherapie zugeführt werden. Auch diejenigen, die fälschlicherweise als krank klassifiziert wurden (falsch Positive), würden dann eine Behandlung erhalten – auch mit dem Risiko der entsprechenden Nebenwirkungen. Weiters könnte es sein, dass manche Symptome (z.B. Energielosigkeit, diffuse Schmerzen, Appetitlosigkeit) fälschlicherweise der psychischen Erkrankung

zugeordnet würden und eine eventuell vorhandene körperliche Krankheit übersehen würde (Ford 1988). Außerdem würde die bei psychisch Gesunden durchgeführte psychiatrische Behandlung unnötige Kosten verursachen und auch die psychisch Gesunden würden unnötigerweise dem Risiko einer Stigmatisierung aufgrund psychischer Krankheit ausgesetzt [US Preventive Services Task Force 2002, Wancata 2002].

Um zu überprüfen, ob der Patient wirklich an einer Depression leidet, ist also unbedingt eine weitere Untersuchung durch den Praktischen Arzt erforderlich.

Um eine unnötige Stigmatisierung der screening-positiven Patienten bzw. deren Belastung durch etwaige Befürchtungen zu vermeiden, ist es unbedingt nötig zuerst abzuklären, ob eine Depression vorliegt, bevor der Patient mit der Diagnose konfrontiert wird oder an spezialisierte Berufsgruppen (z.B. Psychiater, Psychotherapeuten, Psychologen) überwiesen wird. Die diagnostische Abklärung muss also – von einzelnen Ausnahmen abgesehen (z.B. Suizidalität, Auftreten von Komplikationen, Nichtansprechen auf Therapie) – durch den Praktischen Arzt erfolgen.

2.2.2. Hinweise auf potenziell negativen Folgen von Screening aus randomisiert-kontrollierten Studien

Die US Preventive Services Task Force (2002) berichtet, dass die aktuelle Datenlage bezüglich negativer Folgen eines Depressions-Screenings ungenügend sei. Da aus den vorliegenden Studien keinerlei explizite Hinweise auf etwaige negative Folgen eines Screenings auf Depression vorliegen, nahmen die Autoren an, dass Depressions-Screening keine negativen Folgen habe.

Trotz einer umfangreichen Literatursuche und Bearbeitung der Literatur konnten wir keine Studien finden, die die Frage, ob Depressions-Screening negative Folgen haben könne, *explizit* untersuchten. Um zu überprüfen, ob sich *indirekte* Hinweise auf negative Folgen für die Patienten durch das Depressions-Screening finden lassen, wurden die randomisiert-kontrollierten Studien, die die Wirksamkeit von Depressions-Screening untersuchten, einer weiteren Auswertung unterzogen (Tabelle 11).

Es ist anzunehmen, dass Patienten, die sich durch das Screening bzw. die anschließenden psychiatrischen Untersuchungen und die Behandlung belastet fühlen, eine geringere Behandlungszufriedenheit zeigen. In zwei Studien (Reifler et al. 1996, Williams et al. 1999) wurde die Behandlungszufriedenheit untersucht. Beide Studien fanden keine signifikanten Unterschiede zwischen jenen Patienten, die am Depressions-Screening teilnahmen, und jenen, die nicht daran teilnahmen. Dies kann als Indikator gewertet werden, dass zumindest aus subjektiver Sicht der Patienten keine relevanten negativen Folgen auftraten.

Tabelle 11: Randomisiert-kontrollierte Studien: Gibt es Hinweise auf unerwünschte Effekte des Screenings?

Autor, Jahr	Screening-Instrument	Unerwünschte Effekte des Screenings
Callahan et al. 1994	CES-D	--
Callahan et al. 1996	CES-D	--
German et al. 1987	GHQ-28	--
Hoeper et al. 1984	GHQ-28	--
Johnstone & Goldberg 1976	GHQ-60	--
Lewis et al. 1996	GHQ-12; PROQSY	--
Rand et al. 1988	GHQ-28	--
Zung & King 1983	SDS	--
Reifler et al. 1996	SDDS-PC	Kein Unterschied bez. Behandlungszufriedenheit
Williams et al. 1999	1 Frage; CES-D	Kein Unterschied bez. Behandlungszufriedenheit
Linn & Yager 1980	SDS	--
Moore et al. 1978	SDS	--
Dowrick 1995 b	BDI	--
Dowrick 1995 a	BDI	--
Mazonson et al. 1996	SCL-90-R	--
Magruder-Habib et al. 1990	SDS	--
Wells et al. 2000	CES-D	--

Es ist anzunehmen, dass Patienten, die sich durch das Screening bzw. die anschließenden psychiatrischen Untersuchungen und die Behandlung belastet fühlen, dazu neigen, die Teilnahme an der Studie abbrechen. (Auch in vielen Therapie-Studien wird die Rate der Abbrüche als Indikator für negative Effekte der Behandlung gewertet.) Aus zwölf randomisierten Studien lagen keine ausreichenden Informationen über die Raten an Studienabbrechern vor (Tabelle 12). Reifler et al. (1996) berichteten, dass die Anteile der Studienabbrecher ähnlich gewesen seien, ohne jedoch detaillierte Zahlen zu nennen. Aus vier randomisierten Studien lagen ausreichend Daten vor, um weitere Auswertungen über unterschiedliche Raten an Studienabbrechern durchzuführen.

Tabelle 12: Randomisiert-kontrollierte Studien: Welche Informationen liegen vor, die den Vergleich der Ausfälle (= Studienabbrecher) zwischen Interventionsgruppe (Screening durchgeführt) und Kontrollgruppe zulassen

Autor, Jahr	Screening-Instrument	Ausfälle in der Interventionsgruppe	Ausfälle in der Kontrollgruppe
Callahan et al. 1994	CES-D	--	--
Callahan et al. 1996	CES-D	--	--
German et al. 1987	GHQ-28	--	--
Hoeper et al. 1984	GHQ-28	--	--
Johnstone & Goldberg 1976	GHQ-60	--	--
Lewis et al. 1996	GHQ, PROQSY	Zahlen	Zahlen
Rand et al. 1988	GHQ-28	--	--
Zung & King 1983	SDS	Zahlen	Zahlen
Reifler et al. 1996	SDDS-PC	„similar“	„similar“
Williams et al. 1999	1 Frage, CES-D	Zahlen	Zahlen
Linn & Yager 1980	SDS	--	--
Moore et al. 1978	SDS	--	--
Dowrick 1995 b	BDI	--	--
Dowrick 1995 a	BDI	--	--
Mazonson et al. 1996	SCL-90-R	--	--
Magruder-Habib et al. 1990	SDS	--	--
Wells et al. 2000	CES-D	Zahlen	Zahlen

Die aus diesen vier Studien vorliegenden Daten wurden verwendet, um die Raten an Studienabbrechern zu berechnen (Tabelle 13). Die Raten der Studienabbrecher der Interventionsgruppe (Depressions-Screening durchgeführt) und der Kontrollgruppe wurden für die einzelnen Studien und die verschiedenen Untersuchungszeitpunkte berechnet und verglichen. In der Studie von Lewis et al. (1996) fand sich nach 3 Monaten bei jenen Patienten, die mittels GHQ gescreent worden waren, eine signifikant höhere Ausfallsrate (41%) als in der Kontrollgruppe (30,9%). Zu den anderen Untersuchungszeitpunkten (6 Wochen, 6 Monate) fanden sich keine signifikanten Unterschiede. Bei jener Gruppe, die zusätzlich zum GHQ auch mittels PROQSY gescreent worden war (Lewis et al. 1996), fanden sich keine signifikanten Unterschiede zur Kontrollgruppe. In den anderen 3 Studien (Zung et al. 1983, Williams et al. 1999, Wells et al. 2000) fanden sich zu keinem Zeitpunkt signifikante Unterschiede. In den Studien von Zung et al (1983) und von Williams et al. (1999) waren die Raten der Studienabbrecher in der Gruppe, die am Screening teilnahm, sogar etwas geringer als an der Kontrollgruppe, diese Unterschiede erreichten jedoch nicht das Signifikanzniveau.

