



Psychotherapie Psychosomatik Medizinische Psychologie

Herausgeber und Redaktion

Elmar Brähler, Leipzig
Franz Caspar, Bern
Martina de Zwaan, Erlangen
Wolfgang Herzog, Heidelberg
Fritz Hohagen, Lübeck
Uwe Koch, Hamburg
Bernhard Strauß, Jena
Stephan Zipfel, Tübingen

ehemalige Herausgeber

Dieter Beckmann, Gießen
Walter Bräutigam, Berlin
Erdmuthe Fikentscher, Halle
Edgar Heim, Bern
Horst Kächele, Ulm
Margit v. Kerekjarto, Hamburg
Ernst Kretschmer, Tübingen
Wolfgang Senf, Essen
Hubert Speidel, Kiel
Michael v. Rad, München
Walter Th. Winkler, Gütersloh

Beirat

Jürgen Barth, Bern
Uwe Berger, Jena
Manfred E. Beutel, Mainz
Martin Bohus, Mannheim
Sabine Herpertz, Rostock
Stephan Herpertz, Bochum
Peter Herschbach, München
Andreas Hinz, Leipzig
Helmut Kirchmann, Jena
Falk Leichsenring, Gießen
Klaus Lieb, Mainz
Bernd Löwe, Hamburg
Christoph Nikendei, Heidelberg
Ursula Pauli-Pott, Gießen
Christoph Schmeling-Kludas, Rosengarten
Holger Schulz, Hamburg
Ulrich Schweiger, Lübeck
Rolf-Dieter Stieglitz, Basel
Birgit Watzke, Hamburg
Almut Zeeck, Freiburg

Organschaft

**Deutsches Kollegium für
Psychosomatische Medizin**

Verlag

Georg Thieme Verlag KG
Rüdigerstraße 14
70469 Stuttgart
www.thieme.de/fz/ppmp
www.thieme-connect.de/ejournals

Sonderdruck

© Georg Thieme Verlag KG
Stuttgart · New York

Nachdruck nur mit Genehmigung
des Verlags

Diagnostik psychischer Störungen: „Gold“- oder „Blech“-Standard?

Kritische Randbemerkungen zu standardisierten diagnostischen Instrumenten und der Bewertung klinischer Routinediagnosen

Diagnosing Mental Disorders: Gold or Tin Standard?

Critical Comments on Standardized Diagnostic Interviews and Clinical Routine Diagnoses

Autoren

Susanne Knappe, Juliane Runge, Katja Beesdo, Frank Jacobi, Hans-Ulrich Wittchen

Institut

Institut für Klinische Psychologie und Psychotherapie, Technische Universität Dresden

Key words

- CIDI
- gold-standard
- validity
- standardized diagnostic

Zusammenfassung

In einer Stichprobe psychosomatischer Patienten fanden Becker et al. [1] eine geringe Übereinstimmung zwischen klinischen Routinediagnosen und computergestützten DSM-IV-Diagnosen, die durch ein weltweit etabliertes standardisiertes diagnostisches Interview (CIDI, Composite International Diagnostic Interview) erstellt wurden. Sie schlussfolgern, dass eine unzureichende Validität des CIDI im psychosomatischen Kontext vorliege. Der aktuelle Beitrag kommt zu dem Ergebnis, dass die Ergebnisse von Becker et al. eine derartige Schlussfolgerung nicht rechtfertigen. Neben erheblichen methodischen Mängeln wird insbesondere die Annahme der Autoren infrage gestellt, dass klinische Routinediagnosen im psychosomatischen Kontext einen Goldstandard darstellen.

Abstract

In a convenience sample of psychosomatic inpatients, Becker et al. [1] reported poor diagnostic congruence between clinical diagnoses from routine examinations and DSM-IV diagnoses derived from the use of a computer-assisted standardized diagnostic interview (CIDI). The authors conclude that there is insufficient validity of such interviews in psychosomatic settings. This paper emphasizes that the results of Becker et al. do not justify this conclusion. Aside from substantial methodological deficits, we question the authors' assumption that clinical diagnoses could be regarded as an appropriate „gold standard“.

