Published in final edited form as:

J Appl Res Intellect Disabil. 2007 September ; 20(5): 384–390. doi:10.1111/j.1468-3148.2007.00381.x.

Inter-Rater Reliability of the Diagnoses of Psychosis and Depression in Individuals with Intellectual Disabilities

S. Einfeld^{*}, B. Tonge[†], L. Chapman[‡], C. Mohr[†], J. Taffe[†], and S. Horstead^{*}

^{*}*School of Psychiatry, University of New South Wales, Sydney*

[†]*Centre for Developmental Psychiatry and Psychology, Monash University, Clayton, Vic*

[‡]*Prince of Wales Hospital, Randwick, NSW, Australia*

Abstract

Background—There is a history of over-prescription of antipsychotics to individuals with intellectual disability (ID), while antidepressants may be under-prescribed. However, appropriate treatment is best supported when the diagnosis of psychosis or depression is valid and carries good predictive validity. The present authors report a study examining one aspect of validity, namely whether skilled clinicians can agree on whether an individual with an ID is psychotic or depressed.

Materials and Methods—Pairs of clinicians assessed 52 individuals. Agreement was assessed using Cohen's kappa statistic and agreement proportion.

Results—Overall agreement was high for both psychosis and depression. Whether the individual had mild ID or moderate/severe ID did not have a significant impact on agreement.

Conclusions—Experienced clinicians achieved a high level of agreement as to whether a person with ID was psychotic or depressed similar to that found for those without ID. The findings provide some support for treatment interventions based on diagnosis.

Keywords

depression; intellectual disability; inter-rater reliability; psychosis

Introduction

The presence of psychosis in individuals with intellectual disability (ID) has been the subject of considerable comment (Aman & Singh 1986; Aman 1987; Fischbacher 1987; Clarke *et al.* 1990; Deb & Fraser 1994). Much of this literature asserts that people with ID are often prescribed antipsychotic medications without good evidence of a psychotic illness. A number of studies have reported that the rate of prescription of antipsychotic medication far exceeds the expected prevalence of psychoses in people with ID (Wressell *et al.* 1990; Kiernan *et al.* 1995; Emerson *et al.* 1997; Spreat & Conroy 1998; Robertson *et al.* 2000). In contrast, there is a smaller body of literature suggesting that receipt of antidepressant medication is relatively low (Howland 1992; Robertson *et al.* 2000; Sevin *et al.* 2003).

Appropriate pharmacotherapy for mental disorders in persons with ID is enhanced when diagnoses within this population are both accurate and carry predictive validity. There have been a number of papers describing the difficulties of diagnosis of mental disorders in people

with ID (Borthwick-Duffy & Eyman 1990; Einfeld 1992; Sturmey 1993; Gabriel 1994; Moss *et al.* 1997, 2000; Prosser *et al.* 1998; Patel *et al.* 2001). Atypical presentations, maladaptive behaviours, communication and cognitive limitations, different developmental trajectories and limitations in lifestyle, and reliance on diagnostic criteria developed using intellectually normal individuals can lead to under- and mis-diagnoses (Reiss *et al.* 1982; Menolascino *et al.* 1986; Sovner 1986; Borthwick-Duffy & Eyman 1990; Sturmey 1993; Crews *et al.* 1994; Einfeld & Tonge 1999; Moss *et al.* 2000). The Diagnostic and Statistical Manual of Mental Disorders (DSM) and International Classification of Diseases (ICD) diagnostic systems are used extensively in general mental health practice, but they were developed for the non-ID population. The studies which exist examining psychiatric diagnoses in persons with ID have virtually all used modified diagnostic criteria, but have lacked consideration of inter-rater reliability (Sturmey 1993; King *et al.* 1994).

