

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository:<https://orca.cardiff.ac.uk/id/eprint/106959/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Walton, Catherine and Kerr, Michael 2016. Severe intellectual disability: systematic review of the prevalence and nature of presentation of unipolar depression. *Journal of Applied Research in Intellectual Disabilities* 29 (5) , pp. 395-408. 10.1111/jar.12203

Publishers page: <http://dx.doi.org/10.1111/jar.12203>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See <http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



# Severe Intellectual Disability: Systematic Review of the Prevalence and Nature of Presentation of Unipolar Depression

Catherine Walton\* and Mike Kerr†

\*Core Psychiatry Trainee, Cwm Taf University Health Board, Royal Glamorgan Hospital, Pontyclun, Wales, UK;

†Clinical Professor, Institute of Psychological Medicine and Clinical Neurosciences, Cardiff University, Cardiff, UK

**Background** The diagnosis of depression in severe and profound intellectual disability is challenging. Without adequate skills in verbal self-expression, standardized diagnostic criteria cannot be used with confidence. The purpose of this systematic review was to investigate the assessment and diagnosis of unipolar depression in severe and profound intellectual disability. The review aimed to examine the methods used to assess for depression. The secondary aim was to explore the frequency and symptoms of depression.

**Methods** The PRISMA (2009) Checklist for systematic review was followed, and a search of electronic databases was undertaken. Nine studies were included in the qualitative synthesis from over 2000 records identified. Results The quality of the studies was assessed and scored, with a wide range of results. Individual studies scored between 2 and 7 of a maximum possible score of 8. The diagnostic tools utilized by each of the studies were assessed and compared.

**Conclusions** In terms of the methods used to assess for depression, results were varied. This was due to the heterogeneous nature of the individual study designs. The Aberrant Behaviour Checklist consistently showed promise, in particular when combined with other instruments or clinical examination. Qualitative analysis of the selected studies has shown a wide variation in the quality of primary research in this field, with more required to make firm conclusions regarding the diagnosis, frequency and presentation of depression in severe and profound intellectual disability.

## Introduction

Prevalence of mental illness in intellectual disability has been estimated at between 30 and 50%. This figure varies due to the differences in definition of mental illness, the setting of the study and the participants' level of intellectual disability (Borthwick-Duffy & Eyman 1990). Point prevalence of major depressive illness in intellectual disability is 2–7%, and the rate for the general population is 3–5%, illustrating that depression in intellectual disability is not at a dissimilar level to that of the general population and could be higher (Prasher 1999). The prevalence and incidence of depression in those with severe and profound intellectual disability are difficult to accurately define, due to a paucity of good quality published research (Cooper et al. 2007a).

Depression is the most common mental health problem diagnosed in the community (National Collaborating Centre for Mental Health, 2010). It is characterized by its range of symptoms, including loss of positive affect plus related cognitive, emotional, physical and behavioural difficulties. Differentiating between 'normal' mood changes and those that are clinically significant can be challenging (National Collaborating Centre for Mental Health, 2010). These difficulties are compounded in intellectual disability, in particular those with severe intellectual disability who are, by definition, unable to communicate their thoughts and feelings (Matson et al. 1999).

The idea that individuals with intellectual disability suffer from mental illness is relatively novel – early research claimed that those with intellectual disability did not suffer from depression (Earl 1961). However, in the latter part of the 20th century research showed that individuals with intellectual disability were suffering from depressive episodes. The term 'diagnostic overshadowing' was established (Smiley & Cooper 2003). It referred to the idea that once an individual had diagnosis of intellectual disability then all other problems that individual may suffer from would be attributed to the intellectual disability diagnosis. Changes in behaviour are attributed to the intellectual disability as opposed to recognition of mental illness. Diagnostic overshadowing

can lead to a lack of thorough psychiatric assessment and therefore suboptimal treatment and care (Yoo et al. 2012).

Morbidity associated with depression in the general population is high. There is a significant impact upon level of function, physical health and mortality (National Collaborating Centre for Mental Health 2010). Therefore, the recognition and treatment of depression is paramount in all populations. This need is compounded in the intellectual disability population, to narrow existing health inequalities (Cooper et al. 2007b). According to Cooper et al. (2007a,b), literature pertaining to depression in intellectual disability is scarce and can be methodologically flawed. Further knowledge of the prevalence, incidence and presentation of depression in severe and profound intellectual disability will guide the development of effective interventions and services (Cooper et al. 2007b).

Features of depression such as anhedonia, guilt, self-reproach and suicidal thoughts require the individual to both recognize and express innermost thoughts and feelings. These features form part of the diagnosis of depression using standard diagnostic systems such as ICD-10 (World Health Organisation, 1992). Those with severe intellectual disability lack these skills, and therefore, standardized rating scales may not be appropriate. It is proposed that individuals with severe intellectual disability may suffer from a different range of symptoms in comparison with less severe intellectual disability and the general population (Meins 1995).

The purpose of this systematic review was to investigate the assessment and diagnosis of unipolar depression in severe and profound intellectual disability. The review will aim to examine the methods used to assess for depression. The secondary aim was to explore the frequency and nature of presentation of depression in individuals with severe and profound intellectual disability. For the purpose of this review, severe and profound intellectual disability will be referred to collectively as 'severe' intellectual disability.

## **Methods**

The Preferred Reporting Items for Systematic Reviews and Meta-analyses: The PRISMA statement (2009) checklist for systematic review was followed where possible. The PRISMA checklist was developed specifically to improve reporting quality and the assessment of primary research in systematic reviews.

Eligibility criteria

Participants

Individuals with severe or profound intellectual disability were diagnosed by accepted criteria (IQ testing and/or functional ability. This would be IQ 34 and below.)

No limitations for age or gender.

Intervention or test of interest

Primary research investigating the method of diagnosis and rate of unipolar depression in severe intellectual disability.

Exclude bipolar affective disorder.

Exclude challenging behaviour (see Appendix 1 for rationale).

Exclude studies pre-dating 1990 due to volume of studies.

## **Comparison**

The aim was to complete a systematic review and a qualitative analysis of the studies found.

## **Outcome**

Data for the presentation of unipolar depression, methods of diagnosis and rates of depression in individuals with severe intellectual disability.

## **Study design**

Primary research excluding individual case studies.

### Information sources

A computerized search using Cardiff University's Electronic Portal of the following databases was undertaken; Medline, Embase, PsychInfo, Web of Science and CINAHL. This was for English language, peer-reviewed journals published between 1 January 1990 and 30 September 2013.

