A systematic review of instruments to measure depressive symptoms in patients with schizophrenia

Irene M. Lako a,b,1, R. Bruggeman a,b,c,2, H. Knegtering a,d,3, D. Wiersma a,4, R.A. Schoevers a,c,5, C.J. Slooff a,e,6, K. Taxis b,⁎

⁎ Corresponding author. Tel.: +31 50 3638205; fax: +31 50 3632772.
E-mail addresses: i.m.lako@umcg.nl (I.M. Lako), r.bruggeman@umcg.nl (R. Bruggeman), h.knegtering@lentis.nl (H. Knegtering), durk.wiersma@umcg.nl (D. Wiersma), r.a.schoevers@umcg.nl (R.A. Schoevers), cees.slooff@ggzdrenthe.nl (C.J. Slooff), k.taxis@rug.nl (K. Taxis).

Background: Depressive symptoms require accurate recognition and monitoring in clinical practice of patients with schizophrenia. Depression instruments developed for use in depressed patients may not discriminate depressive symptoms from negative psychotic symptoms.

Objective: We reviewed depression instruments on their reliability and validity in patients with schizophrenia.

Methodology: A systematic literature search was carried out in three electronic databases. Psychometric properties were extracted for those instruments of which reliability, divergent, concurrent and predictive validity were reported in one or more publications.

Results: Forty-eight publications described the reliability and validity of six depression instruments in patients with schizophrenia. The only self-report was the Beck Depression Inventory (BDI). The Brief Psychiatric Rating Scale—Depression subscale (BPRS-D), Positive and Negative Syndrome Scale—Depression subscale (PANSS-D), Hamilton Rating Scale for Depression (HAMD), Montgomery Asberg Depression Rating Scale (MADRS) and Calgary Depression Scale for Schizophrenia (CDSS) were clinician rated. All instruments were reliable for the measurement of depressive symptoms in patients with schizophrenia. The CDSS most accurately differentiated depressive symptoms from other symptoms of schizophrenia (divergent validity), correlated well with other depression instruments (concurrent validity), and was least likely to miss cases of depression or misdiagnose depression (predictive validity).

Conclusions: We would recommend to use the CDSS for the measurement of depressive symptoms in research and in daily clinical practice of patients with schizophrenia. A valid self-report instrument is to be developed for the use in clinical practice.

© 2011 Elsevier B.V. All rights reserved.
1. Introduction

Depressive symptoms are highly prevalent (25%) in patients with schizophrenia (Buckley et al., 2008; Siris and Bench, 2003). These comorbid depressive symptoms are associated with a higher burden of disease and more frequent relapses (Conley et al., 2007; Tollefson et al., 1999). Schizophrenia is a lifelong psychiatric disorder and depressive symptoms may occur through all phases of illness: during acute psychosis (Häfner, 2000; Leff et al., 1988) as well as after remission of psychosis (Birchwood et al., 2000). Recent literature suggests that depressive symptoms may also be understood as a dimension within the schizophrenia concept and that individual symptom profiles should guide treatment (Van Os and Kapur, 2009). Furthermore, the upcoming fifth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) advocates to measure psychopathology in terms of quantitative dimensions, instead of solely as discontinuous categories (http://www.dsm5.org). Adequate screening and monitoring of depressive symptoms is required to guide appropriate treatment (Van Os and Kapur, 2009). This systematic review provides an overview of instruments that can be used for the screening on depressive symptoms (further referred to as “depression instruments”). Instruments are compared regarding their divergent validity and other psychometric properties in this patient population. This review may help in choosing a suitable instrument for the measurement of depressive symptoms in research as well as in daily clinical practice of patients with schizophrenia.

2. Methods

2.1. Search procedure

As a first step, titles and abstracts were screened on relevance for the defined topic and, if appropriate, the full paper was examined. Inclusion criteria were: 1) studies assessing psychometric properties of instruments measuring depressive symptoms in a population of patients with schizophrenia or non-affective psychotic disorders, 2) the availability of a validated English translation of the depression instrument and 3) publication in English, German, French or Dutch language. Unidimensional depression instruments (measuring a single dimension, in this case depressive symptoms), as well as multidimensional instruments measuring multiple symptom dimensions providing a subscale for depressive symptoms, were included. We refer to the depression subscale of a multidimensional instrument by the addition of [-D] to discriminate depressive symptoms from other symptom dimensions in schizophrenia (divergent validity) (Allan and Martin, 2009; Fitzgerald et al., 2002). Currently there is no overview of available depression instruments and their psychometric properties in patients with schizophrenia.
the abbreviation of the instrument, for example BPRS-D. We excluded studies describing diagnostic instruments and instruments designed to measure related symptoms, such as anxiety or suicidality.