Einleitung

In ihrem Artikel „Standardisierte Diagnosefindung: Computergestützte (CIDI) – versus klinische Diagnosestellung an einer psychosomatischen Stichprobe“ [1] untersuchten Becker et al. anhand einer Stichprobe psychosomatischer Patienten (n = 230) die Übereinstimmung zwischen klinischen Routinediagnosen und Diagnosen, die durch ein computergestütztes standardisiertes Interview (Composite International Diagnostic Interview [CIDI]; [2]) erstellt wurden. Die Autoren berichteten eine geringe Übereinstimmung der beiden Verfahren und schlussfolgerten, dass die Validität der computergestützten Diagnosen im psychosomatischen Kontext anzuzweifeln ist. Aufgrund vieler methodischer Probleme und Unzulänglichkeiten ist diese Schlussfolgerung zu hinterfragen.

Mit diesem Beitrag möchten wir einige Grundprobleme dieser Arbeit ansprechen und auf einer

breiteren Basis von Informationen einige zentrale Aspekte der Überprüfung der Gütekriterien des CIDs kritisch diskutieren. Dabei gehen wir ein auf:

1. den Einsatz des CIDs im klinischen Kontext,
2. die Berücksichtigung von Komorbiditäten in der klinischen Diagnostik und bei der Validierung standardisierter Verfahren,
3. Reliabilität und Validität diagnostischer Entscheidungen und
4. die Verwendung klinischer Diagnosen als Validierungskriterium im Sinne eines „Goldstandards“.

Einsatz des CIDI

Das CIDI ist ein vollstandardisiertes Interview zur Erfassung psychischer Störungen und zeichnet sich wie ähnliche standardisierte bzw. strukturierte Interviews (z. B. SKID: [3]; SCAN: [4]; DIPS: [5]) dadurch aus, dass eine große Breite an Psychopathologie (und nicht nur die aktuell im Vor-

eingereicht 21.4.2007
akzeptiert 19.7.2007

Bibliografie

DOI 10.1055/s-2007-986197
Online-Publikation:
24. September 2007
Psychother Psych Med 2008;
58: 72 – 75 © Georg Thieme
Verlag KG Stuttgart · New York ·
ISSN 0937-2032

Korrespondenzadresse

Dipl.-Psych. Susanne Knappe
Institut für Klinische Psychologie und Psychotherapie,
Technische Universität Dresden
Chemnitzner Straße 46
01187 Dresden
knappe@psychologie.tu-
dresden.de

dergrund stehende Symptomatik) umfassend exploriert wird. Standardisierte Interviews sind eine Antwort auf die unzulängliche Reliabilität und Validität klassischer „freier“ psychiatrischer Diagnosen. Sie beruhen auf der deskriptiven Formalisierung der neuen diagnostischen Klassifikationssysteme DSM-III, DSM-III-R, DSM-IV [6–8] und ICD-10 [9, 10]. Infolge der damit verbundenen Abkehr von einer nosologisch orientierten Diagnostik hin zu einer deskriptiven Diagnostik psychischer Störungen im Rahmen des DSM-III und DSM-III-R wurde das CIDI 1990 durch die WHO entwickelt mit dem Ziel, psychopathologische Symptome und Störungsbilder möglichst reliabel, valide und vollständig zu erfassen [2, 11]. Voraussetzung für den Einsatz ist ein standardisiertes WHO-Training zur Durchführung und Auswertung des Interviews. Das CIDI wurde für den Einsatz in epidemiologischen Studien konzipiert, wird jedoch auch erfolgreich in klinischen Studien bei nichtpsychotischen Störungen eingesetzt [12]. Es orientiert sich an den Kriterien psychischer Störungen des DSM-IV, erfasst aber auch Kriterien des ICD-10, soweit diese davon abweichen, und leitet von den erfassten Symptomen über ein komplexes Regelsystem die Diagnosen psychischer Störungen ab. Das CIDI diagnostiziert somit auch „psychosomatische“ Störungen, aber nur soweit sie den Diagnosen des DSM-IV bzw. des Kapitel F des ICD-10 entsprechen. Psychische Symptome, die auf einen medizinischen Krankheitsfaktor oder Substanzen zurückzuführen sind, werden allerdings durch das CIDI nur als Residualkategorie erfasst und dementsprechend codiert. Diese wurden jedoch bei Becker et al. [1] allem Anschein nach nur bei den affektiven Störungen in die Auswertung mit einbezogen.