Review articles pertaining to depression and intellectual disability were also searched.

### Search strategy

At this stage, search terms were kept broad, in order for as many studies to be found as possible.

Search terms were as follows.

Intellectual disability Developmental disability Learning dis\*

Learning difficulties Mental retardation

Or Learning disorder

These were searched individually and then combined with the following:

Psychopathology Depression

Mood disorder

Or Affective disorder

## **Study selection**

When the two groups of terms were combined on the individual databases, the resulting titles were reviewed by the author. The bibliographies of relevant review articles were also searched at this stage. Studies selected presented original data pertaining to severe intellectual disability and depression.

### Summary measures and synthesis of results

To aid objective comparison of the selected studies, the system devised by Hermans & Evenhuis (2010) was utilized. Hermans & Evenhuis (2010) developed a system to evaluate screening tools for the identification of depression in the general intellectual disability population. (See Appendix 2 for full Hermans & Evenhuis (2010) measures). An additional column has been added to the table to clearly present any frequency measure of depression (i.e. percentage, prevalence or incidence). This replaces the column assessing for psychopathology of participants with reference to a screening tool, which is not applicable to this review.

## **Results**

A total of 60 articles were found that were deemed relevant to the aims and objectives of this review (Figure 1).

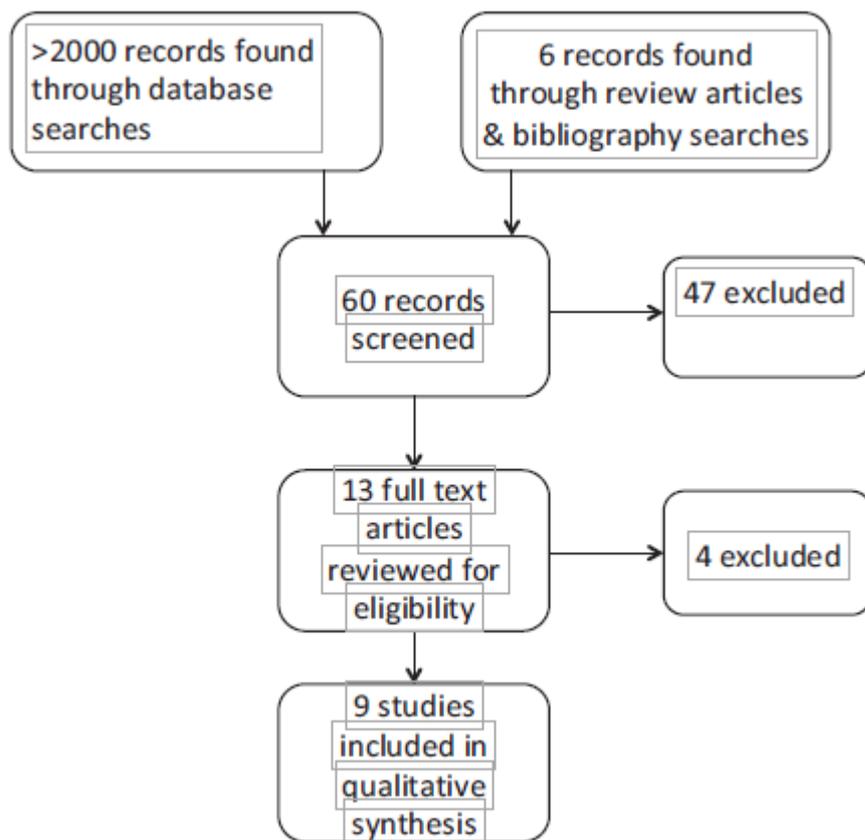


Figure 1 Results of search strategy as per PRISMA Guidance for severe intellectual disability (Moher et al. 2009).

The authors independently reviewed the abstracts of the 60 articles, and 13 articles were selected for detailed review. From these 13 articles, four were excluded. In the case of three of the articles, this was due to the paper not specifically investigating the diagnosis of depression in individuals with severe or profound intellectual disability. The articles were the following: Meins (1995), Marston et al. (1997) and Charlot et al.(1993) One further article, Kozłowski et al. (2011), was excluded at this stage due to its primary focus being on symptom clusters in those with severe intellectual disability instead of diagnosis of depression.

One study caused discussion between the authors. This was Charlot et al.'s (2007) investigation of the validation of the Mood and Anxiety (MASS) interview. The instrument is aimed for use with non-verbal subjects (in particular those with severe and profound intellectual disability). However, the study assessed a range of intellectual disability. It was decided that the inclusion of the study would add further depth to discussion.

#### Study characteristics

Table 1 shows a summary assessment of the quality of individual studies. From a total possible score of 8, it is apparent that there is a wide range in quality as assessed by these criteria. The studies were varied in design and measured outcomes and therefore could not be compared by meta-analysis.