The following search terms were entered in the online databases PubMed, Embase and PsychINFO: (“depression” or “depressive symptoms”) and (“psychosis” or “psychotic”) and (“instrument” or “rating scale” or “scale” or “questionnaire” or “interview”) and (“psychometric” or “reliability” or “validity” or “reproducibility”). The search was carried out in May 2010. All retrieved studies were checked for cross-references.

2.2. General information

General information about the most recent version of each instrument was collected from (original) validation studies and the Handbook of Psychiatric Measures. In order to quantify the recent use of the selected instruments in research, we counted the number of studies published between May 2005 and May 2010. The composition of each depression instrument was explored as follows. Each item of an instrument was categorized under one of the nine diagnostic criteria for a Major Depressive Episode (MDE), defined by the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) (American Psychiatric Association, 1994). Remaining items were categorized under three additional symptom dimensions: “delusional ideas,” “other vital symptoms” and “anxiety.” Four of the nine diagnostic criteria for MDE show overlap with symptom dimensions of schizophrenia, in particular negative symptoms and extrapyramidal symptoms. For each instrument, the number of potentially overlapping items was divided by the total number of items. This illustrated the instruments’ ability to discriminate depressive symptoms from other symptom dimensions of schizophrenia.

Psychometric properties were extracted for those instruments of which reliability, divergent, concurrent and predictive validity were reported in one or more publications. In the next paragraphs we explain these psychometric properties.

2.3. Reliability

Reliability is generally estimated by internal consistency, inter-rater and test-retest reliability. Internal consistency reflects the coherence between items within an instrument. Corresponding Cronbach’s alpha values of 0.60–0.70 are considered acceptable and values of >0.70 as good (Cicchetti, 1994). Good inter-rater and test-retest reliability is reflected by little variation between the scores by different raters and, respectively, by repeated measurements; these are commonly expressed by intra-class coefficients (ICC) >0.70.

2.4. Divergent validity

Divergent (or discriminant) validity refers to the extent that different symptom dimensions are unrelated to each other. Here, an instrument designed to measure depressive symptoms, should not measure negative symptoms, EPS or anxiety as well. Divergent validity is commonly expressed by the Pearson’s product moment correlation (PPMC) between scores on a depression instrument and scores on an instrument measuring another symptom dimension. Absent correlation with negative symptoms or EPS indicates good divergent validity. Nevertheless weak correlations (<0.30) are acceptable, as depressive symptoms tend to occur together with negative symptoms and EPS (Kulhara et al., 1989; Van Putten and May, 1978).

Divergent validity can also be evaluated on the stability of the underlying factor structure of a particular instrument across different samples. For multidimensional instruments, principal component factor analysis (PCA) should identify depressive symptoms as a separate factor from psychotic symptom dimensions. In addition, the content of this depressive factor should remain stable by confirmatory factor analysis in different samples. PCA of unidimensional instruments in a population with schizophrenia should identify factors describing depressive symptom dimensions, but no psychotic symptom dimensions.

2.5. Concurrent validity

Concurrent (or convergent) validity refers to the extent that common symptom dimensions are in fact related. Concurrent validity is high when the scores on two instruments measuring the same symptom dimension correlate well (PPMC). Based on the mean correlation of each possible comparison between two instruments, we calculated a pooled mean correlation over all comparisons for each instrument.

2.6. Predictive validity

Predictive validity represents the accuracy of an instrument to correctly detect a case (here of depression). Included were publications using a validated diagnostic interview such as the Structured Clinical Interview for DSM-IV (First et al., 1995) as gold standard to identify positive cases of depression. Good predictive validity is reflected by high sensitivity (not likely to miss cases of depression) combined with high specificity (not likely to misdiagnose depression) at the optimal cut-off value, i.e. the best balance between sensitivity and specificity determined by area under the receiver operating curve methods (Hanley and McNeil, 1983).