Zusammenfassend kann also gesagt werden, dass bei einem einzigen Untersuchungszeitpunkt die Ausfälle in der Screening-Gruppe signifikant häufiger waren, bei 10 Untersuchungszeitpunkten jedoch keine signifikanten Unterschiede gefunden werden konnten. Dies kann als Indikator gewertet werden, dass Depressions-Screening in diesen Studien keine bzw. vernachlässigbare negative Folgen hatte.

Tabelle 13: Randomisiert-kontrollierte Studien: Statistisch signifikante Unterschiede der Ausfälle (= Studienabbrecher) zwischen Interventionsgruppe (Screening durchgeführt) und Kontrollgruppe (nur Studien mit ausreichender Datenbasis angeführt)

Autor, Jahr	Instrument		Ausfälle in der Interventionsgruppe	Ausfälle in der Kontrollgruppe	Unterschiede bezüglich der Ausfälle (Z-Test)
Lewis et al. 1996	GHQ GHQ + PROQSY	6 Wochen	31,3	30,8	n.s.
		3 Monate	41,0	30,9	p < 0,05
		6 Monate	40,5	40,5	n.s.
		6 Wochen	29,1	30,8	n.s.
		3 Monate	36,1	30,9	n.s.
		6 Monate	41,4	40,5	n.s.
Zung & King 1983	SDS	4 Wochen	8,7	23,1	n.s.
Williams et al. 1999	1 Frage	3 Monate	1,5	2,2	n.s.
	CES-D	3 Monate	0,6	2,2	n.s.
Wells et al. 2000	CES-D	6 Monate	15,4	12,9	n.s.
		12 Monate	17,3	15,3	n.s.

2.2.3. Screening-Negative

Wenn das Screening-Instrument eine vorhandene Depression nicht erfasst (falsch Negative), kann dies zur Folge haben, dass die Depression nicht einer entsprechenden Behandlung zugeführt wird. Außerdem könnte dies den Praktischen Arzt dazu verleiten zahlreiche somatische Untersuchungen und Tests durchzuführen, um eine körperliche Erkrankung zu suchen, die nicht vorhanden ist.

Die bei allen Instrumenten gefundenen hohen Werte für den NPV weisen darauf hin, dass die Zahl der falsch Negativen sehr klein ist. Für die meisten Studienergebnisse liegt der Anteil der falsch Negativen von allen Screening-Negativen bei etwa 1 – 3 %. Dies trifft sowohl für die international gebräuchlichen umfassenderen Screening-Fragebögen (Tabelle 14) als auch die Kurz-Screening-Fragebögen zu. Auch die deutschsprachigen Versionen von Screening-Instrumenten liegen in diesem Bereich (Tabelle 15).

Tabelle 14: Anteil der falsch Negativen von allen Screening-Negativen: Validitätsstudien von international gebräuchlichen Screening-Instrumenten (Research Triangle Institute 2002) und der Kurz-Screening-Fragebögen (Abkürzungen der Screening-Instrumente siehe Anhang B)

Autor, Jahr	Screening-Instrument	Anteil falsch Negativer von allen Screening-Negativen	
		Prävalenz 5%	Prävalenz 10%
Gebräuchliche Screening-Instrumente			
Fechner-Bates et al 1994	CES-D >16	1,5	3,1
Hendrie et al 1995	CES-D >16	1,1	2,4
Zich et al 1990	CES-D >27	0,0	0,0
Klinkman et al 1997	CES-D >16	1,4	2,8
Lyness et al 1997	CES-D >21	0,5	1,0
Myers & Weissman 1980	CES-D >17	2,0	4,1
	Median	1,3	2,6
Holcomb et al 1996	BDI >16	1,0	2,1
Lustman et al 1997	BDI >13	0,9	1,9
Zich et al 1990	BDI >16	0,0	0,0
	Median	0,9	1,9
Lyness et al 1997	GDS-30: >10	0,0	0,0
van Marwijk et al 1995	GDS-30: >11	2,7	5,5
	Median	1,4	2,8
Leon et al 1996	SDDS-PC >2	1,7	3,6
Broadhead et al 1995	SDDS-PC >2	0,7	1,4
	Median	1,2	2,5
Upadhyaya & Stanley 1997	HAD-D: >9	1,8	3,7
Kurz-Screening-Fragebögen			
Spitzer et al. 1994	2 Fragen	1,0	2,0
Whooley et al. 1997	2 Fragen	0,4	0,8
Arroll et al. 2003	2 Fragen	0,2	0,5
Williams et al. 1999	1 Frage	1,2	2,5
	Median	0,7	1,4

Tabelle 15: Anteil der falsch Negativen von allen Screening-Negativen: Validitätsstudien von Screening-Instrumenten in deutscher Sprache, die ausschließlich depressive Störungen untersuchten

Autor, Jahr	Screening-Instrument	Anteil falsch Negativer von allen Screening-Negativen	
		Prävalenz 5%	Prävalenz 10%
Bergant et al. 1998	EPDS	0,2	0,4
Katschnig et al. 1980	GHQ-30-B	1,1	2,4
Muzik et al. 2000	EPDS	0,8	1,6
	SDS	0,7	1,6
	SCL-90-R	1,3	2,7

Bei Verwendung der Kombination aus Kurz-Screening gefolgt von einem längeren Screening-Verfahren ist die Zahl der falsch Negativen ebenfalls sehr klein (Tabelle 16). Die Zahl der falsch Negativen liegt bei der Kombination aus Kurz-Screening gefolgt von einem längeren Screening-Verfahren um etwa 0,5% bis 2% über jener bei Verwendung von einzelnen Screening-Verfahren. Der Praktische Arzt kann also im Normalfall bei allen Varianten davon ausgehen, dass jemand, der im Screening negativ ist, auch wirklich keine Depression hat. Die Vorgangsweise bei einzelnen Fällen, bei denen der Praktische Arzt unsicher ist, ob sie nicht doch unter einer Depression leiden, ist bei der Vorgangsweise bezüglich Überprüfung des screening-positiven Resultates beschrieben.

Tabelle 16: Anteil der falsch Negativen von allen Screening-Negativen: Gegenüberstellung der Validitäts-Kennwerte (Median) der international gebräuchlichen Screening-Instrumente und der Kurz-Screening-Fragebögen mit jenen eines kombinierten Screenings (Kurz-Screening-Fragebogen gefolgt von gebräuchlichen Screening-Instrumenten)

	Anteil falsch Negativer von allen Screening-Negativen	
	Prävalenz 5%	Prävalenz 10%
Kurz-Screening-Fragebögen		
Median	0,7	1,4
Gebräuchliche Screening-Instrumente		
CES-D Median	1,3	2,6
BDI Median	0,9	1,9
Kombination von Kurz-Screening mit längeren Fragebögen:		
Originaldaten		
Kurz-Screening plus CES-D (Median)	1,5	3,0
Kurz-Screening plus BDI (Median)	1,2	2,5
Kombination von Kurz-Screening mit längeren Fragebögen:		
Hypothetische Verminderung der Sensitivität und Spezifität um jeweils 5%		
Kurz-Screening plus CES-D (Median)	1,7	3,6
Kurz-Screening plus BDI (Median)	1,5	3,1

2.3. Überprüfung des Screening-Resultates

Da, wie erwähnt, Screening-Instrumente nur einen Hinweis auf eine erhöhte Wahrscheinlichkeit für das Vorhandensein bzw. die Abwesenheit einer Erkrankung geben, ist es notwendig in einem 2. Schritt abzuklären, ob wirklich eine Depression vorliegt. Die teilweise recht geringen PPV-Werte zeigen, dass dieser 2. Schritt von größter Bedeutung ist.

2.3.1. Vorgangsweise bei Screening-Positiven

Da psychische Erkrankungen und deren behandelnde Berufsgruppen (z.B. Psychiater, Psychotherapeuten, Psychologen) noch immer von Stigma betroffen sind, besteht das Risiko, dass manche Patienten nur schwer akzeptieren können von einer psychischen Erkrankung betroffen zu sein und deshalb Behandlung zu benötigen. Es ist daher nötig, dieses Problem bei der Planung der weiteren Vorgehensweise für screening-positive Patienten zu berücksichtigen. Wenn die Abklärung der Frage, ob wirklich eine Depression vorliegt, durch den Praktischen Arzt selbst erfolgt, wird es häufig gelingen, das Problem einer potenziellen Stigmatisierung von falsch Positiven zu vermeiden. Die diagnostische Abklärung muss also durch den Praktischen Arzt erfolgen. (Nur jene Patienten, wo trotz umfangreicher Bemühungen des Praktischen Arztes eine eindeutige Diagnostik nicht gelingt, sollten an spezialisierte Berufsgruppen weiterverwiesen werden.)