**Table 1** Quality of the selected studies

|                         | Instrument   | n                    | Characteristics  | Depression distribution   | Gold standard  | Report on measures of variability | Quality score (eight maximum) |
|-------------------------|--|----------------------|--|---|--|-----------------------------------|-------------------------------|
| Ross & Oliver (2003a)   | Mood, Interest and Pleasure Questionnaire (MIPQ)   | 53                   | Severe and profound intellectual disability<br>Community sample  | History mental health problems<br>18.9% Referred psychiatrist in last year<br>30.2% On antidepressant medication<br>57% diagnosed using DSM-IV criteria | ABC used as other screening instrument as comparator               | 1                                 | 6                             |
| Matson et al. (1999)    | Diagnostic Assessment for the Severely Handicapped-II (DASH-II)  | 57                   | Severe and profound intellectual disability<br>Residential facility  | 31.5% diagnosed using DSM-IV criteria   | Diagnosis by psychiatrist based on DSM-IV criteria                 | 1                                 | 7                             |
| Paclawski et al. (1997) | Diagnostic Assessment for the Severely Handicapped-II (DASH-II) and Aberrant Behaviour Checklist (ABC) | 233                  | 2.2% mild<br>5.2% moderate<br>15.1% severe<br>75.4% profound<br>Residents of a developmental centre  | -   | Comparison of DASH-II with ABC                                     | 0                                 | 3                             |
| Cherry et al. (1997)    | Diagnostic Assessment for the Severely Handicapped (DASH)  | 168                  | Severe and profound intellectual disability<br>Residential facility  | Young - 21%<br>Old - 25%  | Matson et al. (1991) Criteria for the diagnosis of psychopathology | 1                                 | 7                             |
| Mudford et al. (1999)   | Desamethasone suppression test   | 40                   | Severe and profound intellectual disability<br>Residential facility  | 7/40 DST positive = 17.5%   | DASH   | -                                 | 5                             |
| Cooper et al. (2007b)   | N/A<br>Population-based prospective cohort study   | T1 184<br>T2 134     | Profound intellectual disability<br>Community sample   | 1.1% (DSM-IV-TR) to 3.3% (clinical diagnosis and DC-LD).  | Psychiatric assessment   | 0                                 | 6                             |
| Charlot et al. (2007)   | Mood and Anxiety Semi-structured (MASS) interview.   | 93                   | 52% mild intellectual disability<br>3% moderate intellectual disability<br>13% profound or severe intellectual disability<br>Psychiatric admissions unit | 44% DSM-IV diagnosis<br>59% MASS  | DSM-IV diagnosis on discharge                                      | 0                                 | 6                             |
| Evans et al. (1999)     | Behavioural criteria of DSM-IV-R for depression and Aberrant Behaviour Checklist (ABC)                 | 89                   | Severe and profound intellectual disability<br>Residential facility  | -   | Nurse pairs using DSM-IV-R criteria                                | 0                                 | 2                             |
| Tsiouris (2001)         | Marston 30 symptoms checklist  | 22 selected from 150 | 3 mild intellectual disability<br>4 moderate intellectual disability<br>9 severe intellectual disability<br>six profound intellectual disability         | Preselected   | DSM-III-R criteria   | 0                                 | 2                             |

**Table 2** Psychometric properties of instruments utilized

| Relevant psychometric properties                 |   |   |                         |  |   |  |                           |
|--|---|---|-------------------------|--|---|--|---------------------------|
| Study  | Instrument  | Description   | Test-retest reliability | Inter-rater reliability                          | Internal consistency  | Pearson correlation coefficient (comparison of scales)   | Sensitivity & specificity |
| Ross & Oliver (2003a)                            | Mood, Interest and Pleasure Questionnaire (MIPQ)                | 25-item Likert scale questionnaire split into the two subscales: <ul style="list-style-type: none"> <li>mood &amp; interest</li> <li>pleasure</li> </ul>  | 0.87                    | 0.76   | Cronbach's alpha: Mood: 0.89 Interest & pleasure 0.90 Total score MIPQ 0.94   | MIPQ total score and ABC 'lethargy, social withdrawal' subscale - negative correlation predicted: Pearson correlation coefficient -0.59 ( $P < 0.001$ )  | Not available             |
| Ross & Oliver (2003a)<br>Paclawski et al. (1997) | Aberrant Behaviour Checklist (ABC)                              | 58 items describing behaviour problems that correspond to 5 categories: <ul style="list-style-type: none"> <li>irritability</li> <li>agitation &amp; crying</li> <li>lethargy &amp; social withdrawal, stereotypic behaviour, hyperactivity &amp; non-compliance</li> <li>excessive speech</li> </ul> | Not available           | Not available                                    | Cronbach's alpha 0.95 (Ross & Oliver 2003a)<br>Coefficient alpha - subscales all >0.78 (fair), with lethargy scored 0.92 (excellent). Total 0.94. (Paclawski et al. 1997) | As above - significant negative correlation with the MIPQ total score (Ross & Oliver 2003a)<br>DASH-II depression subscale with ABC: Correlated with: irritability, lethargy, stereotypies, hyperactivity and speech ( $P < 0.001$ ). Pearson correlation coefficient >0.5 (Paclawski et al. 1997) | Not available             |
| Matson et al. (1999)<br>Paclawski et al. (1997)  | Diagnostic Assessment for the Severely Handicapped-41 (DASH-41) | Screening instrument with 13 subscales including anxiety, depression and mania  | Not available           | Not available                                    | Coefficient alpha score for total score 0.87. Average for the subscales was 0.51. (Paclawski et al. 1997)   | See results above. Paclawski et al. (1997)   | Not available             |
| Cherry et al. (1997)                             | Diagnostic Assessment for the Severely Handicapped (DASH)       | Earlier version of DASH   | Not available           | Frequency 0.91<br>Duration 0.95<br>Severity 0.96 | Coefficient alpha: Depression subscale 0.48. Overall 12 of the 13 subscales had values less than 0.8 indicating less than adequate internal consistency                   | Not available  | Not available             |

Table 2 (continued)

| Study                 | Instrument   | Description  | Relevant psychometric properties |   |                      |   | Pearson correlation coefficient (comparison of scales)  | Sensitivity & specificity |
|-----------------------|--|--|----------------------------------|---|----------------------|---|---|---------------------------|
|                       |  |  | Test-retest reliability          | Inter-rater reliability   | Internal consistency |   |   |                           |
| Charlot et al. (2007) | Mood and Anxiety Semi-structured (MASS) interview                                      | Based on behavioural descriptions of DSM-IV-TR symptom criteria of mood and anxiety disorders  | Not available                    | Cohen's kappa coefficient: Major depression subgroup: 0.58 ( $P < 0.001$ ) (moderate agreement) | Not available        | Not available   | MASS interview<br>Sensitivity: 0.93<br>Specificity: 0.67<br>(Compared to clinical diagnosis at discharge) |                           |
| Evans et al. (1999)   | Behavioural criteria of DSM-IV-R for depression and Aberrant Behaviour Checklist (ABC) | Assessment of the use of the behavioural criteria of DSM-IV-R by nurses to recognize depression  | Not available                    | Intraclass correlation coefficient: 0.50 ( $P < 0.05$ ) (fair agreement)                        | Not available        | Correlation with DSM-III-R criteria for MDE and ABC subscale: Irritability (+0.50), lethargy (+0.44) (Significant $P < 0.001$ ) and stereotypy (+0.4) (significant $P < 0.05$ ). Speech – not significant | Not available   |                           |
| Mudford et al. (1995) | Dexamethasone suppression test   | Biochemical test suggested to be specific to depression. Compared to DASH and BSD (Behavioural Symptoms of Depression – Mudford et al. 1995) | Not available                    | Not available   | Not available        | Not available   | DST sensitivity: 20% for DASH 22% BSD<br>DST Specificity: >80% for both DASH & BSD                        |                           |
| Cooper et al. (2007b) | Not applicable: Population-based prospective cohort study                              | -  | -                                | -   | -                    | -   | -   |                           |

Ross & Oliver (2003a) investigated the use of the Mood, Interest and Pleasure Questionnaire (MIPQ) for individuals with severe and profound intellectual disability. Items are based on the definitions of low mood and anhedonia from ‘criteria for major depressive episodes’ in Diagnostic and Statistical Manual of Mental Disorders-IV (DSM-IV) (American Psychiatric Association, 2000) and also on ‘symptomatic behaviours’ (Lowry 1998). The aim of the study was to compare the use of the MIPQ with the subscale ‘lethargy and social withdrawal’ of the Aberrant Behaviour Checklist (ABC) as a test of construct validity. (See Table 2 for further information on the scales).

