

Assessing psychotic symptoms in forensic evaluations of criminal responsibility – a pilot study using Positive And Negative Syndrome Scale

Pia J. Løvgren , Petter Laake , Solveig Klæbo Reitan & Kjersti Narud

To cite this article: Pia J. Løvgren , Petter Laake , Solveig Klæbo Reitan & Kjersti Narud (2020) Assessing psychotic symptoms in forensic evaluations of criminal responsibility – a pilot study using Positive And Negative Syndrome Scale, The Journal of Forensic Psychiatry & Psychology, 31:4, 490-502, DOI: [10.1080/14789949.2020.1771402](https://doi.org/10.1080/14789949.2020.1771402)

To link to this article: <https://doi.org/10.1080/14789949.2020.1771402>



© 2020 Oslo University Hospital and the University of Oslo. Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 31 May 2020.



Submit your article to this journal [↗](#)



Article views: 486



View related articles [↗](#)



View Crossmark data [↗](#)

Assessing psychotic symptoms in forensic evaluations of criminal responsibility – a pilot study using Positive And Negative Syndrome Scale

Pia J. Løvgren ^{a,b}, Petter Laake^{b,c}, Solveig Klæbo Reitan^{d,e} and Kjersti Narud^{a,b}

^aThe Regional Centre for Research and Education in Forensic Psychiatry for the South-Eastern Norway Regional Health Authority, Oslo University Hospital, Ullevål Hospital, Norway;

^bFaculty of Medicine, University of Oslo, Norway; ^cOslo Centre for Biostatistics and Epidemiology, Department of Biostatistics, University of Oslo, Norway; ^dDepartment of Mental Health, St Olav's University Hospital, Trondheim, Norway; ^eDepartment of Mental Health, NTNU Faculty of Medicine and Health Science, Trondheim, Norway

ABSTRACT

Description of symptoms and signs related to psychotic disorders at the time of the crime is essential in forensic evaluations of legal insanity. Knowledge of the content of forensic reports is important to improve and secure their quality. Here we report the findings of a pilot study using PANSS as an instrument to assess descriptions of psychotic symptoms in forensic psychiatric reports. Three experienced psychiatrists assessed 20 forensic reports focusing on forensic experts' descriptions of the defendant's mental state at the time of the observation and at the time of the alleged crime. PANSS was evaluated as a tool for examining relevant psychotic symptoms, and interrater reliability was calculated. Interrater reliability was satisfactory. It varied based on the percentage of symptoms not described in the reports and on the type of symptom. At both times more symptoms were described from the positive scale of PANSS, than from the negative and the general scale. This pilot study shows that PANSS can be used as an instrument for the structured assessment of psychotic symptoms in written forensic reports and indicates that psychotic symptoms at the time of the alleged crime are poorly described.

ARTICLE HISTORY Received 23 May 2019; Accepted 15 May 2020

KEYWORDS Forensic psychiatry; criminal responsibility; psychotic disorders; legal insanity; psychotic symptoms

1. Introduction

To be culpable for a criminal act, a person has to be criminally responsible or legally sane. Certain mental conditions may lead to a person being acquitted from the charges. Though legislation differs between nations, psychotic conditions are the mental conditions most often leading to legal insanity (Cochrane et al., 2001; Gowensmith et al., 2017). Descriptions of psychotic

CONTACT Pia J. Løvgren  pialoe@ous-hf.no

© 2020 Oslo University Hospital and the University of Oslo. Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

conditions are important in forensic psychiatric evaluations of criminal responsibility in most nations.

The forensic conclusion depends on evaluated symptoms of mental illness in the defendant at the time of the crime, and on how these symptoms affected the defendant's behavior and perception of reality at the time of the alleged crime. This is in some sense true regardless of legislation. In Norway, the penal code states that a person has to be evaluated as 'psychotic' as a judicial term to be found legally insane. A relationship between the mental illness and the criminal act is not needed (<https://lovdata.no/lov/2005-05-20-28/S20>). Several countries and states use other principles, which demands an evaluation of how the symptoms of mental illness affected the defendant's behavior at the time of the alleged crime. Despite the differences between legislations, the basis for the forensic evaluation is a mental evaluation of the defendants' symptoms of mental illness.

