Psychiatry Research 259 (2018) 482–487

Contents lists available at ScienceDirect

Psychiatry Research

journal homepage: www.elsevier.com/locate/psychres

Validity of routine clinical diagnoses in acute psychiatric inpatients

Eduard Zander[a,1], Lea Wyder[a,b,1], Martin grosse Holtforth[b,c], Ulrich Schnyder[d], Urs Hepp[e,*], Niklaus Stulz[a,b]

[a] Psychiatric Services Aargau, P.O. Box 432, CH-5201 Brugg, Switzerland
[b] University of Berne, Department of Psychology, Fabrikstrasse 8, CH-3012 Berne, Switzerland
[c] Inselspital, Psychosomatic Competence Center, Haus C.L. Lory, CH-3010 Berne, Switzerland
[d] University Hospital Zurich, Department of Psychiatry and Psychotherapy, University of Zurich, Culmannstrasse 8, CH-8091 Zurich, Switzerland
[e] Integrated Psychiatric Services Winterthur – Zurcher Unterland, Wieshofstrasse 102, P.O. Box 144, CH-8408 Winterthur, Switzerland

ARTICLE INFO

Keywords:
Assessment
Interview
Examination
Psychiatry
ICD-10
DSM-IV
DSM-5

ABSTRACT

Aim: To examine the validity of diagnoses obtained by clinicians during routine clinical examination on acute psychiatric inpatient wards.
Methods: N = 100 inpatients with a broad spectrum of major mental disorders were randomly selected in a mental hospital's department of general psychiatry. Patients were diagnosed by independent assessors within Md = 5 (Range: 1–18) days of admission using the SCID I in order to examine the validity of the diagnoses given by the clinical staff based on routine assessments.
Results: The commonly used clinical examination technique had good overall agreement with the SCID I assessments regarding primary diagnoses at the level of ICD-10 main categories (F2, F30-31, F32-F33, F4; κ = 0.65). However, agreement between routine clinical diagnoses and the SCID I diagnoses tended to be low for some specific mental disorders (e.g., depressive disorders) and for secondary diagnoses.
Conclusions: The validity of routine clinical diagnoses established in acute inpatient settings is limited and should be improved.

1. Introduction

Careful and sound diagnostic assessment of mental disorders according to the criteria of ICD-10 (World Health Organization, 1992), DSM-IV (American Psychiatric Association, 1994) or DSM-5 (American Psychiatric Association, 2013) is of crucial importance in psychiatric research and practice. Psychiatric diagnoses inform treatment decisions suggested by treatment guidelines for specific mental disorders (American Psychiatric Association, 2006), and they facilitate the communication between clinicians, researchers, and other stake holders such as health insurances or governmental health departments. With the implementation of Diagnoses-Related-Groups (DRG), accurate diagnostic procedures also gain importance regarding the reimbursement of (inpatient) mental health care in many countries (Drozd et al., 2006).

Comprehensive structured diagnostic interviews such as the SCID (First et al., 1997, 1994) are widely acknowledged as the "gold standard" for diagnostic assessment. However, they are often considered too time-consuming for everyday clinical practice and are therefore mainly used in the context of research (Rettew et al., 2009). In routine inpatient care, diagnoses are usually obtained by means of unstructured intake interviews. Despite their importance and far reaching implications, only few studies assessed the accuracy and validity of routine clinical diagnoses (Egan et al., 2003; Kashner et al., 2003; Ramirez Basco et al., 2000; Shear et al., 2000). Most related studies were either restricted to specific disorders (e.g., depression) or conducted in outpatient or community mental health settings. To our knowledge, there are only very few studies that examined the accuracy and validity of routinely assessed clinical diagnoses among inpatients in mental hospitals (Andreas et al., 2009; Miller et al., 2001; Steiner et al., 1995). These studies examined rather small samples of 53–56 inpatients, and thus provide limited informative value for agreement within diagnostic subgroups. Furthermore, Andreas et al. (2009) only included female patients from a psychotherapy ward with a markedly different diagnostic distribution compared to an acute inpatient setting.

