

The Concept of Major Depression

II. Agreement Between Six Competing Operational Definitions in 600 Psychiatric Inpatients

Michael Philipp, Wolfgang Maier, and Cynthia D. Delmo

Department of Psychiatry, University of Mainz, Untere Zahlbacher Straße 8, W-6500 Mainz, Federal Republic of Germany

Received June 14, 1990

Summary. Six operational definitions of the concept of major depression were submitted to empirical evaluation in 600 psychiatric inpatients. Special attention was given to the comparison of major depression in DSM-III-R and ICD-10. The data base created by a polydiagnostic interview revealed relevant classificatory differences between the six definitions under study. Sources of different diagnostic base rates were: inclusion or omission of anhedonia as an obligatory mood criterion; minimal number of syndrome criteria required for the syndrome diagnosis; different width and reference points of time criteria; exclusion rules for co-existing schizophrenic symptoms and for previous nonaffective and manic episodes. The empirically evaluated overlap between pairs of diagnostic definitions was less than excellent in most of the diagnostic definitions under study; only the DSM-III and DSM-III-R definitions agreed with each other to a highly comparable degree. The relatively good agreement of the 1989 draft definition of ICD-10 for major depression ("mild depression") with the other five operational definitions ($\kappa = 0.69$) led us to expect that this definition should receive sufficient international acceptance.

Key words: Major depression – Operational diagnosis – ICD-10 – DSM-III-R

Introduction

A descriptive comparison of six competing operational definitions of the concept of major depression (Philipp et al. 1990a) has demonstrated a high degree of similarity as regards the item pool and the syndrome-defining criteria, and – to a lesser degree – the time and exclusion criteria completing the diagnostic algorithms. This descriptive finding led to the expectation that the application of these competing definitions could guarantee the selection of very similar patient samples. Yet it was nec-

essary that this expectation should be examined empirically. We therefore conducted a polydiagnostic study on a large-scale sample of psychiatric inpatients in order to evaluate the effect of the discrete differences of diagnostic definitions on the classificatory agreement of these competing diagnostic systems.

Within the empirical comparison of the competing diagnoses of major depression an effort was made to isolate the effects of the single step of the operational definition process. This seemed to be necessary for estimating the consequences for future developments in defining diagnostic rules for major depression. It is the current development of ICD-10 which stimulated this polydiagnostic study; the high turnover of drafts of ICD-10 in development suggested that we concentrate our interest on the latest changes, which were between 1987 and 1989 (World Health Organization, diagnostic criteria for research, draft April 1987 and draft April 1989). The quick revision of DSM-III (American Psychiatric Association 1980, 1987) provided a reason to look for parallels in the comparison of DSM-III and DSM-III-R. Finally, the last two diagnostic definitions of major depression in the Feighner Diagnostic Criteria (FDC; Feighner et al. 1972) and in the Research Diagnostic Criteria (RDC; Spitzer et al. 1978) stimulated a special comparison of these two definitions, because the development of the RDC was intensively co-authored by the same St. Louis group which had developed the FDC.

Patients and Methods

The study comprised 600 psychiatric inpatients admitted to the Psychiatric Department of the University of Mainz between March 1987 and November 1988. All newly admitted inpatients meeting the the following inclusion and exclusion criteria were eligible for the study: functional psychiatric disturbance; no organic brain syndrome; informed consent for participation in the study; ability to participate in a structured interview within the first 4 weeks after admission.

The first 200 patients were recruited consecutively during a 7-month period in 1987 among all newly admitted patients meeting

the above-mentioned criteria. Four hundred patients were recruited during the following 25 months applying the additional selection criterion of having at least one first-degree relative and willing to participate in a family interview study. Neither part of the total sample was selected by type or severity of symptoms; nevertheless the special structure of this department and the predominant research interest in affective disorders led to a clear-cut prevalence of affective disorders and an unusually low rate of psychotic disorders.

