

The Munich Diagnostic Checklist for the Assessment of DSM-III-R Personality Disorders for Use in Routine Clinical Care and Research

Thomas Bronisch, Diego Garcia-Borreguero, Susan Flett, Rainer Wolf, and Wolfgang Hiller

Max-Planck-Institute of Psychiatry, Clinical Institute, Kraepelinstrasse 10, W-8000 Munich 40, Federal Republic of Germany

Summary. Diagnostic checklists for the assessment of DSM-III-R Axis I diagnoses have proven to be a reliable and feasible instrument in research and routine clinical care. Therefore, a checklist for the assessment of the DSM-III-R Personality Disorders (MDCL-P) has been developed. An English version of the MDCL-P is available. The MDCL-P has been tested for reliability in a test-retest design. The average duration of the interview was 36 min. Of the patients, 48% received a diagnosis of at least one personality disorder. The Kappa value concerning the distinction personality disorder as opposed to no personality disorder was 0.62. The range of Kappa values of specific personality disorders, which were diagnosed at least five times, was from 0.35 to 0.79.

Key words: Diagnostic checklist – Personality disorders – DSM-III-R – Reliability

Introduction

Since the introduction of DSM-III in 1980, instruments for the assessment of personality disorders according to DSM-III and its revised form, DSM-III-R, have been developed (Personality Disorders Questionnaire, PDQ, [11], PDQ-R, [12]; Millon Clinical Multiaxial Inventory, MCMI, [17, 18]; Structured Interview for DSM-III Personality Disorders, SIDP, [21], SIDP-R, [20]; Personality Disorder Examination, PDE, [15, 16]; Structured Clinical Interview for DSM-III Disorders, Axis II, SCID, [24, 25]). Most of these instruments have demonstrated a satisfactory to excellent test-retest reliability. However, a comparison between a self-rating questionnaire (PDQ-R) and a structured interview (SIDP-R, PDE) revealed a high sensitivity but a low specificity of the questionnaire [13, 14]. Furthermore, the structured interviews on which reliability data have been published (SIDP, PDE), seem to be of restricted clinical and scientific value since they last 1.5–4 h or more [16, 20].

To augment diagnostic investigations and the assessment of criteria-related diagnoses in routine clinical care, the Munich Diagnostic Checklists (MDCL) have been developed for most of the common psychiatric disorders according to DSM-III-R [10]. Test-retest studies for the MDCL showed that a satisfactory to excellent diagnostic agreement between two independent interviewers using a test-retest design can be reached for most diagnostic categories [9, 10]. Recently, an additional version of the MDCL has been developed for the assessment of all DSM-III-R personality disorders (MDCL-P). The MDCL-P is used in routine clinical care at the Crisis Intervention Ward at the Max-Planck-Institute of Psychiatry (Munich, FRG) and a test-retest reliability study has been undertaken.

Subjects and Methods

Description of the MDCL-P

The MDCL-P is a 28-page pocket-sized booklet. The front page contains a short description of the main characteristics of personality disorders and summarizes the content of the booklet. On pages 2–27 each of the individual personality disorders given by DSM-III-R as well as the two optional personality disorders (self-defeating personality disorder, sadistic personality disorder) are evaluated. Two pages are reserved for each personality disorder. The last page contains a summary of personality disorders which have been assessed in the interview. Herein one can also code the residual category of personality disorders not otherwise specified, which can be diagnosed if none of the specific personality disorders apply despite the clinical impression that a severe and longstanding disturbance of personality is present in the patient being examined.

The design and structure of the MDCL-P are illustrated in Fig. 1 with the section evaluating the borderline personality disorder (pages 10–11). It can be seen that all the criteria are listed which must be evaluated in order to decide whether a borderline personality disorder

Borderline Personality Disorder

A pervasive pattern of *instability of mood, interpersonal relationships, and self-image*.

The disturbance begins at early adulthood and is present in a variety of contexts.