Inter-rater reliability refers to the consistency with which a particular assessment method comes to the same conclusions when applied by different raters to the same body of information. Reliability is particularly important when there is no 'gold standard' as is the case in making psychiatric diagnoses. Where the result of rater judgement is a classification, two classes of measure of agreement have been used – those based only on the percentage of subjects receiving the same classification from the raters ('percentage agreement') and those incorporating also a 'correction for chance agreement'. This study reports percentage agreement, Cohen's kappa (the most widely used of the chance-corrected measures) and variants of this statistic incorporating adjustments for 'bias' and 'prevalence'. Proportion (percentage) agreement varies between 0 and 1 (0 and 100), and is readily interpretable. Kappa can take negative values, but in practice mostly also lies between 0 and 1. Its interpretation is less straightforward. Byrt *et al.* (1993) showed that kappa may be expressed as a function of percentage agreement, the disparity between 'yeses' and 'noes' among the agreements ('prevalence') and the disparity in the proportions of yeses between the raters ('bias'). Departure from a 50/50 split of yeses and noes among the agreements lowers kappa, while departure from equality of proportions of yeses between raters increases kappa. It is wise therefore to keep prevalence and bias in mind when interpreting kappa.

In considering the inter-rater reliability of psychiatric diagnoses in persons with ID, it is salient to consider the same in persons without communication or cognitive difficulties. Presumably, this provides the upper limit of the inter-rater reliability achievable with persons with ID. Here, the present authors consider the data available relating to the diagnoses of psychosis and depression.

The DSM-IV Field Trial for Schizophrenia and Related Psychotic Disorders (American Psychiatric Association, 1992) examined the reliability and concordance of three alternative sets of options for diagnosing DSM-IV psychotic disorders plus the criteria from DSM-III, DSM-III-R, and ICD-10. In the international field trials for ICD-10, inter-rater agreement based on diagnoses at the 'two character' group level, kappa values of 0.82 for schizophrenic disorders and 0.66 for depressive episode were reported (Sartorius *et al.* 1993, 1995). However subsequent studies using less structured formats have reported lower kappa-values. Way *et al.* (1998) videotaped 30 emergency department psychiatric assessments and these were then re-rated by eight different psychiatrists. For psychosis a kappa-value of 0.64 was reported and for depression a kappa of 0.48.

In a review examining the accuracy of the clinical examination for diagnosing clinical depression Williams *et al.* (2002) found seven studies using the Structured Clinical Interview for DSM Diagnoses (Spitzer *et al.* 1979) in which inter-rater reliability for major depression was evaluated. Study design ranged from multiple clinicians viewing a videotaped interview, to paired clinicians conducting sequential interviews, with training varying from psychology

trainees to experts in the mood disorders field. Kappa values ranged from 0.64 to 0.93. They found a further seven studies evaluating inter-rater reliability of DSM-IV diagnoses using non-standardized interviews and thus more closely simulating clinical practice as in the current study. With the exception of one study using a videotaped interview, the study designs involved paired, generally blinded, interviewers conducting joint or sequential interviews. Here, the kappa-values ranged from 0.55 to 0.74.

In a smaller study, Miller (2001) compared the inter-rater reliability of structured versus unstructured interviews and found that the traditional, unstructured diagnostic assessment gave kappa-values from 0.24 to 0.43, whereas the inter-rater agreement using the SCIDCV, Computer Assisted Diagnostic Interview, gave a kappa of 0.75. A number of other studies (Keller *et al.* 1995; Roy *et al.* 1997; Shear *et al.* 2000; Simpson *et al.* 2002) support the finding that in individuals without ID, the inter-rater agreement for diagnosis of psychosis and depression gave kappas in the range 0.6 to 0.8, with structured assessments having higher values than unstructured.

As previously noted there has been little research on the inter-rater reliability of either the DSM or ICD diagnostic systems when used to assess persons with an ID and there continues to be a relative lack of suitable alternative techniques for detection and diagnosis of psychiatric morbidity in this population. Screening instruments, which tend to lack the depth required for accurate diagnosis include the Psychopathology Instrument for Mentally Retarded Adults (PIMRA) (Matson *et al.* 1984), the Reiss screen (see Sturmey *et al.* 1995), and the Diagnostic Assessment for the Severely Mental Handicapped (DASH) scale (Matson *et al.* 1991). Moss & Goldberg (Moss *et al.* 1993) developed the Psychiatric Assessment Schedule for Adults with a Developmental Disability (PAS-ADD) as a semi-structured interview for use in persons with an ID and is based on diagnostic criteria utilised in ICD and DSM. From the PAS-ADD item set, 13 possible symptom syndromes, such as 'depressed mood' or 'situational anxiety' can be generated by the CATEGO computer algorithm. While a detailed report of the psychometric properties of these measures is outside the scope of this report, none of these has documented sufficient validity to warrant its use as a structured assessment protocol for the diagnosis of psychosis or depression in people with ID.

