informa
healthcare

Validity of the Swedish SCID and ADDIS diagnostic interviews for substance use disorders: Sensitivity and specificity compared with a LEAD golden standard

ARNE GERDNER, JENNY KESTENBERG, MATTIAS EDVINSSON

Gerdner A, Kestenberg J, Edvinsson M. Validity of the Swedish SCID and ADDIS diagnostic interviews for substance use disorders: Sensitivity and specificity compared with a LEAD golden standard. Nord J Psychiatry 2014;64:1–9.

Objective: The study explores agreement on diagnoses and diagnostic criteria for substance use disorders between two structured assessment interviews, the Structured Clinical Interview for the DSM-IV (SCID) and the Alkohol/Drog Diagnos InStrument (ADDIS). Both interviews are compared with a golden standard (GS), based on a LEAD model (Longitudinal, Expert, All Data). *Method:* Patients were interviewed concerning substance use problems by trained interviewers using SCID and ADDIS separately and blind to each other's results. SCID and ADDIS interviews were compared with each other, and both were compared with a GS. *Results:* Satisfactory agreement exists between SCID and ADDIS on criteria as well as final diagnostic suggestions, although ADDIS tended to propose dependence diagnoses somewhat more often than SCID. Agreement between SCID and GS is moderate. Sensitivity of SCID is satisfactory, as is specificity for lifetime diagnoses, while specificity for current diagnoses is perfect. ADDIS demonstrates substantial to perfect agreement with GS on dependence diagnoses and moderate agreement on abuse diagnoses (both lifetime and current), as well as showing excellent to perfect overall sensitivity and specificity. Both instruments are in almost perfect agreement with the GS on severity ratings. *Conclusion:* Both ADDIS and SCID can be used to ensure good standards in the diagnostic assessment of substance use disorders (both alcohol and drugs), with and without psychiatric comorbidity. *Significant outcomes.* Both SCID and ADDIS are in good agreement with the GS based on a LEAD model concerning substance use disorders.

• *ADDIS, SCID, Sensitivity, Specificity, Substance use disorders, Validity.*

Arne Gerdner, Jönköping University, School of Health Sciences, PO Box 1026, SE-55111 Jönköping, Sweden, E-mail: arne.gerdner@hhj.hj.se; Accepted 15 May 2014.

Structured assessment concerning addictions is often lacking in many countries' treatment systems. This situation is contrary to the recommendations of the American Society of Addiction Medicine (1), which stipulates that diagnostic assessment be performed before an individual is admitted to any type of addiction treatment. It is argued that lack of diagnostic practices is both unethical and uneconomical, since treatment planning and goal setting in accordance with evidence-based practice should differentiate those with substance dependence from those with other types of misuse problems (2). Treatment may also benefit from such practice since it has positive cognitive and motivational effects on the patient (3).

The two structured diagnostic instruments most often used in Sweden to assess substance dependence are the Structured Clinical Interview for the DSM-IV (SCID) (4) and the Alkohol/Drog Diagnos InStrument (ADDIS) (5). Both are translated into Swedish from the American originals and demand special training. Prerequisites for SCID training are psychiatric medical specialization or clinical psychology certification. Those in training for ADDIS need to have an academic degree and education in addiction.

SCID was first constructed for DSM-III-R assessment, and later for DSM-IV, with modules for the most common psychiatric diagnostic groups, including psychotic, affective, anxiety and personality disorders (4). One

DOI: 10.3109/08039488.2014.926987

module assesses substance use disorders. SCID is used within psychiatric research and for treatment planning. SCID has been used as a "golden standard" (abbreviated GS) for psychiatric diagnoses in the validation of other psychiatric assessment instruments (6). Studies focusing on specific diagnoses show good inter-rater reliability (7–12), also concerning substance use disorders (13). SCID showed good concurrent, discriminant and predictive validity for substance use disorders with various drug types assessed by a research technician (14). Psychiatric diagnoses derived from SCID interviews by research technicians had better validity than diagnoses by master level clinicians (15). Both versions of SCID (DSM-III-R and DSM-IV) were translated into Swedish by Jörgen Herlofsson. To our knowledge, there is still no validation study on the Swedish version of SCID, and there is no international study on the sensitivity and specificity of SCID's module on substance use disorders.

ADDIS is constructed to diagnose substance use disorders according to both ICD-10 and DSM-IV. It is a translation and Swedish adaption of the American instrument SUDDS, which was developed by Hoffmann and Harrison as an improvement from NIMH-DIS (The National Institute of Mental Health Diagnostic Interview Schedule—Version II), a structured interview for assessment of psychiatric disorders with high validity and reliability (16, 17). Lynn Wickström translated and adapted the SUDDS to the Swedish culture in 1987, and named it ADDIS. Wickström is also responsible for the revisions making ADDIS compatible with ICD-10 and DSM-IV. (The new version, DSM-5, was not yet published when this study began.)