3. Results

3.1. Inclusion of studies

The systematic search generated a total of 2642 articles, of which 57 publications were eligible for further evaluation (Fig. 1). For six depression instruments complete information on psychometric properties in a population with schizophrenia or psychotic disorders was described in forty-nine publications. These included two multidimensional instruments: the Brief Psychiatric Rating Scale, Expanded Version (BPRS) (Lukoff et al., 1986; Overall et al., 1972) and the Positive And Negative Syndrome Scale (PANSS) (Kay et al., 1987), and four unidimensional instruments: the Hamilton Rating Scale for Depression (HAM-D) (Hamilton, 1960), Montgomery Asberg Depression Rating Scale (MADRS) (Montgomery, 1979), Calgary Depression Scale for Schizophrenia (CDSS) (Addington et al., 1993) and Beck Depression Inventory-II (BDI) (Beck et al., 1996). The remaining 8 publications
described depression instruments with incomplete information about their psychometric properties in schizophrenia. For example, no information was available on reliability or divergent validity in schizophrenia for the Brief Symptom Inventory (BSI) (Derogatis and Spencer, 1982) and Center for Epidemiologic Studies Depression Scale (CES-D) (Radloff, 1977).

### 3.2. General characteristics

General characteristics of the six reviewed depression instruments are described in Table 1. Only one instrument was based on self-report. Unidimensional instruments required on average 10 minutes less time to be completed than multidimensional instruments. The number of items per depression instrument varied between 4 and 21. A quantitative investigation of the use of these depression instruments over the past five years showed that the CDSS was most frequently used in research in this period, closely followed by the PANSS-D and the HAMD.

Table 2 illustrates the composition of the depression instruments in the context of schizophrenia. The depressive symptoms covered by the CDSS had minimal overlap with...
other symptom dimensions of schizophrenia. In contrast, about three quarter of the items of the PANSS-D and BPRS-D showed overlap with anxiety and positive symptoms. The HAMD contained many items on delusional symptoms. Almost half of the items of the MADRS and BDI could also be interpreted as negative symptoms.

Table 2
Composition of instruments evaluating depressive symptoms in schizophrenia.

<table>
<thead>
<tr>
<th>Symptom dimensions</th>
<th>Symptoms</th>
<th>MDE criteria</th>
<th>BPRS-D</th>
<th>PANSS-D</th>
<th>HAMD</th>
<th>MADRS</th>
<th>CDSS</th>
<th>BDI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depressive symptoms</td>
<td>Depressed mood</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Changed appetite or weight</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>2</td>
<td>1</td>
<td>-</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Sleeping problems</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Worthlessness</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>2</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Suicidal ideation</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative symptoms</td>
<td>Loss of interest or pleasure</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Fatigue/lack of initiative or motivation</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Indecisiveness/lack of concentration</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Anxiety/tension</td>
<td>-</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>-</td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Other vital symptoms</td>
<td>-</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Items identifying symptoms in non-depressive dimensions</td>
<td>4</td>
<td>4</td>
<td>10</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total number of items of the (sub)scale</td>
<td>96</td>
<td>5</td>
<td>1</td>
<td>7</td>
<td>1</td>
<td>0</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td>% of items identifying non-depressive dimensions</td>
<td>44%</td>
<td>67%</td>
<td>80%</td>
<td>59%</td>
<td>40%</td>
<td>11%</td>
<td>48%</td>
<td></td>
</tr>
</tbody>
</table>

Depressive symptoms (dimensions) potentially overlapping with psychotic symptoms. a) Appeared or perceived depressed mood, including hopelessness, crying, pessimism, irritability and diurnal variation of mood. b) Including self-blame and non-delusional feelings of guilt. c) Including paranoid symptoms, hypochondriacal delusions, feeling criticized by others, poor insight and delusional feelings of guilt or punishment. d) Including loss of libido and somatization. e) Including obsessional and compulsory symptoms. f) Depression subscale of the BPRS as defined by Dingemans et al. (1995); depression subscale of the PANSS as defined by Kay et al. (2000).

Table 3
Aspects of reliability and validity of depression instruments in schizophrenia.

a. Reliability

<table>
<thead>
<tr>
<th>Internal consistency</th>
<th>Inter-rater</th>
<th>Test-retest</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>BPRS-D</td>
<td>0.67</td>
<td>0.74</td>
<td>0.72</td>
</tr>
<tr>
<td>PANSS-D</td>
<td>0.77</td>
<td>0.80</td>
<td>-</td>
</tr>
<tr>
<td>HAMD</td>
<td>0.75 (0.73–0.77)</td>
<td>0.94 (0.93–0.95)</td>
<td>0.75 (0.65–0.80)</td>
</tr>
<tr>
<td>MADRS</td>
<td>0.91</td>
<td>0.81</td>
<td>0.71</td>
</tr>
<tr>
<td>CDSS</td>
<td>0.82 (0.76–0.88)</td>
<td>0.86 (0.73–0.98)</td>
<td>0.83 (0.69–0.93)</td>
</tr>
<tr>
<td>BDI</td>
<td>0.90 (0.88–0.91)</td>
<td>n.a.</td>
<td>-</td>
</tr>
</tbody>
</table>