Participants scored a mean of 34.65/48 (range 10–48) for the mood subscale and 26.87/52 (12–46) for the interest and pleasure subscale. Lower scores were indicative of lower mood, interest and pleasure levels. No comment was made upon how these scores would translate to rates of depression.

Test–retest reliability and inter-rater reliability of the MIPQ were assessed (see Table 2). Test–retest reliability for total scores was 0.87, and inter-rater reliability was 0.76, which, according to Hermans & Evenhuis (2010) ‘guideline to interpret internal consistency and correlation coefficients’, show a high correlation. Ross & Oliver (2003a) concluded that this was evidence that reliability of the MIPQ is ‘highly satisfactory’. The internal consistency of the MIPQ was also assessed using Cronbach’s alpha (see Table 2). The overall score for MIPQ was 0.94 – showing excellent internal consistency according to Hermans & Evenhuis (2010) criteria (Appendix 2).

The association between MIPQ scores and the ABC's lethargy, social withdrawal scale was assessed to investigate the validity of the MIPQ. Significant ( $P < 0.001$ ) negative correlation (predicted) was found with both the lethargy, social withdrawal scale of the ABC and also the stereotypic behaviour ABC subscale. Significant negative correlation ( $P < 0.05$ ) was also observed between the MIPQ mood subscale and the ABC irritably, anger and crying subscale.

Ross & Oliver (2003a) concluded that the preliminary investigation of the MIPQ for the assessment of mood in individuals with severe and profound intellectual disability was promising and comparable to the ABC.

Matson et al. (1999) assessed 57 people with severe and profound intellectual disability in a residential facility using the Diagnostic Assessment for the Severely Handicapped-II (DASH-II). The DASH-II was created in 1991 using the DSM classification system and the authors' past experience. This study focused on the depression subscale. The aim was to evaluate the scale as a preliminary screen for depression in those with severe or profound intellectual disability (Matson et al. 1999).

Fifty-seven individuals were split into groups. Eighteen had a DSM-IV diagnosis of depression, a control group of 19 with autism and 20 with no diagnosis. There was no apparent matching for age or other demographic characteristic. The study showed that of the group with pre-existing depression, 73.3% had a rise in the DASH-II depression subscale. The control group with no DSM-IV diagnosis showed no significant elevation. A repeated-measures analysis of variance (ANOVA) showed a significant difference between the groups ( $P < 0.001$ ). The number of individuals accurately diagnosed with depression was 93.3% (Matson et al. 1999).

The second part of this study investigated the most frequently identified core and associated symptoms in the depressed group. The most commonly identified core symptoms (from 15 on the subscale) were 'restless or agitated' (73.3%), 'responds slowly' (53.4%) and 'cranky or irritable' (53.3%). Other less commonly identified symptoms were 'has difficulty staying awake during the day', 'wakes frequently through the night', 'cries easily or for no obvious reason'. The three most commonly found associated symptoms were 'resists or ignores others attempts to interact with them' (66.7%), 'mood seems totally unrelated to what's going on around him' (53.3%) and 'resists instruction or guidance from family or staff' (53.3%) (Matson et al. 1999). It was noted that 46% of the individuals in the depressed group had increased scores for mania and impulse subscales.

Matson et al. (1999) concluded that the study showed that the DASH-II was a valid indicator of depression according to DSM-IV criteria. The findings of the second part of the study showed that individuals with severe or profound intellectual disability presented with depression in non-verbal ways. The core and associated symptoms revealed disturbances in sleep, psychomotor problems and irritability. Also, there is a change in their ability to engage in their usual manner with the development of stereotypic behaviours, aggression, non-compliance and unresponsiveness to environment (Matson et al. 1999).

Paclawski et al. (1997) investigated the validity of the DASH-II in correlation with the ABC. There was a mix of levels of intellectual disability with the majority (90.5%) being severe or profound. The data were analysed to evaluate convergent validity between the 2 scales (see Table 2). The Pearson correlation coefficient was greater than 0.5 (moderate correlation according to Hermans & Evenhuis 2010 guidance) in irritability and hyperactivity. The correlation coefficient for internal consistency of the scales was variable, but 'acceptable' according to the authors (see Table 2.)

The measure of internal consistency (see Table 2) was compared between the DASH-II and the ABC; it appeared that the ABC subscales were more reliable and consistent than the DASH-II depression subscales. However, the analysis did show good correlation between the two scales.

Table 1 shows that the study by Cherry et al. (1997) scored the highest for quality. It did not score full marks because the 'gold standard' that the patients' outcome on the DASH was compared to was a scale developed by the authors called the 'criteria for the diagnosis of psychopathology', Matson et al. (1991), as opposed to standardized criteria. In comparison with Matson et al.'s 1999 study, this study utilized the DASH (Diagnostic Assessment for the Severely Handicapped), an earlier version of the DASH-II.

Cherry et al. (1997) investigated the incidence of psychopathology in older adults with severe and profound intellectual disability. The mean frequency of depression found in the sample by the DASH depression subscale was 0.21 in young people and 0.25 in older adults. Univariate analysis of variance showed no age effects for depression. Internal consistency was assessed and the depression subscale scored 0.48 – which according to Hermans & Evenhuis (2010) would be interpreted as unacceptable. The authors do state that the coefficient alpha values for 12 of the 13 DASH subscales are below the minimum of 0.8, which is less than adequate.

Of interest, Cherry et al. (1997) did comment that individuals with severe and profound intellectual disability appear to present less often with 'classic forms' of mental illness, yet 'aberrant behaviour disorders' were evident. This would mean that although psychopathology is likely to be prevalent in this population, prevalence rates might be different to the general population for specific disorders. More commonly presenting findings were symptoms of mania and pervasive developmental disorder.