● Evaluate the individual personality traits

- | | No | Probable | Yes |
|---|--------------------------|--------------------------|--------------------------|
| (1) a pattern of unstable and intense <i>interpersonal relationships</i> characterized by alternating between extremes of overidealization and devaluation. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| (2) impulsiveness in at least two areas that are potentially <i>self-damaging</i> (e.g., spending, sex, substance use, shoplifting, reckless driving, binge eating; do <u>not</u> include suicidal or self-mutilating behavior covered in #5). | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| (3) <i>affective instability</i> (i.e., marked shifts from baseline mood to depression, irritability, or anxiety, usually lasting a few hours and only rarely more than a few days). | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| (4) inappropriate, intense <i>anger</i> or lack of control of anger (e.g., frequent displays of temper, constant anger, recurrent physical fights). | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| (5) recurrent <i>suicidal</i> threats, gestures, or behavior, or self-mutilating behavior. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| (6) marked and persistent <i>identity disturbance</i> manifested by uncertainty about at least two of the following: * self-image * sexual orientation * type of friends desired * long-term goals or career choice * preferred values | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |

der can be diagnosed or not. The clinician can code each trait with "No", "Yes", or "Probably" (whenever a trait cannot be assessed definitely but is judged to be present with a sufficient degree of confidence). During exploration, the list can be used to systematically screen for all traits and the diagnostician can decide if the criteria required for the diagnosis of a borderline personality disorder are fulfilled.

The MDCL-P differs from structured interviews in that its use is not restricted to face-to-face explorations. The clinician is also free to include third-party information (e.g. from relatives, friends, or former medical reports of direct observations of the patient's behavior).

The MDCL-P does not require standardized questioning, probing or a fixed order of progression. The diagnostician is free to focus on the most prominent complaints as verbalized by the patient. He can also focus on the most prominent personality features which the patient has demonstrated during the interview. If there are hints for one or more personality disorders the diagnostician can unfold the corresponding part of the booklet and ask for specific criteria of the personality disorder in question.

The completed checklist can be enclosed with an in-patient's clinical records and serves as a standardized and more objective documentation. It also might serve as a teaching tool for residents, students, and nonpsychiatric

- | | No | Probable | Yes |
|--|--------------------------|--------------------------|--------------------------|
| (7) chronic feelings of <i>emptiness</i> or <i>boredom</i> . | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| (8) frantic efforts to avoid real or imagined <i>abandonment</i> (do <u>not</u> include suicidal or self-mutilating behavior covered in #5). | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |

If the principal criterion is met together with at least five from (1) to (8):

Borderline Personality Disorder

3 0 1.8 3

☐ not met

Probable ☐

met ☐

Fig. 1. The Borderline Personality Disorder according to the Munich Diagnostic Checklists for DSM-III-R Personality Disorders (MDCL-P)

clinicians. Another use might be to select specific groups of patients for investigations in psychiatric research or to supplement other more structured instruments that focus on specific diagnoses. If the checklist were incorporated in clinical records, future research using retrospective chart reviews might be facilitated. In addition, a manual for the use of the MDCL-P exists [3]. English versions of the MDCL-P and the manual are available.

Sample Characteristics

The subjects of the MDCL-P test-retest study were patients who had been consecutively admitted to the Crisis Intervention Ward of the Max-Planck-Institute of Psychiatry between July and December 1990. Exclusion criteria were acute or chronic organic mental disorders. Patients were evaluated after the acute symptomatology had subsided (values of > 30 in the Adjective Mood Scale (AMS) [29], no drugs for treatment of the withdrawal syndrome in case of an alcohol or drug dependence). Other inclusion criteria were a minimum age of 21 years and the patient's agreement to participate in the study (i.e. allowing his data to be used for scientific purposes and returning for a second interview with a different diagnostician). All subjects were informed about the reasons for being interviewed twice and were instructed to consider both explorations as independent (i.e. to give

Table 1. Base rates of MDCL-P diagnoses (DSM-III-R)^a

| Diagnosis | No. of patients (%) ^b |
|-------------------------|----------------------------------|
| Paranoid PD | 8 (13) |
| Schizoid PD | 3 (5) |
| Schizotypal PD | 4 (7) |
| Antisocial PD | 1 (2) |
| Borderline PD | 10 (17) |
| Histrionic PD | 3 (5) |
| Narcissistic PD | 1 (2) |
| Avoidant PD | 23 (38) |
| Dependent PD | 14 (23) |
| Obsessive-compulsive PD | 1 (2) |
| Passive-aggressive PD | 4 (7) |
| Any PD | 29 (48) |
| Overall total | 72 (120) |