Probleme der Arbeit von Becker et al.

Gültigkeitsbereich: Bei der Verwendung diagnostischer Verfahren sollte darauf geachtet werden, dass das Verfahren für die vorliegende Stichprobe geeignet und geprüft ist. Das CIDI ist in erster Linie für Erwachsene bis 65 Jahre validiert. Becker et al. verwenden jedoch eine Stichprobe im Alter von 18–77 Jahren. Das Verfahren ist jedoch für über 65-Jährige nicht validiert [13]. Eine differenzierte Analyse der Übereinstimmungsraten für Substichproben von 18–65 und 66–77 Jahren könnte Hinweise darauf geben, ob geringere Übereinstimmungsraten für die ältere Substichprobe vorliegen.

Komorbidität: Becker et al. kritisieren, dass das CIDI im Fall von Mehrfachdiagnosen (Komorbidität) keine Unterscheidung zwischen Haupt- und Nebendiagnosen trifft. Dies ist richtig, denn nach DSM-IV bzw. ICD-10 erfordert eine Einteilung in Haupt- und Nebendiagnosen grundsätzlich das Urteil eines klinischen Experten und ist somit abhängig von versorgungs- und therapiebezogenen Erwägungen, die im Vordergrund der klinischen Aufmerksamkeit der jeweiligen Behandlungseinrichtung zum Zeitpunkt der Interventionsentscheidung stehen [8]. Ziel standardisierter Interviews ist eine möglichst vollständige, DSM-/ICD-10 basierte reliable Beschreibung der Symptome und des Zeitverlaufs, aus der eine objektive (computerisierte) Diagnose abgeleitet wird. Die Ableitung einer „Hauptdiagnose“ ist – nach den basalen Regeln des DSM-IV – eine situations- und kontextabhängige bezogene klinische Bewertung; sie kann nicht von einem Instrument geleistet werden.

Der 1980 eingeführte deskriptive diagnostische Ansatz (DSM-III) sowie die Reduktion diagnostischer Hierarchien (DSM-III, DSM-III-R) führten zwangsläufig zu einer deutlichen Zunahme des Anteils von Personen mit multiplen bzw. komorbiden Diagnosen. Eine Zusammenfassung heterogener Störungskomplexe, wie es früher z. B. bei breiten neurotischen Hauptdiagnosen üb-