The heterogeneous results of these studies point to the need for further validation of commonly used procedures for routine diagnostic assessment of mental disorders in inpatient settings (Andreas et al., 2009). Therefore, the aim of the current study was to analyze the diagnostic agreement between routine clinical practices (techniques) and the diagnoses rendered by structured research interviews for the most

* Corresponding author.
[1] Joint first authors.

https://doi.org/10.1016/j.psychres.2017.11.004
Received 21 March 2017; Received in revised form 26 October 2017; Accepted 1 November 2017
Available online 06 November 2017

common mental disorders in general psychiatric inpatient care.

2. Methods

2.1. Participants

Participants were inpatients on acute general psychiatry wards of a mental hospital in Switzerland. Further inclusion criteria were: aged 18–64 years, and one of the following primary routine clinical diagnoses according to ICD-10 (World Health Organization, 1992): schizophrenia, delusional, or other psychotic disorder (F2), mood (affective) disorder (F3), or anxiety, dissociative, stress-related, and somatoform disorder (F4). These are the most prevalent diagnoses in general psychiatry and restricting our analyses to these diagnoses ensured for sufficient sample sizes in diagnostic subgroups. In Switzerland, health insurances require diagnoses to be coded according to ICD-10, therefore our analyses rest upon F-codes of chapter V of the ICD-10. Exclusion criteria of the current study were secondary diagnoses of organic mental disorders (F0) or intellectual disabilities (F7).

2.2. Measures

The SCID I (First et al., 1997) is a standardized interview consisting of a screening section and a subsequent structured interview to diagnose major (DSM-IV Axis I) mental disorders based on DSM-IV criteria (American Psychiatric Association, 1994), except for personality disorders, which are covered by the SCID II (First et al., 1994). SCID I is a widely used interview schedule (usually considered the "gold standard" for DSM-IV Axis I disorders) with moderate to excellent inter-rater agreement of the Axis I disorders (κ = 0.60–0.83) (Lobbestael et al., 2011).

2.3. Procedures

Participants were randomly drawn from all patients who were admitted to one of seven acute general psychiatry wards (143 beds) of a Swiss mental hospital between January and September 2016. Attending psychiatrists and psychologists assigned clinical diagnoses according to ICD-10 (World Health Organization, 1992) to all patients at the day of admission as part of routine procedures. Clinical diagnoses were based on unstructured clinical interviews during routine clinical examination and, where available, on additional medical data and/or reports by relatives of the patient. A senior psychiatrist supervised clinical assessments by resident psychiatrists and psychologists in routine daily case conferences.

Whenever the independent and blinded SCID I assessors had capacity to perform the next interview they contacted an independent member of the research team. This team member screened all newly admitted patients for eligibility and randomly drew the next SCID participants. On weekdays, these participants were randomly drawn from all new admissions of the prior day, except for Mondays when participants were drawn from all admissions of the last three days (i.e. Friday to Sunday).

A weighted random algorithm that accounted for different prevalence rates (baseline probabilities) of specific mental disorders in the hospital was used for patient selection. This aimed at attaining a disproportionally stratified sample with roughly equal numbers of patients in the following diagnostic groups (primary diagnoses) according to ICD-10: (a) F20, F23-F24; (b) F25, F30-F31; (c) F32-F33; and (d) F4. Patients with a routine clinical primary diagnosis of a bipolar affective disorder (F30-F31), for example, had a higher probability of being drawn for the study sample than patients with an unipolar affective disorder (F32-F33), since bipolar patients had been much less frequent (13.6%) in the past year than unipolar patients (30.9%).

The four diagnostic groups are equivalent to ICD-10 classifications except for group (b). We merged this residual category of schizoaffective (F25) and bipolar disorders (F30-F31) for two reasons: First, diagnostic reliability of schizoaffective disorder is generally poor, and the ongoing debate on nosology of this condition remains inconclusive in terms of an empirically supported allocation to either the psychotic or affective group of disorders (Jäger et al., 2011). Second, prevalence rates of schizoaffective and bipolar disorders are considerably lower than those of, e.g., schizophrenia or depression; through forming an additional subgroup we were able to ensure sufficient cell sizes.