SUDDS has good agreement with the diagnostic assessment of experienced clinicians (Cohen's kappa = 0.71–0.87) and test–retests show high correlation ($R = 0.81$–0.90) (18). Diagnostic proposals based on SUDDS were in excellent agreement with assessment by clinicians (19). Reliability is similar in various ethnic groups, with internal consistency (Cronbach's alpha) for dependence varying between 0.93 and 0.97 and for abuse between 0.84 and 0.90 (20).

The Swedish ADDIS (21) shows good construct validity concerning alcohol in two populations: a clinical population and a DWI (driving while intoxicated) population. It is homogeneous (all factor loadings > 0.40) with acceptable explained variance (R^2 for dependence = 0.46; abuse = 0.40). Separate analyses for the two populations and for women provided similar results. In discriminant analysis, ADDIS could correctly classify 94% of the two samples. Cronbach's alpha is satisfactory or excellent in all analysis.

The ultimate validity test of diagnostic tools, however, should be to compare its use to a GS, i.e. some method that is accepted as the most certain assessment available. By comparing agreement between the instruments and GS, one may estimate sensitivity and specificity. While sensitivity focuses on the instruments' ability to "capture cases", specificity focuses on the instruments ability to exclude "non-cases". There is, however, no commonly accepted GS. As mentioned, SCID has sometimes been used as a GS, possibly since the motive for constructing SCID was to improve diagnostic quality compared with unstructured diagnostic practices (4). ADDIS (SUDDS), built on a previous structured instrument, was also constructed to improve diagnostic quality. Spitzer suggested, in 1983, that a LEAD standard (LEAD = Longitudinal, Expert, All Data) could be constructed based on expert evaluation of all available data from hospital journals (4). In the absence of an established GS, such a LEAD standard may function as a GS for studying sensitivity and specificity.

With that inspiration, this study assumes that both SCID and ADDIS have strengths that could be combined in forming a GS, together with previously collected patient documentation. Such documentation is consulted to solve disagreements between the two assessments. The strength of ADDIS is that all diagnostic criteria are captured by many different specific questions, while the strength of SCID is that it is constructed to categorize psychiatric problems in an overall psychiatric diagnostic approach. Our assumption was that ADDIS should have advantages in sensitivity, while SCID may have advantages in specificity with better possibilities to exclude symptoms of other problems than substance use disorders. The aim here is therefore to study validity of SCID and ADDIS, with regard to DSM-IV diagnoses of substance use disorders, by comparing agreement between the two, as well as the agreement of each instrument with GS, which is created through a re-evaluation of all data, i.e. both interviews combined with previous collected patient documentation. The new version, DSM-5, was not yet published when the study began.

Methods

The study was conducted in two types of treatment facilities: 1) Rockesholm Treatment Centre—a residential institution delivering psychosocial treatment (12-step facilitation, based on the philosophy of Alcoholics Anonymous and Narcotics Anonymous) admitting persons with alcohol as well as drug use problems; and 2) three psychiatric outpatient clinics in Västra Götaland Region. In these facilities, diagnostic interviews were conducted by trained interviewers who were members of the staff. Interviewers were trained by Jörgen Herlofsson (SCID) and Lynn Wickström (ADDIS); thus training was conducted by the two persons responsible for translating and introducing the instruments in Sweden.

Instruments for data collection

Concerning SCID, the E-module (substance use disorders) was used in total, i.e. all SUD questions were asked, and, when applicable after initial screening questions, the modules on anxiety and affective disorders were administered. In about half of the interviews, these two modules were replaced with the Beck Depression Inventory (BDI) and Beck Anxiety Inventory (BAI), se below. The entire ADDIS interview was used. Time spent was not recorded but is estimated to have been about the same for both interviews, i.e. about 45–60 min per interview when only alcohol is used and 15–20 min more when drugs are used. The two interviews were carried out on different days, with the order of interviews shifting. Interviews were carried out with different interviewers, blind to each other's results. The median time between interviews was 3.5 days. Both SCID and ADDIS interviewers documented the interviews and concluded the assessment with a proposed diagnosis—dependence, abuse or no diagnosis—according to respective instrument procedures. After each interview, the interviewer rated the patient's score on Global Assessment of Functioning (GAF) (22), and made note of possible psychiatric problems of depression and anxiety. The data available for these notes included DSM-IV diagnoses from SCID modules and/or screening tests of depression and anxiety; BAI (23) and BDI (24) were used with the cut-offs for moderate problems (BAI: 16 points/BDI: 17 points). Replacement of the SCID modules on depression and anxiety with BAI and BDI was done to save time. The change in study plan was accepted since the study did not aim to validate those other SCID modules. All interview documentation was filed in the patient's journal.

