

The Reliability of Evidence About Psychiatric Diagnosis After Serious Crime: Part I. Agreement Between Experts

Olav Nielsen, MB, BS, MCrim, Gordon Elliott, MB, BS, and Matthew Large, BSc (Med), MB, BS

The purpose of this study was to assess the extent of agreement on psychiatric diagnosis in written evidence provided by experts in serious criminal matters in Australia. We found good or very good inter-rater agreement on the diagnoses of acquired brain injury, schizophrenia-spectrum psychosis, substance-induced psychotic disorder, and intellectual disability. There was moderate agreement on the diagnosis of depressive and personality disorders. Agreement on anxiety disorders, in particular post-traumatic stress disorder, was poor. Agreement on the principal Axis I diagnosis was moderate, and there was a similar probability of agreement within pairs of experts engaged by the same side and those engaged by opposite sides. Concern about bias in expert psychiatric opinion in criminal cases appears to have been overstated. There was little evidence to suggest that experts' adversarial roles influenced their opinions on psychiatric diagnosis.

J Am Acad Psychiatry Law 38:516–23, 2010

The evidence presented by experts to support the presence of psychiatric disorders has been described as biased^{1,2} or based on an inadequate scientific standard.³ However, most of the criticism of the work of mental health expert witnesses has been anecdotal² and has focused on possible causes rather than empirical evidence of bias.

There are several potential sources of bias and perceived bias in adversarial legal proceedings. A lawyer may be aware of the opinions expressed by several experts in previous matters and may choose the expert who is most likely to support the client's case. The lawyer selects the documents to be examined and can instruct the expert to limit evidence to specific areas. The expert may feel sympathy for the

client and write a report that is helpful, and, in some jurisdictions, an unfavorable report does not have to be presented in court. Factors that reduce the likelihood of bias include the expert's duty to be truthful and impartial and the prospect of cross-examination that could harm the expert's reputation.

Expert evidence can be excluded from proceedings if it is shown to be biased or unreliable. If the *Daubert* standard, which requires that scientific evidence have a known error rate, is strictly applied, the absence of empirical studies demonstrating that a branch of science is reliable may affect the admissibility of expert evidence.^{4,5}

The reliability of the diagnosis is central to the science of psychiatry and to psychiatric evidence, because few psychiatric disorders have specific biological markers. Concern about the reliability of psychiatric diagnosis was the main reason for the adoption of specific definitions of mental disorders in the Diagnostic and Statistical Manual (DSM) classification system.^{6–9} The DSM-III and the International Classification of Diseases (10th edition; ICD-10) field trials used methods approximating ordinary clinical practice and reported satisfactory levels of agreement

Dr. Nielsen is Conjoint Senior Lecturer, University of New South Wales, and Psychiatrist, St. Vincent's Hospital, Sydney, NSW, Australia. Dr. Elliott is Psychiatrist, St. Vincent's Hospital, and New South Wales Justice Health, Sydney, NSW, Australia. Dr. Large is Conjoint Senior Lecturer, University of New South Wales, and Psychiatrist, The Prince of Wales Hospital, Sydney, Australia. Address correspondence to: Matthew Large, BSc (Med), MB, BS, The Euroa Centre, The Prince of Wales Hospital, Randwick, NSW, 2031, Australia. E-mail: mmbl@bigpond.com.

Disclosures of financial or other potential conflicts of interest: None.

on the diagnosis of severe psychiatric illnesses, such as schizophrenia and bipolar disorder.^{8,10} However, agreement on the definitions of common outpatient conditions, including anxiety disorders and the less severe forms of depression, was less satisfactory, and neither trial reported the level of inter-rater agreement on the new diagnosis of post-traumatic stress disorder (PTSD).^{8,10}

Subsequent studies of the inter-rater reliability or agreement on psychiatric diagnoses have involved structured and semistructured interviews; for example, the Structured Clinical Interview for DSM Disorders was used in the assessment of inter-rater reliability in the DSM-IV trials.¹¹ Although structured diagnostic interviews have been found to improve levels of inter-rater agreement,¹² they have not been widely used in forensic evaluations because of doubts about the validity of diagnoses generated when a list of symptoms is suggested to the person being evaluated.^{13,14} Hence, methods similar to those applied in forensic evaluations have seldom been used in research examining the reliability of psychiatric diagnoses.