SCID I interviews were conducted by a senior psychiatrist (E.Z.) and a postgraduate psychologist (L.W.). On average, the structured interviews were conducted within $Md = 4$ days (Range: 1–11 days, 75th percentile: < 5 days) of the routine clinical assessment at intake. To accommodate patient resources, in 24 cases the structured assessment was split into two sessions. On average, these interviews were completed within $Md = 5$ days (Range: 1–18 days, 75th percentile: < 6 days) of the initial routine clinical assessment.

Both interviewers had received extensive training on the SCID I prior to the study. Inter-rater reliability of the two SCID raters was calculated as follows: Nine interviews were videotaped and rated by the other interviewer. Due to the small number of cases, we analyzed the inter-rater reliability using the percentage of agreement for assigned primary SCID diagnoses, which resulted in an overall inter-rater agreement of P_O = 77.8%. This figure complies with previously reported inter-rater reliability of the SCID (P_O = 82%) (Ventura et al., 1998).

The current study was embedded into a larger home treatment research program (ClinicalTrials.gov: NCT02322437). It was approved by the local ethics committee and conducted according to the declaration of Helsinki. All patients gave written informed consent.

2.4. Data analysis

The structured research diagnoses were assessed using the SCID I (First et al., 1997) according to the criteria of DSM-IV (American Psychiatric Association, 2013), and the resulting diagnoses were subsequently translated into ICD-10 codes (World Health Organization, 1992) according to the rules given in the SCID I manual. The primary outcome of this study was the agreement between clinical diagnoses and structured research diagnoses, which was calculated using Cohen's κ (Cohen, 1960). Cohen's κ has proved the most preferred measure to determine diagnostic agreement, as it takes the agreement by chance into account by incorporating base rates of categories (Rettew et al., 2009). The κ coefficient determines agreement between two nominal scores (e.g. presence or absence for a diagnosis of schizophrenia for the same individual derived from one research and one clinical assessment). Cohen's κ ranges from −1.00 to +1.00; κ values are classified as poor ($\kappa \leq 0.40$), fair (0.41–0.59), good (0.60–0.74), and excellent ($\geq$ 0.75) agreement (Landis and Koch, 1977). Analyses were performed using SPSS, Version 21 (SPSS Inc, 2009).

3. Results

One hundred (57.1%) of the 175 randomly selected patients were interviewed using the SCID I. Reasons for non-participation were: patients' refusal to participate (n = 39), an acute mental health state that did not allow for a structured clinical assessment (n = 25), hospital discharge before the SCID could be performed (n = 19), and other defined reasons (e.g., patients previously known to the interviewers; n = 9). Participants did not differ significantly from non-participants with respect to sex (42.4% vs. 49.3% female; $\times^2 = 0.822$, $p = 0.365$), but they were younger than non-participants (M = 38.1 vs. M = 43.0 years; $U = 2926.00$, $Z = -2.391$. $p = 0.017$), and were less often diagnosed with psychotic disorders based on routine clinical diagnoses (28.0% vs. 52.0%; $\times^2 = 11.15$, $p = 0.011$). The final, disproportionally stratified sample consisted of N = 100 patients (age: M

Table 1
Coefficients of agreement for grouped primary diagnoses (level 1) between nonstructured routine clinical interviews vs. SCID. (Diagnostic groups with n ≥ 5 cases in both interview types only.)

Diagnosis (ICD-10)	2 × 2 Table		Diagnostic sensitivity	Diagnostic specificity	PPV	NPV	Overall agreement	κ (95% CI)
	a c	b d						
Psychotic disorders (F2)	25 2	3 70	0.93	0.96	0.89	0.97	0.95	0.88 (0.77–0.98)
Bipolar mood disorders (F30-F31)	16 2	3 79	0.89	0.96	0.84	0.98	0.95	0.83 (0.69–0.98)
Unipolar depressive disorders (F32-F33)	19 14	7 60	0.58	0.90	0.73	0.81	0.79	0.50 (0.31–0.68)
Anxiety and stress-related disorders (F4)	14 7	13 66	0.67	0.84	0.52	0.90	0.80	0.45 (0.25–0.66)

Notes: a = SCID (+) and routine clinical interview (+); b = SCID (−) and routine clinical interview (+); c = SCID (+) and routine clinical interview (−); d = SCID (−) and routine clinical interview (−); PPV = positive predictive value; NPV = negative predictive value.