We sought to add to this information by measuring the extent to which experienced clinicians agreed about a diagnosis of depression or psychosis in individuals with ID.

Materials and Methods

Recruitment of subjects

Subjects for this study were recruited in two ways. They were chosen as the most expeditious method for recruiting sufficient numbers. The first approach was to invite psychiatrists with particular experience and expertise in the assessment of individuals with ID to refer subjects whom they thought may have depression of any type or psychosis of any type. Secondly, the present authors asked general practitioners who cared for the general health of persons with ID, to refer to the study subjects whom they thought to be suffering from either psychosis or depression. That is, the first group was seen by experts in mental health and ID, and the second group by non-experts. For the first group, the subjects referred by psychiatrists with experience in ID (Rater 1) were seen by a psychiatrist member of the research team (Rater 2) within 3 weeks of the referring psychiatrists' assessment, and the assessments of the referring psychiatrist and the research psychiatrist were examined for their inter-rater agreement. The subjects referred by the general practitioners were assessed by a member of the research team (Rater 1) and that assessment simultaneously observed by another member of the team (Rater 2). The team comprised experienced psychologists or psychiatrists specializing in psychopathology in persons with ID, and the assessments of these two research team members

were examined for their inter-rater agreement. Thus, although the two groups were recruited differently inter-rater agreement was compared between experts for all cases. Thus, although each rater was blind to the diagnosis reached by the other rater, all raters were aware that subjects were suspected of having possible depression or psychosis.

Method of assessment

As described earlier, given the lack of validated assessment tools, the present authors used a traditional psychiatric interview approach for the inter-rater assessments. That is, they interviewed and obtained a history from persons attending with the subject and where possible from the subject themselves. In addition, they also examined the mental state of the subject. The referring psychiatrists and the research assessors also completed a DSM-IV criteria checklist for schizophrenia and major depression as a semi-structured aid in assessment. For the purpose of this study, psychosis was defined as the presence of delusions or hallucinations in the past 2 years. Depression was defined as any DSM-IV depressive disorder in the last 12 months. The raters scored the diagnoses as present, maybe present, or absent. Three of the authors (SE, BT, CM) each completed 14 interviews and another 13 specialist clinicians completed three interviews on average.

Measures and analysis

There were three rating classifications – ‘yes’, ‘no’ and ‘maybe’. The ‘maybe’ ratings represent indecision between ‘yes’ and ‘no’ rather than a third category about which agreement was being investigated. It is reasonable to regard them as temporary results which would have been determined as ‘yes’ or ‘no’ after further information or observation. The present authors present parallel analyses. The first is based only on the yes and no ratings, ignoring any person about whom there was at least one maybe rating. The second is based on ‘resolving’ the maybes according to the following rule: yes and maybe resolves to yes and yes, no and maybe to no and no, and maybe and maybe to yes and no or no and yes, equally.

The level of agreement between raters is measured by simple percentage or proportion of agreement and by Cohen’s kappa. Bias-adjusted and prevalence-and-bias- adjusted versions of kappa are also presented.

Results

The gender and level of ID of the subjects is shown in Table 1. The mean age of the participants was 31 years (age range = 11–63 years).

Depression

The results of ratings for depression are shown in Table 2.

Table 3 shows parallel results for all subjects, for those subjects with mild ID and for those with moderate or severe ID, first ignoring any person receiving at least one maybe classification and then with the maybes resolved according to the rule given above. The proportion of agreements ranges from 0.81 to 0.89. Bias-adjusted kappa (Byrt *et al.* 1993) is identical to kappa to second-decimal-place accuracy, and kappa differs for prevalence-adjusted-bias-adjusted kappa by at most 2 units in the second decimal place.

The lack of difference between bias-adjusted and raw kappa values for depression ratings shows that the raters do not disagree about the prevalence of depression among the young people with ID. The prevalence effects on kappas for agreement about depression, measured by the differences between bias-adjusted and prevalence- adjusted-bias-adjusted kappa values, are very small. The net result is that for measuring inter-rater agreement for depression nothing

is gained by using kappa, which is difficult to interpret, in preference to proportions (or percentages) of agreement, which are transparent.