^a Total number of positive diagnoses; cases with disagreement were counted as positive

^b Percentages were computed by dividing the number of patients with the particular diagnosis by the total number of patients (since multiple diagnoses could be given, the total percentage exceeds the percentage of any diagnosis)

complete information in both occasions and not to regard the second interview as a continuation of the first one).

The sociodemographic characteristics of the sample at the time of investigation were as follows: (1) age: 35.6 + 8.6 years (mean + SD) with a range of 21–57 years; (2) sex distribution: 19 (32%) males, 41 (68%) females; (3) marital status: single 31 (52%), married 16 (27%), divorced 10 (17%), separated 2 (3%), widowed 1 (2%); (4) educational level: 25 (42%) primary school with or without graduation, 16 (27%) high school, 8 (13%) college, 11 (28%) university.

The following DSM-III-R Axis I diagnoses were given the 60 patients: schizophrenia 2 (3%); schizophreniform disorder 2 (3%); manic episode 1 (2%); major depression 12 (20%); dysthymic disorder 1 (2%); adjustment disorder with depressed mood 4 (7%); panic disorder 7 (12%); agoraphobia with panic attacks 3 (5%); social phobia 2 (3%); obsessive-compulsive disorder 3 (5%); generalized anxiety disorder 1 (2%); alcohol dependence 34 (57%); drug dependence 9 (15%); anorexia nervosa 1 (2%); bulimia nervosa 2 (3%). A total of 84 diagnoses was obtained, since multiple diagnoses could be made according to the DSM-III-R concept of comorbidity.

Table 1 shows the base rates of MDCL-P diagnoses for the 60 patients of our sample. A total of 72 personality disorder diagnoses was obtained for only 29 (48%) patients who received at least one personality disorder diagnosis (representing a mean of 2.5 diagnoses for each patient).

Test-retest Procedure

The MDCL-P was administered to each of the 60 patients on two separate occasions by one of four independent diagnosticians. The average duration time of the interview was 36 min (range: 20–150 min). The time interval between the two examinations was 1–4 days.

We intended to minimize, as far as possible, any bias due to the distribution and assignment of individual interviewers. One interviewer (T. B.) performed 60 interviews, each of the other three interviewers (D. G.; S. F.; R. W.) 20 interviews. In the one half of their interviews the diagnosticians were test and in the other half retest interviewers.

All four interviewers were psychiatrists, one (T. B.) with more than 10 years experience, two (D. G.; R. W.) with 4 years experience, and one (S. F.) with 1 year experience. Three out of four interviewers (D. G.; S. F.; R. W.) were not involved in the development of the MDCL-P and had never been engaged in personality research or in research with regard to diagnostic issues. Whenever a specific patient was examined for the second time (re-test), the interviewer did not know the result of the first investigation. Furthermore, both interviewers did not use information from clinical records. Before the study began the diagnosticians had performed three interviews with patients. Afterwards, there were two case conferences where one interviewer (T. B.) examined a patient in the presence of the other three interviewers who also filled out the MDCL-P. After the examination any discrepancies were discussed.

Statistical Analyses

Agreement between diagnosticians on diagnoses was calculated using k , a chance-corrected measure of congruence for binary ratings [6, 7]. The values of k range from -1.0 to 1.0 , with higher values representing higher levels of agreement. We performed a test of significance for each k value (one-tailed on the 5% level of error), but an interpretation of the magnitude of k was considered to be more important (since k may easily become statistically significant despite unsatisfactory agreement).

The k statistic currently represents the standard method of assessing diagnostic agreement in psychiatry [23], but it has been criticized because it is strongly influenced by the base rate of the diagnosis under study [5, 8, 26]. However, Spitznagel and Helzer [26] have proposed Yule's Y as a measure independent of base rates. Y is closely related to k since it can be regarded as an approximation to maximum k across all possible base rates. We additionally employed this measure to analyse our data. A pseudo-Bayes estimation was applied whenever a single cell of the fourfold classification table became 0 (otherwise, Y would have reached the endpoint value of 1.0 despite incomplete congruence; cf. [1]).

