

**B. Löwe, R.L. Spitzer,
S. Zipfel & W. Herzog:
PHQ-D. Gesundheitsfrage-
bogen für Patienten.**

2. Auflage, Pfizer GmbH,
Karlsruhe, 2002, kostenlos

1 Testart

Der „Gesundheitsfragebogen für Patienten (PHQ-D)“ ist ein klinisch-psychologisches Fragebogenverfahren zur syndromalen Screening-Diagnostik von somatoformen Störungen, depressiven Störungen, Angststörungen, Essstörungen und Alkoholmissbrauch.

2 Testmaterial

Der PHQ-D besteht aus einem Testmanual (TM), einer Kurzanleitung, 20 Fragebogen (Langform), 20 Fragebogen (Kurzform), 20 Diagnose- und Therapiebögen und drei Auswertungsschablonen. Zusätzlich wird ein Schreibgerät benötigt.

3 Testgliederung

Die Langform des PHQ-D umfasst 89 Items, zu denen mit einer Ausnahme („Was belastet Sie zur Zeit in Ihrem Leben am meisten?“) Antwortkategorien vorgegeben sind. Die Kategorienformate variieren in den verschiedenen Fragebogenbereichen, im einfachsten Fall ist zwischen „Ja“ oder „Nein“ zu wählen. Der Bogen ist in 16 verschiedene, zum Teil nur ein Item umfassende, Abschnitte gegliedert. Einige der Bereiche sollen bei Verneinung einer Eingangsfrage übersprungen werden (Tabelle 1).

Die Fragen zu den einzelnen Störungen erfassen weitestgehend die DSM-IV-Kriterien für die jeweilige Diagnose. Weitere Items beschäftigen sich mit der Schwere der Beeinträchtigung durch die abgefragten Beschwerden, der psychosozialen Funktionsfähigkeit, möglichen Stressoren und kritischen Lebensereignissen. Sechs Items am Ende des Bogens erfragen (nur für Frauen) einige Aspekte von Monatsblutung, Schwangerschaft und Geburt.

Die Kurzform des PHQ-D (15 Items, 1 Seite) enthält die Fragen zu depressiven Störungen und Angststörungen so-

wie das Item zur Schwere der Beeinträchtigung in identischer Weise wie in der Langversion. Im TM (S. 5) sind Möglichkeiten angeführt, um die verschiedenen Bereiche von Kurz- und Langform zu kombinieren.

4 Grundkonzept

Der PHQ-D bzw. dessen Kurzform sind die autorisierten deutschen Versionen des amerikanischen „Prime MD Patient Health Questionnaire (PHQ)“ bzw. des „Prime MD Brief Patient Health Questionnaire (Brief PHQ)“ (Spitzer, Kroenke & Williams, 1999), die aus dem „Prime MD“ (Spitzer, Williams, Kroenke, Linzer, DeGruy, Hahn, Brody & Johnson, 1994) hervorgegangen sind. Die Skalen zu Depression und Somatisierung sind auch als PHQ-9 (Kroenke, Spitzer & Williams, 2001) bzw. PHQ-15 (Kroenke, Spitzer & Williams, 2002) publiziert.

Ziel des PHQ-D ist es, „die Erkennung und die Diagnostik der häufigsten psychischen Störungen in der Primärmedizin zu erleichtern“ (TM, S. 4). Als Anwender sind damit ausdrücklich vorrangig praktisch tätige Mediziner angesprochen. Hintergrund dieser Zielgruppenwahl ist, dass auch bei psychischen Problemen zunächst oft ein (niedergelassener) Arzt konsultiert wird oder bei somatischen Erkrankungen häufig eine psychische Komorbidität besteht. Untersuchungen zeigen jedoch, dass psychische Störungen durch Mediziner oft nicht oder nicht richtig erkannt werden (vgl. z.B. Thompson, Kinmonth, Stevens, Peveler, Stevens, Ostler, Pickering, Baker, Henson, Preece, Cooper & Campbell, 2000). Um eine adäquate Behandlung auch der psychischen Probleme des Patienten zu sichern, soll mit dem PHQ-D ein Instrument bereitgestellt werden, das hilft, diesen Zustand zu verändern. Der Fragebogen ermöglicht das Erkennen von somatoformen Störungen, depressiven Störungen, Angststörungen, Essstörungen und Alkoholmissbrauch/-abhängigkeit. Die im Fragebogen ermittelten Daten sollten jedoch im Gespräch vertieft und untersetzt werden (TM, S. 5, 7). Mit dem PHQ-D können Diagnosen nach ICD-10 oder DSM-IV vergeben werden, dazu liegt eine Tabelle im TM (S. 9) vor. Der PHQ-D ist zur Status- und Verlaufsdagnostik sowohl für den Praxiseinsatz als auch bei For-