Considering only those decisions about which both raters are definite, raters are in agreement in 86% of cases, and this changes marginally to 87% when 'maybe' ratings are resolved in a plausible way.

The agreement percentages within the mild ID and moderate-to-severe ID groups differ significantly neither from the overall figures nor from each other.

Chance correction has zero (to two decimal place accuracy) influence on kappas for depression when maybes are resolved.

Psychosis

Tables 4 and 5 show the results of the corresponding analysis of ratings of psychosis.

Considering only those decisions about which both raters are definite, raters are in agreement in 85% of cases, and this changes marginally to 83% when 'maybe' ratings are resolved.

The agreement statistics within the mild ID and moderate- to-severe ID groups differ significantly neither from the overall figures nor from each other.

There is no observable bias between raters, as was observed for agreements about depression. Kappas are slightly lower than their prevalence-adjusted-bias-adjusted equivalents – the lower prevalence of psychosis (compared to depression) among young people with ID has resulted in a small deflationary chance correction.

Agreement figures are slightly, but not significantly, higher for mild ID than for moderate-to-severe ID.

Discussion

This study is one of the few to attempt to systematically assess whether experienced specialist clinicians can agree in making psychiatric diagnoses for individuals with ID. Overall, the present authors found that percentage agreement was high and similar to what has generally been found in studies of persons without ID, particularly, where structured assessment methods were used. Presumably, the difficulties in eliciting psychopathological phenomena described previously were not such as to lead to greater disagreement, at least when assessments were conducted by specialists in mental health and ID. Agreement in this study may have been even higher without heterogeneity of assessors, but this heterogeneity is in keeping with practice in service settings.

Where raters registered uncertainty by giving a 'maybe' response the present authors sought to resolve the uncertainties in a plausible way, to obtain a reasonable estimate of what the state of agreement between the raters might have been had they resolved their own uncertainties. For a person about whom the ratings were one definite yes or no and one maybe they saw the likely result as agreement, as the weight of evidence about the young person was towards the definite judgement. Counterbalancing this resolution, where both raters were uncertain they resolved the maybes as disagreements in both directions, there being no weight of evidence in either. This 'minimalist' procedure implies no change in bias and little change in prevalence (the effects of both of which were in any case very small) from the definite-judgements-only situation. The resultant changes to agreement statistics, both percentage agreement and kappa, were very marginal: an increase for depression and a decrease for psychosis. Our comparisons of unadjusted and adjusted versions of kappa show that bias (disparity between the two raters'

proportions of yeses) is virtually absent in our raters' judgements. Further, the prevalence effect (due to disparity between the overall proportions of yeses and noes) is not in evidence except in judgements about psychosis in those with mild ID. With negligible bias and prevalence effects kappa reduces to twice the agreement proportion less 1, as Byrt *et al.* (1993) have shown. In these circumstances the transparent concept of percentage agreement is to be preferred over kappa, which contains no more information and does not lend itself to any clear interpretation.

The study has made us aware of the difficulty in conducting inter-rater reliability studies of psychiatric diagnoses. There are many challenges to be overcome in arranging for sufficient numbers of subjects with sufficiently high rates of suspected disorders to be interviewed by independent raters of sufficient expertise in comparable circumstances. It will take a considerable effort by researchers to amass the data required to canvass the diagnostic reliability of other psychiatric disorders.

Preliminary work towards the development of a diagnostic system specific for ID, such as the "Diagnostic criteria for psychiatric disorders for use with adults with learning disabilities" (DC-LD) (Royal College of Psychiatrists, 2001) creates the conditions necessary for the development of structured diagnostic instruments. These instruments will need to rely on observed behaviours and carer reports augmented if possible by self reflection by the person with ID. Until then, this study provides some confidence in the application, by specialist clinicians, of existing diagnostic criteria for depression and psychosis to adults with ID.

The finding that overall reliability was substantial is a necessary, though insufficient condition for concluding that the diagnoses of depression and psychosis are valid in persons with ID. Ultimately, the clinical importance of such validity rests largely in its capacity to predict treatment outcome. Therefore, one might argue that since the advent of antidepressants and antipsychotics with fewer side-effects it is justified to undertake a therapeutic trial on the basis of a probable diagnosis. However, newer psychopharmacological do have troublesome, if less toxic side-effects. A systematic approach to usage of psychotherapeutic medications in the face of uncertain diagnostic reliability is described in detail in Einfeld (2001). Cautious prescribing for psychiatric diagnoses will continue to be needed for individuals with ID for the foreseeable future, and will thus be supported by diagnoses as valid as possible within the constraints of reliability limits.