The congruence measures presented here refer to the reliability of probably or certain diagnoses. This corresponds to the usual clinical situation where an intervention seems to be justified even if the diagnosis is still provisional, or if it has to be complemented by additional information. However, the MDCL-P also makes it possible to evaluate agreement for certain diagnoses only (i.e. by contrasting certain and probable diagnoses). We have conducted this analysis, but the results are almost identical with the data presented here, and no systematic bias could be found.

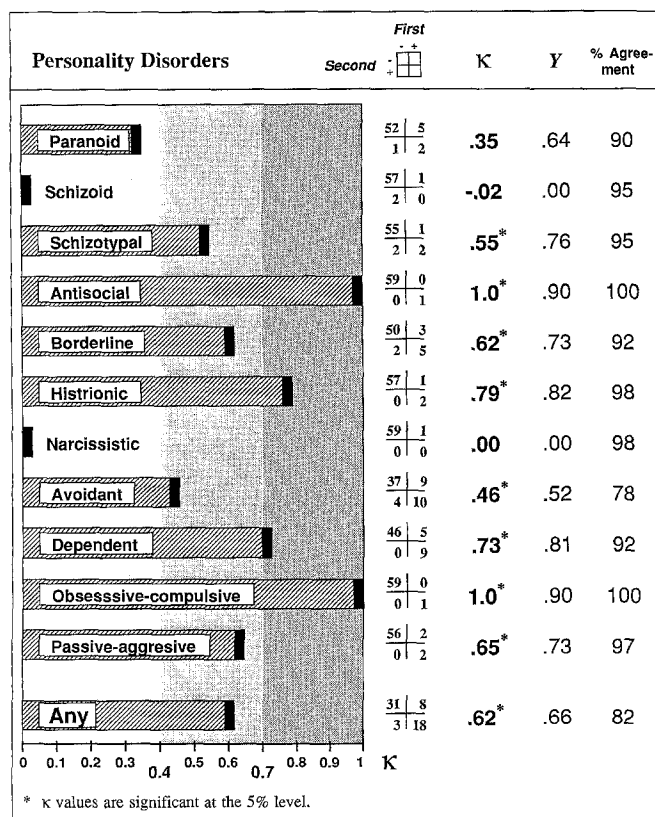


Fig. 2. Test-retest reliability of DSM-III-R Personality Disorders diagnosed by means of the MDCL-P

Results

The reliability of the MDCL-P diagnoses are presented in Fig. 2. Each row gives, for a single category of a personality disorder, a fourfold table with the exact distribution of congruently and incongruently assigned patients, k , Yule's Y , and the overall percentage agreement. Values of k are illustrated graphically, showing more clearly the different levels of interpretation (k above 0.40 as acceptable, k above 0.70 as excellent; cf. [7]).

With the exception of the personality disorders (PDS) paranoid, schizoid and narcissistic, all k values in Fig. 2 were significant at the 5% level, including the differentiation between no personality disorder versus probable or certain personality disorder. However, most of the specific personality disorders were diagnosed fewer than five items by one of the interviewers. Only the paranoid PD, schizotypal PD, borderline PD, avoidant PD and dependent PD were diagnosed five times or more: The range of k values was from 0.35–0.79. The duration (min) of the interview for patients with no personality disorder was $\times 32.8$, SD 11.8 and for patients with probable of certain personality disorder $\times 42.1$, SD 21.4 ($P \leq 0.05$, t -test, two tailed). The overall duration was 36 min (range 20–150).

Comparison with Other Studies

There are two structured interviews (SIDP, [28]; PDE, [16]) where k values have been published. Loranger et al. [16] reported a k value of 0.80 for the differentiation of

personality disorder versus no personality disorder using only one interview. The range of specific personality disorders which were diagnosed at least five times or more by one interviewer was 0.70–0.96. Stangl et al. [28] reported a k value of 0.66 for the differentiation of personality disorder versus no personality disorder in 20 patients in a test-retest design. The range of specific personality disorders was from 0.62–0.90 in 63 patients; however, in two thirds of the interviews both raters were present.