lich war, wurde als Hauptursache der vollkommen unzulänglichen Reliabilität und Validität grundsätzlich vermieden. Dieser Ansatz sollte einerseits zu einer Verbesserung der Reliabilität führen, andererseits sollte er ermöglichen empirisch zu prüfen, welche Symptomstrukturen und daraus abgeleitete diagnostische Algorithmen „nützlicher“ sind (z. B. in Bezug auf das Ansprechen auf eine Therapie oder den Erkrankungsverlauf) [10]. Wie auch Becker et al. berichten, ist die Anzahl klinischer Diagnosen bei einem Patienten zumeist wesentlich geringer, als die bei einer standardisierten Diagnostik. Ursache kann u. a. sein, dass Kliniker „ähnliche“ Diagnosen, wie z. B. mehrere Angststörungen, zusammenfassen [1] oder eine hypothesengeleitete und dadurch unvollständige Exploration der Psychopathologie, z. B. durch voreilige Festlegung auf eine Hauptdiagnose ohne dass andere bzw. differenzialdiagnostische Aspekte noch abgeprüft werden. Dies führt nicht nur zu einer geringeren Anzahl diagnostizierter Störungen, sondern erhöht auch die Gefahr von Fehlklassifikationen, da bei einer Reihe von Störungen Überschneidungen einiger Symptome bestehen können [14]. Dies wird dem substanzwissenschaftlichen Anliegen und den Regeln von DSM-IV und ICD-10 nicht gerecht. Eine Reihe von Studien [15, 16] belegen z. B. eine höhere Anzahl von Angststörungen im Rahmen der standardisierten Diagnostik, selbst wenn klinische Diagnostiker explizit darauf hingewiesen wurden, alle möglichen, auch Lebenszeitdiagnosen, neben den aktuellen Diagnosen anzugeben. Dies wird darauf zurückgeführt, dass es erstens Klinikern ohne Instrument unmöglich ist, alle 500 Kriterienseiten des DSM-IV präsent zu haben; zweitens benutzt ein Kliniker immer Heuristiken, die sein Explorationsverhalten effizient gestalten, ungeachtet der Richtigkeit in Bezug auf DSM-IV und ihrer Angemessenheit.

Reliabilität und Validität des CIDI. Das CIDI, wie auch andere standardisierte diagnostische Verfahren, erwies sich in bisherigen Studien für die meisten Diagnosen (Ausnahmen: akute Psychosen, akute bipolare Störungen, Demenzerkrankungen) als äußerst reliabel im Hinblick auf die zeitliche Stabilität bei der Untersuchung durch zwei voneinander unabhängige Interviewer [2, 17–19] – eine notwendige Voraussetzung für eine hohe Validität des Instruments. Eine derart hohe Übereinstimmung ließ sich unseres Wissens nicht mit klinischen Diagnosen erzielen [2, 11]. Die Validität kann aber nie höher als die Reliabilität sein. Bezüglich der Übereinstimmung zwischen klinischen und standardisierten Diagnosen führen Becker et al. einige Studien mit geringen Übereinstimmungsraten an. Das mag so sein, aber wer hat recht? Für eine kritische Diskussion ist zunächst zu berücksichtigen, dass für *beide* zum Vergleich herangezogene Verfahrensweisen zunächst die Reliabilität vergleichend zu prüfen ist. Ist diese hinreichend – und benutzt sowohl für klinische wie auch standardisierte Entscheidungen entsprechende Methoden, erhält man üblicherweise durchaus zufriedenstellende und höhere Übereinstimmungsquoten, z. B. zwischen CIDI und klinischen Urteilen [4, 16] oder zwischen dem CIDI und anderen standardisierten Interviews oder Checklisten [20, 21].

Klinische Diagnosen als Validierungskriterium. Bei der Wahl klinischer Diagnosen als „Goldstandard“ ist zu beachten, dass ihre Reliabilität verglichen mit standardisierten Interviews in der Regel gravierend eingeschränkt ist [11, 12]. Deshalb werden klinische Diagnosen mitunter auch als „Blech“-Standard bezeichnet. Gründe für eine mangelnde Reliabilität liegen dabei auf allen Ebenen des diagnostischen Prozesses: der Befunderhebung, der Symptombenennung und der Zuordnung von Symptomen zu Syndromen und Diagnosen [12] im Sinne unterschiedlicher Fra-

gen oder einer unvollständigen Exploration der Psychopathologie (s.o.) sowie einer abweichenden Interpretation diagnostischer Kriterien. Eine hohe Reliabilität ist jedoch Voraussetzung der Validität. Zur Steigerung der Güte klinischer Diagnosen forderte Meehl [22] bereits 1954 explizite diagnostische Kriterien und standardisierte Messmethoden für die Diagnostik psychischer Störungen.