Mudford et al. (1995) investigated the use of the dexamethasone suppression test (DST) for the diagnosis of depression in severe LD. Forty participants were selected from a residential facility. The participants were assessed using the ABC, the DASH and a questionnaire on Behavioural Symptoms of Depression (BSD). Five participants fulfilled possible criteria for depression using the DASH depression subscale. This increased to nine participants by scoring on the BSD scale. DST-positive results (i.e. non-suppression) were detected in seven participants.

Sensitivity for the DST using the DASH depression subscale was 20 or 22% using the BSD. Specificity fared better with 86% comparing with the DASH depression subscale and 87% for the BSD (see Table 2). Diagnostic confidence for the DST using DASH depression subscale was 17 or 33% for BSD. When atypical symptoms of depression were investigated using the ABC profiles, none of the ABC subscales differentiated DST positive from DST negative (Mudford et al. 1995). Table 1 shows that this study received five of a possible score of 8. No measures of variability or consistency were quoted. No recognized diagnostic 'gold standard' was utilized.

A population-based prospective cohort study was used by Cooper et al. (2007b) to investigate mental ill health in adults with profound intellectual disability. The community-based sample had the largest sample size of all the selected studies (T1 184 participants, T2 134 participants). Depression was assessed using a full psychiatric assessment including clinical examination, assessment using Diagnostic Criteria for psychiatric disorders for adults with Learning Disabilities (DC-LD)(Royal College of Psychiatrists, 2001), Diagnostic Criteria for Research ICD-10 (DCR-ICD-10) (World Health Organisation 1992) and Diagnostic and Statistical Manual of Mental Disorders-IV Text Revision (DSM-IV-TR) (American Psychiatric Association 2000), plus the PAS-ADD (Psychiatric Assessment Schedule for Adults with a Developmental Disability) (Moss et al. 1993) which is a screening tool for mental ill health in the intellectual disability population.

At T1, the point prevalence of affective disorder was found to range between 1.1% (DSM-IV-TR) and 3.3%(clinical diagnosis and DC-LD). The 2-years incidence of affective disorder in the cohort was found to range between 1.5% (DSM-IV-TR) and 6.1% (clinical diagnosis).

Charlot et al. (2007) aimed to validate the MASS (Mood and Anxiety Semi-Structured) interview for individuals with intellectual disability – comparing scores with the DSM-IV diagnosis of individuals on discharge from an inpatient psychiatric facility and the HDRS (Hamilton Depression Rating Scale). It is aimed at individuals without expressive language skill therefore uses informant history.

As is apparent from Table 1 not all of the 93 participants had severe or profound intellectual disability. The sample was from an inpatient admissions unit. Fifty-five of 93 participants were diagnosed with depression by MASS interview, in comparison with 41 of 93 participants by DSM-IV. Charlot et al. (2007) found that the sensitivity of the MASS in comparison with clinical diagnosis at discharge was 0.93 and specificity was 0.67. This showed a reasonably good degree of agreement between the clinical diagnosis and the MASS interview. Table 2 shows that Cohen's kappa coefficient for the major depression subgroup was 0.58. This equates to moderate inter-rater agreement as per Landis and Koch's guidance (1977), in Charlot et al. 2007). The aim of Evans et al. (1999) was to assess the level of agreement between nurse pairs of their assessment of an individual with severe intellectual disability. The nurses assessed the individuals over 4 weeks according to the

behavioural criteria for major depressive episode in DSM-III-R. Inter-rater agreement showed that 82% of pairs had agreement >70 and 62% of pairs had agreement >80% (using adjacent/lenient agreement). The inter-rater reliability was also calculated using intraclass correlation coefficient. Calculating the nurse pairs as a group, the intraclass correlation coefficient was 0.50 ( $P < 0.01$ ). According to Cicchetti & Sparrow (1981)'s criteria, this was a 'fair' level of agreement.

Evans et al. (1999) also investigated the correlation between total scores for DSM-III-R criteria for major depressive episode and the scores for the subscale on the ABC. It was found that the nurse pairs' scoring using the DSM-III-R checklist did correlate positively with the subscales of the ABC (see Table 2). The correlation would be considered low to moderate according to Hermans & Evenhuis (2010) guidance.

Tsiouris (2001) investigated the use of the Marston 30 Symptoms Checklist for detecting depression in a preselected group who had already been diagnosed with major depression using DSM-III-R criteria by a psychiatrist (Tsiouris 2001). The Marston 30 Symptoms checklist was developed following work by Meins (1995) who reported higher incidences of aggression, tantrums, screaming, crying and stereotypes in people with severe intellectual disability (Tsiouris 2001). The term 'behavioural depressive equivalents' was introduced.

From the 150 investigated for depression, 22 were selected. Not all had severe intellectual disability (see Table 1). In the severe subgroup, the signs and symptoms most commonly observed were depressive affect or irritable mood, sleep disturbances, appetite disturbances, loss of interest, social isolation, psychomotor agitation, self-injury and aggression (Tsiouris 2001). This was taken from the 15 with severe intellectual disability, with at least 10 of the 15 showing these symptoms. No further statistical analysis was undertaken which reflects the quality score of the study – 2 (Table 1).

## **Synthesis of Results**

### **Methods of diagnosis of depression**

The definition of 'depression' in the selected studies is varied. Not all utilized an accepted 'gold standard' definition to lay claim to the success of their screening or diagnostic instrument. In terms of quality, 'clinical diagnosis' of depression is at the higher scoring end of the range. This suggests that diagnoses may be missed if standard criteria for the general population are applied (i.e. DSM-IV or ICD-10) without the combination with clinical assessment by a psychiatrist. This raises questions about the presentation of affective disorders in severe intellectual disability and whether the acceptance of the use of standardized criteria for the diagnosis of depression as a 'gold standard' may well in itself introduce bias into the review. Perhaps, instead, studies utilizing clinical diagnosis should be afforded a greater weighting.