Für die Durchführung der primären Diagnostik und Behandlung durch den Praktischen Arzt sprechen noch zwei weitere Argumente:

1. Unabhängig vom ausgewählten Screening-Instrument wird eine beträchtliche Zahl von Personen ein screening-positives Resultat zeigen, die alle einer diagnostische Abklärung bedürfen. Eine Weiterleitung all dieser Personen an spezialisierte Berufsgruppen wäre in jedem Fall mit zusätzlichen Kosten verbunden, die häufig vermeidbar sind.
2. Studien weisen darauf hin, dass viele Patienten oft über lange Zeit beim selben Praktischen Arzt in Behandlung stehen. Durch diese langjährige Behandlung entwickelt sich häufig ein ausgeprägtes Vertrauensverhältnis zum Praktischen Arzt, das eine wichtige Grundlage für die Behandlung psychischer Erkrankungen ist. Durch die langjährige Behandlung kennt der Praktische Arzt oft die Geschichte seiner Patienten sehr gut und kann auf dieses Wissen (über frühere Probleme, Belastungen und Krankheiten) auch für die Diagnostik der Depression zurückgreifen (Zintl-Wiegand et al. 1980).

Um die Frage zu klären, ob aufgrund der wissenschaftlichen Literatur bei Screening-Positiven einer ganz bestimmten Vorgangsweise gegenüber anderen Verfahren der Vorzug zu geben sei, wurden alle randomisierten kontrollierten Studien, die als ersten Schritt ein Screening auf Depression durchgeführt hatten, durchgesehen.

Die US Preventive Services Task Force (2002) kam aufgrund ihrer Meta-Analyse zu dem Schluss, dass Studien, die in einem 2. Schritt zur Ergänzung des Screenings Schulungsprogramme oder praktische Empfehlungen für Praktische Ärzte durchführten, bessere Resultate erzielten. Sie gaben jedoch keine genauen Empfehlungen bezüglich der Art des 2. Schrittes zur Überprüfung des Screening-Resultates ab. In Tabelle 17 ist beschrieben, welche Interventionen durchgeführt worden waren, um den Hinweis auf das Vorhandensein einer Depression zu bestätigen. Bei diesen Interventionen handelte es sich im Wesentlichen um die Weitergabe allgemeiner Richtlinien zur Diagnostik und Behandlung von depressiven Erkrankungen und um strukturierte Interviews zur Diagnose einer Depression. Wesentlich erscheint weiters, dass Patienten mit Suizid-Gedanken sofort zu einem Facharzt weitergeleitet wurden. Auch regelmäßige Kontakte mit dem Patienten und die Information des Patienten über seine Erkrankung scheinen empfehlenswert.

Tabelle 17: Was wurde in den randomisiert-kontrollierten Studien bei den Screening-Positiven in einem 2. Schritt durchgeführt, um den Hinweis auf das Vorhandensein einer Depression zu bestätigen (-- meint: diese Studie beinhaltet keine Untersuchung des 2. Schrittes)

Autor, Jahr	Screening-Instrument	Was wurde bei Screening-Positiven in einem 2. Schritt durchgeführt
Callahan et al. 1994	CES-D	<ul style="list-style-type: none"> • Info-Blatt für Arzt (Med. die Depression verursachen können) • Info-Blatt über Depression für Patienten • allg. Richtlinien bez. Behandlung) • alle Pat. mit Hinweis auf Suizid-Gedanken wurden sofort zu Facharzt transferiert
Callahan et al. 1996	CES-D	<ul style="list-style-type: none"> • Info-Blatt für Arzt (Med. die Depression verursachen können) • Info-Blatt über Depression für Patienten • allg. Richtlinien bez. Behandlung) • alle Pat. mit Hinweis auf Suizid-Gedanken wurden sofort zu Facharzt transferiert
German et al. 1987	GHQ-28	--
Hoeper et al. 1984	GHQ-28	--
Johnstone & Goldberg 1976	GHQ-60	--
Lewis et al. 1996	GHQ-12	PROQSY (PC-gestütztes Interview zur Diagnose psychischer Erkrankungen): besseres Ergebnis als Kontrollgruppe
Rand et al. 1988	GHQ-28	--
Zung & King 1983	SDS	Behandlung mit dem Psychopharmakon Alprazolam (wirksam im Vergleich zur Kontrollgruppe)
Reifler et al. 1996	SDDS-PC self-rating	Strukturiertes psychiatrisches Interview (SDDS-PC-2nd-stage-module)
Williams et al. 1999	1 Frage; CES-D	Kopie allgemeiner Richtlinien zur Behandlung der Depression
Linn & Yager 1980	SDS	--
Moore et al. 1978	SDS	--
Dowrick 1995 b	BDI	--
Dowrick 1995 a	BDI	--
Mazonson et al. 1996	SCL-90-R	Allgemeine Information für Arzt über: psychiatrische Dienste und Behandlung
Magruder-Habib et al. 1990	SDS	--
Wells et al. 2000	CES-D	Schulung des Arztes und Ordinationshilfe, Feedback an Arzt bez. Behandlung; monatliche Kontakte der Ordinationshilfe mit Pat. bez. Einnahme Medikation

Aus den Ergebnissen dieser randomisierten kontrollierten Studien allein ist es nicht möglich bestimmten Interventionen den Vorzug über anderen zu geben. Mehrere internationale und nationale Expertengruppen haben jedoch auf Basis einer umfassenden Bearbeitung der wissenschaftlichen Literatur Richtlinien zur Behandlung depressiver Krankheiten ausgearbeitet, wobei einige speziell die Situation beim Praktischen Arzt berücksichtigten. Diese Richtlinien und Empfehlungen von Expertengruppen sollen in Form eines Praxis-Manuals für Praktische Ärzte in Österreich zur Verfügung stehen.

2.3.2. Psychiatrische Interviews für Praktische Ärzte

In den letzten Jahren wurden zwei unterschiedliche psychiatrische Interviews speziell für Praktische Ärzte entwickelt und getestet. Es muss aber berücksichtigt werden, dass auch diese Interviews nicht zu absolut exakten psychiatrischen Diagnosen führen, da sie ja von Ärzten durchgeführt werden müssen, die psychiatrisch nur wenig oder nicht ausgebildet sind. Sie stellen jedoch einen wichtigen Fortschritt in der Diagnose psychiatrischer Krankheiten im Setting des Praktischen Arztes dar.

Bei diesen psychiatrischen Interviews handelt es sich um das „Symptom Driven Diagnostic System für Primary Care“ (SDDS-PC; Weissman et al. 1995) und die „Primary Care Evaluation of Mental Disorders“ (PRIME-MD, Spitzer et al. 1994). Beide Interviews beinhalten Module zur Diagnose von Depressionen. Es ist zu erwähnen, dass beide Interviews mit Unterstützung der pharmazeutischen Industrie entwickelt wurden – sie wurden aber in international anerkannten wissenschaftlichen Fachzeitschriften publiziert. Da beide Interviews in den USA entwickelt wurden, sind sie primär auf das US-amerikanische Diagnose-System DSM (American Psychiatric Association 1994) ausgerichtet, das sich in einigen Punkten vom in Europa verwendeten Diagnose-System ICD-10 (World Health Organization 1993, Dilling et al. 1993) unterscheidet. Während für das SDDS-PC unseres Wissens weder eine Adaptierung für ICD-10-Diagnosen vorgenommen wurde noch eine Übersetzung auf Deutsch erfolgte, geschah dies für das PRIME-MD. In deutscher Sprache existieren zwei Versionen: PRIME-MD für DSM-IV und TRIPS für ICD-10 („Training für interaktives psychiatrisches Screening“, modifizierte Kurz-Version von H. Katschnig und F. Gföllner).