b. Divergent validity

<table>
<thead>
<tr>
<th>Negative symptoms</th>
<th>References</th>
<th>EPS</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>BPRS-D</td>
<td>0.00 (−0.11–0.10)</td>
<td>(2,20,21)</td>
<td>0.14 (0.07–0.21)</td>
</tr>
<tr>
<td>PANSS-D</td>
<td>0.19 (−0.11–0.41)</td>
<td>(3,15,21,24–27)</td>
<td>0.07 (0.01–0.20)</td>
</tr>
<tr>
<td>HAMD-D</td>
<td>0.18 (0.02–0.45)</td>
<td>(2,15,21,23,24,26–31)</td>
<td>0.40 (0.02–0.79)</td>
</tr>
<tr>
<td>MADRS</td>
<td>0.36 (0.12–0.51)</td>
<td>(3,15,23,25)</td>
<td>0.52 (0.16–0.86)</td>
</tr>
<tr>
<td>CDSS</td>
<td>0.10 (−0.24–0.54)</td>
<td>(9,11–15,21,23,24,26,27,32,33)</td>
<td>0.26 (0.07–0.42)</td>
</tr>
<tr>
<td>BDI</td>
<td>0.10 (−0.11–0.21)</td>
<td>(19,26,31,34)</td>
<td>0.23</td>
</tr>
</tbody>
</table>

c. Concurrent validity

<table>
<thead>
<tr>
<th>BPRS-D</th>
<th>PANSS-D</th>
<th>HAMD</th>
<th>MADRS</th>
<th>CDSS</th>
<th>BDI</th>
<th>Pooled mean</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.23</td>
<td>0.66</td>
<td>0.66</td>
<td>0.79</td>
<td>0.64</td>
<td>0.60 (0.17–0.87)</td>
<td>(1,2,5,21,23,28,35,36)</td>
<td></td>
</tr>
<tr>
<td>0.62</td>
<td>0.66</td>
<td>0.49</td>
<td>0.54</td>
<td>0.17–0.87</td>
<td>(3,7,9,11,13,15,19,21,24–27,36)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.80</td>
<td>0.57</td>
<td>0.68</td>
<td>0.26–0.90</td>
<td>(1,2,5,7,8,13–15,21,23,24,26–28,30,31,34,37)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.81</td>
<td>0.56–0.90</td>
<td>(3,7,8,11,14,15,23,25)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.83</td>
<td>0.77</td>
<td>0.26–0.90</td>
<td>(1,7,9,11,15,21,23,24,26,27,37)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.63</td>
<td>0.44–0.90</td>
<td>(1,5,12,19,26,30,31,34,35)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

d. Predictive validity

<table>
<thead>
<tr>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Cut-off value</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>BPRS-D</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>PANSS-D</td>
<td>78% (74–81%)</td>
<td>85% (79–90%)</td>
<td>≥ 5; ≥ 10</td>
</tr>
<tr>
<td>HAMD</td>
<td>79% (67–91%)</td>
<td>83% (81–84%)</td>
<td>≥ 12</td>
</tr>
<tr>
<td>MADRS</td>
<td>81%</td>
<td>81%</td>
<td>≥ 11</td>
</tr>
<tr>
<td>CDSS</td>
<td>88% (67–100%)</td>
<td>88% (74–97%)</td>
<td>≥ 5; ≥ 6; ≥ 9</td>
</tr>
<tr>
<td>BDI</td>
<td>72%</td>
<td>77%</td>
<td>≥ 25</td>
</tr>
</tbody>
</table>
3.3. Reliability

The internal consistency of the BPRS-D was acceptable and good for the remaining instruments in schizophrenia (Table 3a). The inter-rater and test-retest reliability was good for all instruments, especially the inter-rater reliability of the HAMD.

3.4. Divergent validity

The MADRS correlated with negative symptoms and the HAMD with EPS, whereas the other reviewed instruments neither showed substantial correlation with negative nor extrapyramidal symptom dimensions (Table 3b). The following instruments were used for the rating of negative and extrapyramidal symptoms: Affective Flattening Scale (AFS) (Andreasen, 1979), Scale for the Assessment of Negative Symptoms (SANS) (Andreasen, 1982), negative subscale of the PANSS (Kay et al., 1987), negative subscale of the BPRS (Lukoff et al., 1986; Overall et al., 1972), Psychomotor Retardation Scale (Widlocher, 1983), and Rating Scale for Extrapyramidal Side Effects (Simpson and Angus, 1970).