The DASH and DASH-II were the only tools designed specifically for individuals with severe or profound intellectual disability. Matson et al. (1999) showed that the depression subscale of the DASH-II was able to select those with depression in 93% of cases. This was, however, a small sample size drawn from a residential facility. Also, as pointed out by Ross & Oliver (2003a), some items on the measure rely on self-report from the individual. On further analysis of the DASH-II in terms of correlation with the ABC (Paclawski et al. 1997), the depression subscale measure of internal consistency was rated poorly (overall, the scale was rated as acceptable). This fitted with analysis of the predecessor of the DASH-II, the DASH, where the measure of internal consistency of the scale was deemed 'unacceptable'. In comparison, the work by Paclawski et al. (1997) with the ABC scale showed more promising scores of internal consistency. The authors concluded that individuals scoring highly on any of the ABC subscales warranted further assessment of mood.

Ross & Oliver (2003a) aimed to design a questionnaire completed by a carer and not reliant on any verbal skills from the individual being assessed. The MIPQ did fare well in terms of internal consistency of the scale. It was also significantly correlated with the ABC. This preliminary data show that the combination of assessments could be a useful tool in the assessment of mood in severe and profound intellectual disability.

The MASS interview (Charlot et al. 2007) was based on behavioural symptoms from the DSM-IV-R. It proved to have fairly poor sensitivity, yet the specificity was better. The subjects were recruited from an inpatient

psychiatric unit. This could be considered a weakness of the study, that is selection bias due to poor generalisability to the rest of the population.

Overall, the studies included were small, which will introduce bias due to a lack of statistical power. Another weakness of most of the studies (except Cooper et al. 2007b) was the selection of participants. This was generally from large long-term residential units or psychiatric units, increasing the potential for selection bias. The differing aims and methodologies of the studies mean that there is not one firm conclusion to be drawn about tools for screening for, or the diagnosis of, depression in this population. It is apparent from Paclawski et al. (1997) and Ross & Oliver (2003a) that the ABC has scored better in terms of measures of reliability and consistency than other scales. Also, it appears regularly as a comparator, which perhaps reflects a popular choice for ease of use and perceived reliability by researchers.

#### Prevalence and incidence of depression

Not all of the studies specifically aimed to measure the prevalence of depression. Cooper et al. (2007b) found that prevalence of affective disorder varied dependent upon the method of diagnosis. For example, at T1, 1.1% (DSM-IV-TR) in comparison with 3.3% (clinical diagnosis and DC-LD).

The 2-years incidence of affective disorder in the cohort was between 1.5% (DSM-IV-TR) and 6.1% (clinical diagnosis). However, 50 subjects were no longer part of Cooper et al.'s (2007b) study by point T2 with the possibility of attrition bias.

In comparison with the community sample of Cooper et al. (2007b), Charlot et al. (2007) utilized a population from a psychiatric admissions unit. The prevalence of depression in this sample was markedly higher (44% DSM-IV diagnosis and 59% MASS). This illustrates the difference between the two populations and the importance of sample selection. This result is reflected by Mudford et al. (1995) using the DST with a point prevalence of depression of 35%.

Cherry et al. (1997) compared the frequency of depression different age groups using the depression subscale of the DASH. No age effects were found for depression.

#### Presentation of depression

Matson et al. (1999) showed that individuals with severe or profound intellectual disability present with depressive symptoms in non-verbal ways including disturbed sleep, psychomotor changes, increased irritability and a change in the ability of the individual to engage in their usual manner.

Tsiouris (2001) investigated the diagnosis of depression using the Marston 30 Symptoms Checklist. This checklist was based on the idea of 'behavioural depressive equivalents'. This was a small study in which depressed individuals were preselected. There was little statistical analysis of the results. However, of some value was the list of most commonly observed features, that is, depressive affect, irritable mood, sleep disturbances and appetite disturbance. A positive point was the manner in which identification of depressed individuals took place. This used a multifaceted approach – including assessment by a psychiatrist, DSM criteria, casenote review, discussion with carers and behavioural observation.

#### Discussion

Various measures of psychopathology have been assessed in this review. These are the DASH, the DASH-II, the MIPQ, the DST, the MASS, the Marston 30 Symptoms Checklist, the ABC and the behavioural criteria for DSM-IV-R. Other scales of general psychopathology which include depression subscales not included in this review (due to a lack of studies found investigating severe or profound intellectual disability) include the Reiss Screen for Maladaptive Behaviour, the PAS-ADD (Psychiatric Assessment Schedule for Adults with a Developmental Disability) (Moss et al. 1993) and the PIMRA (Psychopathology Inventory for Mentally Retarded Adults) (Senatore et al. 1985). Not all are designed specifically for severe intellectual disability.

According to Paclawski et al. (1997), individuals with severe intellectual disability are less likely to receive a formal psychiatric diagnosis than those with mild or moderate intellectual disability. This is due to the difficulty

in self-expression to report their subjective feelings when assessed and the difference in the presentation of the clinical features of the depressive disorder. There is little literature specifically investigating severe intellectual disability in comparison with general intellectual disability.

This review has given some insight into the presentation of depression in severe intellectual disability. As illustrated by Cooper et al.'s (2007b) cohort study, clinical diagnosis appears to diagnose affective disorders at a higher rate than standardized (non-intellectual disability) criteria. This shows that there is a need for the development of screening and diagnostic tools specifically designed for the severe intellectual disability group. These need to be valid, consistent and accurate – a stage perhaps not yet reached by some of the instruments reviewed, for example the DST (Mudford et al. 1995). However, the ABC (perhaps in combination with another tool, such as the MIPQ) could be of use in day-to-day clinical work. The use of these scales as screening tools could prompt clinical examination by a clinician. As illustrated by Matson et al. (1999) and Tsiouris (2001), the descriptive symptoms of depression in severe intellectual disability are valuable to current clinical practice. Matson et al. (1999) found that individuals tended to present in non-verbal ways such as changes in sleep patterns, psychomotor problems and irritability. Cherry et al. (1997) commented upon 'aberrant behaviour disorders'. With knowledge of possible presenting symptoms, it is possible to have a high index of suspicion for depression in this group. The diagnosis is certainly not straightforward and does not appear to present in the classical manner according to DSM or ICD guidelines. However, there are aspects of presentation that do tally, and these must be the 'red flag' signs for clinicians working with the severe and profound intellectual disability population.

This review has encountered and commented upon weaknesses in the design of studies for depression in severe intellectual disability. Ross & Oliver (2003b) succinctly summarize these difficulties, which are due to differences in the following:

- Populations studied (i.e. community versus institutional sample)
- Sampling (i.e. clinical sample versus support network)
- Diagnostic criteria
- Prevalence rates quoted and lack of clarity about lifetime risk or point prevalence quoted.
- Diagnostic approach in people with little or no expressive language
- Inclusion of behavioural disorder, which increases prevalence rates for general psychopathology.