Da die von Expertengruppen erarbeiteten Richtlinien zur Behandlung depressiver Krankheiten nicht auf alle Sub-Typen der Depression eingehen, scheint es nicht sinnvoll, alle im PRIME-MD bzw. im TRIPS enthaltenen Sub-Klassifikationen zu berücksichtigen. Aus diesem Grund wird hier ein gekürztes und vereinfachtes Interview vorgeschlagen, das einerseits an das TRIPS und andererseits an die Formulierung der ICD-10-Diagnosekriterien der WHO für praktische Ärzte (Leitlinien zur Diagnostik in der Primärversorgung [Müßigbrodt et al. 1996]) angelehnt ist.

Die Vorgangsweise bei der diagnostischen Abklärung und Behandlung wurde in verschiedenen internationalen, aber auch österreichischen Expertenstatements und Behandlungsrichtlinien beschrieben, von denen einige auch die spezielle Vorgangsweise beim Praktischen Arzt berücksichtigten (Veterans Health Administration 2000, American Psychiatric Association Work group 2003, Institute for Clinical Systems Improvement 2003 a, 2003 b, 2003 c, Härter et al. 2003, Aschauer et al. 2001, Snow et al. 2000).

2.3.3. Vorgangsweise bei Screening-Negativen

Auch wenn die bei allen Instrumenten gefundenen hohen Werte für den NPV zeigen, dass nur sehr selten Patienten mit einer Depression durch die Screening-Instrumente nicht erfasst werden, kann es doch vereinzelt vorkommen. Wenn der Praktische Arzt bei einzelnen Patienten unsicher ist, ob sie nicht doch unter einer Depression leiden, kann er ein Interview so wie bei Screening-Positiven durchführen oder an einen Facharzt für Psychiatrie zur genauen Abklärung weiterverweisen.

2.4. Stehen Möglichkeiten einer effektiven Behandlung mit geringen Nebenwirkungen zur Verfügung?

2.4.1. Behandlung mit Antidepressiva

Als Basistherapie der Depression gilt heute die Behandlung mit Antidepressiva, insbesondere bei mittelschweren und schweren Depressionen sowie wenn die Depressionen einen chronischen Verlauf zeigen (DGPPN 2000, US Preventive Services Task Force 2002, Aschauer et al. 2001). Zahlreiche internationale Studien haben die Wirksamkeit von Antidepressiva zur Behandlung der Depression bestätigt. Ein klares Überwiegen des Nutzens über das Risiko für potenzielle Schäden (Nebenwirkungen) ist die Voraussetzung für die Zulassung aller Medikamente sowohl in den USA als auch in allen europäischen Ländern. Es gilt als erwiesen, dass Antidepressiva nicht nur von Fachärzten für Psychiatrie, sondern auch von Praktischen Ärzten effektiv und sicher angewendet werden können (z.B. DGPPN 2000, US Preventive Services Task Force 2002, Veterans Health Administration 2000, Härter et al. 2003).

2.4.2. Psychotherapie

Eine Behandlung mittels spezifischer psychotherapeutischer Methoden gilt bei leichten bis mittelgradigen Depressionen als indiziert. Auch bei bestehender Kontraindikation gegen Antidepressiva oder der Ablehnung von Psychopharmaka durch den Patienten wird empfohlen, eine psychotherapeutische Behandlung zu erwägen (Härter et al. 2003). Diese spezifischen psychotherapeutischen Interventionen sind durch (ärztliche oder nicht-ärztliche) Psychotherapeuten durchzuführen, die die entsprechende Ausbildung und Kompetenz besitzen. Internationale Expertengruppen (US Preventive Services Task Force 2002, Härter et al. 2003) gehen davon aus, dass die Wirksamkeit spezifischer psychotherapeutischer Methoden zur Behandlung der Depression ausreichend belegt ist, um sie empfehlen zu können. In ähnlicher Weise wird davon ausgegangen, dass der Nutzen größer ist als das Risiko für potenzielle Schäden (Nebenwirkungen). Wissenschaftlich am besten abgesichert sind die kognitive Verhaltenstherapie und die interpersonelle Psychotherapie (Härter et al. 2003). Hinsichtlich der tiefenpsychologisch-orientierten Kurzzeittherapie gibt es Hinweise, dass diese ähnliche Effekte wie die kognitive Verhaltenstherapie hat (Härter et al. 2003).

Darüber hinaus gilt eine psychotherapeutische *Basisbehandlung*, die von jedem Praktischen Arzt durchzuführen ist, als zentraler Bestandteil jedes Gesamtbehandlungsplans von Depressionen (Härter et al. 2003, Aschauer et al. 2001). Dies umfasst den Aufbau einer empathischen und vertrauensvollen Beziehung zwischen Praktischem Arzt und Patienten (= „supportive“ Psychotherapie). Eine solch vertrauensvolle Beziehung ist die Voraussetzung, dass der Patient erfahren kann, dass der Behandler dem Patienten in einer schwierigen Situation stützend zur Seite steht. Diese psychotherapeutische Basisbehandlung gilt als Voraussetzung, um beim Patienten Verständnis für die Krankheit und Compliance mit der Behandlung zu erzielen. Es wird davon ausgegangen, dass der Nutzen dieser psychotherapeutischen Basisbehandlung größer ist als das Risiko für potenzielle Schäden (Härter et al. 2003, Aschauer et al. 2001).

2.4.3. Soziotherapie

Eine Behandlung mittels soziotherapeutischer Methoden gilt bei schweren (und chronisch verlaufenden) Depressionen als indiziert, während bei leichteren und mittelgradigen Depressionen derartige Interventionen nur selten notwendig sind (DGPPN 2000). Soziotherapeutische Interventionen umfassen unter anderem die Einbeziehung der Angehörigen, die konkrete Vermittlung von Hilfen für die Alltagsbewältigung sowie Hilfen im

Beruf. Soziotherapeutische Interventionen können weder die psychotherapeutischen noch die medikamentösen Interventionen ersetzen, sondern sind nur in Ergänzung sinnvoll. Es wird davon ausgegangen, dass der Nutzen dieser soziotherapeutischen Basisbehandlung größer ist als das Risiko für potenzielle Schäden (DGPPN 2000).

2.5. Effektivität von Screening, gefolgt von diagnostischer Abklärung und Behandlung

Die gesamte Intervention (Screening, gefolgt von diagnostischer Abklärung und Behandlung) soll für die Betroffenen bessere Ergebnisse (Verringerung der Symptomatik bzw. der Krankheitsfolgen) bringen im Vergleich zu keiner Intervention. Diese Fragestellung wurde gerade in den letzten Jahren durch mehrere großangelegte randomisierte, kontrollierte Studien untersucht.

Während frühere Übersichtsarbeiten und Studien (z.B. Gilbody et al. 2001) immer wieder zu dem Schluss kamen, dass nicht ausreichend Studien vorliegen, die eine Routine-Screening auf Depressionen rechtfertigen, kam die neueste Meta-Analyse (US Preventive Services Task Force 2002) zu dieser Fragestellung zu einem positiven Ergebnis. Die neueste Meta-Analyse zu dieser Fragestellung (US Preventive Services Task Force 2002) schloss auch die neuesten verfügbaren randomisierten, kontrollierten Studien ein. In dieser Meta-Analyse zeigte sich, dass Personen, die auf Depressionen gescreent wurden, über alle Studien hinweg in Summe ein besseres Ergebnis (d.h. Verringerung der Symptomatik bzw. der Krankheitsfolgen) zeigten als jene, die nicht gescreent wurden.

Aufgrund ihrer Meta-Analyse kam die US Preventive Services Task Force (2002) zu dem Schluss, dass routinemäßiges Screening auf Depressionen im Primary Care Setting zu empfehlen sei. Allerdings betonte die US Preventive Services Task Force (2002), dass in einem 2. Schritt zur Ergänzung des Screenings eine genaue diagnostische Abklärung durch Praktische Ärzte (unterstützt durch praktische Empfehlungen oder Schulungsprogramme) erforderlich sei.