= 38.1 years, *SD* = 13.0 years; 42% female), with roughly equal numbers of patients in the following clinically-derived diagnostic groups: schizophrenia, delusional or brief psychotic disorder (n = 26; 26%); schizoaffective or bipolar affective disorders (n = 21; 21%); depressive disorders (n = 26; 26%); and anxiety or stress-related disorders (n = 27; 27%). Of these N = 100 patients, 74 were admitted on weekdays (M = 14.8 per day) and 26 were admitted on weekends (M = 13.0 per day). The average number of primary and secondary diagnoses per patient given after unstructured routine clinical interviews was M = 1.36 (SD = 0.87) and M = 1.73 (SD = 0.81) pursuant to SCID I assessments ($Z = -3.491$; $p < 0.001$).

All diagnoses were specified on two levels, representing different degrees of diagnostic accuracy: on level 1, specific diagnoses were grouped on a relatively high level of abstraction based on shared predominant symptoms; e.g. schizophrenia, delusional, and schizoaffective disorders were grouped into psychotic disorders (Table 1). On level 1, overall diagnostic agreement between primary clinical diagnoses and primary SCID diagnoses was good (κ = 0.65, 95% CI = 0.54–0.77), with kappa values ranging from fair (κ = 0.45) for anxiety and stress-related disorders to excellent (κ=0.88) for psychotic disorders (Table 1). The overall diagnostic agreement was only slightly higher (κ = 0.71, 95% CI = 0.60–0.81) if SCID assessors had used additional information from medical records to potentially modifiy their diagnostic classification. We therefore report SCID diagnoses without medical record information only. When considering both primary and secondary diagnoses, diagnostic agreement within level 1 diagnostic groups ranged from κ = 0.29 for anxiety and stress-related disorders to κ = 0.88 for psychotic disorders (Table 2). Note that analyses within diagnostic groups were performed only for disorders that were diagnosed at least five times by both the clinicians and the SCID assessors (Lobbestael et al., 2011; Steiner et al., 1995).

At level 2, diagnostic agreement was analyzed for more specific categories of mental disorders; e.g., a single depressive episode was distinguished from a recurrent depressive disorder (Table 3). At this second level, the overall diagnostic agreement between clinically-derived and structured diagnoses was still fair (κ = 0.59, 95% CI = 0.49–0.70) for primary diagnoses, and it was again only slightly increased if SCID assessors had considered additional information from medical records (κ = 0.65, 95% CI = 0.55–0.75). Except for single depressive episodes (κ = 0.39), diagnostic agreement was still fair (κ = 0.55) to excellent (κ = 0.83) for all remaining specific primary disorders at level 2 (Table 3). When also taking into account secondary diagnoses, diagnostic agreement between unstructured and structured clinical interviews ranged from κ = 0.42 for depressive episodes and for reaction to severe stress and adjustment disorders, respectively, to κ = 0.83 for bipolar mood disorders (Table 4).

4. Discussion

This study examined the agreement between clinical diagnoses derived from unstructured clinical interviews commonly used under routine inpatient conditions and the diagnoses generated by the SCID I in recently admitted psychiatric inpatients with a broad spectrum of specific mental disorders (ICD-10: F2-F4). Routine clinical interviews demonstrated a good overall agreement with SCID I assessments by independent raters regarding the main type of primary diagnosis (κ = 0.65). Within these main diagnostic groups, diagnostic accuracy ranged from fair (κ = 0.45) for anxiety and stress-related disorders (ICD-10:

Table 2
Coefficients of agreement for grouped primary and secondary diagnoses (level 1) between nonstructured routine clinical interviews vs. SCID. (Diagnostic groups with n ≥ 5 cases in both interview types only.)

Diagnosis (ICD-10)	2 × 2 Table		Diagnostic sensitivity	Diagnostic specificity	PPV	NPV	Overall agreement	κ (95% CI)
	a c	b d						
Substance use disorders (F1)	12 14	2 72	0.46	0.97	0.86	0.84	0.84	0.51 (0.31–0.71)
Psychotic disorders (F2)	26 2	3 69	0.93	0.96	0.90	0.97	0.95	0.88 (0.77–0.98)
Bipolar mood disorders (F30-F31)	16 2	3 79	0.89	0.96	0.84	0.98	0.95	0.83 (0.69–0.98)
Unipolar depressive disorders (F32-F33)	24 11	8 57	0.69	0.88	0.75	0.84	0.81	0.57 (0.40–0.74)
Anxiety and stress-related disorders (F4)	19 23	10 48	0.45	0.83	0.66	0.68	0.67	0.29 (0.11–0.48)

Notes: a = SCID (+) and routine clinical interview (+); b = SCID (−) and routine clinical interview (+); c = SCID (+) and routine clinical interview (−); d = SCID (−) and routine clinical interview (−); PPV = positive predictive value; NPV = negative predictive value.