As psychotic conditions are the most frequent diagnostic classifications to give legal insanity, the identification and rating of symptoms of psychosis are among the most important premises for the diagnostic and legal conclusion in forensic evaluations (Fuger et al., 2014; Gowensmith et al., 2013; Kois & Chauhan, 2018). Disagreement on the assessment of psychotic symptoms may lead to disagreement on the diagnosis (Aboraya, Rankin, France, El-Missiry & John, 2006), as well as disagreement on the legal conclusions of criminal responsibility (Fuger et al., 2014). Making the premises for the diagnostic conclusion testable and open for review is essential (Gowensmith et al., 2017). It is crucial that psychotic symptoms are described in a reliable and testable way. To secure that forensic evaluations have an adequate quality has been a concern for a long time (Borum & Grisso, 1996), but to our knowledge there exist no studies that examine forensic experts' registration of symptoms and signs of psychosis in forensic reports regarding legal insanity.

A major challenge forensic psychiatric experts face when assessing legal insanity is that the evaluation of a defendant's mental state at the time of the crime is retrospective (Gowensmith et al., 2013; Kacperska et al., 2016). Furthermore, conditions regarded as a psychotic disorder, either medically or in diagnostic systems (like ICD and DSM), show a large variation in symptomatology. With increasing possibilities for relief of symptoms through treatment and support, people with a psychosis diagnosis may not always have symptoms of a degree the legislation demands for legal insanity.

The retrospective focus of the forensic reports, their written form and the large variety of expressions of psychotic disorders increase the demand for concise and systematic registration and reporting on the individual symptoms and signs of psychosis.

In this article, we report the results of a pilot study using the assessment instrument PANSS (The Positive and Negative Syndrome Scale (Kay et al., 1987)) to assess how forensic experts described symptoms and signs of

psychosis in forensic evaluations of legal insanity. The forensic experts evaluated the defendants at the time of the mental observation and retrospectively at the time of the alleged crime. Three assessors evaluated 20 forensic reports independently, and agreement between the three assessors was calculated. Our primary hypothesis was that experts with extended clinical and forensic experience are able to reach the acceptable agreement when assessing descriptions of psychotic symptoms in forensic reports based on PANSS.

2. Methods

2.1 Material

Twenty forensic psychiatric reports sent to the courts in Norway in 2013 were selected from the archives of the Norwegian Forensic Medical Board (NFMB). Regarding indictment, murder or attempted murder were set as criteria for inclusion in the study. This was done in an attempt to assure that only complete forensic assessments were included.

Reports are structured, consisting of mainly four parts, starting with the appointment and mandate from the court. A description of the indictment made by the court comes next, together with a brief recollection of the police investigation documents of relevance for the psychiatric evaluation (part one). A summary of the examination of the defendant along with a clinical mental status (part two) is followed by relevant documentation from patient files collected from the health-care system, if approval for this is given by the defendant (part three). Finally, the diagnostic and legal considerations together with the conclusion answering the mandate are reported (part four).

2.2 Measurement of symptoms

The Positive and Negative Syndrome Scale (PANSS) is an instrument for rating the severity of psychotic symptoms in schizophrenia (Kay et al., 1987). It is considered a valid and reliable instrument for rating change in symptoms of psychosis in clinical settings and for research purposes and is widely used.

PANSS uses a rating scale from 1 to 7 on each item. In this study, we only considered the items to be present ('yes') or not ('no'), as proposed by Kay et al. (1987) as a first step. All 30 items from PANSS were recorded. These items are organized on a positive scale (7 items), a negative scale (7 items) and a general psychopathology scale (16 items). The items represent symptoms or signs associated with psychotic disorders and will be called 'symptoms' hereafter.

The assessors were instructed to record the symptoms as 'yes' if it was described to be present in the defendant, as 'no' if it was not present in the

defendant, and as 'no information' if the symptom was not described at all in the relevant sections of the report.