schungsfragestellungen geeignet. Die amerikanische Version hat breiten Einsatz in verschiedenen Studien gefunden, wie zahlreiche Publikationen zeigen (vgl. z.B. Goodwin & Olfson, 2002; Rollman, Hanusa, Lowe, Gilbert, Kapoor & Schulberg, 2002).

5 Durchführung

Die Durchführung ist im Einzel- oder Gruppenversuch möglich. Das Ausfüllen erfolgt durch den Patienten selbständig. Für die Langform werden etwa 10 bis 15 und für die Kurzform etwa fünf Minuten benötigt.

6 Auswertung

Die Auswertung (Diagnostik) der Seiten 1 bis 3 der Langform erfolgt mittels beigefügter Schablonen oder alternativ durch Auszählung entsprechend der im TM (S. 6) und in der Kurzanleitung dargestellten Algorithmen. Dafür werden ca. drei Minuten benötigt. Die Items auf Seite 4 sind als Zusatzinformationen gedacht. 10 Fragen davon können zu einem Score für Stress addiert werden. Für Depressivität und somatische Symptome können ebenfalls Skalensummenwerte gebildet werden, die als Schweregradeinschätzungen herangezogen werden können, z.B. zur Beurteilung eines Therapieverlaufs.

Für die Auswertung der Kurzform existiert keine Schablone. Die Auswertungsrichtlinie hierfür ist im TM (S. 6), in der Kurzanleitung und in der Fußzeile des Bogens dargestellt. Der Zeitbedarf beträgt etwa eine Minute.

Die Diagnose(n) können auf dem Diagnose- und Therapiebogen erfasst und um zusätzliche Informationen (z.B. Medikation, weitere Diagnosen) ergänzt werden.

Im TM wird mehrfach darauf hingewiesen, dass zur Absicherung der Diagnose (Störungsebene) weitere Informationen (z.B. Anamnesegespräch) hinzugezogen werden müssen.

7 Gütekriterien

7.1 Objektivität

Auswertungs-, Durchführungs- und Interpretationsobjektivität sind durch Fragebo-

Tabelle 1: Aufbau des PHQ-D (Langform)

Bereich (Items)	Störungen	ICD-10 Klassifikation
Somatoforme Störungen (13 Items)	Somatisierungsstörung	F 45.0
	Undifferenzierte Somatisierungsstörung	F 45.1
	Somatoforme autonome Funktionsstörung	F 45.3
Depressive Störungen (9 Items)	Major Depressives Syndrom	F 32
	leicht	F 32.0
	mittel	F 32.1
	schwer	F 32.2
	Rezidivierende depressive Episode	F 33
	leicht	F 33.0
	mittel	F 33.1
	schwer	F 33.2
	Minor Depression/depressive Episode	F 32.9
	Minor Depression/ Rezidivierende depressive Episode	F 33.9
Dysthymia	F 34.1	
Angststörungen (22 Items)	Panikstörung	F 41.0
	Agoraphobie mit Panikstörung	F 40.01
	Generalisierte Angststörung	F 41.1
	Nicht näher bezeichnete Angststörung	F 41.9
Essstörungen (8 Items)	Bulimia nervosa	F 50.2
	Binge-Eating Disorder	F 50.9
Alkoholsyndrom (6 Items)	Alkoholmissbrauch	F 10.1
	Alkoholabhängigkeit	F 10.2

genform, Auswertungsschablonen bzw. -richtlinien und Auswertungsbeispiel im TM (S. 8) gesichert.

7.2 Reliabilität

Im TM finden sich dazu keine Angaben. Die Zuverlässigkeit des englischsprachigen „Patient Health Questionnaire“ konnte mehrfach gezeigt werden (vgl. z.B. Kroenke, Spitzer & Williams, 2001, 2002). Die interne Konsistenz der Depressivitätsskala des deutschen PHQ-D wird mit $\alpha = 0,88$ angegeben (Gräfe, Quenter, Buchholz, Wild, Zipfel, Herzog & Löwe, 2001; Löwe, Spitzer, Gräfe, Kroenke, Quenter, Zipfel, Buchholz, Witte & Herzog, 2002).