Table 3
Coefficients of agreement for specific primary diagnoses (level 2) between nonstructured routine clinical interviews vs. SCID. (Diagnostic groups with n ≥ 5 cases in both interview types only.)

Diagnosis (ICD-10)	2 × 2 Table		Diagnostic sensitivity	Diagnostic specificity	PPV	NPV	Overall agreement	κ (95% CI)
	a c	b d						
Schizophrenia (F20)	18 1	6 75	0.95	0.93	0.75	0.99	0.93	0.79 (0.07–0.65)
Bipolar mood disorders (F31)	16 2	3 79	0.79	0.96	0.84	0.98	0.95	0.83 (0.69–0.98)
Depressive episode (F32)	4 4	6 86	0.50	0.93	0.40	0.96	0.90	0.39 (0.09–0.69)
Recurrent depressive disorders (F33)	13 12	3 72	0.52	0.96	0.81	0.86	0.85	0.55 (0.35–0.74)
Reaction to severe stress, and adjustment disorders (F43)	12 3	11 74	0.80	0.87	0.52	0.96	0.86	0.55 (0.35–0.74)

Notes: a = SCID (+) and routine clinical interview (+); b = SCID (−) and routine clinical interview (+); c = SCID (+) and routine clinical interview (−); d = SCID (−) and routine clinical interview (−); PPV = positive predictive value; NPV = negative predictive value.

F4) to excellent (κ = 0.88) for psychotic disorders (F2).

If more specific primary diagnoses or secondary diagnoses were considered, however, the diagnostic validity of unstructured routine interviews was found to be poor for single depressive episodes (F32; κ = 0.39–0.42) and for anxiety and stress-related disorders (F4; κ = 0.29–0.45). These results are in line with previous findings (Andreas et al., 2009; North et al., 1997; Shear et al., 2000, Steiner et al., 1995). The particularly low diagnostic sensitivity of routine clinical interviews to detect depressive episodes (sensitivity = 0.50) and recurrent depressive disorders (sensitivity = 0.52) suggests that some of the most prevalent conditions in psychiatric inpatients as well as in the general population are missed by unstructured diagnostic assessment techniques in almost half of the cases. In fact, the most frequent diagnostic discrepancy involved cases where clinicians in routine assessments diagnosed an adjustment disorder, while structured interviews resulted in a diagnosis of a depressive disorder (remarkably, more often recurrent than one single depressive episode). Hence, it seems that for patients presenting with depressive symptoms after a stressful life event, the reactive component of the mental disturbance often outweighed symptom severity in clinicians' judgment of the present primary diagnosis. Here, our results may reflect the more cross-sectional nature of clinical interviews under routine conditions, where the circumstances leading to hospital admission and clarification of treatment objectives are the focus of interest. By contrast, in structured interviews both current and past diagnostic information is extensively explored regardless of their immediate significance for treatment. As a result, e.g., past depressive episodes may be detected more reliably when using a structured interview.

Similarly, co-morbid anxiety and stress-related disorders (sensitivity = 0.45) were also frequently missed by routine interviews. One further possible explanation for these findings could be the ubiquity of anxiety and depressive symptoms, which tend to occur in and are shared with various other mental disorders. Thus, without the guidance of a structured clinical interview, clinicians seem to frequently miss distinct anxiety and depressive disorders. Likewise in our sample, clinicians often missed co-morbid substance use disorders, and particularly cannabis use disorder, when relying on unstructured interviews (sensitivity = 0.46). Overall, SCID assessments rendered significantly more diagnoses per patient ($M = 1.73$) than did unstructured clinical assessments ($M = 1.36$, $p < 0.001$).