Acknowledgements

This study was supported by NHMRC Australia Grant 113844. Doctors who contributed psychiatric assessments were Bruce Chenoweth, Michael Fairley, Terry Heins, Peter Wurth, Jim Friend, Sophie Kavanagh, Julian Davis, David Moseley, Robert Davies and Jennifer Torr.

References

- Aman MG. Overview of pharmacotherapy: current status and future directions. *Journal of Mental Deficiency Research* 1987;31:121–130. [PubMed: 2887660]
- Aman MG, Singh NN. A critical appraisal of recent drug research in mental retardation: the Coldwater studies. *Journal of Mental Deficiency Research* 1986;30:203–216. [PubMed: 2878082]
- American Psychiatric Association. Report from the DSM-IV Field Trial for Schizophrenia and Related Psychotic Disorders. American Psychiatric Association; University of Iowa, Iowa City, IA: 1992.
- Borthwick-Duffy SA, Eyman RK. Who are the dually diagnosed? *American Journal of Mental Retardation* 1990;94:586–595. [PubMed: 2340136]
- Byrt T, Bishop J, Carlin JB. Bias, prevalence and kappa. *Journal of Clinical Epidemiology* 1993;46:423–429. [PubMed: 8501467]

- Clarke DJ, Kelley S, Thinn K, Corbett JA. Psychotropic drugs and mental retardation: 1. Disabilities and the prescription of drugs for behaviour and for epilepsy in three residential settings. *Journal of Mental Deficiency Research* 1990;34:385–395. [PubMed: 2266549]
- Crews WD Jr, Bonaventura S, Rowe F. Dual diagnosis: prevalence of psychiatric disorders in a large state residential facility for individuals with mental retardation. *American Journal of Mental Retardation* 1994;98:724–731. [PubMed: 8054201]
- Deb S, Fraser W. The use of psychotropic medication in people with learning disability: towards rational prescribing. *Human Psychopharmacology* 1994;9:259–272.
- Einfeld SL. Clinical assessment of psychiatric symptoms in mentally retarded individuals. *Australian and New Zealand Journal of Psychiatry* 1992;26:48–63. [PubMed: 1580886]
- Einfeld S. Systematic management approach to pharmacotherapy for people with learning disabilities. *Advances in Psychiatric Treatment* 2001;7:43–49.
- Einfeld SL, Tonge BJ. Observations on the use of the ICD-10 guide for mental retardation. *Journal of Intellectual Disability Research* 1999;43:408–412. [PubMed: 10546965]
- Emerson, E.; Alborz, A.; Kiernan, C.; Mason, H.; Reeves, D.; Swarbrick, R.; Mason, L. Report 5: The Treatment and Management of Challenging Behaviour. The University of Manchester; Manchester: 1997. The HARC Challenging Behaviour Project.
- Fischbacher E. Prescribing in a hospital for the mentally retarded. *Journal of Mental Deficiency Research* 1987;31:17–29. [PubMed: 2884324]
- Gabriel SR. The developmentally disabled, psychiatrically impaired client. Proper treatment of dual diagnosis. *Journal of Psychosocial Nursing & Mental Health Services* 1994;32:35–39. [PubMed: 7799264]
- Howland RH. Fluoxetine treatment of depression in mentally retarded adults. *Journal of Nervous & Mental Disease* 1992;180:202–205. [PubMed: 1588340]
- Keller MB, Klein DN, Hirschfeld RM, Kocsis JH, McCullough JP, Miller I, First MB, Holzer CP 3rd, Keitner GI, Marin DM. Results of the DSM-IV mood disorders field trial. *American Journal of Psychiatry* 1995;152:843–849. [PubMed: 7755112]
- Kiernan C, Reeves D, Alborz A. The use of antipsychotic drugs with adults with learning disabilities and challenging behaviour. *Journal of Intellectual Disability Research* 1995;39:263–274. [PubMed: 7579984]
- King BH, DeAntonio C, McCracken JT, Forness SR, Ackerland V. Psychiatric consultation in severe and profound mental retardation. *American Journal of Psychiatry* 1994;151:1802–1808. [PubMed: 7977889]
- Matson JL, Kazdin AE, Senatore V. Diagnosis and drug use in mentally retarded, emotionally disturbed adults. *Applied Research in Mental Retardation* 1984;5:513–519. [PubMed: 6151823]
- Matson JL, Gardner WI, Coe DA, Sovner R. 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* 1991;159:404–409. [PubMed: 1958951]
- Menolascino FJ, Gilson SF, Levitas AS. Issues in the treatment of mentally retarded patients in the community mental health system. *Community Mental Health Journal* 1986;22:314–327. [PubMed: 3829606]
- Miller PR. Inpatient diagnostic assessments: 2. Inter-rater reliability and outcomes of structured vs. unstructured interviews. *Psychiatry Research* 2001;105:265–271. [PubMed: 11814545]
- Moss S, Patel P, Prosser H, Goldberg D, Simpson N, Rowe S, Lucchino R. Psychiatric morbidity in older people with moderate and severe learning disability. I: development and reliability of the patient interview (PAS-ADD). *British Journal of Psychiatry* 1993;163:471–480. [PubMed: 8252286]
- Moss S, Ibbotson B, Prosser H, Goldberg D, Patel P, Simpson N. Validity of the PAS-ADD for detecting psychiatric symptoms in adults with learning disability (mental retardation). *Social Psychiatry & Psychiatric Epidemiology* 1997;32:344–354. [PubMed: 9299929]
- Moss S, Emerson E, Kiernan C, Turner S, Hatton C, Alborz A. Psychiatric symptoms in adults with learning disability and challenging behaviour. *The British Journal of Psychiatry* 2000;177:452–453. [PubMed: 11060000]