The underlying factor structure of the multidimensional instruments (BPRS and PANSS) generally consisted of one factor for depression and two to four other factors. The depression factor was comprised of the three items “depression,” “guilt” and “anxiety” (Lykouras et al., 2000; McMahon et al., 2002), but additional items loading on the depression factor were “tension” (Lindenmayer et al., 1995; Van Der Gaag et al., 2006; White et al., 1997; Wolthaus et al., 2000), “somatic concern” (Eisenberg et al., 2009; El Yazaji et al., 2002; Lee et al., 2003; Loas et al., 1997) and “suicidality” (Kopelowicz et al., 2008; Ruggeri et al., 2005), or other combinations including “self neglect” (Dingemans et al., 1995) or “motor retardation” (Alves et al., 2005). Inspection of the factor structure of the MADRS and the CDSS did not lead to separate factors for negative symptoms (Lee et al., 2003; Maggini and Raballo, 2006; Wolthaus et al., 2000). The BDI consisted of three factors, including one for “psychosomatic symptoms” (Chemerinski et al., 2008). Of note, no publications reported the factor structure of the HAMD in patients with schizophrenia.

3.5. Concurrent validity

The concurrent validity of the depression instruments in schizophrenia is described in Table 3c. Concurrent validity has been assessed for almost every possible combination of

Notes to Table 3:
a) Reliability was expressed by mean Cronbach's alpha and ICC values; n.a. = not applicable. b) Mean correlation (R²) with either a negative symptom scale or extra-pyramidal symptoms rating scale. c) Average correlation for each comparison of two depression instruments and the pooled mean correlation indices for each instrument. d) Mean sensitivity and specificity values at the optimal cut-off point. References:

1. Addington et al. (1993)
2. Craig et al. (1985)
3. Wolthaus et al. (2000)
5. Baynes et al. (2000)
6. Addington et al. (1996)
8. Lee et al. (2003)
10. Schuetze et al. (2001)
13. Xiao et al. (2009)
15. Liu et al. (2009)
16. Müller et al. (1999)
17. Kontaxakis et al. (2000a)
18. Kaneda et al. (2000)
20. Kuck et al. (1992)
23. Lançon et al. (2000)
24. Collins et al. (1996)
25. Fitzgerald et al. (2002)
27. El Yazaji et al. (2002)
28. Goldman et al. (1992)
31. Norman et al. (1998)
32. Addington et al. (1994)
33. Müller et al. (2002)
34. Mösler et al. (2006)
35. Huppert et al. (2002)
36. Lindenmayer et al. (1992)
37. Müller et al. (2006)
the six instruments. The HAMD was most frequently investigated (by 19 comparative studies), followed by the CDSS, PANSS-D, BDI, BPRS-D and MADRS. The highest concurrent validity indices were found for the CDSS and MADRS.

3.6. Predictive validity

Four studies evaluated whether the six depression instruments adequately predicted the presence of MDE in patients with schizophrenia. Table 3d illustrates that the highest ranges for sensitivity and specificity were found for the CDSS. Of note, the optimal cut-off values obtained for the CDSS and PANSS-D varied widely between studies.

4. Discussion

4.1. Summary of results

We identified five clinician-rated instruments and only one self-report with tested reliability and validity for the measurement of depressive symptoms in patients with schizophrenia.

4.2. Reliability

The reliability of the reviewed depression instruments was good in populations with schizophrenia and comparable to populations with depressed patients or healthy subjects (Müller et al., 2005; Rush et al., 2008). In other words, patients with schizophrenia can reliably be assessed on the presence of depressive symptoms by interview or self-report.

4.3. Validity

The instruments differed in their accuracy to distinguish depressive symptoms from other symptoms of schizophrenia (divergent validity). Correlation studies and factor analysis showed that the CDSS measures nearly no other symptoms of schizophrenia. Inspection of the items of the CDSS supported that the overlap with negative symptoms or EPS was minimal compared to the other depression instruments. The high divergent validity of the CDSS is in line with the fact that this instrument has especially been developed for this population (Addington et al., 1993). For example, “lack of interest” was not included, as this is both a symptom of depression and part of the negative symptoms of schizophrenia (Kulhara et al., 1989; Montgomery, 1979; Romney and Candido, 2001). Divergent validity of the other (older) instruments may be hampered as they are based on several items about anxiety or somatic concern (Snaith, 1993), albeit anxiety-like symptoms do not belong to the current DSM-IV diagnostic criteria for depression.