The exclusion of primary research that included challenging behaviour may have weakened the conclusions of this review. However, the evidence for an association between challenging behaviour and the diagnosis of depression is inconclusive (see Appendix 1 for rationale). Studies that have concluded that there is a demonstrable link that has been small scale and underpowered (Hayes et al. 2011). It must be acknowledged that evidence does exist for a relationship between aggressive behaviour and impulse control disorder and bipolar disorder (Tsiouris et al. 2011); therefore, further larger-scale research in all mood disorders is likely to be necessary. Tsiouris et al. (2011) also observed the under diagnosis of depression in individuals with severe and profound intellectual disability who present with self-injurious behaviour; therefore, future reviews may consider widening inclusion criteria in this direction.

A further weakness of this review remains the lack of evidence found specifically for individuals with severe intellectual disability. There is likely to be much sound research in papers referring to general intellectual disability, which may have subcategories investigating severe intellectual disability that will have been missed by the strict inclusion criteria of this review. The existing evidence is heterogeneous, and therefore, one is unable to analyse and compare directly by, for example, meta-analysis. Instead, a qualitative comparison had to be undertaken.

In conclusion, this aim of this systematic review was to assess the existing evidence regarding the methods of assessment and diagnosis of unipolar depression in severe and profound intellectual disability. Secondary outcomes of interest were the frequency and presentation of depression in these individuals. The rate of depression in severe and profound intellectual disability is likely to be close to that of those with less severe impairment. However, it is probable that it is under diagnosed. Through qualitative analysis of the studies found, it has been shown that there are multiple scales and tools utilized to aid in diagnosis. The relative

merits of these scales have been discussed; however, the take-home message from the evidence appears to be that clinical diagnosis by an experienced clinician continues to be superior to the use of rating scales alone. Although the strict inclusion criteria of this review may have limited the number of studies included, it is fair to say that there is relatively little good quality research in this area. Further investigation is needed. A starting point might be assessing the reliability of the combination of two scales. There is also scope for investigation of the combination of an initial screening instrument plus clinical examination, which is closer to the realities of clinical practice, as suggested by Paclawski et al. (1997).

### **Correspondence**

Any correspondence should be directed to Catherine Walton, Mental Health Directorate, Royal Glamorgan Hospital, Pontyclun, Wales, CF72 8XR, UK. (e-mail: catherine.walton@wales.nhs.uk)

### **References**

- American Psychiatric Association (2000) Diagnostic and Statistical Manual of Mental Disorders, 4th edn. Washington, DC, American Psychiatric Association.
- Borthwick-Duffy S. A. & Eyman R. K. (1990) 'Who are the dually diagnosed? American Journal on Mental Retardation 36, 586–595.
- Charlot L. R., Doucette A. C. & Mezzacappa E. (1993) Affective symptoms of institutionalized adults with mental retardation. American Journal on Mental Retardation 98, 408–416.
- Charlot L., Deutsch C., Hunt A., Fletcher K. & McIlvane W. (2007) Validation of the Mood and Anxiety Semi-structured (MASS) Interview for people with intellectual disabilities. Journal of Intellectual Disability Research 51, 821–834.
- Cherry K. E., Matson J. L. & Paclawski T. R. (1997) Psychopathology in older adults with severe and profound mental retardation. American Journal on Mental Retardation 101, 445–458.
- Cicchetti D. V. & Sparrow S. A. (1981) Developing criteria for establishing interrater reliability of specific items: applications to assessment of adaptive behavior. American Journal of Mental Deficiency 86, 127–137.
- Cicchetti D. V. & Sparrow S. A. (1990) Assessment in adaptive behavior in young children. In Developmental Assessment in Clinical Child Psychology: A Handbook (eds J. J. Johnson & J. Goldman), pp. 173–196. Pergamon Press, New York.
- Cooper S. A., Smiley E., Morrison J., Williamson A. & Allan L. (2007a) An epidemiological investigation of affective disorders in a population based cohort of 1023 adults with intellectual disabilities. Psychological Medicine 37, 873–882.
- Cooper S. A., Smiley E., Finlayson J., Jackson A., Allan L., Mantry D. & Morrison J. (2007b) The prevalence, incidence and factors predictive of mental ill-health in adults with profound intellectual disabilities. Journal of Applied Research in Intellectual Disabilities 20, 493–501.
- Earl C. J. C. (1961) Subnormal Personalities. Balliere, Tindall and Cox, London.
- Evans K. M., Cotton M. M., Einfeld S. L. & Florio T. (1999) Assessment of depression in adults with severe or profound intellectual disability. Journal of Intellectual and Developmental Disability 24, 147–160.
- Hayes S., McGuire B., O'Neill M., Oliver C. & Morrison C. (2011) Low mood and challenging behaviour in people with severe and profound intellectual disabilities. Journal of Intellectual Disability Research 55, 182–189.
- H2011) Low mood and challenging behaviour in people with severe and profound intellectual disabilities. Journal of Intellectual Disability Research 55, 182–189.
- Instruments screening for depression in adults with intellectual disabilities: systematic review. Research in Developmental Disabilities 31, 1109–1120.