In a study of reports about psychiatric injury after motor vehicle accidents, we found that experts' diagnoses generally favored the side that engaged them and that experts engaged by the same side were significantly more likely to agree on the presence of a psychiatric disorder.^{15,16} The levels of agreement on the diagnosis of PTSD and other anxiety and depressive conditions were generally unsatisfactory, with kappa values of between -0.12 and 0.38 .¹⁶ We were unable to locate any studies of the reliability of psychiatric diagnoses in reports prepared for criminal cases.

In the present study, we used methods similar to those in our study of civil cases¹⁶ to determine the extent of agreement on the psychiatric diagnoses made by experts engaged to examine clients in criminal matters. Considering our findings in the civil cases, we expected to find low levels of agreement on the diagnosis of anxiety and depressive disorders and a greater likelihood that experts engaged by the same side would agree about the psychiatric diagnosis than experts from opposite sides. Our first objective was to establish the level of agreement on diagnostic categories. We then tested two specific hypotheses:

First, that pairs of reports by experts from the same side are more likely to agree on the principal

diagnosis and that pairs of reports of experts from opposing sides are less likely to agree. An association between same side report pairs and agreement on the diagnosis suggests that bias arises from the experts' adversarial roles, and a lack of association suggests an absence of the effect of role.

Second, that pairs of reports by experts in the same profession (psychiatrist or psychologist) are more likely to agree on the principal diagnosis and that agreement is less likely in pairs of reports by experts in different professions. An association between same profession report pairs and agreement on the diagnosis suggests an effect of differences in the experts' training and experience, and a lack of association suggests an absence of effect of profession.

We also examined the association of age, sex, marital status, employment status, history of criminal convictions, and seriousness of the charges with agreement on the principal psychiatric diagnosis, because of the possibility that agreement on diagnosis is associated with the defendant's demographic or criminologic variables.

Methods

The Sample of Reports

Copies of reports from a consecutive series of 110 criminal cases concluded between 2005 and 2007, in which there were two or more reports written by psychiatrists, psychologists, or both, were made available by the Office of the Director of Public Prosecutions (ODPP) in New South Wales (NSW), Australia. There were 270 reports, 226 of which were written by 30 psychiatrists and 44 of which were submitted by 15 psychologists. Defense experts wrote 148 of the reports and prosecution experts wrote 122.

All the cases involved serious indictable offenses that were dealt with in the higher courts, including 30 charges of murder or attempted murder (homicide offenses), 35 of malicious wounding or serious assault, 14 of sexual assault, 12 of offenses against property, 10 of drug-related offenses, and 9 of other offenses, including fraud, kidnapping, arson, and firearms offenses.

Permission to perform the study was obtained from the NSW Justice Health Research and Ethics

Committee and the NSW Director of Public Prosecutions.

Data Collection

Data were collected on the following variables:

whether the expert was engaged by the defense or by the prosecution;

whether the expert was a psychiatrist or a psychologist;

the sex of the defendant and the age, marital status, and employment status at the time of the offense;

the most serious current charge and history of convictions;

and the psychiatric diagnoses.

All three authors independently rated reports from 28 cases, and there was no inter-rater disagreement on the experts' diagnoses or the other variables recorded. The remaining 82 cases were rated by M.L. and O.N. or by M.L. and G.E., with no disagreements.

Coding of the Diagnoses

Australian psychiatrists generally adhere to the DSM diagnostic system, and the diagnoses made by the expert witnesses in this study were generally consistent with DSM-IV diagnoses. The diagnoses were coded according to categories that were consistent with DSM-IV chapter headings, and methods described in earlier research⁸ were used in the analysis of agreement. Hence schizophrenia, schizophreniform disorder, schizoaffective disorder, delusional disorder, and psychosis not otherwise specified were grouped together for the analysis of agreement on the presence of schizophrenia-spectrum disorder. Major depressive disorder, dysthymia, and adjustment disorder with depressed mood were grouped in a depressive disorders category. PTSD was included with anxiety disorders such as obsessive-compulsive and panic disorders. Non-DSM categories included acquired brain injury, intellectual disability, and psychosis. The category of acquired brain injury included traumatic brain injury, alcohol-related brain damage, and the dementias. Intellectual disability was coded if the expert made a diagnosis of mental disability of any severity, including borderline low intelligence. A further analysis was conducted of cases that fell in the broader category of psychotic

disorders, which included schizophrenia-spectrum psychosis, substance-induced psychotic disorder (SIPD), psychotic depression, and mania, because the diagnosis of any of these disorders often results in a finding of reduced criminal responsibility.