7.3 Validität

Das TM enthält dazu keine detaillierten Aussagen. Bei einer Stichprobe von N=528 Patienten (ambulante Klinikpatienten und Hausarztpatienten) wurden SKID-I (Wittchen, Zaudig & Fydrich, 1997) und PHQ-D verglichen und eine „gute Übereinstimmung“ (TM S. 9) gefunden. Validitätshinweise für die ameri-

kanischen Vorläufer liegen mehrfach vor (Kroenke, Spitzer & Williams, 2001, 2002). Löwe et al. (2002) berichten für den Depressionsteil des PHQ-D Korrelationen von 0,74 zur HADS (Herrmann, Buss & Snaith, 1995) und 0,73 zum WBI-5 (WHO, 1998). In ihrer Studie mit N = 501 Patienten wurde eine Sensitivität von 98 % und Spezifität von 80 % ermittelt, wobei als Kriterium Ergebnisse des SKID dienen.

7.4 Normierung

Die Überprüfung des PHQ-D wurde bei insgesamt 1619 Personen (ambulante Klinikpatienten, Hausarztpatienten) durchgeführt. Referenzwerte sind im TM nicht enthalten. Für die individuelle Diagnosestellung sind jedoch solche auch nicht erforderlich. Für den Skalensummenwert für Depressivität sind im Manual (S. 7) Wertebereiche zur Einordnung des Schweregrades vorgeschlagen.

8 Kritik

8.1 Der PHQ-D als ein ausdrücklich für Ärzte geschaffener Test berührt eine

wichtige „Gretchenfrage“: Dürfen (Allgemein-)Mediziner (ohne entsprechende Zusatzausbildung) psychologische Tests einsetzen? Die Gründe *dagegen* sind vielgestaltig (z.B. die fehlende fachliche Qualifikation) und m.E. oft berechtigt. *Dafür* sprechen jedoch im vorliegenden Falle auch einige Argumente, wie etwa die zahlreichen Fehl- bzw. Nichtdiagnosen, die (Haus-)Ärzte bei psychischen Störungen treffen. Ärzte sind jedoch auch bei psychischen Problemen meist primäre Ansprechpartner des Patienten. Im Sinne einer optimalen Patientenversorgung kann daher der PHQ-D den MedizinerInnen helfen, psychische Störungen zu erkennen und Entscheidungen zu treffen, z.B. eine weitergehende psychologische Diagnostik und ggf. Psychotherapie zu empfehlen. Der PHQ-D ist auch nicht das erste psychologische Testverfahren, das Ärzte als Nutzer anspricht. Die im TM benannte Zielgruppe Mediziner schließt zudem eine Nutzung durch Psychologen nicht aus. Insofern stellt der PHQ-D einen Test an der Schnittstelle zwischen Medizin und Psychologie, ein medizinpsychologisches Verfahren dar.

8.2 Die Entwicklung des deutschen und amerikanischen PHQ-D wurde durch ein Pharmaunternehmen (Pfizer) gefördert, ohne dass sich dies m.E. nachteilig auf die Gestaltung des Bogens auswirkt, etwa durch sehr auffällige Werbung. Im Gegenteil: Der Fragebogen kann kostenlos abgegeben werden.

8.3 Die Verbreitung des Bogens über die Testautoren oder die finanzierende Firma birgt auch Gefahren. Eine Kontrolle, dass Unbefugte den Test benutzen, ist m.E. nicht vollständig gewährleistet. Bei psychologischen Tests, die etwa über die Testzentrale (www.testzentrale.de) vertrieben werden, ist dies deutlich besser sichergestellt.

8.4 Der PHQ-D besticht durch eine graphisch exzellente, gut strukturierte, übersichtliche und hochwertige Gestaltung, die Vorbild für viele psychologische Tests sein kann.

8.5 Besonders gut gefiel mir die Kurzanleitung, die die wesentlichsten Fakten über den Test und die Auswertungsrichtlinien auf einer Doppelseite präsentiert und so eine sehr schnelle erste Orientierung und Einschätzung ermöglicht.