Preparing study protocols for analyses

After both interviews were completed, the interview documentation was transformed into a study protocol, with a code number replacing the patient's name. Alcohol and other drugs, if used, generated separate protocols, i.e. a person could generate more than one protocol. The protocol included gender, age, dates of interviews, date of the LEAD re-evaluation and who was responsible for each of these. Following this were three columns, one column for ADDIS, one for SCID and one for the re-evaluation. Each column included fulfilment of each of the 11 DSM-IV criteria (seven dependence and four abuse criteria) both lifetime and current (last 12 months). At the end of each column, GAF scores and outcomes of psychiatric screenings or psychiatric diagnoses, when applicable, were recorded. (Study protocol is available from the corresponding author on request.) The Research Ethics Committee of Mid Sweden University reviewed the study plan (28 November 2008) and raised no objections from an ethical point of view.

In total, 42 individuals were interviewed, and from these interviews 55 protocols were completed and form the basic data for this study. Thirty-five of these protocols (28 alcohol, seven drugs) came from Rockesholm, while 20 protocols (14 alcohol, six drugs) came from the psychiatric outpatient clinics. In four ADDIS and 16 SCID protocols there are altogether seven missing ADDIS items and 49 missing SCID items (i.e. 56 of the 2420 items: 11 criteria $\times$ two interviews $\times$ 2 time-spans $\times$ 55 protocols), i.e. 2.3% missing data.

For 34 individual cases (45 protocols), a LEAD re-evaluation was done. At the Rockesholm Treatment Centre this was carried out by one of the two consultant medical doctors, while at the psychiatric outpatient clinic it was carried out as a unanimous decision by the two psychiatrists (co-authors JK and ME) who were also responsible for the previous interviews. This re-evaluation is regarded as the best available assessment and therefore treated as GS. For eight individuals (10 protocols) at Rockesholm, it was not possible to get the doctors re-evaluation.

The mean age of the interviewees was 36.5 years (standard deviation = 13.5 years) and 65% were men. Forty-two of the protocols concerned alcohol, while 13 concerned other drugs (amphetamine six protocols, benzodiazepines three, cannabis two, solvents one and mixed drugs one). GAF was scored in both interviews and highly correlated (Pearson $R = 0.52$, $P > 0.001$) although moderately in agreement ($\gamma = 0.42$, $P = 0.002$). When ratings differed, these were re-evaluated in similar procedures as the substance use diagnoses, and when re-evaluation was lacking, the mean of available GAF scores were used. These "combined" scores varied between a minimum of 41 points (i.e. having serious psychiatric or behavioural symptoms or dysfunctional problems concerning social contacts, work or school) and a maximum of 67 (i.e. some minor symptoms or social problems but mostly functioning and with established relations to significant others), with a mean of 55 points (i.e. moderate symptoms or functional problems in relations, work or school). There was a relatively high prevalence of anxiety and depression: (lifetime: 57% and 67%, respectively; current: 55% both).

Statistical analyses

The following statistical methods were applied: absolute agreement is the percentage having identical outcome on the particular criterion/diagnosis. McNemar (or McNemar–Bowker in case of more than two categories) is a test of symmetry or a tendency in disagreements and tested for significance with the null hypotheses of no symmetry. Cohen's kappa (K) estimates agreement when corrected for random agreement, varying from 1 in case of perfect agreement, to 0 which is not more than random

agreement. Negative values exist when agreement is less than random. Κ can be interpreted as follows: <0.00 = "poor", 0.00–020 = "slight", 0.21–0.40 = "fair", 0.41–0.60 = "moderate", while 0.61–0.80 is "substantial" and >0.81 is "almost perfect" (25). Agreement on diagnoses are analysed both as dichotomies (dependent/not dependent) and as trichotomies (no diagnosis/abuse/dependence). Since the latter are ordinal scales, systematic correlation is tested by gamma (γ), here interpreted in the same way as Κ. Both Κ and γ are tested for significance, with the null hypotheses of no agreement or systematic correlation more than random, respectively.

Sensitivity is the percentage of "true" cases (i.e. based on GS) captured by the instrument, while specificity is the percentage of "true" non-cases, excluded by the instrument. "Area under the curve" (AUC) is a measure that can be used when the predictor is a continuous variable, which is not the case here. In comparisons of dichotomies, however, the mean of sensitivity and specificity corresponds to AUC (26). AUC is sometimes used as an alternative agreement measure, since Κ in skewed populations may be very restrictive, despite a high total agreement (27). Positive predictive value (PPV) estimates the likelihood that the person predicted as having the diagnosis actually has it, while negative predictive value (NPV) estimates the likelihood that a person predicted as not having a diagnosis, does not in reality have it. Efficiency corresponds to absolute agreement. Absolute agreement (efficiency), sensitivity, specificity, PPV and NPV are here interpreted as follows: <0.70 = "poor", 0.70–0.80 = "moderate", 0.80–0.90 = "satisfactory", 0.90–0.99 = "excellent" and 1.00 = "perfect". AUC is interpreted in a similar way (25). In addition to γ, the product moment correlation (Pearson *R*) is used for correlations on numeric variables.