Psychopathological Assessments. Structured assessment and documentation of psychopathology was done by means of the Polydiagnostic Interview (PODI; Philipp and Maier 1986). For the purpose of this study version 2.1 was developed by additionally including DSM-III-R criteria and ICD-10 diagnostic criteria for research. The PODI-based assessment of psychopathology included structured interviews of patients and systematic evaluation of other sources of information (written case reports, observation of on-ward behaviour), combined with a final correction of codings by the treating psychiatrist.

The PODI interviews were done by a total of 16 research assistants, all senior medical students, who had completed a 30-h structured PODI training. Each PODI interview was conducted and coded by pairs of two research assistants with one person asking the questions and the other observing. After completion of the interview discrepant codings were identified and discussed. A consensus had to be found between the two interviewers, including re-interviewing the patient because of unclear items. Information from written case reports and observed on-ward behaviour was used to correct the item codings of the PODI. The coded consensus interview was then discussed with the treating psychiatrist for a final correction of codings. The PODI interviews were done within the first 4 weeks after admission, as soon as the patient was able to participate in this 2- to 3-h interview. The reference time frame of the PODI was always the worst period of the current psychiatric episode; therefore postponing the interview for a maximum of 4 weeks after admission did not lead to a loss of diagnostic information.

Stepwise Calculation and Comparison of Operational Diagnoses. Diagnostic evaluation of the PODI-based psychopathology data was done by application of a Turbo-Pascal computer program running on MS-DOS personal computers (Delmo 1988; Delmo and Philipp 1989). Version 2.1 of the computer program was expanded twice during the study adding DSM-III-R algorithms and a first version of ICD-10 algorithms in 1988 and a second version of ICD-10 algorithms in 1989. Operational diagnoses of major depression were compared for six diagnostic systems: FDC (Feighner et al. 1972; Feighner 1981), RDC (Spitzer et al. 1978), Diagnostic and Statistical Manual of Mental Disorders, third edition (DSM; American Psychiatric Association 1980), Diagnostic and Statistical Manual of Mental Disorders, third edition, revised (DSR; American

Psychiatric Association 1987), ICD-10 Diagnostic Criteria for Research, draft April 1987 (I87; WHO 1987), and ICD-10 Diagnostic Criteria for Research, draft April 1989 (I89, WHO 1989); the diagnostic algorithms are described and compared in the first of this series of papers (Philipp et al. 1990a). Two additional definitions of major depression were introduced: a narrow definition (in the following abbreviated NAR) was met if all six of the above-mentioned operational definitions were fulfilled; and a wide definition (WID) was met if at least one of the six operational definitions was fulfilled.

All competing operational definitions of major depression were calculated stepwise on every level of the operational definition process: items to criteria to syndrome scores, syndrome scores to categorical syndrome decisions; syndrome decisions were then completed by time criteria, and in the last step full diagnoses were generated by adding exclusion criteria. This paper presents comparisons at the last four levels of the operational definition process: (a) obligatory mood inclusion criterion; (b) pure syndrome definition without time and exclusion criteria; (c) syndrome definition with time criterion included; (d) syndrome definition with time criterion and exclusion of schizophrenic symptoms; (e) full diagnosis applying additional history-related exclusion criteria.

At each of the five levels a–d diagnostic base rates were counted for each of the six competing definitions of major depression. Additionally, syndrome scores were calculated at levels b, c and d. Overlap was calculated for each pair of diagnostic systems at each of levels a–d, applying Cohen's kappa (1960) as a measure of chance corrected agreement for each pair of diagnoses, averaging the kappa coefficients for each diagnosis.

Results

Diagnostic Base Rates

Looking at the frequency of patients meeting the obligatory mood inclusion criterion FDC (76.7%), RDC (75.8%), DSM (75.7%) and DSR (76.0%) display nearly identical base rates (Table 1); those of I87 (70.2%) and I89 (69.2%) are more than 5% lower. The base rates of patients fulfilling the pure syndrome definition are very similar: 63.3% (I87) being the lowest and 70.2% (DSR) the highest (Table 1). The base rates of DSM and DSR are nearly identical (difference of 0.2%); the same is true for FDC and RDC. In contrast, I87 and I89 differ by 5%.