8.6 Das TM zeichnet sich durch eine hohe Lesbarkeit aus und kommt ohne Fachchinesisch daher. Insbesondere die Testdurchführung und -auswertung sind exzellent beschrieben. Das TM beschränkt sich auf nur 6 Textseiten, daher vermissen ich leider (ausführlichere) bedeutsame Informationen über das Verfahren, wie etwa zu dem amerikanischen Vorbild und zur Generierung der deutschen Version.

8.7 Nahezu unverzeihlich ist aus psychologischer Sicht der Verzicht auf die differenzierte Angabe jeglicher Gütekriterien (Reliabilität, Validität) und von Referenzwerten im TM.

8.8 Die Gütekriterien können nur bei den Items zur Depression (die dem amerikanischen PHQ-9 entsprechen) als gesichert angesehen werden. Hier konnten allerdings beeindruckende Ergebnisse gezeigt werden (Löwe et al., 2002): Der PHQ-D schnitt deutlich besser ab als andere Screeninginstrumente (HADS und WBI-5) und noch wesentlich besser als ärztliche Diagnosen. Für die Bereiche somatoforme Störungen, Angststörungen, Essstörungen und Alkoholmissbrauch steht eine Überprüfung (bzw. deren Veröffentlichung) der psychometrischen Kriterien bislang aus, sodass deren Anwendung nur mit Vorbehalt erfolgen sollte.

8.9 Im Vergleich zu den zur Validierung herangezogenen Instrumenten WBI-5 und HADS ist ein Vorteil des PHQ-D, dass er eine (syndromale) Diagnosestellung depressiver Störungen ermöglicht, ohne wesentlich umfangreicher zu sein.

8.10 Der PHQ-D ist dank seiner Konzentration auf die DSM-IV-Kriterien ein sehr ökonomisches Screeninginstrument, dem daher auch eine hohe Nützlichkeit zuzuschreiben ist. Laut Gräfe et al. (2001) stößt er auf eine hohe Akzeptanz bei Ärzten.

8.11 Insbesondere für Forschungsfragen ist die Bereitstellung von Normwerten für verschiedene Patiententypen und Normalpersonen zu fordern.

8.12 Das Vorliegen der amerikanischen und auch einer spanischen (Diez-Quevedo, Rangil, Sanchez-Planell, Kroenke & Spitzer, 2001) Version machen den PHQ-D für Studien in internationalen Forschungsverbänden interessant.

8.13 Der PHQ-D erfordert, dies mag sowohl ein Vor- als auch ein Nachteil sein, beim Anwender (Testleiter) nur wenig Voraussetzungen, wie etwa Kenntnisse über die Diagnosekriterien definierter psychischer Störungen.

8.14 Beim Patienten dagegen wird eine große Auskunftsbereitschaft, die Fähigkeit zur Selbstreflexion und eine entsprechend hohe Motivation vorausgesetzt.

8.15 Die im TM (S. 4) empfohlene Anwendung in der klinischen Praxis bei allen neu aufgenommenen Patienten und danach einmal im Jahr halte ich für fraglich.

8.16 Die Bezeichnung „Gesundheitsfragebogen“, die eine Übersetzung des amerikanischen „Health Questionnaire“ darstellt, ist irreführend, da mit der Konzentration auf einige psychische Störungen nur ein Teil der gemäß WHO-Definition zum Konstrukt „Gesundheit“ gehörenden Bereiche abgefragt wird.

8.17 Für die Kurzform gibt es keine Auswertungsschablone, was im Sinne der Objektivitätssicherung zu wünschen wäre. Die (m.E. nicht besonders komplizierte) Auswertung ist u. a. in der Fußzeile des Fragebogens beschrieben und damit auch für den Patienten sofort einsichtig und evtl. nachvollziehbar, trotz Umschreibungen wie "Maj Dep Syn", "And Dep Syn" bzw. "Pan Syn". Damit besteht eine (wenn auch minimale) Gefahr der Verfälschbarkeit der Ergebnisse.

8.18 Der Vergleich der nun vorliegenden zweiten Auflage mit der ersten aus dem Jahre 2001 zeigt nur wenige Veränderungen, sodass die ältere Version uneingeschränkt weiter verwendet werden kann. Fragebogen und Auswertungsschablonen wurden gar nicht modifiziert. Im Manual sind nun 43 statt 26 Literaturnachweise aufgenommen. Eine Tabelle zu Spezifität, Sensitivität und Trefferquote des amerikanischen Prime MD ist weggefallen. Wichtigste Änderung ist, dass neben einem Skalensummenwert für Depressivität nun auch für Stress und somatische Symptome die Bildung eines solchen Scores vorgeschlagen wird.