2.3 The assessors

The three assessors in this study were consultant psychiatrists with long experience in clinical and forensic psychiatric practice. They were instructed to read only the parts of the reports describing the present mental status of the defendants and the parts describing the diagnostic and forensic evaluations (the parts 2 and 4 as described in 2.1). The assessors were instructed to assess the reports' descriptions of symptoms at the time of the alleged crime and at the time of the mental observation. This gave two different sets of data of recorded symptoms.

2.4 Statistical methods

Interrater reliability (agreement between the three assessors) was calculated with Gwet's AC_1 (Gwet, 2012). The degree of interrater reliability is usually assessed by Cohen's kappa or some of its variants (Kraemer et al., 2012). A kappa measure is to estimate the agreement beyond that for a chance. Fleiss' kappa is an extension of Cohen's kappa for the case of more than two assessors (Fleiss, 1971). However, both Cohen's kappa and Fleiss' kappa are influenced by the marginal distribution, for instance, when there is a high degree of agreement in one category. As Gwet's AC_1 does not have this undesirable property, it should be preferred when there are high rates of agreement (Wongpakaran et al., 2013).

Landis and Koch (1977) proposed a classification of kappa values that is applicable regardless of how the interrater agreement is quantified. A value less than 0.20 is considered slight agreement, a value between 0.21 and 0.40 is fair agreement, a value between 0.41 and 0.60 is moderate agreement, between 0.61 and 0.80 is substantial agreement, and finally, above 0.81 is considered almost perfect agreement. Kraemer et al. (2012) argue for a lower level of adequate agreement classifications, where the upper range is almost miraculous, values between 0.6 and 0.8 are very good, and values between 0.4 and 0.6 are the most realistic.

The percentage of symptoms from PANSS that was not described by the experts was reported, both at the time of the alleged crime and at the time of the mental observation ('no information').

The association between the Gwet's AC_1 measures and the percentage of undescribed symptoms was studied.

Data were analyzed using SPSS v 25 and Stata16. AgreeStat 2015.6 was used for the interrater reliability analysis (Advanced Analytics, Gaithersburg, MD).

2.5 Ethics

The Regional Ethics Committee for Medical Research Ethics in South-Eastern Norway Regional Health Authority (REC) has evaluated the pilot study outside the scope of the Health Research Act (2014/539). The Office of the Attorney General and the Council of confidentiality and research in the Ministry of Justice have approved the pilot study. The NMBF recommended access to its records to the Ministry of Justice. In accordance with the Public Administration Act § 13 d and Section 63 of the Courts Act, permission was given to inspect the reports. The Data Protection Officer at Oslo University Hospital has given its recommendation to the pilot project (case number 2014/7784).

No personally identifiable data on the defendants were registered. The statements and scoring forms were given a corresponding ID number, and the data are stored anonymized in Oslo University Hospital's research server.

3. Results

In [Table 1](#), the second and fifth columns show the percentage of symptoms from the PANSS instrument, which are not described at the time of the alleged crime (second column) and at the time of the mental observation (fifth column). The table also shows the agreement between the assessors in rating the symptoms from PANSS, calculated by percentage agreement scores (third and sixth columns) and interrater reliability scores (Gwet's AC_1) (fourth and seventh columns).

Interrater reliability of symptoms assessed at the time of the alleged crime varied from Gwet's score 0.295 (G12 Lack of insight) to 0.966 (N5 Difficulty in abstract thinking, N7 Stereotyped thinking, G1 Somatic concern). The interrater reliability at the time of the mental observation varied from 0.326 (P7 Hostility) to 0.822 (P4 Excitement).

[Table 2](#) shows mean and median interrater reliability scores, with range. There is, in general, a higher interrater reliability at the time of the alleged crime (median Gwet's AC_1 0.851) than at the time of the mental observation (median Gwet's AC_1 0.531).