Diagnostic base rates are reduced for all diagnoses by additionally introducing time criteria. This reduction is smallest in FDC (drop of 5.8%) and RDC (drop of 5.9%); it is larger in I89 (drop of 12.3%) and much larger in the

Table 1. Diagnostic base rates in six competing definitions of major depression

Diagnostic system	Obligatory mood inclusion criterion (% pos. patients)	Pure syndrome without time criterion (% pos. patients)	Syndrome with time criterion (% pos. patients)	Syndrome and time with exclusion of schizophrenic symptoms ^a (% pos. patients)	Full diagnosis ^b (% pos. patients)
FDC	76.7%	65.0%	59.2%	47.2%	33.7%
RDC	75.8%	65.2%	59.3%	44.7%	
DSM	75.7%	70.0%	52.0%	46.5%	
DSR	76.0%	70.2%	51.8%	46.2%	
I87	70.2%	63.3%	40.8%	38.7%	30.5%
I89	69.2%	68.3%	56.0%	45.8%	42.7%

^a The study design excluded patients with somatic or toxic origin of the depressive syndrome

^b FDC additionally excludes patients with a history of nonaffective disorders; I87 and I89 exclude bipolar patients

Table 2. Distribution of syndrome criteria met by 600 patients in six definitions of major depression

Patients meeting mood criterion and additional syndrome criteria Number of pure syndrome criteria (without time criterion)										
	0	1	2	3	4	5	6	7	8	9
FDC	5	2	9	22	23	//	83	108	129	79
RDC	4	3	12	13	32	//	71	114	132	74
DSM	2	4	10	19	//	25	52	94	132	117
DSR	0	2	4	11	18	//	27	54	92	132
I87	1	0	6	7	27	//	55	106	132	87
I89	0	5	11	//	42	73	107	104	73	

Rearrangement of patient distribution according to the similarity of the effect of cut-off definitions on the diagnostic base rates										
	0	1	2	3	4	5	6	7	8	9
FDC	5	2	9	22	23	//	83	108	129	79
RDC	4	3	12	13	32	//	71	114	132	74
DSM	2	4	10	19	//	25	52	94	132	117
DSR	0	2	4	11	18	//	27	54	92	132
I87	1	0	6	7	27	//	55	106	132	87
I89		0	5	11	//	42	73	107	104	73

remaining diagnoses (I87: drop of 22.5%; DSR: drop of 18.4%; DSM: drop of 18.0%) (Table 1). The range between the widest diagnosis (RDC: 59.3%) and the narrowest diagnosis (I87: 40.8%) is now nearly 20%.

The opposite effect is seen when introducing schizophrenic symptoms as exclusion criteria: RDC (drop of 14.6%), FDC (drop of 12.0%) and I89 (drop of 10.2%) now display the highest reductions in diagnostic base rates, while DSM (8.7%) and DSR (5.6%) show a smaller reduction, and I87 is only reduced by 2.1% (Table 1). Simultaneously, the range between the most restrictive diagnosis I87 (38.7%) and the now least restrictive diagnosis FDC (47.2%) becomes smaller (less than 10%).

FDC, I87 and I89 additionally exclude patients because of a history of nonaffective psychiatric disorders (FDC) or a history of manic disorders (I87 and I89). Both exclusion criteria lead to an additional drop in base rates, with I87 staying the most restrictive diagnosis of major depression (30.5%) and the unchanged DSM diagnosis now being the least restrictive diagnosis (46.5%) (Table 1); by this, the range of diagnostic base rates increases again to 16%.