Allerdings weist die US Preventive Services Task Force (2002) darauf hin, dass für Kinder und Jugendliche nicht ausreichend Daten vorliegen, um ein Screening zu empfehlen. Weiters wurde hervorgehoben, dass die erwähnten Ergebnisse nicht auf andere Settings wie Krankenhäuser oder Pflegeheime übertragbar seien.

Da die Studien, die in diese Meta-Analyse eingeschlossen wurden, nicht nur in den USA, sondern teilweise auch in europäischen Ländern durchgeführt wurden, kann man davon ausgehen, dass die Ergebnisse auch auf Österreich zutreffen. Somit kann aufgrund der internationalen Literatur ein Screening auf Depressionen als Teil der Vorsorgeuntersuchung in Österreich empfohlen werden.

2.6. In welchen Abständen sollte ein Screening auf Depressionen durchgeführt werden?

Es liegen weder Studien noch evidenzbasierte Empfehlungen vor, die sich mit dieser Fragestellung beschäftigen.

Um zu einer Einschätzung zu kommen, in welchen Abständen eine Screening-Prozedur sinnvoll sein könnte, können die folgenden Aspekte hilfreich sein:

1. Die durchschnittliche Dauer einer unbehandelten Krankheitsepisode beträgt etwa 4-6 Monate (Härter et al 2003).
2. Bei zwei Drittel der Erkrankten ist ohne Behandlung nicht mit einer vollständigen und anhaltenden Remission zu rechnen (Härter et al 2003).
3. Eine möglichst frühzeitige Behandlung verbessert die längerfristige Prognose (US Preventive Services Task Force 2002).

Unter Berücksichtigung dieser Aspekte dürfte ein Screening auf Depressionen im Abstand von etwa 1 bis 2 Jahren sinnvoll sein.

2.7. Welche Maßnahmen zur Qualitätssicherung könnten nützlich sein?

In der wissenschaftlichen Literatur konnten wir keine evidenzbasierte Empfehlungen finden, die beschreiben, welche Maßnahmen zur Qualitätssicherung unter den Bedingungen, die mit der österreichischen Vorsorgeuntersuchung vergleichbar sind, zu empfehlen sind.

Unter Berücksichtigung ausländischer Literatur (z.B. Kratz et al. 2003, Schneider et al. 2003) wären die folgenden Maßnahmen zur Qualitätssicherung in Österreich zu überlegen:

- Wird die Screening-Prozedur von den Patienten als akzeptabel erlebt? (z.B. erfasst mittels Fragebogen-Erhebung)
- Wird das diagnostische und therapeutische Vorgehen bei screening-positiven Resultaten von den Patienten als akzeptabel erlebt? (z.B. erfasst mittels Fragebogen-Erhebung)
- Erfolgt die Aufklärung über das Vorhandensein einer Depression auf eine Weise, die von den Patienten als akzeptabel erlebt wird? (z.B. erfasst mittels Fragebogen-Erhebung)
- Werden die verordneten Psychopharmaka auch wirklich eingenommen? Werden die verordneten Psychopharmaka in der verordneten Dosis und Regelmäßigkeit eingenommen? (z.B. erfasst mittels Fragebogen-Erhebung oder Tiefeninterviews)
- Wird eine etwaige Zuweisung zur Psychotherapie angenommen? Wird die Psychotherapie in der vorgeschlagenen Häufigkeit in Anspruch genommen? (z.B. Vergleich der Zuweisungen mit der Inanspruchnahme)
- Führt die angebotene Behandlung innerhalb eines gewissen Zeitraums zu keiner Verminderung der Symptomatik? (z.B. Ursachen erfasst mittels Experten-Interviews, Review der Krankenakten, Fragebögen zu Medikamenteneinnahme und Psychotherapie)
- Wie hoch ist die Rate jener, die *keine* Depression haben, vom praktischen Arzt aber einer (nicht notwendigen) Behandlung zugeführt werden? (z.B. erfasst mittels Experten-Interviews)
- Wird das zur Verfügung gestellte Praxis-Manual zur Diagnostik und Behandlung von den Praktischen Ärzten als hilfreich empfunden? (z.B. erfasst mittels Fragebogen-Erhebung)

3. SCHLUSSFOLGERUNGEN UND ZUSAMMENFASSUNG

Im Falle einer routinemäßigen Implementierung eines Screenings auf Depressionen im Rahmen der Vorsorgeuntersuchung müssen die einleitend gestellten Fragen zufriedenstellend beantwortet werden können.

3.1. Ad 1: Screening-Instrumente sollen möglichst exakt vorhersagen, ob jemand an einer Depression leidet (= Kriteriums-Validität), und gleichzeitig zeitökonomisch in der Anwendung sein.

Die neueren Kurz-Screening-Fragebögen mit 1 oder 2 Items zeigen deutlich schlechtere Kennwerte der Kriteriums-Validität als die umfassenderen Screening-Fragebögen. In deutscher Sprache liegen nur wenige Daten zur Kriteriums-Validität der umfassenderen Screening-Fragebögen vor, die aber deutlich bessere Ergebnisse als die englischsprachigen Kurz-Screening-Fragebögen haben. Leider finden sich unter den deutschsprachigen Studien auch Instrumente, die eher für spezielle Zielgruppen (z.B. EPDS für Frauen nach der Geburt) als für die Patienten von Praktischen Ärzten entwickelt worden waren und daher für die vorliegende Aufgabe eher nicht geeignet scheinen.

Unter Berücksichtigung der Kennwerte der Kriteriums-Validität scheint das Beck-Depression-Inventory (BDI) bzw. die Center for Epidemiological Studies Depression Scale (CES-D) geeignet. Beide Instrumente wurden auf deutsch übersetzt, jedoch gelang es uns nicht Daten zur Kriteriums-Validität der deutschsprachigen Versionen zu finden. (Die Durchführung einer Studie zur Validität des BDI bzw. der CES-D im Setting von Praktischen Ärzten in Österreich wäre vor dem Einsatz im Rahmen der Vorsorgeuntersuchungen ernsthaft in Erwägung zu ziehen.)

Bei Verwendung der neueren Kurz-Screening-Fragebögen mit 1 oder 2 Items müsste eine sehr große Zahl von screening-positiven Patienten in einem zweiten Schritt untersucht werden, ob sie wirklich unter einer Depression leiden. Bei Verwendung des BDI oder der CES-D wäre die Zahl screening-positiver Patienten deutlich geringer, was eine geringere Anzahl von Untersuchungen erfordern würde. Obwohl auch bei der Geriatric Depression Scale (GDS-30) nur eine relativ kleine Zahl screening-positiver Patienten in einem zweiten Schritt untersucht werden müsste, muss man berücksichtigen, dass für die GDS-30 ausschließlich Daten über geriatrische Patienten vorliegen. Unter dem Aspekt Arbeitersparnis wären also das BDI oder die CES-D am brauchbarsten.

Um Screening-Instrumente im Rahmen der Vorsorgeuntersuchung einsetzen zu können, ist es erforderlich, dass sie in möglichst kurzer Zeit angewandt werden können. Auch wenn die Kurz-Screening-Fragebögen mit 1 oder 2 Items weniger als eine Minute Zeit zum Beantworten benötigen, sparen die längeren Instrumente mit etwa 20 Items (2-4 Minuten zum Ausfüllen benötigt) insgesamt mehr Zeit, da die Praktischen Ärzte deutlich weniger Patienten einer genaueren Untersuchung unterziehen müssen.

Sowohl CES-D als auch BDI waren auch im Rahmen von großen randomisierten Studien verwendet worden, die die Effektivität des Screenings auf Depressionen nachwiesen. Unter Berücksichtigung all dieser Aspekte scheinen das BDI oder die CES-D am ehesten geeignet im Rahmen der Vorsorgeuntersuchung eingesetzt zu werden.

Mögliche Kombination aus Kurz-Screening gefolgt von längerem Screening

Es wäre zu überlegen, dass jeder Patient im Rahmen der Vorsorgeuntersuchung den Kurz-Screening-Fragebogen mit 2 Items ausfüllt. Diejenigen, die im Kurz-Screening positiv sind, werden gebeten einen der üblichen längeren Screening-Fragebögen auszufüllen. Nur

diejenigen, die auch im längeren Screening positiv sind, werden vom Praktischen Arzt auf das Vorhandensein einer Depression untersucht.