Results

The first analysis concerns the 11 specific diagnostic criteria, which are the bases of diagnostic categories of substance use disorders in DSM-IV, seven criteria on dependence and four on abuse.

Agreement between SCID and ADDIS on criteria level

Table 1 explores agreement in the 55 protocols between the two interviews (SCID vs. ADDIS) concerning the 11 criteria (lifetime).

Thus, there is in general satisfactory absolute agreement, and Κ is significant for all criteria. The mean Κ is moderate (0.49), with different criteria varying from fair to almost perfect (0.21–0.84). There are significant systematic tendencies, deviating from symmetry, as shown by McNemar's test concerning disagreements in five of 11 criteria (D3, D5, D7, A2 and A4), in which ADDIS captures cases that are not as often captured in SCID. The differences may occur due to higher sensitivity or lower specificity in ADDIS compared with SCID. Similar tendencies (not shown here) were found concerning current diagnoses.

Agreement between SCID and ADDIS on diagnostic level

The analyses hitherto concern the criteria by which DSM-IV diagnoses are constructed. The next step is to explore agreement in terms of diagnoses based on these criteria. The dependence diagnosis requires three or more of the D1–D7 criteria to be met during 12 months. The abuse diagnosis requires at least one of the A1–A4 criteria to be met, only if the dependence diagnosis is not fulfilled for a given drug. Thus, diagnostic evaluation categorises the individuals in one of three current or lifetime

Table 1. Comparisons between Structured Clinical Interview for the DSM-IV (SCID) vs. Alkohol/Drog Diagnos InStrument (ADDIS) based on the 55 protocols on criteria for dependence (D1–D7) and abuse (A1–A4).

Lifetime	*n*	Absolute agreement	K	McNemar's test: *P*	Disagreements
D1: Tolerance	54	0.85	0.51***	0.73	Ad = 5/Sc = 3
D2: Withdrawal	53	0.87	0.69***	0.45	Ad = 2/Sc = 5
D3: Used more/longer than planned	52	0.81	0.37***	0.002	Ad = 10/Sc = 0
D4: Incapable of cutting down	53	0.81	0.43**	0.75	Ad = 6/Sc = 4
D5: Excessive time spent using/recovering from use	54	0.74	0.21*	0.002	Ad = 13/Sc = 1
D6: Giving up activities	53	0.79	0.49***	0.07	Ad = 9/Sc = 2
D7: Use despite health problems	52	0.81	0.42***	0.002	Ad = 10/Sc = 0
A1: Neglect of responsibilities	50	0.84	0.62***	0.29	Ad = 6/Sc = 2
A2: Repeated harmful use	49	0.80	0.48***	0.002	Ad = 10/Sc = 0
A3: Legal problems due to use	51	0.92	0.84***	0.13	Ad = 4/Sc = 0
A4: Reoccurring social problems	50	0.76	0.36***	0.000	Ad = 12/Sc = 0
Mean agreements		0.82	0.49		

Agreement is explored as absolute agreement between interviews and as measured with Cohen's kappa. Differences are explored for tendencies of either of the instruments being more sensitive, using McNemar's test. *$P<0.05$; **$P<0.01$; ***$P<0.001$. Ad = ADDIS captures criterion, not captured in SCID; Sc = SCID captures criterion, not captured in ADDIS.

categories: dependence, abuse or no diagnosis. Since there were too few cases using specific drugs other than alcohol, these were pooled together as "drug use disorder". Diagnostic proposals based on SCID and ADDIS are compared for all interviewed individuals in Table 2.

As shown, K varies from only fair (alcohol, current) and moderate (alcohol, lifetime; drugs, lifetime) to almost perfect (drugs, current). Absolute agreement varies from moderate to satisfactory, and γ is almost perfect or perfect. Of these tests, K is the most conservative. The McNemar–Bowker tests on alcohol diagnoses indicate that there is a systematic tendency corresponding to the tendency shown on criteria level. That tendency is significant concerning lifetime diagnosis: ADDIS tends to result in lifetime alcohol dependence diagnosis more often than SCID. Which of these is more correct was explored when compared with the GS for both lifetime and current diagnoses. Table 3 presents agreement with GS (K, sensitivity, specificity, AUC, PPV, NPV and efficiency) based on dichotomies (dependent/not dependent).

This table demonstrates that SCID has acceptable validity for substance dependence diagnosis, with moderate K on alcohol and almost perfect on drugs. AUC is close to 90%. Sensitivity is substantial or almost perfect, while specificity (and PPV) is perfect. The only problem is poor NPV on alcohol, i.e. a prediction of non-dependence on alcohol has only 44% chance of being correct. Thus, the SCID is better in confirming cases of alcohol dependence than in confirming non-cases. ADDIS, on the other hand, seems to have almost perfect (or perfect) validity, K, sensitivity, specificity, AUC, PPV, NPV and efficiency on both alcohol and drug dependence.