Statistical Analysis

Analysis of multiple rating pairs for each subject presents a methodological dilemma arising from the number of degrees of freedom when three or more pairs are examined.¹⁶⁻¹⁸ The statistical methods used are described in our earlier study¹⁶ and include the use of generalized estimating equations (GEEs) to examine the association of the variables with agreement on the principal Axis I diagnosis.¹⁶

The kappa statistic^{19,20} was used to assess the level of agreement on the presence or absence of each diagnosis between experts on opposite sides and those on the same side. The ranges of kappa statistics showing agreement have been defined as poor, <0.2; fair, 0.2 to 0.4; moderate, 0.4 to 0.6; good, 0.6 to 0.8; and very good, 0.8 to 1.0.²¹ Nonparametric tests (the χ^2 test, or Fisher's exact test when the number of cases was less than five in any cell) were used to examine possible associations between the experts' roles and specific diagnoses, where agreement was rated as poor or fair.

Overall agreement in both the kappa and GEE analyses was determined by first applying a diagnostic hierarchy to the principal Axis I diagnosis similar to that described in a previous study.¹⁵ In this hierarchy, the diagnosis of acquired brain injury was considered first. If the diagnosis was made in one of the reports, the other report was examined for agreement or disagreement. If neither report contained a diagnosis of brain injury, the reports were examined for a diagnosis of schizophrenia-spectrum psychosis and then SIPD, other psychotic disorders (including in this study two cases of mania and one case of alcohol withdrawal delirium), depressive disorders, and anxiety disorders. Thus, the highest disorder in the hierarchy of acquired brain injury, schizophrenia spectrum psychosis, SIPD, other psychotic disorders, depressive disorders, and anxiety disorders was coded as the principal Axis I diagnosis. Substance dependence and abuse disorders were not considered in the determination of agreement on the principal Axis I diagnosis, in part because a proportion of the reports recorded a history of substance abuse in the body of the report but did not include a diagnosis of a sub-

stance abuse disorder in the conclusions. Axis II diagnoses of personality disorder were not considered in the GEE analysis of overall agreement because of the low number of cases and because a rating of agreement or nonagreement could be made on the basis of an Axis I diagnosis in every report pair. There were no pairs of reports with a disagreement on personality disorder without an accompanying disagreement on the Axis I diagnosis.

The use of GEEs allowed the examination of factors associated with agreement on the principal Axis I diagnosis among the complete sample of report pairs. We used univariate GEEs to examine the associations between the independent variables and the binary dependent variable of agreement on the principal Axis I diagnosis. The individual case was regarded as the subject variable, and the particular report pair was the within-subject variable. All factors were also included in a multivariate GEE main-effects model.

A power analysis of the proportions of these variables (for example, same or opposite expert role) and the dependent variable (agreement or disagreement on the principal Axis I diagnosis) was performed prospectively, to determine the sample size needed for the GEEs on the basis of the following assumptions: that report pairs in which there was an agreement on the Axis I diagnosis would be twice as likely to include reports written by experts from the same side and that there would be an equal number of defense and prosecution reports. We calculated that 108 report pairs were necessary for an 80 percent chance of finding a true association ($p < .05$) between agreement and report pairs with experts from the same side.²² The statistics were performed with SPSS for Windows, version 15.0.

Results

Pairs of reports by experts engaged by the prosecution and the defense were available in 105 of the 110 cases. In 60 cases, there were two reports, in 38 cases there were three reports, in 5 cases there were four reports, and in 2 cases there were five reports. Thus there were 224 unique pairs of reports: $60 + (3 \times 38) + (5 \times 6) + (2 \times 10)$. These comprised 147 report pairs by experts from opposite sides and 77 report pairs by experts engaged by the same side (Table 1). A psychiatric diagnosis equivalent to a DSM-IV diagnostic category was made, or a statement that no mental disorder was present, was re-

corded in every report. Most report writers explicitly used DSM-IV categories, and the criteria relied on to make the diagnoses were generally referred to in the conclusions.