In Summe kann man auf Basis der verfügbaren Daten annehmen, dass diese Kombination von zwei unterschiedlichen Screening-Verfahren Vorteile im Vergleich zur Verwendung einzelner Screening-Fragebögen hat. (Die Durchführung einer Studie zur Validität dieses kombinierten Verfahrens im Setting von Praktischen Ärzten in Österreich wäre vor dem routinemäßigen Einsatz empfehlenswert.)

Für welche Personengruppen liegen ausreichend Daten vor, die ein Screening auf Depression rechtfertigen?

Für Patienten von Praktischen Ärzten beiderlei Geschlechts und aller Altersgruppen über 20 Jahren liegen ausreichend Daten vor. Aus praktischen Gründen scheint es sinnvoll die folgenden Personen auszuschließen:

- Patienten mit ausgeprägten kognitiven Defiziten
- Personen, die aufgrund von Sprachschwierigkeiten den Fragebogen nicht verstehen
- Personen mit ausgeprägter Schreib- bzw. Leseschwäche

Für Personen mit ausgeprägten Sehbehinderungen dürfte es hilfreich sein, spezielle Fragebögen mit großer Schrift zu entwickeln oder die einzelnen Fragen dem Patienten vorzulesen.

Ad 2: Die Verwendung von Screening-Instrumenten und die anschließende diagnostische Abklärung soll für die Betroffenen keine nachteiligen Folgen (potenzielle Schäden) haben.

Aus der vorhandenen Literatur lassen sich *indirekte* Hinweise darauf finden, dass Depressions-Screening keine relevanten negativen Folgen für die Patienten hat.

- Die Zahl jener Personen, die die Teilnahme an Depressions-Screening-Programmen abbrach, war mit einer einzigen Ausnahme nicht signifikant erhöht.
- Die Zufriedenheit mit der Behandlung war bei jenen, die an Depressions-Screening-Programmen teilnahmen, nicht schlechter als in der Vergleichsgruppe.

Obwohl diese Frage bislang noch nicht *explizit* untersucht wurde, kann aus den beschriebenen Ergebnissen der Schluss gezogen werden, dass Depressions-Screening wahrscheinlich keine relevanten negativen Folgen für die Patienten hat. Allerdings muss berücksichtigt werden, dass in all diesen Studien die diagnostische Abklärung durch den Praktischen Arzt und nicht durch Psychiater, Psychotherapeuten oder Psychologen erfolgte. (Einige der vorgeschlagenen Maßnahmen zur Qualitätssicherung können weitere Hinweise auf diese Fragestellung geben.)

Ad 3: Überprüfung des Screening-Resultates: Wenn das Screening-Resultat auf das Vorhandensein einer Depression hinweist, soll möglichst einfach geklärt werden können, ob wirklich eine Depression vorliegt.

Um das unnötige Risiko einer etwaigen Stigmatisierung screening-positiver Patienten zu vermeiden, ist es erforderlich zuerst abzuklären, ob eine Depression vorliegt, bevor der Patient mit der Diagnose konfrontiert wird oder an spezialisierte Berufsgruppen wie Psychiater, Psychotherapeuten oder Psychologen überwiesen wird. Die diagnostische Abklärung muss also durch den Praktischen Arzt erfolgen. (Nur jene Patienten, wo trotz

umfangreicher Bemühungen des Praktischen Arztes eine eindeutige Diagnostik nicht gelingt, sollten an Spezialisten weiterverwiesen werden.)

Auch aufgrund der internationalen Literatur zu diesem Thema dürfte die Überprüfung eines positiven Screening-Resultates am besten durch den Praktischen Arzt erfolgen. Die internationale Literatur weist außerdem darauf hin, dass Richtlinien zur Diagnostik und Behandlung und strukturierte Interviews zur Diagnose einer Depression für die Praktischen Ärzte hilfreich sind. Aus diesem Grund sollen die vorhandenen Richtlinien und Empfehlungen von Expertengruppen in einem Praxis-Manual für Praktische Ärzte in Österreich zur Verfügung stehen.

Ad 4: Es sollen Möglichkeiten einer effektiven Behandlung zur Verfügung stehen, wobei die verfügbaren Behandlungstechniken im Vergleich zu ihrem Nutzen ein möglichst geringes Risiko für potenzielle Schäden (Nebenwirkungen) haben sollen.

Sowohl für Antidepressiva als auch spezifische psychotherapeutische Methoden liegen ausreichend Studien vor, die nachweisen, dass sie effektiv in der Behandlung von Depressionen sind, und zeigen, dass ihr Nutzen deutlich größer als ihre potenzielle Schäden (Nebenwirkungen) ist.

Ad 5: Die gesamte Intervention (Screening, gefolgt von diagnostischer Abklärung und Behandlung) soll für die Betroffenen bessere Ergebnisse (Verringerung der Symptomatik bzw. der Krankheitsfolgen) bringen im Vergleich zu keiner Intervention.

Große randomisierte, kontrollierte Studien wurden in den letzten Jahren zu dieser Frage durchgeführt. Die neueste Meta-Analyse zu dieser Fragestellung (US Preventive Services Task Force 2002) kommt zu dem Schluss, dass Personen, die auf Depressionen gescreent werden, ein besseres Ergebnis (d.h. Verringerung der Symptomatik bzw. der Krankheitsfolgen) zeigen als jene, die nicht gescreent werden.

Ad 6: Gibt es Hinweise darauf, in welchen Abständen eine derartige Screening-Prozedur sinnvoll ist?

Um einschätzen zu können, in welchen Abständen eine Screening-Prozedur sinnvoll ist, wurden verschiedene klinische Aspekte berücksichtigt. Dies führte zu der Einschätzung, dass ein Screening auf Depressionen im Abstand von etwa 1 bis 2 Jahren sinnvoll sein dürfte. (Allerdings muss berücksichtigt werden, dass weder Studien noch evidenzbasierte Empfehlungen zu dieser Fragestellung vorliegen.)

Ad 7: Welche Maßnahmen zur Qualitätssicherung könnten nützlich sein?

Unter Berücksichtigung ausländischer Literatur wurden einige Maßnahmen vorgeschlagen, die zur Qualitätssicherung im Rahmen der Vorsorgeuntersuchungen in Österreich zu überlegen wären.

Zusammenfassung

Zwei Screening-Instrumente, die aus etwa 20 Fragen bestehen (BDI, CES-D), sagen ausreichend exakt vorher, ob jemand an einer Depression leidet, und scheinen vom zeitlichen Aufwand akzeptabel. (Da diese Daten aus internationalen Studien stammen, wäre aber die Durchführung einer Studie zur Validität von BDI bzw. CES-D im Setting von Praktischen Ärzten in Österreich empfehlenswert.) Aus der vorhandenen Literatur lassen sich indirekte Hinweise darauf finden, dass die Verwendung von Screening-Instrumenten und die anschließende diagnostische Abklärung für die Betroffenen keine nachteiligen Folgen (potenzielle Schäden) hat. Die diagnostische Abklärung sollte durch den Praktischen Arzt an Hand eines speziellen Manuals erfolgen. Es stehen Möglichkeiten einer effektiven Behandlung (Antidepressiva, Psychotherapie) zur Verfügung. Eine neue internationale Meta-Analyse kommt zu dem Schluss, dass Screening auf Depressionen beim Praktischen Arzt zu empfehlen ist. Ein solches Screening dürfte im Abstand von etwa 1 bis 2 Jahren sinnvoll sein.