The upper half of Table 2 illustrates the effects of the different cut-offs for the diagnostic base rates on the level of pure syndrome definition. FDC, RDC, DSR and I87 require at least 5 syndrome criteria (with DSR offering a catalogue of 9 eligible criteria, whereas the other three diagnoses only offer 8); DSM and I89 are different by requiring 4 out of 8 (DSM) or 2 out of 7 (I89) criteria; the DSM cut-off (4 out of 8) equals exactly that of DSR (5 out of 9); the I89 cut-off (2 out of 7) provides similar results. This is more apparent in the rearrangement shown in the lower part of Table 2. The cut-off definitions now create two groups of diagnoses: one group with a higher

Table 3. Matrix of agreement (kappa) between six definitions of major depression

	FDC	RDC	DSM	DSR	I87	I89
Obligatory mood inclusion criterion						
FDC	–					
RDC	0.97	–				
DSM	0.96	1.00	–			
DSR	0.97	1.00	0.99	–		
I87	0.83	0.86	0.86	0.85	–	
I89	0.80	0.83	0.83	0.83	0.85	–
Mean	0.91	0.93	0.93	0.93	0.85	0.83
Syndrome without time criterion						
FDC	–					
RDC	0.92	–				
DSM	0.89	0.88	–			
DSR	0.88	0.89	1.00	–		
I87	0.93	0.90	0.85	0.85	–	
I89	0.81	0.82	0.87	0.87	0.81	–
Mean	0.89	0.88	0.90	0.90	0.87	0.84
Syndrome with time criterion						
FDC	–					
RDC	0.74	–				
DSM	0.62	0.79	–			
DSR	0.62	0.79	0.95	–		
I87	0.55	0.64	0.78	0.78	–	
I89	0.63	0.78	0.75	0.75	0.68	–
Mean	0.63	0.75	0.78	0.78	0.69	0.72
Syndrome with time and exclusion of schizophrenic symptoms ^a						
FDC	–					
RDC	0.77	–				
DSM	0.60	0.71	–			
DSR	0.59	0.71	0.95	–		
I87	0.52	0.59	0.76	0.77	–	
I89	0.63	0.74	0.70	0.71	0.66	–
Mean	0.62	0.70	0.74	0.75	0.66	0.69

^a Secondary depressions in FDC and bipolar depressions in I87 and I89 were not excluded here

cut-off (FDC, RDC and I87) and another one with a lower cut-off (DSM, DSR and I89). I87 and I89 belong to different groups, while FDC and RDC on the one hand, and DSM and DSR on the other hand belong to the same group.

Overlap Between Diagnoses

The amount of agreement corrected by chance (kappa) between all diagnoses being compared pairwise is shown in Table 3. For the mood inclusion criterion the agreement is nearly perfect (0.96–1.00) between FDC, RDC, DSM and DSR; it is still excellent, but lower between I87 and I89 (0.85) and between both ICD-10 definitions and the other four diagnostic systems (0.80 ± 0.86).

Table 4. Overlap between narrow and wide definition of major depression

	Narrow definition							
	Obligatory mood criterion		Pure syndrome definition		Syndrome with time criterion		Syndrome + time + exclusions of schizophrenic symptoms ^a	
	-	+	-	+	-	+	-	+
Wide	-139	0	-168	0	-188	0	-227	0
Defini-	+ 62	399	+ 71	361	+183	229	+198	175
tion	kappa = 0.75		kappa = 0.74		kappa = 0.44		kappa = 0.40	

^a For FDC, I87 and I89 the additional exclusion of secondary and bipolar patients was omitted

These differences are diminished on the level of pure syndrome definition; mean kappa coefficients are still excellent ($> = 0.80$) for all comparisons; yet perfect agreement (1.00) exists only between DSM and DSR. The lowest kappa values are found between I89 and I87 (0.81), between I89 and FDC (0.81), and between I89 and RDC (0.82); the mean kappa value of I89 also is lowest (0.84), while those of DSM and DSR are highest (both 0.90).

Mean agreement between diagnoses is considerably reduced when introducing time criteria; mean kappa values are now between 0.63 (FDC) and 0.78 (DSM and DSR). Only DSM and DSR are still in excellent agreement (kappa = 0.95); I87 and I89 agree well, but there is no longer excellent agreement (kappa 0.68). FDC has the lowest kappa values in all comparisons; the agreement with RDC is reduced from 0.92 to 0.74, that with I87 from 0.93 to 0.55 (Table 3).