Wenige kleine Fehler wurden bei der Überarbeitung nicht korrigiert: So ist etwa bei der Schablone zur Auswertung

der Seite 2 des PHQ-D die umrahmte Seitenzahl (Anlegemerkmal) falsch platziert.

9 Empfehlung

Die Fragen des PHQ-D (Kurz- und Langform) zur Diagnostik depressiver Störungen sind sehr empfehlenswert. Ansonsten besitzt der PHQ-D ein m.E. überzeugendes Konzept, sollte aber aufgrund bislang ausstehender Validitätsbeweise der deutschen Version nur unter entsprechender Vorsicht eingesetzt werden.

Literatur

Diez-Quevedo, C., Rangil, T., Sanchez-Planell, L., Kroenke, K. & Spitzer, R. L. (2001). Validation and utility of the patient health questionnaire in diagnosing mental disorders in 1003 general hospital Spanish inpatients. *Psychosomatic Medicine*, 63, 679–686.

Goodwin, R. & Olfson, M. (2002). Self-perception of poor health and suicidal ideation in medical patients. *Psychological Medicine*, 32, 1293–1299.

Gräfe, K., Quenter, A., Buchholz, C., Wild, B., Zipfel, S., Herzog, W. & Löwe, B. (2001). Der Gesundheitsfragebogen für Patienten (PHQ-D) – Wie gut diagnostiziert ein Screeningfragebogen psychische Störungen? *Nervenheilkunde*, 20 (Suppl. 3), S99.

Herrmann, C., Buss, U. & Snaith, R.P. (1995). HADS-D: *Hospital anxiety and depression scale – Deutsche Version*. Bern, Huber.

Kroenke, K., Spitzer, R.L. & Williams, J.B.W. (2002). The PHQ-15: Validity of a New Measure for Evaluating the Severity of Somatic Symptoms. *Psychosomatic Medicine*, 64, 258–266.

Kroenke, K., Spitzer, R.L. & Williams, J.B.W. (2002). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, 16, 606–613.

Löwe, B., Spitzer, R.L., Gräfe, K., Kroenke, K., Quenter, A., Zipfel, S., Buchholz, C., Witte, S. & Herzog, W. (2002). Comparative validity of three screening questionnaires for DSM-IV depressive disorders and physicians diagnoses. *Journal of Affective Disorders*, in press, (<http://www.elsevier.com/locate/jad>)

Rollman, B.L., Hanusa, B.H., Lowe, H.J., Gilbert, T., Kapoor, W.N. & Schulberg, H.C. (2002). A randomized trial using computerized decision support to improve treatment of major depression in primary care. *Journal of General Internal Medicine*, 17, 493–503.

Spitzer, R.L., Kroenke, M.D., Williams, J.B. & The patient health questionnaire primary care study group (1999). Validation and Utility of a self-report version of PRIME-MD. The PHQ primary care study. *Journal of the American Medical Association*, 282, 1737–1744.

Spitzer, R.L., Williams, J.B., Kroenke, K., Linzer, M., DeGruy, F.V., Hahn, S.R., Brody, D., & Johnson, J.G. (1994). Utility of a new procedure

re for diagnosing mental disorders in primary care: the PRIME-MD 1000 study. *Journal of the American Medical Association*, 272, 1749–1756.

Thompson, C., Kinmonth, A.L., Stevens, L., Peveler, R.C., Stevens, A., Ostler, K.J., Pickering, R.M., Baker, N.G., Henson, A., Preece, J., Cooper, D. & Campbell, M.J. (2000). Effects of a clinical-practice guideline and practice-based education on detection and outcome of depression in primary care: Hampshire Depression Project randomised controlled trial. *Lancet*, 355, 185–191.

Wittchen, H.U., Zaudig, M. & Fydrich, T. (1997). *Strukturiertes klinisches Interview für DSM-IV (SKID), Achse I*. Göttingen, Hogrefe.

World Health Organization (WHO) (1998). *Wellbeing Measures in Primary Health Care.: The DepCare Project*. Copenhagen, WHO.

Hendrik Berth, Dresden