In Table 4, agreement between the two interviews vs. the GS is explored concerning substance use diagnoses as trichotomies, i.e. with all three possible categories (no diagnosis/abuse/dependent).

The table shows that SCID has moderate systematic agreement (K), satisfactory absolute agreement and almost perfect systematic ordinal correlation (γ) for alcohol use disorder. Thus, there is better agreement concerning presence of some alcohol use disorder diagnosis, than in providing the exact diagnosis. Concerning drug use disorders, SCID produces diagnostic proposals not only in excellent to almost perfect systematic ordinal correlation (γ) with GS, but also in satisfactory systematic agreement (K). Thus, SCID shows better diagnostic precision on drugs than on alcohol. ADDIS demonstrates perfect or almost perfect agreement on all measures. Therefore, ADDIS provides diagnostic proposals that are quite precise on both alcohol and drug use disorders.

Analysis of misclassifications

Misclassifications were further analysed as to what factors may predict them. Multivariate logistic regression models were tried but no significant model was found. Thus, misclassifications in SCID as well as in ADDIS showed no relation to setting, age, gender, GAF score or psychiatric problems. Both instruments thus seem to be robust in relation to these factors.

Agreement on severity

The number of criteria met can be interpreted as severity. Severity ratings could include only dependence criteria and vary from 0 to 7, or they can include both dependence and abuse criteria and vary from 0 to 11. The latter is more interesting since the abuse/dependence dichotomy has been abolished in the DSM-5, seeing

Table 2. Agreement between Structured Clinical Interview for the DSM-IV (SCID) and Alkohol/Drog Diagnos InStrument (ADDIS) at the level of substance use disorder diagnoses—trichotomies ($n = 42$).

ADDIS	SCID: No diagnosis	SCID: Abuse	SCID: Dependence	Agreement measures
Alcohol use disorder (Lifetime)				
No diagnosis	2	0	0	K = 0.62; $P < 0.001$
Abuse	0	3	0	γ = 1.00; $P = 0.009$
Dependence	0	5	32	Absolute agreement = 0.88; McNemar–Bowker: $P = 0.025$
Alcohol use disorder (Current)				
No diagnosis	2	0	0	K = 0.55 ; $P < 0.001$
Abuse	1	2	0	γ = 0.98 ; $P = 0.008$
Dependence	1	4	32	Absolute agreement = 0.86; McNemar–Bowker: $P = 0.112$
Drug use disorder (Lifetime)				
No diagnosis	26	0	0	K = 0.75; $P < 0.001$
Abuse	0	0	1	γ = 0.98; $P < 0.001$
Dependence	1	1	5	Absolute agreement = 0.74; McNemar–Bowker: $P = 0.607$
Drug use disorder (Current)				
No diagnosis	28	0	0	K = 0.92; $P < 0.001$
Abuse	0	0	0	γ = 1.00; $P < 0.001$
Dependence	0	1	5	Absolute agreement = 0.79; McNemar could not be tested*

*McNemar and McNemar–Bowker need the same categories in both variables. Here ADDIS produced a dichotomy, and SCID a trichotomy.

Table 3. Agreement between Structured Clinical Interview for the DSM-IV (SCID) and Alkohol/Drog Diagnos InStrument (ADDIS), respectively, vs. the golden standard (GS) at the level of dependence diagnoses—dichotomies (dependent vs. not dependent; $n = 34$).

Diagnosis	Interview prevalence %	GS prevalence %	True-Positives (TP)	False-Positives (FP)	False-Negatives (FN)	True-Negatives (TN)	κ	Sensitivity (SN) = TN/ (TN+FP)	Specificity (SP) = TP/ (TP+FN)	AUC (SN+SP)/2	Positive predictive value (PPV) = TP/ (TP+ FP)	Negative predictive value (NPV) = TN/ (TN+ FN)	Efficiency (TP+ TN)/n
SCID: Lifetime alcohol dependence	74%	88%	25	0	5	4	0.54	0.83	1.00	0.92	1.00	0.44	0.85
SCID: Current alcohol dependence	74%	88%	25	0	5	4	0.54	0.83	1.00	0.92	1.00	0.44	0.85
SCID: Lifetime drug dependence	18%	24%	6	0	2	26	0.82	0.75	1.00	0.88	1.00	0.93	0.94
SCID: Current drug dependence	15%	18%	5	0	1	28	0.89	0.83	1.00	0.92	1.00	0.97	0.97
ADDIS Lifetime alcohol dependence	88%	88%	30	0	0	4	1.00	1.00	1.00	1.00	1.00	1.00	1.00
ADDIS Current alcohol dependence	88%	88%	30	0	0	4	1.00	1.00	1.00	1.00	1.00	1.00	1.00
ADDIS: Lifetime drug dependence	21%	24%	7	0	1	26	0.92	0.96	1.00	0.98	1.00	0.96	0.97
ADDIS: Current drug dependence	18%	18%	6	0	0	28	1.00	1.00	1.00	1.00	1.00	1.00	1.00

addiction as a continuum based on a severity rating from 0 to 11. The 11 criteria used in DSM-IV are not identical to the 11 criteria in DSM-5, but differences are few: Craving is now included, and problems with law enforcement are excluded, as such problems are culturally dependent. It is therefore relevant to explore the correlations (γ and Pearson R) of severity scorings from ADDIS and SCID to severity scorings of GS (Table 5).