The exclusion of patients with schizophrenic symptoms influences the mean agreement between diagnoses far less (Table 3): mean kappa values range between 0.62 (FDC) and 0.75 (DSR). The agreement between FDC and I87 is still lowest (kappa = 0.52); that between FDC and RDC increases a little (0.77). DSM and DSR are still in excellent agreement (kappa = 0.95), while that between I87 and I89 is still only good (0.66) (Table 3). No cross-tabulations were calculated for full diagnoses of FDC, I87 and I89. The additional exclusion of secondary depressions in FDC and of bipolar depressions in I87 and I89 creates incomparabilities, necessarily leading to a drop in agreement with other diagnoses, which do not use these exclusion criteria.

Cross-tabulations of NAR and WID are presented in Table 4. An excellent agreement could be found neither for the obligatory mood criterion (kappa = 0.75), nor for the pure syndrome definition (kappa = 0.74); the introduction of time criteria leads to a major drop in agreement (kappa = 0.44), while the additional exclusion of schizophrenic symptoms only reduces agreement to a minor degree (kappa = 0.40). The distribution of patients in Table 4 shows that at the level of full diagnosis there are more patients with discrepant classification ($n = 198$) than patients with concordant positive classification ($n = 175$).

Discussion

The empirical comparison of the six competing definitions of major depression allows us to demonstrate clearly which differences in the operational definitions are negligible and which lead to relevant discrepancies. By this method of empirical comparison, the main question of our descriptive comparison (Philipp et al. 1990a) can be answered: the differences in diagnostic algorithms are *not* negligible; major discrepancies result, which render most of the competing operational definitions incompatible. The empirical relevance of descriptively small differences can be demonstrated for nearly every detail of the complex operational definition process.

Definition of the Obligatory Mood Criterion

The descriptive comparison of diagnostic algorithms (Philipp et al. 1990a) had already shown that both I87 and I89 are more restrictive in defining the obligatory mood inclusion criterion than the other diagnostic systems: I87 requires not only depressed mood but also anhedonia; I89 requires at least two out of a catalogue of three symptoms, in which depressed mood and anhedonia are supplemented by decreased energy. The empirical evaluation now shows that this difference is relevant: it leads to a reduction of base rates of more than 5% in comparison with the other four diagnostic systems, which do not explicitly require anhedonia. This comparative analysis can not decide as to the validity of these alternative definitions; yet, if the results of the forthcoming validation studies should find arguments for a more restrictive or for a less restrictive definition of major depression, this would influence the decision for a choice of either a wide mood criterion (leaving depressed mood and anhedonia as alternatives) or for a narrow mood criterion (requiring both symptoms).

As for the wide definition of the obligatory mood criterion our data show no difference whether the definition requires only depressed mood (as in the FDC) or whether depressed mood and anhedonia are defined as alternative symptoms, of which at least one has to be present (as in the RDC, DSM and DSR). Whether this finding is only valid for psychiatric inpatients or whether it also applies to outpatients will be examined in an outpatient sample (manuscript in preparation).

Definition of the Syndrome Cut-off

The descriptive comparison of syndrome cut-offs (Philipp et al. 1990a) had shown four different definitions: 2 out of 7 criteria in I89, 4 out of 8 criteria in DSM, 5 out of 8 criteria in the FDC, RDC and I87, and 5 out of 9 criteria in DSR. The empirical evaluation of patient distributions according to the number of criteria being met showed that in approximation these differences could be reduced to two groups: one group with a higher cut-off requiring 5 out of 8 criteria (FDC, RDC, I87) and another group with a lower cut-off, corresponding to a requirement of 4 out of 8 criteria (DSM, DSR, I89); this classification is immediately reflected by the base rates on the level of

the pure syndrome definition, where the group with the higher cut-off has about 5% lower base rates than the group with the lower cut-off. Validation studies will have to decide whether a lower or a higher cut-off is more valid, taking into consideration inpatient and outpatient samples.