[Table 2](#) also shows the mean and median percentage of symptoms not described for all PANSS scales, and for the positive, negative and general scales. We note that more symptoms from all three scales of PANSS are described at the time of the mental observation (median 29.2% not described) than at the time of the alleged crime (median 90.9% not described). Symptoms from the negative scale have a higher percentage of not being described than symptoms from a positive scale at both time points. At the time of the mental observation, 13.3% of the positive symptoms were not described, 23.3% of the negative symptoms and 35.8% of the general symptoms were not

Table 1. Percentage agreement, percentage not described, and Gwet's AC₁ at the time of the crime and time of the observation.

Item	Time of alleged crime			Time of mental observation		
	Percentage not described	Percentage agreement	Gwet's AC ₁	Percentage not described	Percentage agreement	Gwet's AC ₁
P1	53.3	76.7	0.683	13.3	76.7	0.665
P2	76.7	76.7	0.713	11.7	66.7	0.562
P3	65.0	80.0	0.732	11.7	66.7	0.526
P4	86.7	93.3	0.924	13.3	86.7	0.822
P5	80.0	83.3	0.835	51.7	66.7	0.524
P6	78.3	73.3	0.596	30.0	58.3	0.376
P7	86.7	76.7	0.697	35.0	51.7	0.326
N1	95.0	90.0	0.895	21.7	66.7	0.508
N2	90.0	83.3	0.797	21.7	58.3	0.384
N3	95.0	93.3	0.926	11.7	78.3	0.694
N4	91.7	83.3	0.819	23.3	71.7	0.591
N5	98.3	96.7	0.966	63.3	53.3	0.368
N6	96.7	95.0	0.948	26.7	56.7	0.369
N7	98.3	96.7	0.966	66.7	68.4	0.583
G1	98.3	96.7	0.966	73.3	68.3	0.603
G2	81.7	76.7	0.667	28.3	65.0	0.483
G3	98.3	96.7	0.966	51.7	46.7	0.248
G4	95.0	90.0	0.890	40.0	76.7	0.676
G5	96.7	93.3	0.931	38.3	71.7	0.624
G6	85.0	88.3	0.865	26.7	61.7	0.471
G7	93.3	93.3	0.924	20.0	80.0	0.717
G8	91.7	86.7	0.867	16.7	75.0	0.679
G9	63.3	60.0	0.475	26.7	68.3	0.528
G10	78.3	75.0	0.695	10.0	73.3	0.670
G11	95.0	90.0	0.890	31.7	75.0	0.636
G12	68.3	60.0	0.295	33.3	66.7	0.505
G13	91.7	85.0	0.837	61.7	60.0	0.450
G14	73.3	75.0	0.686	55.0	68.3	0.556
G15	95.0	90.0	0.895	45.0	66.7	0.522
G16	90.0	83.3	0.817	58.3	66.7	0.535

P1 = Delusions, P2 = Conceptual disorganization, P3 = Hallucinations, P4 = Excitement, P5 = Grandiosity, P6 = Suspiciousness/persecution, P7 = Hostility, N1 = Blunted affect, N2 = Emotional withdrawal, N3 = Poor rapport, N4 = Passive/apathetic, N5 = Difficulty in abstract thinking, N6 = Lack of spontaneity and flow of conversation, N7 = Stereotyped thinking, G1 = Somatic concern, G2 = Anxiety, G3 = Guilt feelings, G4 = Tension, G5 = Mannerisms and posturing, G6 = Depression, G7 = Motor retardation, G8 = Uncooperativeness, G9 = Unusual thought content, G10 = Disorientation, G11 = Poor attention, G12 = Lack of judgment and insight, G13 = Disturbance of volition, G14 = Poor impulse control, G15 = Preoccupation, G16 = Active social avoidance.

described. At the time of the alleged crime, 78.3% of the positive symptoms were not described, 95.0% of the negative symptoms, and 91.7% of the general symptoms, where all percentages are median values.

The associations between Gwet's index and the percentage of symptoms not described at the time of the alleged crime and at the time of the mental observation are shown in Figure 1. At the time of the alleged crime (white dots), the interrater reliability is higher when the symptoms are undescribed. At the time of the mental observation (red squares), no such clear pattern is seen. The association is more profound at the time of the crime ($p < 0.001$, $r = 0.80$) than at the time of the mental observation ($p = 0.068$, $r = -0.34$).