The agreement on severity ratings are, as indicated by the extremely high correlations, almost perfect or, indeed, perfect. Thus both SCID and ADDIS provide reliable ratings on severity.

Prediction of missing data

Even though the low number of missing data (2.3%) does not seem problematic, it is relevant to explore what factors might be associated with data loss. The missing data in the protocols for each of the two interviews were analysed in multivariate regression models (backward deletion) with the following variables as predictors: age, gender, type of drug (alcohol vs. drugs), setting, GAF score, prevalence of depression or anxiety. For ADDIS, no predictor reached significance. For SCID, the optimal model was significant ($P = 0.022$) and explained a modest share of the variation (adjusted $R^2 = 0.14$). It included the following two items: younger age ($\beta = 0.31$; $P = 0.040$) and lower GAF score ($\beta = 0.26$; $P = 0.083$). Thus, persons of young age and low social function may have some more problems in completing SCID.

Discussion

One obvious weakness of the study is the relatively small data sample due to lack of funding. The cooperating institutions carried out the interviews and re-evaluations as a quality improvement study on their own budgets. The findings, however, are clear and consistent. Systematic agreement (κ) between SCID and ADDIS on the criteria level (lifetime) is moderate, and on the diagnosis level it varies from moderate to almost perfect. ADDIS tends to produce dependence diagnosis more often than SCID. The differences were explored by comparing the results with the re-evaluation based on all data, here used as a GS.

Agreement (κ) between SCID and GS is moderate concerning alcohol dependence and satisfactory concerning drug dependence. Alternative agreement measures, i.e. AUC and efficiency (total agreement), are almost perfect. Specificity is perfect and sensitivity is satisfactory, while the NPV is poor. Thus, the SCID is better in finding and confirming cases, and in excluding non-cases of alcohol dependence rather than in confirming non-cases. Agreement between ADDIS and GS is perfect (or close to), as is sensitivity and specificity.

Table 4. Agreement between Structured Clinical Interview for the DSM-IV (SCID) and Alkohol/Drog Diagnos InStrument (ADDIS), respectively, vs. the golden standard (GS) at the level of substance use disorder diagnoses—trichotomies ($n = 34$).

	No diagnosis (GS)	Abuse (GS)	Dependence (GS)	Agreement measures
Alcohol use disorder (SCID—lifetime)				
No diagnosis	1	0	0	K = 0.55; $P = 0.020$
Abuse	0	3	5	γ = 1.00; $P < 0.001$
Dependence	0	0	25	Absolute agreement = 0.85
Alcohol use disorder (SCID—current)				
No diagnosis	1	1	1	K = 0.47; $P < 0.001$
Abuse	0	2	4	γ = 0.96; $P = 0.018$
Dependence	0	0	25	Absolute agreement = 0.82
Drug use disorder (SCID—lifetime)				
No diagnosis	26	0	1	K = 0.83; $P < 0.001$
Abuse	0	0	1	γ = 1.00; $P < 0.001$
Dependence	0	0	6	Absolute agreement = 0.94
Drug use disorder (SCID—current)				
No diagnosis	28	0	0	K = 0.90; $P < 0.001$
Abuse	0	0	1	γ = 1.00; $P = 0.001$
Dependence	0	0	5	Absolute agreement = 0.97
Alcohol use disorder (ADDIS—lifetime)				
No diagnosis	1	0	0	K = 1.00; $P < 0.001$
Abuse	0	3	0	γ = 1.00; $P = 0.017$
Dependence	0	0	30	Absolute agreement = 1.00
Alcohol use disorder (ADDIS—current)				
No diagnosis	1	0	0	K = 1.00; $P < 0.001$
Abuse	0	3	0	γ = 1.00; $P = 0.017$
Dependence	0	0	30	Absolute agreement = 1.00
Drug use disorder (ADDIS—lifetime)				
No diagnosis	26	0	0	K = 0.92, $P < 0.001$
Abuse	0	0	1	γ = 1.00; $P < 0.001$
Dependence	0	0	7	Absolute agreement = 0.97
Drug use disorder (ADDIS—current)				
No diagnosis	28	0	0	K = 1.00; $P < 0.001$
Abuse	0	0	0	γ = 1.00; $P = 0.001$
Dependence	0	0	6	Absolute agreement = 1.00

Agreement with GS was also explored using all three diagnostic classifications, i.e. dependence, abuse and no diagnosis. Here, SCID's agreement with GS is moderate concerning alcohol and excellent concerning drug use disorders, while ADDIS is perfect or almost perfect in agreement on all measures. Therefore, ADDIS provides precise diagnostic proposals on both alcohol and drug use disorders.