Discussion

Since the introduction of DSM-III, i.e., the introduction of a separate axis (Axis II) for personality disorders, interest in clinical practice and research of personality disorders has increased. There is good evidence from treatment studies [22] that the additional diagnosis of a personality disorder to an Axis I diagnosis leads to a worse course and outcome. However, the demarcation of symptoms and traits, of clinical syndromes and personality disorders, especially in the realm of affective and anxiety disorders, remain uncertain [4].

Therefore there is a need of assessment of the Axis II diagnoses in addition to the clinical syndromes in routine clinical care as well as in research. Whereas self-rating questionnaires (PDQ-R, [12]; MCMI, [18]) revealed a high sensitivity but a low specificity in test-retest studies [13, 14], the structured interviews (SIDP, [28]; PDE, [16]) demonstrated an acceptable to excellent interrater reliability in a test-retest design. However, these structured interviews last 2–4 h or more and, therefore, do not seem to be feasible in routine clinical care or in research projects which have an extended battery of assessment instruments.

Before the development of the MDCL-P, experience was gained in the use of checklists assessing the most important DSM-III-R Axis I diagnoses [9, 10]. Acceptably high levels of agreement were indicated for most disorders using several statistics (including k) [9]. In comparison with one of the two structured interviews (SIDP, [28]), from which k values were obtained, the k values of the MDCL-P were similar (agreement personality disorder versus no personality disorder, SIDP: $k = 0.66$; MDCL-P: $k = 0.62$; agreement specific personality disorders: SIDP: $k = 0.62$ –0.90 with only one interview; MDCL-P: $k = 0.35$ –0.79). However, the k values from the other structured interview (PDE, [16]) were excellent (personality disorder versus no personality disorder: $k = 0.80$; specific personality disorders: $k = 0.70$ –0.96) with only one interview.

It has to be taken into account that, in our study, three out of four interviewers were not involved in the development of the MDCL-P and had no experience in structured interviews concerning psychopathology and personality assessment. Furthermore, there was only very limited training before the reliability study. On the other hand, the PDE, used by clinicians who were not involved in the development of the PDE, revealed lower k values of 0.41–0.78 [27]. Finally, in contrast with our test-retest design Loranger et al. [16] and Stangl et al. [28] con-

ducted only one interview with two raters which could enhance the agreement of the two interviewers [2]. Despite the fact that O'Boyle and Self [19] found considerable discrepancies between the two structured interviews (SIDP, PDE), the MDCL-P will be tested against a structured interview (PDE, [16]) and against a self-rating questionnaire (PDQ-R, [12]). It would also be of interest to know if simply providing a form on which to code each criterion would improve routine clinical assessments.