- Patel NC, Crismon ML, Rush AJ, Frances A. Practitioner versus medication-expert opinion on psychiatric pharmacotherapy of mentally retarded patients with mental disorders. *American Journal of Health-System Pharmacy* 2001;58:1824–1829. [PubMed: 11596698]
- Prosser H, Moss S, Costello H, Simpson N, Patel P, Rowe S. Reliability and validity of the Mini PAS-ADD for assessing psychiatric disorders in adults with intellectual disability. *Journal of Intellectual Disability Research* 1998;42:264–272. [PubMed: 9786440]
- Reiss S, Levitan GW, Szyszko J. Emotional disturbance and mental retardation: diagnostic overshadowing. *American Journal of Mental Deficiency* 1982;86:567–574. [PubMed: 7102729]
- Robertson J, Emerson E, Gregory N, Hatton C, Kessissoglou S, Hallam A. Receipt of psychotropic medication by people with intellectual disability in residential settings. *Journal of Intellectual Disability Research* 2000;44:666–676. [PubMed: 11115021]
- Roy MA, Lanctot G, Merette C, Cliche D, Fournier JP, Boutin P, Rodrigue C, Charron L, Turgeon M, Hamel M, Montgrain N, Nicole L, Pirés A, Wallot H, Ponton A-M, Garneau Y, Dion C, Lavalée J-C, Potvin A, Szatmari P, Maziade M. Clinical and methodological factors related to reliability of the best-estimate diagnostic procedure. *American Journal of Psychiatry* 1997;154:1726–1733. [PubMed: 9396953]
- Royal College of Psychiatrists. Diagnostic Criteria for Psychiatric Disorders for Use with Adults with Learning Disabilities/Mental Retardation OP 48. Gaskell; London: 2001.
- Sartorius N, Kaelber CT, Cooper JE, Roper MT, Rae DS, Gulbinat W, Ustun TB, Regier DA. Progress toward achieving a common language in psychiatry. Results from the field trial of the clinical guidelines accompanying the WHO classification of mental and behavioral disorders in ICD-10. *Archives of General Psychiatry* 1993;50:115–124. [PubMed: 8427551]
- Sartorius N, Ustun TB, Korten A, Cooper JE, van Drimmelen J. Progress toward achieving a common language in psychiatry, II: Results from the international field trials of the ICD-10 diagnostic criteria for research for mental and behavioral disorders. *American Journal of Psychiatry* 1995;152:1427–1437. [PubMed: 7573580]
- Sevin JA, Bowers-Stephens C, Crafton CG. Psychiatric disorders in adolescents with developmental disabilities: longitudinal data on diagnostic disagreement in 150 clients. *Child Psychiatry & Human Development* 2003;34:147–163. [PubMed: 14617947]
- Shear MK, Greeno C, Kang J, Ludewig D, Frank E, Swartz HA, Hanekamp M. Diagnosis of nonpsychotic patients in community clinics. *American Journal of Psychiatry* 2000;157:581–587. [PubMed: 10739417]
- Simpson SG, McMahon FJ, McInnis MG, MacKinnon DF, Edwin D, Folstein SE, DePaulo JR. Diagnostic reliability of bipolar II disorder. *Archives of General Psychiatry* 2002;59:736–740. [PubMed: 12150650]
- Sovner R. Limiting factors in the use of DSM-III criteria with mentally ill/mentally retarded persons. *Psychopharmacology Bulletin* 1986;22:1055–1059. [PubMed: 3809370]
- Spitzer RL, Forman JB, Nee J. DSM-III field trials: I. Initial interrater diagnostic reliability. *American Journal of Psychiatry* 1979;136:815–817. [PubMed: 443467]
- Spreat S, Conroy J. Use of psychotropic medications for persons with mental retardation who live in Oklahoma nursing homes. *Psychiatric Services* 1998;49:510–512. [PubMed: 9550242]
- Sturmey P. The use of DSM and ICD diagnostic criteria in people with mental retardation. A review of empirical studies. *Journal of Nervous & Mental Disease* 1993;181:38–41. [PubMed: 8419513]
- Sturmey P, Burcham KJ, Perkins TS. The Reiss Screen for Maladaptive Behaviour: its reliability and internal consistencies. *Journal of Intellectual Disability Research* 1995;39:191–195. [PubMed: 7640488]
- Way BB, Allen MH, Mumpower JL, Stewart TR, Banks SM. Interrater agreement among psychiatrist in psychiatric emergency assessments. *American Journal of Psychiatry* 1998;155:1423–1428. [PubMed: 9766775]
- Williams JW Jr, Noel PH, Cordes JA, Ramirez G, Pignone M. Is this patient clinically depressed? *JAMA* 2002;287:1160–1170. [PubMed: 11879114]
- Wressell SE, Tyrer SP, Berney TP. Reduction in antipsychotic drug dosage in mentally handicapped patients: a hospital study. *British Journal of Psychiatry* 1990;157:101–106. [PubMed: 1975757]