Table 5. Correlations between Structured Clinical Interview for the DSM-IV (SCID) and Alkohol/Drog Diagnos InStrument (ADDIS), respectively, and golden standard (GS) on severity ratings, i.e. number of criteria (0–11) for alcohol and drug problems met ($n = 34$).

	Alcohol		Drugs	
	γ	*R*	γ	*R*
SCID: Lifetime	0.90	0.83	0.98	0.93
SCID: Current	0.88	0.87	0.97	0.94
ADDIS: Lifetime	0.97	0.96	1.00	1.00
ADDIS: Current	0.87	0.94	1.00	1.00

All $P \leq 0.001$.

Both instruments demonstrate good quality and can be used as diagnostic tools. Still, ADDIS has better sensitivity compared with SCID, while having the same excellent specificity. Muthén (28) pointed out low sensitivity to be the most common problem in diagnostic instruments. The high sensitivity in ADDIS is therefore important. The difference in sensitivity was expected, since ADDIS uses more questions to capture each criterion. Often, such strategies may lead to lower specificity, but since these questions are very precise, they do not corrupt specificity. The expectation that SCID might have an advantage over ADDIS in specificity therefore proved wrong.

The problems detected in SCID are: 1) a somewhat lower sensitivity compared with ADDIS, and 2) a poor NPV—both problems concern only alcohol use disorders, not drug use disorders. These problems with the SCID are, however, not alarming. The problem is in precision, i.e. differentiating dependence from abuse. This dichotomy is not used in the new version, DSM-5, which instead uses a continuum approach based on number of criteria, very similar to the severity rating tried here. As

that severity rating indicates almost perfect or perfect correlations to GS, the problem will be less important in the future.

The problem of a low predictive value in SCID must also be evaluated in relation to the strengths also shown in SCID. There is still an acceptable sensitivity, and an excellent specificity, within a broad-spectrum instrument that also covers a range of other psychiatric disorders. SCID was designed to sort out a variety of psychiatric problems from each other, by having few "gate questions" to open up new modules, and then develop these problems more in depth in those modules. In that situation, SCID could not have too many questions in each module, but if the object—for clinical or research reasons—is to have precise diagnosis on the substance use disorder, including alcohol dependence and abuse as distinct criteria, ADDIS should be preferred. Despite more questions on substance use problems, ADDIS was no more problematic to administer. On the contrary, there was less data loss.

Both instruments can be used in populations with various level of functioning and with some psychiatric comorbidity—here concerning depression and anxiety. The few misclassifications were not related to age or gender, neither to low social functioning nor to psychiatric problems. Both instruments seem to be robust enough to handle such cases.

In conclusion, both ADDIS and SCID can be used to ensure good standards in the diagnostic assessment of substance use disorders, with and without psychiatric comorbidity, and concerning alcohol as well as drug use problems.

Acknowledgements—The authors are grateful to the anonymous patients who agreed to participate, and to the persons responsible for carrying out interviews and re-evaluations in Rockesholm Treatment Centre: Stefan Keventer, Karl Hildonen, Ann Fredmark (ADDIS interviewers), Christina Unge, Bill Engman, Tuulikki Kinnunen (SCID interviewers), MDs Lars Gäwert and Eva Carlgren Rosendal (final diagnostic assessments). The authors also want to thank Marianne Fast, School of Health Science, Jönköping University, for administering the protocols and recording them in the SPSS database.