For schizophrenia and bipolar mood disorders, the agreement between routine and structured clinical assessments was excellent (κ = 0.79–0.83). The very salient and more distinctive symptomatology of these very severe mental disorders seems to facilitate the establishment of a valid diagnosis during routine clinical examinations. The diagnostic accuracy for schizophrenia and bipolar mood disorders tended to be even higher in our study than in previous research (North et al., 1997; Shear et al., 2000, Steiner et al., 1995). This may be explained by the fact that the majority of patients were emergency referrals presenting with manifest and very acute symptomatology at intake, potentially

Table 4
Coefficients of agreement for specific primary and secondary diagnoses (level 2) between nonstructured routine clinical interviews vs. SCID. (Diagnostic groups with n ≥ 5 cases in both interview types only.)

Diagnosis (ICD-10)	2 × 2 Table		Diagnostic sensitivity	Diagnostic specificity	PPV	NPV	Overall agreement	κ (95% CI)
	a c	b d						
Mental and behavioural disorders due to use of cannabinoids (F12)	7 9	2 82	0.44	0.98	0.78	0.90	0.89	0.50 (0.13–0.25)
Schizophrenia (F20)	19 1	6 74	0.95	0.93	0.76	0.99	0.93	0.80 (0.66–0.94)
Bipolar mood disorders (F31)	16 2	3 79	0.89	0.96	0.84	0.98	0.95	0.83 (0.69–0.98)
Depressive episode (F32)	5 3	7 85	0.63	0.92	0.42	0.97	0.90	0.42 (0.13–0.70)
Recurrent depressive disorders (F33)	14 12	4 70	0.54	0.95	0.78	0.85	0.84	0.54 (0.34–0.73)
Reaction to severe stress, and adjustment disorders (F43)	12 7	13 68	0.63	0.84	0.48	0.91	0.80	0.42 (0.21–0.63)

Notes: a = SCID (+) and routine clinical interview (+); b = SCID (−) and routine clinical interview (+); c = SCID (+) and routine clinical interview (−); d = SCID (−) and routine clinical interview (−); PPV = positive predictive value; NPV = negative predictive value.

making diagnostic classification easier for clinicians than in previous studies where patients were typically recruited in less acute inpatient or outpatient settings (Andreas et al., 2009; Shear et al., 2000; Steiner et al., 1995).

Structured clinical interviews such as the SCID (Spitzer et al., 1992) are widely acknowledged as the "gold standard" for diagnostic assessment of mental disorders, particularly in research settings. Their usefulness for routine clinical practice has also been criticized, however. They usually are too time consuming for everyday clinical practice, they still pose limitations regarding the interpretation of symptoms, and their strictly prescribed sequence of questions has been criticized to jeopardize the therapeutic alliance (Spitzer, 1983). Spitzer (1983) therefore proposed the *Longitudinal Expert Evaluation using all available Data* (LEAD) standard in order to maximize the validity of psychiatric diagnoses. In fact, previous findings suggest that combining the SCID with additional information from medical records yields more accurate diagnoses than the structured interview alone (Ramirez Basco et al., 2000). In our study, however, agreement rates with unstructured clinical interviews did not improve markedly when the SCID diagnoses were refined after additionally considering all available information from the patient's medical history. For instance, the overall agreement regarding the primary diagnosis increased only slightly from $\kappa = 0.59$ (95% CI = 0.49–0.70) to $\kappa = 0.65$ (95% CI = 0.55–0.75) when enriching SCID data with information from medical records.

One practical way to increase diagnostic validity in routine care would be for clinicians to insist on a standardized acquisition of additional information from interviews with e.g. relatives, teachers or colleagues who could provide valuable external data on prodromal signs, stress triggers, hereditary factors, and disturbance duration. However, gathering this information is often time consuming and the availability of the mentioned persons is not always guaranteed. Future studies should investigate whether adding external information improves the validity of diagnoses in routine care significantly.

Beyond the accuracy of the diagnostic process in mental health care, psychiatric diagnoses themselves have sometimes been criticized for being too deficit-oriented (Baumann et al., 2015). The recently emerging *person-centred integrative diagnosis model* argues for a broader concept of diagnosis and covers both ill health and positive health (Salloum and Mezzich, 2011). This provides a step towards a more resource oriented approach, as the foundation of treatment planning.