Die maximal zu ermittelnde Validität des CIDI ist abhängig von der Güte des Validitätskriteriums, in diesem Fall klinischer Urteile. Die Validität klinischer Urteile lässt sich durch spezifische Methoden, wie z. B. die LEAD-Methode [23], verbessern, bei der dem Diagnostiker möglichst viele längsschnittliche Informationen vorliegen, und auf deren Basis eine Konsensusdiagnose getroffen wird. Die Übereinstimmung zwischen derart gewonnenen Diagnosen und computergestützten Diagnosen ist deutlich höher (vgl. [16]).

Aufgrund der Komplexität des DSM-IV und ICD-10 im Sinne der Symptomvielfalt und Codierungsregeln ist eine geringere Güte klinischer Diagnosen zwangsläufig, wenn keine standardisierten Hilfen oder – weniger befriedigend – Checklisten eingesetzt werden. Eine stärkere Ausrichtung auf spezifische diagnostische Kriterien kann die Reliabilität klinischer Diagnosen deutlich erhöhen [15,16]. Daher ist im Rahmen der Verwendung klinischer Urteile als Validitätskriterium die Absicherung ihrer größtmöglichen Reliabilität und Validität unerlässlich.

Becker et al. geben keine Informationen zur Reliabilität ihrer klinischen Diagnosen, sodass eine Abschätzung ihrer Güte unmöglich ist. Aufgrund einer möglicherweise mangelhaften Ausrichtung der Diagnosen auf die diagnostischen Kriterien des DSM-IV bzw. ICD-10 und fehlender anderer Verfahren zur Steigerung der Validität ist eine geringe Reliabilität und Validität der klinischen Urteile anzunehmen.

Eine weitere notwendige Voraussetzung einer Validierung ist, dass sich die Konstrukte des Validierungskriteriums und –objekts möglichst gut entsprechen. Im Rahmen des Vergleichs klinischer und standardisierter Diagnosen ist darauf zu achten, dass sich beide auf dieselben diagnostischen Kriterien (z. B. DSM-IV) und dieselbe zeitliche Dimension (12-Monats-Diagnosen vs. aktuelle Diagnosen) beziehen. Auch diese Bedingung ist bei Becker et al. nicht erfüllt. An dieser Stelle ist die von Becker et al. getroffene Annahme der Stabilität von Diagnosen in der psychosomatischen Behandlung über 2–8 Wochen Behandlungszeit infrage zu stellen. Studien zeigen, dass psychische Störungen oftmals Fluktuationen in der Symptomatik und im Schweregrad unterliegen [24–26], und auch bereits kurze Zeit nach Therapiebeginn deutliche Verbesserungen der Symptomatik vorliegen können [27]. Bei Becker et al. betrug das Zeitintervall zwischen der klinischen und der computergestützten Diagnosefindung bis zu acht Wochen. CIDI-Diagnosen zu Beginn der Therapie können daher deutlich von den (Entlassungs-)Diagnosen der klinischen Urteiler abweichen.

Schlussfolgerungen

In Anbetracht der aufgeführten Schwierigkeiten und Unzulänglichkeiten bei Becker et al. sowie weiterer methodischer Probleme und einer unvollständigen, teilweise nicht nachvollziehbaren oder fehlerhaften Darstellung der Ergebnisse, erscheint uns die Interpretation der geringen Übereinstimmung zwischen klinischen und computergestützten Diagnosen im Sinne einer mangelnden Validität des CIDI als falsch. Der Nutzen des CIDI

als Prototyp standardisierter Diagnoseverfahren für den klinischen Einsatz wird dadurch nicht infrage gestellt. In einer kaum überschaubaren Anzahl von Studien der letzten 20 Jahre hat sich das CIDI als reliables und in Hinblick auf Außenkriterien relativ valides Diagnostikum bewährt. Dabei ist das Ausmaß der Reliabilität und Validität keineswegs befriedigend, und Gegenstand kontinuierlicher empirisch fundierter Verbesserungsbemühungen. Im Vergleich zu allen Alternativen wird das CIDI manchmal auch als „Einäugiges unter Blinden“ eingeordnet. Klinische Diagnosen ohne Strukturierung sind hingegen notorisch unreliabel, und somit zwangsläufig auch in ihrer Validität (bezüglich Indikation, Verlauf, Therapie) hochgradig eingeschränkt. Zudem lässt sich klinische Routinediagnostik nicht „empirisch“ verbessern. Zur Überwindung dieser Mängel wurde die standardisierte und computergestützte Diagnostik entwickelt. Dass diese nun anhand klinischer Diagnosen validiert wird, ist paradox und ein Rückschritt in die 70er-Jahre. Sinnvoll wäre es gewesen, andere standardisierte oder strukturierte Verfahren als Validierungskriterium heranzuziehen und so einen Beitrag zur Verbesserung der diagnostischen Standards zu leisten.