Definition of Time Criteria

Three diagnostic systems (DSM, DSR and I87) define very strict time criteria by requiring that a symptom has to be present nearly every day during a 2-week period; these diagnostic systems exclude nearly one-third of all patients meeting the pure syndrome criterion but not the strict time criterion. In the other three definitions (FDC, RDC and I89) only the total depressive episode or the obligatory mood inclusion criterion has to be present for at least 2 weeks (4 weeks in FDC); this reduces the diagnostic base rates by between 6% and 12%. The base-rate-reducing effect of strict time criteria is apparently much higher than that of a higher syndrome cut-off. Up to now, there are no empirical studies which evaluate the validity differences of both ways of strict and mild time criteria. Should future validation studies show that the current definitions of major depression are too narrow, the elimination of strict time criteria would be the most effective way for widening the diagnostic concept.

Definition of Exclusion Criteria

The same lack of empirical data is to be found for the exclusive power of coinciding schizophrenic symptoms. The FDC, RDC and I89 have the most simple definition: the simultaneous existence of depressive and schizophrenic symptoms is sufficient to exclude patients from the diagnosis of major depression. Our data show that this leads to a base rate drop of about 15%. DSM, DSR and I87 require schizophrenic symptoms to be present in the absence of depressive symptoms to be an exclusion criterion; the simple coincidence of schizophrenic and depressive symptoms is not sufficient. In consequence, the reduction of diagnostic base rates is much lower for DSM, DSR and I87 (between 2% and 6%). Only the predictive validity of these differences in definition can decide between these possibilities; our data only stress that this difference is relevant for the restrictivity of the diagnostic definition.

Restricted Comparability of the Competing Diagnostic Definitions

In empirical research, operational definitions of diagnostic concepts are normally used only at the level of full definition including syndrome, time and exclusion criteria. Therefore the comparisons of full diagnoses are of particular relevance. In fact, comparability is nearly excluded if one diagnostic system accepts bipolar or secondary depressions and others like the FDC, I87 and I89 do not. The structure of our data processing method enables us to concentrate on modifications of FDC, I87 and I89, where a history of nonaffective psychiatric dis-

orders of mania does not exclude major depression. Yet the strict application of the original diagnostic rules would not have allowed such a modification; a comparison of ICD-10 definitions of major depression with any of the other diagnostic systems would then have been virtually impossible. For comparability with already existing diagnostic systems, the final version of the ICD-10 should offer a general category of major depression, which would then be subdivided into unipolar and bipolar forms, rather than keeping the current concept of primarily distinguishing between unipolar and bipolar forms of depression and then offering different concepts of subdividing depression within the unipolar and the bipolar group (unipolar being subdivided into three degrees of severity, bipolar only into two).

The final conclusions will be drawn on the level of comparable exclusion criteria neglecting the exclusion of secondary and bipolar patients. Whereas the simple inspection of base rates does not show large differences between the most restrictive diagnosis (I87: 38.7%) and the least restrictive diagnosis (FDC: 47.2%), the comparison of NAR and WID reveals such an impressive difference that the comparability of the competing diagnostic systems can at once be questioned. Comparing patient samples which are classified as major depression by different diagnostic systems requires that NAR and WID should not be too different. Applying the criteria that Bech and Clemmessen (1983) proposed to qualify the kappa coefficients in reliability studies, only DSM and DSR overlap to an excellent amount ($\text{kappa} > 0.80$); for all other pairs of comparison, the overlap is not excellent ($\text{kappa} < 0.80$) with some of the comparisons (FDC and I87, RDC and I87) being just acceptable ($\text{kappa} > 0.40$), but still far from being good ($\text{kappa} > 0.60$).