Table 2. Mean and median (range) for Gwet's AC₁ and percentage symptoms not described at the time of the alleged crime and at time of mental observation.

Item	Time of alleged crime			Time of mental observation		
	Mean	Gwet's AC ₁ Median (range)	Percentage symptoms not described Mean Median (range)	Mean	Gwet's AC ₁ Median (range)	Percentage symptoms not described Mean Median (range)
PANSS total	0.805	0.851 (0.295–0.966)	86.2 90.9 (53.3–98.3)	0.540	0.531 (0.248–0.822)	34.1 29.2 (10.0–73.3)
Positive scale	0.740	0.713 (0.596–0.924)	75.2 78.3 (53.3–86.7)	0.543	0.526 (0.326–0.822)	23.8 13.3 (11.7–51.3)
Negative scale	0.902	0.926 (0.797–0.906)	95.0 95.0 (90.0–98.3)	0.500	0.508 (0.368–0.694)	33.6 23.3 (11.7–66.7)
General scale	0.792	0.806 (0.295–0.966)	87.2 91.7 (63.3–98.3)	0.556	0.546 (0.248–0.717)	38.3 35.8 (10.0–73.3)

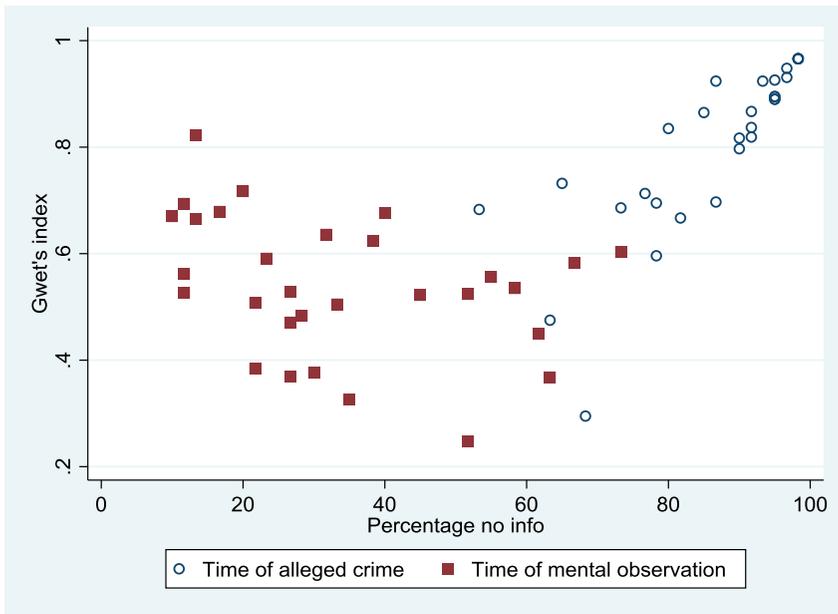


Figure 1. Association between Gwet's index and percent no information at the time of the alleged crime and at time of observation.

Figure 2 shows the main ICD-10 diagnoses given to the defendants in the 20 reports (blue columns), together with the forensic conclusions of the reports (red columns). One defendant was not given an ICD-10 diagnosis. As seen in the figure, a total of 12 reports (55%) concluded with legal insanity (forensically psychotic in the Norwegian legislation). Of these, 11 (92%) concluded with diagnoses in the psychotic spectrum (F2-chapter in ICD-10), and one concluded with a major depression with psychotic symptoms (F3-chapter in ICD-10).

4. Discussion

As we see in Table 2, all the mean and median Gwet's AC_1 scores lie in the range from moderate to almost perfect agreement according to Landis and Koch's classification. The individual items have large variations, as shown in Table 1. Of all 60 variables, no item has only slight agreement according to Landis and Koch's classification, seven items have fair agreement, while 53 items reach moderate, substantial and almost perfect agreement. We consider this to be a satisfactory agreement and conclude that our primary hypothesis was confirmed.