4. VERWENDETE LITERATUR

American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders: Fourth Edition. American Psychiatric Association, Washington DC, 1994

American Psychiatric Association Work group on Major Depression: Practice Guideline for the Treatment of Patients with Major Depression. 2003, www.psych.org

Arolt V: Psychische Störungen bei Krankenhauspatienten. Berlin, Springer, 1997

Arroll B., Khin N., Kerse N.: Screening for depression in primary care with two verbally asked questions: cross sectional study. *BMJ* 327, 2003, 1144-1146

Aschauer H, et al.: Depression: Medikamentöse Therapie – State of the Art 2001. *Clinicum Sonderheft* April 2001

Beck AT: The depression inventory. Philadelphia: Center for Cognitive Therapy. 1978

Bergant AM, Nguyen T, Heim K, Ulmer H, Dapunt O: Deutschsprachige Fassung und Validierung der "Edinburgh postnatal depression scale". *Dt Med Wochensh* 1998; 123: 35-40

Broadhead WE, Leon AC, Weissman MM, et al. Development and validation of the SDDS-PC screen for multiple mental disorders in primary care. *Arch Fam Med*. 1995;4:211-219

Callahan CM, Dittus RS, Tierney WM. Primary care physicians' medical decision making for late-life depression. *J Gen Intern Med*. 1996;11:218-225

Callahan CM, Hendrie HC, Dittus RS, Brater DC, Hui SL, Tierney WM. Improving treatment of late life depression in primary care: a randomized clinical trial. *J Am Geriatr Soc*. 1994;42:839-846

Cox J, Holden J, Sagovsky R: Detection of postnatal depression: development of the 10-item Edinburgh Postnatal Depression Scale. *Br. J. Psychiatry* 1987; 150: 782-786

Derogatis L: SCL-90-R administration, scoring and procedures manual-II. Towson, MD: Clinical Psychometric Research, 1992

DGPPN – Deutsche Gesellschaft für Psychiatrie, Psychotherapie und Nervenheilkunde (Hrsg.): Behandlungsleitlinie Affektive Erkrankungen. Steinkopff-Verlag, Darmstadt 2000

Dilling H., Mombour W., Schmidt M. (Hg.): Internationale Klassifikation psychischer Störungen – ICD-10 Kapitel V (F). Verlag Hans Huber, Bern 1993

Dowrick C, Buchan I. Twelve month outcome of depression in general practice: does detection or disclosure make a difference? *BMJ* 1995 a; 311: 1274-1276

Dowrick C. Does testing for depression influence diagnosis or management by general practitioners? *Fam Pract*. 1995 b;12:461-465

Fechner-Bates S, Coyne JC, Schwenk TL. The relationship of self-reported distress to depressive disorders and other psychopathology. *J Consult Clin Psychol* . 1994;62:550-559

Ford D. Principles of screening applied to psychiatric disorders. *Gen Hosp Psychiatry* 1988; 10: 177-188

German PS, Shapiro S, Skinner EA. Detection and management of mental health problems of older patients by primary care providers. *JAMA* 1987; 257: 489-496

Gilbody SM, House AO, Sheldon TA: Routinely administered questionnaires for depression and anxiety: Systematic review. *Brit Med J* 2001; 322: 406-409

Goldberg D: Use of the general health questionnaire in clinical work. *Brit Med J* 1986; 293: 1188-1189

Goldberg D, Huxley P.: *Mental illness in the community: the pathway to psychiatric care.* Tavistock, London, 1980

Goldberg DP, Gater R, Sartorius N, Ustun TB, Piccinelli M, Gureje O, Rutter C: The validity of two versions of the GHQ in the WHO study of mental illness in general health care. *Psychol Med* 1997; 27: 191-7

Goldberg DP, Oldehinkel T, Ormel J: Why GHQ threshold varies from one place to another. *Psychol Med* 1998; 28: 915-21

Härter M, Bermejo I, Schneider F, Kratz S, Gaebel W, Hegerl U, Niebling W, Berger M: Versorgungsleitlinien zur Diagnostik und Therapie depressiver Störungen in der hausärztlichen Praxis. *Zeitschrift ärztliche Fortbildung und Qualitätssicherung* 2003; 97 (Suppl. IV): 16-35

Hautzinger M, Bailer M, Worall H, Keller F: 1994 -Depressions-Inventar (BDI). Bern: Huber Beck

Hendrie HC, Callahan CM, Levitt EE, et al. Prevalence rates of major depressive disorders: The effects of varying the diagnostic criteria in an older primary care population. *Am J Geriatr Psychiatry* . 1995;3:119-131

Hoeper EW, Nycz GR, Kessler JD, Pierce WE. The usefulness of screening for mental illness. *Lancet* 1984; 1: 33-35

Holcomb WL Jr, Stone LS, Lustman PJ, Gavard JA, Mostello DJ. Screening for depression in pregnancy: characteristics of the Beck Depression Inventory. *Obstet Gynecol.* 1996;88:1021-1025

Institute for Clinical Systems Improvement: *Health Care Guideline: Major depression in Adults for Mental Health Care Providers.* ICSI, Bloomington 2003 a

Institute for Clinical Systems Improvement: *Health Care Guideline: Major depression in Adults in Primary Care.* ICSI, Bloomington 2003 b

Institute for Clinical Systems Improvement: *Health Care Guideline for Patients and their Families: Major depression in Adults in Primary Care.* ICSI, Bloomington 2003 c

Johnstone A, Goldberg D. Psychiatric screening in general practice. A controlled trial. *Lancet.* 1976;1:605-608.

Katschnig H, Berner W, Haushofer M, Barfuss M, Seelig P. Psychiatric case identification in general practice: self-rating versus interview. *Acta Psychiat Scand* 1980; 62 (Suppl 285): 164-175

- Klinkman MS, Coyne JC, Gallo S, Schwenk TL. Can case-finding instruments be used to improve physician detection of depression in primary care? *Arch Fam Med*. 1997;6:567-573
- Kratz S, Härter M, Bermejo I, Berger M, Schneider F, Gaebel W: Beratungsanlass und Diagnosestellung bei depressiven Patienten in der hausärztlichen Praxis. *Zeitschrift ärztliche Fortbildung und Qualitätssicherung* 2003; 97 (Suppl. IV): 50-56
- Kühner C: Gender differences in unipolar depression: an update of epidemiological findings and possible explanations. *Acta Psychiat. Scand*. 2003; 108: 1-12
- Leon AC, Olfson M, Weissman MM, et al. Brief screens for mental disorders in primary care. *J Gen Intern Med*. 1996;11:426-430
- Lewis G, Sharp D, Bartholomew J, Pelosi AJ. Computerized assessment of common mental disorders in primary care: effect on clinical outcome. *Fam Pract*. 1996;13:120-126.
- Linn LS, Yager J. The effect of screening, sensitization, and feedback on notation of depression. *J Med Educ*. 1980; 55:942-949.
- Lustman PJ, Clouse RE, Griffith LS, Carney RM, Freedland KE. Screening for depression in diabetes using the Beck Depression Inventory. *Psychosom Med*. 1997;59:24-31
- Lyness JM, Noel TK, Cox C, King DA, Conwell Y, Caine ED. Screening for depression in elderly primary care patients. A comparison of the Center for Epidemiologic Studies-Depression Scale and the Geriatric Depression Scale. *Arch Intern Med*. 1997;157:449-454
- Magruder-Habib K, Zung WW, Feussner JR. Improving physicians' recognition and treatment of depression in general medical care. Results from a randomized clinical trial. *Med Care*. 1990;28:239-250.
- Mazonson PD, Mathias SD, Fifer SK, Beusching DP, Patrick DL. The mental health patient profile: does it change primary care physicians practice patterns? *J Am Board Fam Pract* 1994; 9: 336-345
- Moore JT, Silimperi DR, Bobula JA. Recognition of depression by family medicine residents: the impact of screening. *J Fam Pract*. 1978;7:509-513.
- Mulrow CD, Williams JW Jr, Gerety MB, Ramirez G, Montiel OM, Kerber C. Case-finding instruments for depression in primary care settings. *Ann Intern Med*. 1995;122:913-921
- Müßigbrodt H, Kleinschmidt S, Schürmann A, Freyberger H, Dilling H: Psychische Störungen in der Praxis – Leitfaden zur Diagnostik und Therapie in der Primärversorgung nach dem Kapitel V (F) der ICD-10. Verlag Hans Huber, Bern 1996
- Muzik M, Klier C, Rosenblum K, Holzinger A, Umek W, Katschnig H: Are commonly used self-report inventories suitable for screening postpartum depression and anxiety disorders? *Acta Psychiat Scand* 2000; 102: 71-73
- Myers JK, Weissman MM. Use of a self-report symptom scale to detect depression in a community sample. *Am J Psychiatry*. 1980;137:1081-1084
- Radloff L: The CES-D scale: a self-report depression scale dor research in the general population. *Applied Psychological Measurement* 1977; 1: 385-401
- Rand EH, Badger LW, Coggins DR. Toward a resolution of contradictions. Utility of feedback from the GHQ. *Gen Hosp Psychiatry*. 1988;10:189-196.