There was a good or very good level of agreement on the presence of SIPD ($\kappa_{\text{same role}} = 0.795$, $\kappa_{\text{opposite roles}} = 0.819$) and the presence of intellectual disability ($\kappa_{\text{same role}} = 0.751$, $\kappa_{\text{opposite roles}} = 0.845$). The agreement on schizophrenia-spectrum psychosis ($\kappa_{\text{same role}} = 0.737$, $\kappa_{\text{opposite roles}} = 0.630$), personality disorder ($\kappa_{\text{same role}} = 0.555$, $\kappa_{\text{opposite roles}} = 0.607$), acquired brain injury ($\kappa_{\text{same role}} = 0.623$, $\kappa_{\text{opposite roles}} = 0.653$), and substance misuse ($\kappa_{\text{same role}} = 0.514$, $\kappa_{\text{opposite roles}} = 0.643$) was also classified as good.

There was moderate agreement on the diagnosis of depressive disorder ($\kappa_{\text{same role}} = 0.381$, $\kappa_{\text{opposite roles}} = 0.476$). Prosecution experts diagnosed depression on 19 occasions and defense experts on 28, a difference that did not reach statistical significance ($\chi^2 = 1.357$, $p = .241$). However, there was agreement in fewer than half (20/55) of the pairs of reports in which this diagnosis was raised.

Agreement on the diagnosis of anxiety disorders was poor ($\kappa_{\text{same role}} = 0.187$, $\kappa_{\text{opposite roles}} = 0.144$), PTSD was diagnosed on 16 occasions by defense experts and once by a prosecution expert (Fisher's exact test, $p < .001$). It was the most common anxiety disorder and was diagnosed by at least one expert in 26 of the 224 pairs of reports. There was one case in which a defense and prosecution expert agreed on the presence of PTSD and one other in which two defense experts agreed. In the remaining 24 pairs, only the defense expert made the diagnosis of PTSD. There was one case of obsessive-compulsive disorder ($\kappa = 1.0$ in both groups) and two cases of mania ($\kappa = 1.0$ in both groups). The diagnoses of pathological grief (three pairs), malingering (three pairs), and attention deficit-hyperactivity disorder (two pairs) all had a kappa of 0 in both groups.

There was good agreement on the presence of psychosis ($\kappa_{\text{same role}} = 0.802$, $\kappa_{\text{opposite roles}} = 0.702$) and moderate agreement on the principal Axis I diagnosis ($\kappa_{\text{same role}} = 0.491$, $\kappa_{\text{opposite roles}} = 0.453$).

Kappa values were generally not higher in the group of reports written by experts from the same side compared with pairs of reports by experts on opposite sides, indicating that psychiatric diagnosis was not influenced by the expert's adversarial role. There was no association between report pairs of ex-

Reliability of Diagnostic Evidence: Agreement Between Experts

Table 1 Agreement About the Psychiatric Diagnosis Between Experts From the Same and Opposite Sides

Role of the Expert Witnesses	Agree		Disagree	Kappa	95% CI
	Present	Absent			
Acquired brain injuries					
Same role	12	55	10	0.623	0.400–0.845
Opposite roles	14	121	12	0.653	0.493–0.814
Schizophrenia-spectrum psychoses					
Same role	21	47	9	0.737	0.514–0.959
Opposite roles	45	76	26	0.630	0.469–0.791
Substance-induced psychotic disorder					
Same role	22	48	7	0.795	0.572–1.00
Opposite roles	10	133	4	0.819	0.657–0.980
Depressive disorders					
Same role	6	59	13	0.381	0.154–0.603
Opposite roles	14	111	22	0.476	0.321–0.630
Anxiety disorders					
Same role	2	63	11	0.187	–0.037–0.411
Opposite roles	2	129	16	0.144	–0.0138–0.302
Substance dependence or abuse					
Same role	16	44	17	0.514	0.291–0.737
Opposite roles	45	77	25	0.643	0.481–0.805
Intellectual disability					
Same role	15	55	7	0.751	0.529–0.974
Opposite roles	20	121	6	0.845	0.684–1.00
Personality disorder					
Same role	2	72	3	0.555	0.355–0.755
Opposite roles	6	134	7	0.607	0.445–0.768
Principal Axis I diagnosis					
Same role	42	17	18	0.491	0.281–0.713
Opposite roles	74	35	38	0.453	0.296–0.609
Any psychosis*					
Same role	24	46	7	0.802	0.578–1.00
Opposite roles	61	64	22	0.702	0.542–0.862