Fazit für die Praxis

Die Interpretation der geringen Übereinstimmung zwischen klinischen und computergestützten Diagnosen im Sinne einer mangelnden Validität des CIDI erscheint uns als falsch. Der Nutzen des CIDI als Prototyp standardisierter Diagnoseverfahren für den klinischen Einsatz wird dadurch nicht infrage gestellt. In einer kaum überschaubaren Anzahl von Studien der letzten 20 Jahren hat sich das CIDI als reliables und in Hinblick auf Außenkriterien relativ valides Diagnostikum bewährt. Dabei ist das Ausmaß der Reliabilität und Validität keineswegs befriedigend, im Vergleich zu allen Alternativen wird das CIDI manchmal auch als „Einäugiges unter Blinden“ eingeordnet. Klinische Diagnosen ohne Strukturierung sind hingegen notorisch unreliabel, und somit zwangsläufig auch in ihrer Validität (bezüglich Indikation, Verlauf, Therapie) hochgradig eingeschränkt.

Literatur

- 1 Becker J, Kocalevent R-D, Rose M et al. Standardisierte Diagnosefindung: Computergestützte (CIDI) – versus klinische Diagnosestellungen an einer psychosomatischen Stichprobe. *Psychother Psych Med* 2006; 56: 5–14
- 2 Wittchen HU, Pfister H. DIA-X-Interview: Manual für Screening-Verfahren und Interview; Interviewheft Längsschnittuntersuchung (DIA-X-Lifetime); Ergänzungsheft (DIA-X-Lifetime); Interviewheft Querschnittsuntersuchung (DIA-X-12 Monate); Ergänzungsheft (DIA-X-12 Monate); PC-Programm zur Durchführung des Interviews (Längs- und Querschnittsuntersuchung); Auswertungsprogramm. Frankfurt: Swets & Zeitlinger, 1997
- 3 Wittchen H-U, Zaudig M, Fydrich T. SKID. Strukturiertes Klinisches Interview für DSM-IV. Achse I und II. Handanweisung. Göttingen: Hogrefe, 1997
- 4 World Health Organization. Schedules for Clinical Assessment in Neuropsychiatry (SCAN). Geneva: World Health Organization, 1992
- 5 Schneider S, Margraf J. DIPS für DSM-IV. Diagnostisches Interview bei psychischen Störungen. 3. Auflage. Heidelberg: Springer, 2006
- 6 American Psychiatric Association. DSM-III. Diagnostic and statistical manual of mental disorders (3rd ed.). Washington, DC: Author, 1980
- 7 American Psychiatric Association. DSM-III-R. Diagnostic and statistical manual of mental disorders (3rd ed. rev.). Washington, DC: Author, 1987