Research Triangle Institute: Screening for Depression - Systematic Evidence Review. Task No.3, University of North Carolina 2002

Reifler DR, Kessler HS, Bernhard EJ, Leon AC, Martin GJ. Impact of screening for mental health concerns on health service utilization and functional status in primary care patients. *Arch Intern Med.* 1996;156:2593-2599

Schmitz N, Kruse J, Heckrath C, Alberti L, Tress W: Diagnosing mental disorders in primary care: the General Health Questionnaire (GHQ) and the Symptom Checklist (SCL-90-R) as screening instruments. *Soc Psychiatry Psychiatr Epidemiol* 1999 a; 34: 360-366

Schmitz N, Kruse J, Tress W: Psychometric properties of the General Health Questionnaire (GHQ) in a German primary care sample. *Acta Psychiatr Scand* 1999 b; 100: 462-468

Schneider F, Härter M, Kratz S, Bermejo I, Mulert C, Hegerl U, Gaebel W, Berger M: Unzureichender subjektiver Behandlungsverlauf bei depressiven Patienten in der haus- und nervenärztlichen Praxis. *Zeitschrift ärztliche Fortbildung und Qualitätssicherung* 2003; 97 (Suppl. IV): 57-66

Snow V, Lascher S, Mottur-Pilson C (for the American College of Physicians): Pharmacologic treatment of acute major depression and dysthymia. *Ann. Intern. Med.* 2000; 132: 743-756

Spitzer R, Williams J, Kroenke K, et al.: Utility of a new procedure for diagnosing mental disorders in primary care: the PRIME-MD 1000 study. *JAMA* 1994; 272: 1749-1756

Upadhyaya AK, Stanley I. Detection of depression in primary care: comparison of two self-administered scales. *Int J Geriatr Psychiatry.* 1997;12:35-37

US Preventive Services Task Force: Screening for depression: recommendations and rationale. *Ann. Intern Med* 2002; 136: 760-764

Üstün T, Sartorius N: *Mental illness in general health care: an international study.* Chichester, John Wiley & Sons, 1995

van Marwijk HW, Wallace P, de Bock GH, Hermans J, Kaptein AA, Mulder JD. Evaluation of the feasibility, reliability and diagnostic value of shortened versions of the geriatric depression scale. *Br J Gen Pract.* 1995;45:195-199

Veterans Health Administration, Department of Veterans Affairs, and Health Affairs, Department of Defense, Evidence Based Clinical Practice Guideline Working Group: *Management of Major Depressive Disorder in Adults in the Primary Care Setting.* Washington, DC: May 2000. Office of Quality and Performance publication 10Q-CPG/MDD-00

Wancata J. Stigma – ein Kommentar aus gerontopsychiatrischer Perspektive. *Neuropsychiatrie* 2002; 16: 115-116

Wancata J, Alexandrowicz R, Meise U, Krautgartner M. General Health Questionnaire in general hospitals: selecting a set of items using a stepwise hierarchical procedure. *Intern J Meth Psychiat Res* 2001; 10: 108-117

Wancata J, Benda N, Bach M, Meise U: Recognition of psychiatric disorders in nonpsychiatric hospital wards. *J Psychosom Res* 2000; 48: 149-155

Weissman M, Olfson M, Leon A, et al.: Brief diagnostic interviews (SDDS-PC) for multiple mental disorders in primary care: a pilot study. *Arch. Fam. Med.* 1995; 4: 220-227

Wells KB, Sherbourne C, Schoenbaum M, et al. Impact of disseminating quality improvement programs for depression in managed primary care: a randomized controlled trial. *JAMA.* 2000;283:212-220.

Whooley, Mary A., Avins, Andrew L., Miranda, Jeanne & Browner, Warren S.: Case-Finding Instruments for Depression: Two Questions Are as Good as Many. *Journal of General Internal Medicine* 1997; 12: 439-445

Williams JW Jr, Mulrow CD, Kroenke K, et al. Case-finding for depression in primary care: a randomized trial. *American Journal of Medicine.* 1999; 106: 36-43

World Health Organization: The ICD-10 Classification of Mental and Behavioural Disorders: Diagnostic Criteria for Research. WHO, Geneva, 1993

Zich JM, Attkisson CC, Greenfield TK. Screening for depression in primary care clinics: the CES-D and the BDI. *Int J Psychiatry Med.* 1990;20:259-277

Zintl-Wiegand A., Cooper B., Krumm B. (1980): Psychisch Kranke in der ärztlichen Allgemeinpraxis: eine Untersuchung in der Stadt Mannheim. Beltz, Weinheim

Zung W: A self-rating depression scale. *Arch. Gen. Psychiatry* 1965; 12: 63-70

Zung WW, King RE. Identification and treatment of masked depression in a general medical practice. *J Clin Psychiatry.* 1983;44:365-368.

ANHANG A:

BESCHREIBUNG DER VERWENDETEN KENNWERTE DER KRITERIUMS-VALIDITÄT

Kriteriums-Validität

Validität meint die Gültigkeit der getroffenen Aussage – also ob bzw. inwieweit ein Test auch tatsächlich das misst, was er messen soll. Bei Screening-Instrumenten sollen Zeichen einer Krankheit erfasst werden: diese Krankheit ist das Kriterium, an dem das Screening-Verfahren gemessen wird. Daher spricht man in diesem Fall von Kriteriums-Validität. Die folgenden Kennzahlen beschreiben, in welchem Ausmaß die Ergebnisse eines Screening-Verfahrens mit dem Kriterium Krankheit übereinstimmen, wobei das Kriterium Krankheit üblicherweise unabhängig mit einem psychiatrischen Forschungsinterview erfasst wird.

Sensitivität:

Die Sensitivität meint den Prozentsatz jener Kranken, die durch das Screening-Instrument korrekt erkannt wurden. Eine hohe Sensitivität bedeutet, dass nur wenige Kranke vom Screening-Instrument übersehen werden (d.h. wenige falsch Negative).

Spezifität:

Die Spezifität meint den Prozentsatz jener Gesunden, die durch das Screening-Instrument korrekt erkannt wurden. Eine hohe Spezifität bedeutet somit, dass nur wenige Gesunde vom Screening-Instrument fälschlich als krank klassifiziert werden (d.h. wenige falsch Positive).

Positive Predictive Value (=PPV):

Der Positive Predictive Value (= positive Korrektheit) meint den Prozentsatz wirklich Kranker von allen Screening-Positiven. Ein hoher PPV bedeutet, dass die meisten Screening-Positiven auch wirklich krank sind.

Negative Predictive Value (=NPV):

Der Negative Predictive Value (= negative Korrektheit) meint den Prozentsatz wirklich Gesunder von allen Screening-Negativen. Ein hoher NPV bedeutet, dass die meisten Screening-Negativen auch wirklich gesund sind.

Overall Misclassification Rate (=OMR):

Die Overall Misclassification Rate meint den Prozentsatz all jener, die durch das Screening falsch zugeordnet wurden, d.h. alle falsch Positiven und alle falsch Negativen bezogen auf alle Untersuchten. Eine niedrige OMR bedeutet, dass ein hoher Prozentsatz durch das Screening-Instrument richtig zugeordnet wird.

ANHANG B:**DEPRESSIONS-SCREENING-INSTRUMENTE: VERWENDETE ABKÜRZUNGEN**

Abkürzung	Screening-Instrument
BDI	Beck Depression Inventory;
CES-D	Center for Epidemiologic Study Depression Screen;
SDDS-PC	Symptom Driven Diagnostic System – Primary Care;
Zung SDS	Zung Self-Assessment Depression Scale;
GDS-30	Geriatric Depression Scale (30 items);
HAD-D	Depressions-Subskala der Hospital Anxiety and Depression Scale ;
GHQ-30	General Health Questionnaire (30 items);
EPDS	Edinburgh Postnatal Depression Scale; S
SCL-90-R	revidierte Form der Symptom Checklist mit 90 Items mit Depressions-Subskala