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 necessary 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 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 outward 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 outward 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 these five 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

<table>
<thead>
<tr>
<th>Diagnostic system</th>
<th>Obligatory mood inclusion criterion (% pos. patients)</th>
<th>Pure syndrome without time criterion (% pos. patients)</th>
<th>Syndrome with time criterion (% pos. patients)</th>
<th>Syndrome with time &amp; exclusion of schizophrenic symptoms (% pos. patients)</th>
<th>Full diagnosis (% pos. patients)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FDC</td>
<td>76.7%</td>
<td>65.0%</td>
<td>59.2%</td>
<td>47.2%</td>
<td>33.7%</td>
</tr>
<tr>
<td>RDC</td>
<td>75.8%</td>
<td>65.2%</td>
<td>59.3%</td>
<td>44.7%</td>
<td>44.7%</td>
</tr>
<tr>
<td>DSM</td>
<td>75.7%</td>
<td>70.0%</td>
<td>52.0%</td>
<td>46.5%</td>
<td>46.5%</td>
</tr>
<tr>
<td>DSR</td>
<td>76.0%</td>
<td>70.2%</td>
<td>51.8%</td>
<td>46.2%</td>
<td>46.2%</td>
</tr>
<tr>
<td>I87</td>
<td>70.2%</td>
<td>63.3%</td>
<td>40.8%</td>
<td>38.7%</td>
<td>30.5%</td>
</tr>
<tr>
<td>I89</td>
<td>69.2%</td>
<td>68.3%</td>
<td>56.0%</td>
<td>45.8%</td>
<td>42.7%</td>
</tr>
</tbody>
</table>

*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