The mean kappa value of all pairwise comparisons could perhaps be considered as a measure of relative comparability. DSM and DSR would then be the best operational definitions of major depression, showing mean kappa values of 0.74 and 0.75; I89 would follow as the second best ($\text{kappa} = 0.69$). This measure of mean kappa values could, on the other hand, be considered as a measure of possible international acceptance, with those diagnostic systems being the most acceptable ones which present the highest mean kappa values. In this case, DSM and DSR would have to be considered as the most acceptable operational definitions of major depression, and both ICD-10 definitions are more deviant definitions. Of course, this is a very speculative interpretation; yet it might help to decide whether the current ICD-10 definition of major depression (I89) should really become the final one.

This paper does not comment on the complicated differentiation of I89 between mild, moderate and severe forms of I89 depressions; this will be the subject of a forthcoming paper. For this comparison we chose the I89 definition of mild depression, since the analysis of all three severity subtypes of I89 major depression would have complicated the matter considerably. We feel that a differential comparison of all three severity subtypes would reduce even further the comparability with other definitions of major depression; we will question the use-

fulness of this subtyping in a forthcoming publication (manuscript in preparation).

Findings according concurrent validity of these six definitions for the clinical ICD-9 diagnosis will be presented in the third paper of this series (Philipp et al. 1990b).

Acknowledgements. The study was conducted with the financial support of the Ministry of Health of the Federal Republic of Germany. We thank the following research assistants who performed part of the PODI ratings in the context of their doctoral theses: J. Böhm, M. Kancnyr, K. Lakatos, M. R. Lappas, C. Lesch, B. Wessels, S. Wustinger, M. Zwingenberger.

References

- American Psychiatric Association (1980) Diagnostic and Statistical Manual of Mental Disorders, 3rd edn (DSM-III). APA, Washington, D.C.
- American Psychiatric Association (1987) Diagnostic and Statistical Manual of Mental Disorders, 3rd edn, revised (DSM-III-R). APA, Washington, D.C.
- Bech P, Clemmessen L (1983) The diagnosis of depression: 20 years later. *Acta Psychiatr Scand* [Suppl] 310:9–30
- Cohen J (1960) A coefficient of agreement of nominal scales. *Educ Psychol Measurement* 20:37–46
- Delmo CD (1988) Weiterentwicklung des Polydiagnostischen Interviews. Medical dissertation, University of Mainz Medical School, Department of Psychiatry
- Delmo CD, Philipp M (1989) Computer program for calculating criteria, syndromes and diagnoses for the Polydiagnostic Interview PODI, version 2.1. Department of Psychiatry, University of Mainz
- Feighner JP (1981) Nosology of primary affective disorders and application to clinical research. *Acta Psychiatr Scand* 63 [Suppl 190]:29–41
- Feighner JP, Robins E, Guze SB, Woodruff RA, Winokur G, Munoz R (1972) Diagnostic criteria for use in psychiatric research. *Arch Gen Psychiatr* 26:57–63
- Philipp M, Maier W (1986) The Polydiagnostic Interview: a structured interview for polydiagnostic classification of psychiatric patients. *Psychopathology* 19:175–185
- Philipp M, Maier W, Delmo CD (1990a) The concept of major depression. I. Descriptive comparison of six competing operational definitions including ICD-10 and DSM-III-R. *Eur Arch Psychiatry Neurol Sci* 240:229–236
- Philipp M, Maier W, Delmo CD (1990b) The concept of major depression. III. Concurrent validity of six competing operational definitions for the ICD-9 diagnosis. *Eur Arch Psychiatry Neurol Sci* 240:243–249
- Spitzer R, Endicott J, Robins E (1978) Research Diagnostic Criteria: rationale and reliability. *Arch Gen Psychiatry* 35:773–782
- World Health Organization (1987) April 1987 draft of the ICD-10 diagnostic criteria for research, chapt. V: mental, behavioral, and developmental disorders (F00–F99). WHO, Geneva
- World Health Organization (1989) April 1989 draft of the ICD-10 diagnostic criteria for research, chapt V: mental, behavioral, and developmental disorders (F00–F99). WHO, Geneva