- 8 *American Psychiatric Association*. DSM-IV. Diagnostic and statistical manual of mental disorders (4th ed.). Washington, DC: Author, 1994
- 9 *World Health Organization*. Tenth Revision of the International Classification of Diseases Chapter V (F): Mental and Behavioral Disorders (including disorders of psychological development). Clinical descriptions and diagnostic guidelines. Geneva: World Health Organization, 1991
- 10 *Wittchen H-U*. Epidemiology of mental illness. In: Smelser NJ, Baltes PB (eds): International encyclopedia of the social and behavioral sciences. Oxford: Pergamon, 2001: 9654–9661
- 11 *Wittchen H-U*. Reliability and validity studies of the WHO-Composite International Diagnostic Interview (CIDI): A critical review. *J Psych Res* 1994; 28 (1): 57
- 12 *Wittchen H-U, Lachner G*. Klassifikation. In: Ehlers A, Hahlweg K (Hrsg): Enzyklopädie der Psychologie. Themenbereich D Praxisgebiete, Serie 2 Klinische Psychologie (Vol. 1, pp. 3–67). Göttingen: Hogrefe, 1996
- 13 *Knäuper B, Wittchen H-U*. Heuristics used by older respondents to answer standardized mental health questions. In: Warnecke R (ed): Health survey research methods. Hyattsville, MD: U. S. Department of Health and Human Services, 1996
- 14 *Andrews G, Peters L*. The psychometric properties of the Composite International Diagnostic Interview. *Soc Psychiatry Psychiatr Epidemiol* 1998; 33: 80–88
- 15 *Peters L, Andrews G, Cottler LB, Chatterji S*. The Composite International Diagnostic Interview post-traumatic stress disorder module: Preliminary data. *Int J Methods Psychiatr Res* 1996; 6 (3): 167
- 16 *Reed V, Gander F, Pfister H et al*. To what degree the Composite International Diagnostic Interview (CIDI) correctly identifies DSM-IV disorders? Testing validity issues in a clinical sample. *Int J Methods Psychiatr Resh* 1998; 7: 142–155
- 17 *Lachner G, Wittchen H-U, Perkonig A et al*. Structure, content and reliability of the Munich Composite International Diagnostic Interview (M-CIDI) substance use sections. *Eur Addict Res* 1998; 4 (1–2): 28–41
- 18 *Rubio-Stipec M, Peters L, Andrews G*. Test-retest reliability of the computerized CIDI (CIDI-auto): Substance abuse modules. *Substance Abuse* 1999; 20 (4): 263
- 19 *Wittchen HU, Kessler RC, Üstün TB*. Properties of the Composite International Diagnostic Interview (CIDI) for measuring mental health outcome. In: Tansella M, Thornicroft G (eds): Mental health outcome measures (2nd ed). London: Gaskell, 2001: 212–227
- 20 *Jordanova V, Wickramesinghe C, Gerada C, Prince M*. Validation of two survey diagnostic interviews among primary care attendees: A comparison of CIS-R and CIDI with SCAN ICD-10 diagnostic categories. *Psychol Med* 2004; 34 (6): 1013
- 21 *Haro JM, Arbabzadeh-Bouchez A, Brugha TS et al*. Concordance of the Composite International Diagnostic Interview version 3.0 (CIDI 3.0) with standardized clinical assessments in the WHO world mental health surveys. *Int J Methods Psychiatr Res* 2006; 15 (4): 167–180
- 22 *Meehl PE*. Clinical versus statistical prediction; a theoretical analysis and a review of the evidence: University of Minnesota, Press Minneapolis, 1954
- 23 *Spitzer RL*. Psychiatric diagnosis: Are clinicians still necessary? *Compr Psychiatry* 1983; 24: 399–411
- 24 *Balon R*. Anxiety across the life span: Epidemiological evidence and treatment data. *Depress Anxiety* 2001; 13: 184–189
- 25 *Wittchen H-U, Lieb R, Pfister H, Schuster P*. The waxing and waning of mental disorders: Evaluating the stability of syndromes of mental disorders in the population. *Compr Psychiatry* 2000; 41 (2, Suppl 1): 122–132
- 26 *Arnold IA, de Waal MWM, Eekhof JAH, van Hemert AM*. Somatoform disorder in primary care: Course and the need for cognitive-behavioral treatment. *J Psychosom Res* 2006; 47 (6): 498
- 27 *Ilardi SS, Craighead WE*. The role of nonspecific factors in cognitive-behavior therapy for depression. *J Clin Psych* 1994; 1 (2): 138–156