Table 1

Intellectual disability (ID) level and gender of participants

ID level	Mild	Moderate	Severe	Total
Male	14	8	5	27
Female	12	7	6	25
Total	26	15	11	52

Table 2
Ratings for depression

	Rater 2			
	Yes	No	Maybe	Total
Rater 1				
Yes	11	3	5	19
No	2	19	5	26
Maybe	5	0	2	7
Total	18	22	12	52

Table 3

Agreement statistics for depression

Proportion of agreements	κ	Bias adjusted κ	Prevalence adjusted bias adjusted κ	n
Yes/no only				
All	0.86	0.70	0.71	35
Mild ID	0.81	0.61	0.63	16
Moderate or severe ID	0.89	0.77	0.79	19
Maybes resolved				
All	0.87	0.73	0.73	52
Mild ID	0.85	0.69	0.69	26
Moderate or severe ID	0.88	0.77	0.77	26

ID, intellectual disability.

Table 4
Ratings for psychosis

	Rater 2			Total
	Yes	No	Maybe	
Rater 1				
Yes	11	1	1	13
No	5	24	3	32
Maybe	3	1	3	7
Total	19	26	7	52

Table 5

Agreement statistics for psychosis

	Proportion of agreements	κ	Bias adjusted κ	Prevalence adjusted bias adjusted κ	n
Yes/no only					
All	0.85	0.68	0.67	0.71	41
Mild ID	0.95	0.89	0.89	0.89	19
Moderate or severe ID	0.77	0.47	0.45	0.55	22
Maybes resolved					
All	0.83	0.63	0.63	0.65	52
Mild ID	0.88	0.76	0.76	0.77	26
Moderate or severe ID	0.77	0.50	0.49	0.54	26

ID, intellectual disability.