This wide variation of symptom dimensions covered by the reviewed instruments may explain the modest intercorrelations between most depression instruments. The low concurrent validity between instruments may even be overestimated by the halo-effect. Ideally raters are not influenced by knowledge of the subject’s scores on other instruments (Nisbett and DeCamp Wilson, 1977). However, in some studies multiple instruments for depressive symptoms were rated by a single rater (Lançon et al., 2000), or the distribution of tasks among raters was unclear (El Yazaji et al., 2002; Kim et al., 2006).

The sensitivity and specificity to detect cases of depression in schizophrenia was highest for the CDSS, even though the CDSS did not cover all diagnostic criteria for depression as outlined above. Among the relatively scarce reports of predictive validity we noticed inconsistencies in the reported cut-off values for the PANSS-D and CDSS. Nevertheless we were able to compare the instruments on their predictive validity as we included only those studies with standardized procedures to obtain the optimal cut-off value (area under the curve methods).

4.4. Practical considerations

Practical issues such as time investment may also be important when choosing an instrument, apart from the psychometric aspects discussed above. The amount of training and time to complete the interview of the CDSS was comparable to the HAMD and MADRS. In contrast, the multidimensional instruments BPRS and PANSS may need more time and training to complete the interview, although an advantage may be that besides depressive symptoms, other psychotic symptoms can be evaluated at the same time.

4.5. Future research

An important finding was the lack of self-report instruments for the measurement of depressive symptoms in this population. The concurrent and predictive validity of the only reviewed self-report here BDI was rather poor. Especially for routine outcome monitoring of depressive symptoms in clinical practice, self-report may save time and costs compared to a clinical interview. Although filling out questionnaires may be difficult for patients with considerable cognitive problems (Addington et al., 1993; Müller et al., 2006; Norholm and Bech, 2006) and observable signs of depression could be missed by self-report (Möller, 2009), self-report may provide more independent information on the patients’ experience of depression in schizophrenia than interview-based assessments (Lindenmayer et al., 1992). The literature search identified several other self-report questionnaires for depressive symptoms, such as the CES-D and the BSI (a short version of the Symptom Checklist-90). Evaluation of the composition of the BSI showed that only one of the six items of the depression subscale had potential overlap with negative symptoms [data not shown]. Future research is needed to develop and validate a self-report comparable to the CDSS with respect to reliability and validity in schizophrenia.

4.6. Recommendations and conclusions

In most of the reviewed studies the CDSS outperformed other depression instruments in terms of reliability and validity in patients with schizophrenia. Nevertheless the other depression instruments are still applied in schizophrenia research (Freudenreich et al., 2008; Heald et al., 2008; Möser et al., 2006; Saarni et al., 2010; Schennach-Wolf et al., 2010). This is in accordance to a survey under psychiatrists demonstrating the popularity of the HAMD, BDI and BPRS-D in daily practice (Siris et al., 2001). The current review may aid clinicians and
researchers to choose a well-validated instrument that selectively measures the symptoms of interest.

In summary, the CDSS was most reliable and valid for the measurement of depressive symptoms of schizophrenia. We recommend to use the CDSS in research as well as in daily clinical practice. Patients with a high score should be re-assessed using a diagnostic interview. As self-report is more expedient for the use in routine clinical practice, further research is needed to develop a self-reporting instrument with psychometric properties comparable to the CDSS.

Role of funding source
This study was supported by an unconditional grant from the Christian Fellowship of Care for Mental and Neurological Disorders (VCVZG) and the Dutch Foundation for Mental Health.

Conflict of interest
The authors have not transmitted any conflicts of interest.

L.M.L. received funding and speaker fees by AstraZeneca for a project unrelated to the current manuscript.
R.B. received speaker fees from AstraZeneca, Eli Lilly, and Janssen-Cilag.
H.K. received unconditional research grants from AstraZeneca, Eli Lilly, Janssen-Cilag and Bristol-Myers Squibb and has been working in advisory boards of Eli Lilly and Janssen-Cilag.
C.J.S. received an unconditional grant from Bristol-Myers Squibb for initiating the disease management program.
R.S. and D.W. declared not to have any conflict of interest.

Acknowledgements
The authors wish to thank Truus van Ittersum for methodological advice.

References