The interrater reliability depends on several factors. First, it depends on how the assessors read the reports and evaluate the written material

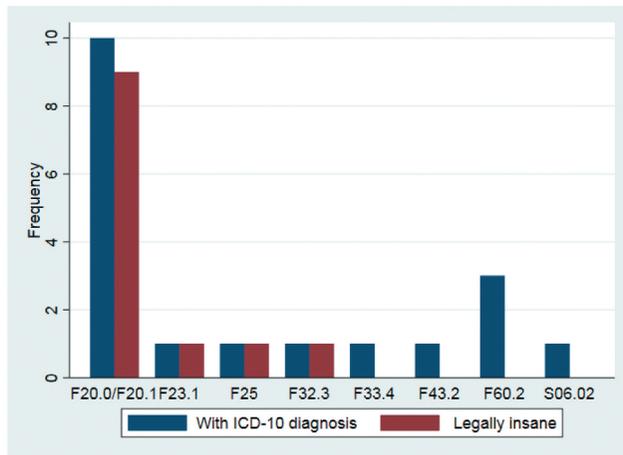


Figure 2. The main diagnoses in the reports (19 out of 20) together with the conclusion of legal insanity (12 out of 20 reports). F20.0 = Paranoid Schizophrenia, F20.1 = Hebephrenic schizophrenia, F23.1 = Acute polymorphic psychotic disorder with symptoms of schizophrenia, F25 = Schizoaffective disorder, F32.3 = Severe depressive episode with psychotic symptoms, F33.4 = Recurrent depressive disorder, currently in remission, F43.2 = Adjustment disorders, F60.2 = Dissocial personality disorder, S06.02 = Diffuse brain injury

compared to the symptom descriptions in PANSS. Some variation can be compensated for by securing the assessors have an almost equal understanding of the descriptions, by scoring several reports together. Second, the forensic experts' ability to write in a clear and concise way can also influence the assessors' ability to recognize the symptoms. Some descriptions can be ambiguous and be interpreted differently by the assessors. This is also partly compensated for by letting the assessors co-assess some reports.

The interrater reliability was high not only when the forensic experts described a symptom well but it was even higher when the experts did not describe the symptom at all. This was an unexpected finding.

Tables 1 and 2 show that a higher percentage of symptoms were not described at the time of the alleged crime than at the time of the mental observation. Certain symptoms were more likely not to be described in both time points (P5 Grandiosity, N5 Difficulty in abstract thinking, N7 Stereotyped thinking, G1 Somatic concern, G13 Disturbance of volition, G14 Poor impulse control).

Symptoms from the positive subscale are more often described than symptoms from the negative or general subscale. At the time of the crime, P1 Delusions, P3 Hallucinations, G9 Unusual thought content and G12 Lack of insight were most often described. At the time of the mental observation, G10 Disorientation, P2 Conceptual thought disorder, P3 Hallucinations, N3 Poor rapport and P1 Delusions were most often described.

The percentage of symptoms described at the time the mental observation and at the time of the alleged crime were both lower than expected, and lower at the time of the alleged crime.

The percentage of symptoms not described had a profound effect on the interrater reliability scores. The interrater reliability is highest when symptoms are less often described, but also high when they are very often described. There is an association between Gwet's index and symptoms with a high percentage not described at the time of the crime, and between Gwet's index and the symptoms with a low percentage not described at the time of the mental observation. This is shown in [Figure 1](#), where symptoms that are often described at the time of the mental observation have a high interrater reliability as have symptoms that are rarely described at the time of the crime.

The percentage of reports concluding with a forensic conclusion of legal insanity was 55% (12 out of 20). Of these, 11 concluded with a diagnosis in the psychotic spectrum, schizophrenic types, and one with an affective psychosis. The percentage with a conclusion of legal insanity for all murder and murder attempts for the year 2013 was 36% (personal communication from the secretary of the Norwegian Forensic Medical Board, 12.18.2019). This means that there are more persons evaluated as in an active psychotic state at the time of the crime in our study than in the national statistics of the year 2013. Because a high proportion of the reports in our study conclude with a diagnosis in the psychotic spectrum, it could be expected that more psychotic symptoms would be described both from the positive and negative scales of the PANSS in our sample than in a sample with fewer psychotic diagnostic conclusions. If this is the case, other samples could show even fewer symptoms described in the reports from both time points.