- Hinkle D., Wiersma W. & Jurs S. (2002). *Applied Statistics for the Behavioral Sciences*. 5th edn. Houghton Mifflin, Boston.
- Holden B. & Gitlesen J. P. (2003) Prevalence of psychiatric symptoms in adults with mental retardation and challenging behaviour. *Research in Developmental Disabilities* 24, 323–332.
- Kozlowski A. M., Matson J. L., Sipes M., Hattier M. A. & Bamburg J. W. (2011) The relationship between psychopathology symptoms clusters and the presence of comorbid psychopathology in individuals with severe to profound intellectual disability. *Research in Developmental Disabilities* 32, 1610–1614.
- Lowry M. A. (1998) Assessment and treatment of mood disorders in persons with developmental disabilities. *Journal of Developmental and Physical Disabilities* 10, 387–405.
- Marston G. M., Perry D. W. & Roy A. (1997) Manifestations of depression in people with intellectual disability. *Journal of Intellectual Disability Research* 41, 476–480.
- Marstson J. L., Gardner W. I., Coe D. A. & Sovner R. (1991) A scale for evaluating emotional disorders in severely and profoundly mentally retarded persons. Development of the Diagnostic Assessment for the Severely Handicapped (DASH) scale. *British Journal of Psychiatry* 159, 404–409.
- Matson J. L., Rush K. S., Hamilton M., Anderson S. J., Bamburg J. W. & Baglio C. S. (1999) Characteristics of depression as assessed by the diagnostic assessment for the severely handicapped-II (Dash-II). *Research in Developmental Disabilities* 20, 305–313.
- Meins W. (1995) Symptoms of major depression in mentally retarded adults. *Journal of Intellectual Disability Research* 39, 41–45.
- Moher D., Tetzlaff J. & Altman D. G.; The PRISMA Group (2009) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Medicine* 6: e1000097. doi:10.1371/journal.pmed.1000097.
- M(2009) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Medicine* 6: e1000097. doi:10.1371/journal.pmed.1000097.
- Mowe S. & Luccino R. (1993) Psychiatric morbidity in older people with moderate and severe learning disability (mental retardation): Part 1. Development and reliability of the patient interview (the PASS-ADD). *British Journal of Psychiatry* 163, 471–480.
- Mudford O. C., Barrers F. J., Murray A., Boundy K., Caldwell K. & Goldberg B. (1995) The dexamethasone suppression test and the diagnosis of depression in adults with severe and profound developmental disabilities. *Journal of Intellectual Disability Research* 39, 275–283.
- National Collaborating Centre for Mental Health (2010) *Depression: The NICE Guidelines on the Treatment and Management of Depression in Adults (Updated Edition)*. Paclawski T. R., Matson J. L., Bamburg J. W. & Baglio C. S. (1997) A comparison of the diagnostic assessment for the severely handicapped-II (dash-II) and the Aberrant Behaviour Checklist (ABC). *Research in Developmental Disabilities* 18, 289–298.
- Prasher V. (1999) Presentation and management of depression in people with learning disability. *Advances in Psychiatric Treatment* 5, 447–454.
- Ross E. & Oliver C. (2003a) Preliminary analysis of the psychometric properties of the Mood, Interest & Pleasure Questionnaire (MIPQ) for adults with severe and profound learning disabilities. *British Journal of Clinical Psychology* 42, 81–93.
- Ross E. & Oliver C. (2003b) The assessment of mood in adults who have severe or profound mental retardation. *Clinical Psychology Review* 23, 225–245.
- Royal College of Psychiatrists (2001) *DC-LD (Diagnostic Criteria for Psychiatric Disorders for Use with Adults with Learning Disabilities/Mental Retardation)*. Gaskell Press, London.

Senatore V., Matson J. L. & Kazdin A. E. (1985) An inventory to assess psychopathology in the mentally retarded. *American Journal of Mental Retardation* 89, 459–466.

Smiley E. & Cooper S. A. (2003) Intellectual disabilities, depressive episode, diagnostic criteria and diagnostic criteria for psychiatric disorders for use with adults with learning Disabilities/Mental Retardation (DC-LD). *Journal of Intellectual Disability Research* 47,62–71.

Sturmey P., Laud R. B., Cooper C., Matson J. L. & Fodstad J. C. (2010) Challenging behaviors should not be considered depressive equivalents in individuals with intellectual disabilities. II. A replication study. *Research in Developmental Disabilities* 31, 1002–1007.

Tsiouris J. A. (2001) Diagnosis of depression in people with severe/profound intellectual disability. *Journal of Intellectual Disability Research* 45, 115–120.

Tsiouris J. A., Kim S. Y., Brown W. T. & Cohen L. (2011) Association of aggressive behaviours with psychiatric disorders, age, sex and degree of intellectual disability: A large-scale survey. *Journal of Intellectual Disability Research* 55,636–649.

World Health Organisation (1992) *The ICD-10 Classification of Mental and Behavioural Disorders: clinical Descriptions and Diagnostic Guidelines*. World Health Organisation, Geneva.

Yoo J. H., Valdoninos M. G. & Schroeder S. R. (2012) The epidemiology of psychopathology in people with intellectual disability: a forty year review. *International Review of Research in Developmental Disabilities* 42,31–54.

#### **Appendix 1: Rationale for the exclusion of challenging behaviour**

When any individual is assessed as per ICD-10 guidance for depression, the observation of changes in behaviour is a diagnostic requirement. One must be clear that there is a difference between behavioural changes (how an individual perceives and interacts

#### **Appendix 2: Evaluation of Study Quality**

Table A1 Evaluating methodological quality of studies with their environment plus eating and sleeping) and challenging behaviour. Meins (1995) concluded that new behaviour problems were observed to start equally in both mild-to-moderate intellectual disability and severe intellectual disability when the groups were compared. It was, however, noted that an increase in pre-existing behaviour issues was observed in the severely disabled group. Conversely, Holden & Gitlesen (2003) found no association between the emergence of depression and the onset of challenging behaviour. This was supported by Sturmey et al. (2010) who found little evidence for challenging behaviour being associated with depression. It is apparent that debate exists in this area, and it is beyond the scope of this review to assess for psychiatric morbidity presenting as new challenging behaviour.

| Item                                | Criteria  | Score                 |
|-------------------------------------|---|-----------------------|
| Number of participants              | >100<br>>30-100<br><30  | 2<br>1<br>0           |
| Characteristic of participants      | Group representing target population of the instrument (screening for depression)   | 0/1                   |
| Psychopathology of the participants | >20-50% of participants had depression<br>10-20% or >50-90% had depression<br><10% or >90% of the participants had depression<br>Unclear  | 2<br>1<br>1<br>0      |
| Gold standard                       | Clinical diagnosis by a psychiatrist or psychologist based on standard diagnostic system<br>Clinical diagnosis by a psychiatrist or psychologist<br>Other depression screening instrument used as reference standard<br>All other<br>Not applicable | 2<br>1<br>1<br>0<br>0 |
| Report on measures of validity      | Standard deviation or standard error, or confidence interval is reported  | 0/1                   |

Reproduced from Hermans & Evenhuis (2010).

Measures of internal consistency and correlation coefficients (when available) were interpreted following the guidance from Hermans & Evenhuis (2010a, 2010b).

**Table A2** Guidance to interpret internal consistency and correlation coefficients

|   | Interpretation |
|---|----------------|
| Internal consistency (Cicchetti & Sparrow 1990) |                |
| <0.70   | Unacceptable   |
| 0.70-0.79                                       | Fair           |