4.1. Limitations and further perspectives

Some limitations of our study merit attention: First, no validated German version of the SCID for DSM-5 (American Psychiatric Association, 2013) was available at the time when the study was conducted. We therefore used the SCID I for diagnosing DSM-IV disorders. DSM-IV diagnoses were then transferred to ICD-10 codes following manualized rules, as routine clinical diagnoses in the hospital were coded according to ICD-10. While this may be considered an important limitation of our study, there are no significant differences between DSM-IV and DSM-5 or between DSM-IV and ICD-10, respectively, regarding the diagnostic criteria of the major mental disorders examined in our study.

Second, we restricted our analyses to the most prevalent major mental disorders of general psychiatry (i.e., schizophrenia, affective disorders, and anxiety and stress-related disorders). Examining the diagnostic validity of routine diagnostic assessment techniques for e.g. substance use disorders or personality disorders should be the subject of further research.

Third, for some patients ($n = 13$), no psychiatric diagnosis had been documented in their medical records at intake. These patients were excluded from our study since we aimed at examining the validity of the routine clinical diagnoses given at intake. Some patients ($n = 27$) were excluded since they presented with very ambiguous symptoms at intake, which prevented resident physicians from making a diagnosis on the day of intake. The exclusion of these patients, as well as the exclusion of patients for whom no structured assessment was possible within seven days of admission (e.g., due to a very severe mental disturbance) ($n = 25$), might have biased our findings.

Fourth, whereas the diagnostic spectrum included in our study was representative for acute inpatients in general psychiatry, participants were significantly younger than non-participants. This may also limit the generalizability of our findings.

Finally, though the spreading of information about details of the ongoing study on diagnostic accuracy within the hospital was tried to keep as limited as possible, clinicians could still have become aware of it. This might have affected their behavior when assessing clinical diagnoses and hence may have impacted on the study results.

4.2. Conclusions

In summary, unstructured clinical interviews as usually conducted under routine inpatient conditions seem to provide valid diagnoses in terms of the main type of the primary mental disorder (F2, F3 and F4). However, regarding more specific primary diagnoses (e.g., F32 vs. F33) or secondary diagnoses, routine diagnostic assessment techniques tend to have poor agreement with SCID I assessments for some mental disorders. As sound and valid diagnoses may be clinically relevant for treatment courses and outcomes, further research should aim at refining strategies to improve the diagnostic process in acute inpatient settings.

Conflict of interests

None.

Funding

This study was funded by a grant from the Hugo & Elsa Isler foundation.

References

American Psychiatric Association, 2013. Diagnostic and Statistical Manual of Mental Disorders, fifth ed. Washington, DC.
American Psychiatric Association, 2006. Practice Guidelines for the Treatment of Psychiatric Disorders. Arlington, VA.
American Psychiatric Association, 1994. Diagnostic and Statistical Manual of Mental Disorders, fourth ed. Washington, DC.
Andreas, S., Theisen, P., Mestel, R., Koch, U., Schulz, H., 2009. Validity of routine clinical DSM-IV diagnoses (Axis I/II) in inpatients with mental disorders. Psychiatry Res. 170, 252–255. http://dx.doi.org/10.1016/j.psychres.2008.09.009.
Baumann, P.S., Marion-Veyron, R., Bardy, S., Solida, A., Conus, P., 2015. Beyond clinical staging of psychiatric disorders. Lancet Psychiatry 2, e27. http://dx.doi.org/10.1016/S2215-0366(15)00377-6.
Cohen, J., 1960. A coefficient of agreement for nominal scales. Educ. Psychol. Meas. 20, 37–46. http://dx.doi.org/10.1177/001316446002000104.
Drozd, E., Cromwell, J., Gage, B., Maier, J., Greenwald, L.M., Goldman, H.H., 2006. Patient casemix classification for medicare psychiatric prospective payment. Am. J. Psychiatry 163, 724–732. http://dx.doi.org/10.1176/appi.ajp.163.4.724.
Egan, S., Nathan, P., Lumley, M., 2003. Diagnostic concordance of ICD-10 personality and comorbid disorders: a comparison of standard clinical assessment and structured interviews in a clinical setting. Aust. N. Z. J. Psychiatry 37, 484–491. http://dx.doi.org/10.1046/j.1440-1614.2003.01226.x.
First, M.B., Spitzer, R.L., Gibbon, M., Williams, J.B., 1997. Structured Clinical Interview for DSM-IV Axis I Disorders (SCID I). New York.
First, M.B., Spitzer, R.L., Gibbon, M., Williams, J.B., Benjamin, L., 1994. Structured Clinical Interview for DSM-IV Axis II Personality Disorders (SCID II). New York.
Jäger, M., Haack, S., Becker, T., Frasch, K., 2011. Schizoaffective disorder - an ongoing challenge for psychiatric nosology. Eur. Psychiatry 26, 159–165. http://dx.doi.org/10.1016/j.eurpsy.2010.03.010.
Kashner, M.T., Rush, A.J., Suris, A., Biggs, M.M., Gajewski, V.L., Hooker, D.J., et al., 2003. Impact of structured clinical interviews on physicians' practices in community mental health settings. Psychiatr. Serv. 54, 712–718.
Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. Biometrics 33, 159–174.
Lobbestael, J., Leurgans, M., Arntz, A., 2011. Inter-rater reliability of the structured clinical interview for DSM-IV Axis I disorders (SCID I) and axis II disorders (SCID II). Clin. Psychol. Psychother. 18, 75–79. http://dx.doi.org/10.1002/cpp.693.