The high interrater reliability of the three assessors for symptoms that are seldom described indicates that the assessors perform equally well at distinguishing which symptoms are not described in the reports. This is an important result, as exploring which symptoms are not described by forensic experts and at what time, will be a main task in the full study.

The forensic experts in Norway are asked to evaluate the defendant's mental state at the time of the observation in addition to at the time of the alleged crime. The mental state at the time of the crime is determinative for the forensic conclusion, and psychotic states most often lead to legal insanity. Psychotic symptoms present at the time of the alleged crime are therefore important premises for the diagnostic and forensic conclusions.

The evaluation of the defendants' mental state at the time of the crime is always retrospective in nature, and the evaluation of symptoms present in the past may be difficult to assess. It is easier for the experts to describe symptoms at the present time than in a historic setting. It may be assumed that if a symptom is not described it is, or was, not present. However, there is

a possibility that the clinician has overseen or forgotten to report the symptom. In a diagnostic evaluation of psychosis, there should ideally not be any symptoms that are not described as either present or absent. In clinical practice, this is rarely the case.

For the conclusions in the reports to be open for review, the premises on which they are based must be clear to the reader. That psychotic symptoms are so often not described by the experts in forensic reports, is an alarming finding that needs to be explored further.

4.1 Weaknesses and strengths

One weakness in the study is the use of an instrument (PANSS) not designed for analyzing only written material. However, PANSS is very well established as a symptom scoring instrument in clinical practice as well as in scientific studies. We have increased interrater reliability by constructing a manual for recording the items from PANSS in the reports. Also, we have not found any other instrument or method better suited for analyzing psychotic symptoms in a large number of reports, as we intend to do in the main study.

Another objection to the pilot study might be that reports from 2013 are analyzed. There is a continuous work on improving the quality of forensic reports in Norway, with increased focus after the trial of the terror attacks 22 July 2011. Thus, it may be interesting to analyze newer reports and also comparing the quality of reports before and after this event. This will be a focus in the main study.

We have no access to the persons subject to forensic psychiatric examination. This means it is not possible to control the clinical validity of the forensic examinations, i.e., assessing whether the symptoms described really was present in the defendant or not.

The strengths of this study are first that we are studying the forensic experts' descriptions of psychotic symptoms, which to our knowledge have not been studied before. Another strength is that we study reports already written, so that we have access to the real practice of the forensic experts, making this a clinically relevant study. In addition, we study the clinical descriptions of the defendants, which make the study relevant for evaluations of criminal responsibility regardless of the nature of the insanity regulations in the jurisdiction where the report is written.

5. Conclusion

In this pilot study, we wanted to find a method to assess psychotic symptoms as they are described in forensic evaluations of criminal responsibility, as a means to explore the quality of symptom description. We studied 20 Norwegian forensic psychiatric reports from 2013 regarding persons accused of murder

or murder attempt. We hypothesized that it would be possible to obtain acceptable interrater reliability when experienced forensic psychiatrists assess forensic reports with the instrument PANSS, that the experts would describe the symptoms consistently in the reports, and we expected that interrater reliability would be higher when there were good symptom descriptions in the reports.

When we used PANSS as an instrument for describing relevant and important symptoms and signs of psychotic states in the reports, we found satisfactory interrater reliability between three experienced psychiatrists who assessed the reports independently. We also found a high proportion of symptoms not described in the reports, and that the agreement between assessors varied differently for the time of the crime and the time of the mental observation.

Our study indicates there might be an underreporting of psychotic symptoms from PANSS in forensic reports of criminal responsibility, in particular at the time of the crime. Negative symptoms, connected to the schizophrenic disorders that are most often linked to legal insanity, are even less often reported than positive or general symptoms.

We will select a total number of 500 reports regarding criminal responsibility from 5 different years, to see if we can find the same pattern of underreporting psychotic symptoms at the time of the crime. We have shown that agreement is especially high when identifying symptoms the experts do not describe in the reports. This will be used in the full study, where we will rate symptoms not described in reports, and compare symptom descriptions over the years, over the professions of the experts, and over the assessment instruments used by the experts.