References

1. Bishop YMM, Feinberg SE, Holland PW (1975) Discrete multivariate analysis: theory and practice. MIT Press, Cambridge, Mass
2. Bronisch T (1992) Diagnostik von Persönlichkeitsstörungen nach den Kriterien aktueller internationaler Klassifikationssysteme. *Verhaltenstherapie* 2: 140–150
3. Bronisch T, Hiller W, Zaudig M, Mombour W (in press) MDCL-P Münchner Diagnose-Checklisten für die DSM-III-R/ICD-10 Persönlichkeitsstörungen (MDCL-P). Hans Huber, Bern
4. Bronisch T, Klerman GL (1991) Personality functioning: change and stability in relationship to symptoms and psychopathology. *J Pers Dis* 5: 307–317
5. Carey C, Gottesman II (1979) Reliability and validity in binary ratings. Areas of common misunderstanding in diagnosis and symptom ratings. *Arch Gen Psychiatry* 35: 1454–1459
6. Cohen J (1960) A coefficient of agreement for nominal scales. *Educ Psychol Meas* 20: 37–46
7. Fleiss JL (1981) Statistical methods for rates and proportions, 2nd edn. Wiley, New York
8. Grove WM, Andreasen NC, McDonald-Scott P, Keller MB, Shapiro RW (1981) Reliability studies of psychiatric diagnosis. Theory and practice. *Arch Gen Psychiatry* 38: 408–413
9. Hiller W, Bose M von, Dichtl G, Agerer D (1990) Reliability of checklist-guided diagnoses for DSM-III-R affective and anxiety disorders. *J Affect Dis* 20: 235–247
10. Hiller W, Zaudig M, Mombour W (1990) Development of diagnostic checklists for use in routine clinical care. A guideline designed to assess DSM-III-R diagnoses. *Arch Gen Psychiatry* 47: 782–784
11. Hyler SE, Rieder RO, Spitzer RL (1978) Personality Diagnostic Questionnaire (PDQ). New York State Psychiatric Institute, New York
12. Hyler SE, Rieder RO, Williams JB, Spitzer RL, Hendler J, Lyons M (1988) The personality diagnostic questionnaire: development and preliminary results. *J Pers Dis* 2: 229–237
13. Hyler SE, Skodol AE, Kellmann HD, Oldham JM, Rosnick L (1990) Validity of the personality diagnostic questionnaire – revised: comparison with two structured interviews. *Am J Psychiatry* 147: 1043–1048
14. Hyler SE, Skodol AE, Oldham JM, Kellman DH, Doidge N (1992) Validity of the Personality Diagnostic Questionnaire – Revised: A replication in an outpatient sample. *Compr Psychiatry* 33: 73–77
15. Loranger AW (1988) Personality Disorder Examination (PDE) Manual. DV Communications, Yonkers, NY
16. Loranger AW, Lehmann Susman V, Oldham JM, Russakoff LM (1987) The personality disorder examination: a preliminary report. *J Pers Dis* 1: 1–13
17. Millon T (1983) Millon Clinical Multiaxial Inventory – Manual (3rd ed). National Computer Systems, Minneapolis
18. Millon T (1987) Millon Clinical Multiaxial Inventory – II (manual). National Computer Systems, Minneapolis
19. O'Boyle M, Self D (1990) A comparison of two interviews for DSM-III-R personality disorder. *Psychiatry Res* 32: 85–92
20. Pfohl B, Blum, N, Zimmerman M, Stangl D (1989) Structured Interview for DSM-III-R Personality SIDP-R. Department of Psychiatry, University of Iowa, Iowa City, Iowa
21. Pfohl B, Stangl D, Zimmermann M (1982) The Structured Interview for DSM-III Personality Disorders (SIDP). Department of Psychiatry, University of Iowa, Iowa City, Iowa
22. Reich JH, Green AI (1991) Effect of personality disorders on outcome of treatment. *J Nerv Ment Dis* 179: 74–82
23. Shrout PE, Spitzer RL, Fleiss JL (1987) Quantification of agreement in psychiatric diagnosis revisited. *Arch Gen Psychiatry* 44: 172–177
24. Spitzer RL, Williams JBW (1986) Structured clinical interview for DSM-III personality disorders (SCID II, 2/1/86). Biometric Research Department, New York State Psychiatric Institute, New York
25. Spitzer RL, Williams JBW, Gibbon M, First M (1990) User's Guide for the Structured Clinical Interview for DSM-III-R. American Psychiatric Association Press, Washington, DC
26. Spitznagel EL, Helzer JE (1985) A proposed solution to the base rate problem in the kappa statistic. *Arch Gen Psychiatry* 42: 725–728
27. Standage K, Ladha N (1988) An examination of the reliability of the personality disorder examination and a comparison with other methods of identifying personality disorders in a clinical sample. *J Pers Dis* 2: 267–271
28. Stangl D, Pfohl B, Zimmermann M, Bowers W, Corenthal C (1985) A structured interview for the DSM-III personality disorders. A preliminary report. *Arch Gen Psychiatry* 42: 591–596
29. Zerssen D von (1986) Clinical self-rating scales (CSR-S) of the Munich Psychiatric Information System (Psychis München). In: Sartorius N, Ban TA (eds) *Assessment of Depression*. Springer, Berlin Heidelberg New York, 270–303