Miller, P.R., Dasher, R., Collins, R., Griffiths, P., Brown, F., 2001. Inpatient diagnostic assessments: 1. Accuracy of structured vs. unstructured interviews. Psychiatry Res. 105, 255–264. http://dx.doi.org/10.1016/S0165-1781(01)00317-1.

North, C.S., Pollio, D.E., Thompson, S.J., Ricci, D.A., Smith, E.M., Spitznagel, E.L., 1997. A Comparison of clinical and structured interview diagnoses in a homeless mental health clinic. Community Ment. Health J. 33, 531–543. http://dx.doi.org/10.1023/A:1025052720325.

Ramirez Basco, M., Bostic, J.Q., Davies, D., Rush, A.J., Witte, B., Hendrickse, W., et al., 2000. Methods to improve diagnostic accuracy in a community mental health setting. Am. J. Psychiatry 157, 1599–1605. http://dx.doi.org/10.1176/appi.ajp.157.10.1599.

Rettew, D.C., Lynch, A.D., Achenbach, T.M., Dumenci, L., Ivanova, M.Y., 2009. Meta-analyses of agreement between diagnoses made from clinical evaluations and standardized diagnostic interviews. Int. J. Methods Psychiatr. Res. 18, 169–184. http://dx.doi.org/10.1002/mpr.289.

Salloum, I.M., Mezzich, J.E., 2011. Outlining the bases of person-centred integrative diagnosis. J. Eval. Clin. Pract. 17, 354–356. http://dx.doi.org/10.1111/j.1365-2753.2010.01581.x.

Shear, M.K., Greeno, C., Kang, J., Ludewig, D., Frank, E., Swartz, H.A., et al., 2000. Diagnosis of nonpsychotic patients in community clinics. Am. J. Psychiatry 157, 581–587.

Spitzer, R.L., 1983. Psychiatric diagnosis: are clinicians still necessary? Compr. Psychiatry 24, 399–411.

Spitzer, R.L., Williams, J.B., Gibbon, M., First, M.B., 1992. The structured clinical interview for DSM-III-R (SCID). I: history, rationale, and description. Arch. Gen. Psychiatry 49, 624–629.

SPSS Inc, 2009. PASW Statistics for Windows.

Steiner, J.L., Tebes, J.K., Sledge, W.H., Walker, M.L., 1995. A comparison of the structured clinical interview for DSM-III-R and clinical diagnoses. J. Nerv. Ment. Dis. 183, 365–369.

Ventura, J., Liberman, R.P., Green, M.F., Shaner, A., Mintz, J., 1998. Training and quality assurance with the structured clinical interview for DSM-IV (SCID-I/P). Psychiatry Res. 79, 163–173.

World Health Organization, 1992. The ICD-10 Classification of Mental and Behavioural Disorders. Geneva.