The presence and the severity of psychotic symptoms are often determinative for the diagnostic and forensic conclusions in the reports and thereby for the final sentencing by the court. A reliable and testable symptom description open for external evaluation is very important but hasto our knowledge not been studied before. As the study gives an indication of underreporting of psychotic symptoms, our results may lead to suggested changes in the forensic experts' way of working.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Pia J. Løvgren  <http://orcid.org/0000-0002-3230-926X>

References

- Aboraya, A., Rankin, E., France, C., El-Missiry, A., & John, C. (2006). The reliability of psychiatric diagnosis revisited: The clinician's guide to improve the reliability of psychiatric diagnosis. *Psychiatry (Edgmont (Pa.: Township))*, 3(1), 41–50. <https://www.ncbi.nlm.nih.gov/pubmed/211349>
- Borum, R., & Grisso, T. (1996). Establishing standards for criminal forensic reports: An empirical analysis. *The Bulletin of the American Academy of Psychiatry and the Law*, 24(3), 297–317. <https://www.ncbi.nlm.nih.gov/pubmed/8889131>
- Cochrane, R. E., Grisso, T., & Frederick, R. I. (2001). The relationship between criminal charges, diagnoses, and psycholegal opinions among federal pretrial defendants. *Behavioral Sciences & the Law*, 19(4), 565–582. <https://doi.org/10.1002/bsl.454>
- Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. 76, 378–382. <https://doi.org/10.1037/h0031619>
- Fuger, K. D., Acklin, M. W., Nguyen, A. H., Ignacio, L. A., & Gowensmith, W. N. (2014). Quality of criminal responsibility reports submitted to the Hawaii judiciary. *International Journal of Law and Psychiatry*, 37(3), 272–280. <https://doi.org/10.1016/j.ijlp.2013.11.020>
- Gowensmith, W. N., Murrie, D. C., & Boccaccini, M. T. (2013). How reliable are forensic evaluations of legal sanity? *Law and Human Behavior*, 37(2), 98–106. <https://doi.org/10.1037/lhb0000001>
- Gowensmith, W. N., Sessarego, S. N., McKee, M. K., Horkott, S., MacLean, N., & McCallum, K. E. (2017). Diagnostic field reliability in forensic mental health evaluations. *Psychological Assessment*, 29(6), 692–700. <https://doi.org/10.1037/pas0000425>
- Gwet, K. L. (2012). *Handbook of inter-rater reliability: The definitive guide to measuring the extent of agreement among raters* (3rd ed. ed.). Advanced Analytics.
- Kacperska, I., Heitzman, J., Bak, T., Lesko, A. W., & Opio, M. (2016). Reliability of repeated forensic evaluations of legal sanity. *International Journal of Law and Psychiatry*, 44, 24–29. <https://doi.org/10.1016/j.ijlp.2015.08.028>
- Kay, S. R., Fiszbein, A., & Opler, L. A. (1987). The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophrenia Bulletin*, 13(2), 261–276. <https://doi.org/10.1093/schbul/13.2.261>
- Kois, L. E., & Chauhan, P. (2018). Criminal responsibility: Meta-analysis and study space. *Behavioral Sciences & the Law*, 36(3), 276–302. <https://doi.org/doi:10.1002/bsl.2343>
- Kraemer, H. C., Kupfer, D. J., Clarke, D. E., Narrow, W. E., & Regier, D. A. (2012). DSM-5: How reliable is reliable enough? *American Journal of Psychiatry*, 169(1), 13–15. <https://doi.org/10.1176/appi.ajp.2011.11010050>
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174. <https://doi.org/10.2307/2529310>
- Wongpakaran, N., Wongpakaran, T., Wedding, D., & Gwet, K. L. (2013). A comparison of Cohen's Kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: A study conducted with personality disorder samples. *BMC Medical Research Methodology*, 13(1), 61. <https://doi.org/10.1186/1471-2288-13-61>