Declaration of interest The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

References

1. Mee-Lee D, editor. ASAM Patient Placement Criteria for the Treatment of Substance-Related Disorders, 2nd edition, revised (PPC-2R). Chevy Chase, MD: American Society of Addiction Medicine, Inc. (ASAM); 2001.
2. Walitzers KS, Connors GJ. Treating problem drinking. Alcohol Res Health 1999;23:138–43.
3. Erickson CK. The science of addiction—From neurology to treatment. New York: W.W. Norton & Company, Inc.; 2007.
4. First MB, Gibbon M, Spitzer RL, Williams J, Benjamin LS. Handbok—SCID-I och SCID-II för DSM-IV. Svensk bearbetning av Jörgen Herlofsson. Danderyd: Pilgrim Press; 1999.
5. ADDIS. Alkohol/Drog Diagnos InStrument, Manual, version 09/2011. Åre: Dahl & Dahl AB; 2011.
6. Sheehan DV, Lecrubier Y, Sheehan KH, Janavas J, Weiller E, Keskiner A, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry 1998;59:22–33.
7. Segal DL, Hersen M, Van Hasselt VB, Kabacoff RI, Roth L. Reliability of diagnosis in older psychiatric patients using the Structured Clinical Interview for DSM-III-R. J Psychopathol Behav Assess 1993;15:347–56.
8. Segal DL, Hersen M, Van Hasselt VB. Reliability of the Structured Clinical Interview for DSM-III-R: An evaluative review. Compr Psychiatry 1994;35:316–27.
9. Segal DL, Kabacoff RI, Hersen M, Van Hasselt VB, Ryan CF, et al. Update on the reliability of diagnosis in older psychiatric patients using the Structured Clinical Interview for DSM-III-R. J Clin Geropsychol 1995;1:313–21.
10. Strakowski SM, Tohen M, Stoll AL, Faedda GL, Mayer PV, Kolbrener ML, et al. Comorbidity in psychosis at first hospitalization. Am J Psychiatry 1993;150:752–7.
11. Strakowski SM, Keck Jr PE, Elroy SL, Lonczak HS, West SA. Chronology of comorbid and principal syndromes in first-episode psychosis. Compr Psychiatry 1995;36:106–12.
12. Stukenberg KW, Dura JR, Kiecolt-Glaser JK. Depression screening scale validation in an elderly, community dwelling population. Psychol Assess 1990;2:134–8.
13. Martin CS, Pollock NK, Bukstein OG, Lynch KG. Inter-rater reliability of the SCID alcohol and substance use disorders section among adolescents. Drug Alcohol Depend 2000;59:173–6.
14. Kranzler HR, Kadden RM, Babor TF, Tennen H, Rounsaville BJ. Validity of the SCID in substance abuse patients. Addiction 1996;91:859–68.
15. Kranzler HR, Kadden RM, Burlesson JA, Babor TF, Apter A, Rounsaville BJ. Validity of psychiatric diagnoses in patients with substance use disorders: Is the interview more important than the interviewer? Compr Psychiatry 1995;36:278–88.
16. Robins LN, Helzer JE, Ratcliff KS, Seyfried WS. Validity of the Diagnostic Interview Schedule, Version II-DSM-III diagnoses. Psychol Med 1982;12:855–70.
17. Harrison PA, Hoffmann NG. SUDDS: Substance Use Disorder Diagnostic Schedule. St. Paul, MN: Ramsey Clinic Associates; 1985.
18. Davis LJ Jr., Hoffmann NG, Morse RM, Luehr J. Substance use disorder diagnostic schedule and an interviewer-administered format. Alcoholism: Clin Exp Res 1992;16:250–4.
19. Hoffmann NG, Harrison PA. SUDDS-IV (Substance use disorder diagnostic Schedule-IV) Manual; 1996.
20. Hoffmann NG, Hoffmann TD. Construct validity for alcohol dependence as indicated by the SUDDS-IV. Substance Use Misuse 2003;38:293–306.
21. Gerdner A. Diagnosinstrument för beroende och missbruk—Granskning av ADDIS validitet och interna konsistens gällande alkoholproblem. Nordisk Alkohol- Narkotikatidskrift 2009;26: 265–76.
22. APA. Global funktionsskattningsskala (GAF). In: MINI-D IV—Diagnostiska kriterier enligt DSM-IV-TR. Stockholm: Pilgrim Press & American Psychiatric Association (APA) 2002; p. 34–5.
23. Beck AT, Steer RA. Beck Anxiety Inventory (BAI). Manual, svensk version. Stockholm: Pearson Assessment; 2005.
24. Beck AT, Steer RA, Garbin MG. Psychometric properties of the Beck Depression Inventory Twenty-five years of evaluation. Clin Psych Rev 1988;8:77–100.
25. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977:33:159–74.
26. Hanley JA, Lipman-Hand A. If nothing goes wrong, is everything all right? Interpreting zero numerators. JAMA 1983;249:1743–5.

27. Sheehan DV, Sheehan KH, Shytle RD; Janays J, Bannon Y, Rogers JE, et al. Reliability and validity of the Mini International Neuropsychiatric Interview for Children and Adolescents (MINI-KID). J Clin Psychiatry 2010;71:313–26.
28. Muthén B. Psychometric evaluation of diagnostic criteria: Application to a 2-dimensional model of alcohol abuse and dependence. Drug Alcohol Depend 1996;41:101–12.

Arne Gerdner, Ph.D., professor, School of Health Sciences, Jönköping University, Jönköping, Sweden.
Jenny Kestenberg, M.D., Clinic of Psychiatry, NU Hospital Group, Region Västra Götaland, Trollhättan, Sweden.
Mattias Edvinsson, M.D., Clinic of Psychiatry, NU Hospital Group, Region Västra Götaland, Trollhättan, Sweden.