* Diagnoses include schizophrenia-spectrum psychosis, SIPD, psychotic depression, mania, and delirium tremens.

perts from the same side and agreement on the principal Axis I diagnosis in the univariate or multivariate GEEs. Agreement on the principal diagnosis was significantly lower in homicides compared with other offenses, when examined with univariate and multivariate GEEs (Table 2). A *post hoc* examination of homicides suggested that this finding was largely due to disagreement on emerging schizophrenia-spectrum psychosis. There were 11 pairs of reports about homicides in which one expert had made a diagnosis of first-episode schizophrenia and the other had not. In these pairs, the other expert had diagnosed SIPD (three pairs), psychotic depression (three pairs), PTSD (one pair), cognitive disorder (one pair), prodromal schizophrenia (one pair), malingered psychosis (one pair), and no psychiatric disorder (one pair).

There was a lower probability of agreement on the principal diagnosis of married than of unmarried de-

fendants in the univariate analysis and a lower probability of agreement on male than on female defendants in the multivariate analysis.

Discussion

We found good agreement on the diagnosis of common psychiatric disorders among defendants in criminal matters. In contrast to the findings of our study of civil reports,¹⁶ there was little evidence of bias arising from whether the expert was engaged by the defense or by the prosecution. Agreement on the principal Axis I diagnosis was moderate, reflecting the narrower definition of inter-rater agreement used in our analysis.

We found low levels of agreement for the diagnosis of anxiety disorders, and to a lesser degree, depressive disorders. The higher level of agree-

Table 2 Generalized Estimating Equation Analysis of Factors Associated With Agreement Between Experts About the Principal Axis I Diagnosis

Factor	B	SE	Wald 95% CI		Hypothesis Test		
			Lower	Upper	Wald χ^2	df	P
Univariate							
Age	-0.003	0.0018	-0.006	0.001	2.035	1	0.154
Male	-0.112	0.0805	-0.270	0.046	1.937	1	0.164
Employed	-0.036	0.0953	-0.223	0.151	0.143	1	0.705
Married	-0.222	0.0946	-0.408	-0.037	5.530	1	0.019
Prior convictions	0.014	0.0956	-0.173	0.202	0.022	1	0.881
Homicide matter	-0.274	0.0871	-0.444	-0.103	9.867	1	0.002
Same expert role	-0.083	0.0550	-0.191	0.025	2.293	1	0.130
Same profession	0.059	0.0842	-0.106	0.224	0.485	1	0.486
Multivariate							
Age	0.000	0.0021	-0.003	0.005	0.226	1	0.635
Male	-0.203	0.0904	-0.026	-0.381	5.069	1	0.024
Employed	-0.017	0.799	-0.173	0.140	0.044	1	0.834
Married	-0.189	0.0995	-0.384	0.006	3.592	1	0.058
Prior convictions	-0.036	0.825	-0.198	0.126	0.191	1	0.662
Homicide matter	-0.304	0.0775	-0.456	-0.152	15.402	1	0.000
Same expert role	-0.046	0.0496	-0.143	0.051	0.860	1	0.356
Same profession	0.105	0.0798	-0.051	0.262	1.737	1	0.188
Intercept	-0.312	0.1439	0.030	0.594	0.4696	1	0.030

ment on psychotic illness may have been due to the presence of objective signs and a documented history of treatment in more of the defendants with those diagnoses than in those with diagnoses of PTSD and depression.

The diagnosis of PTSD is often of little relevance in criminal proceedings, because it usually does not reduce criminal responsibility or carry great weight in mitigation, unless the offending behavior was a direct consequence of some form of trauma. Hence, some experts may not have considered whether the accused had PTSD. However, disorders such as substance abuse and personality disorder are also rarely considered in trials and those diagnoses had moderate or good levels of inter-rater agreement. The poor level of agreement on the diagnosis of PTSD in a study that also found little evidence of bias and good levels of agreement in the diagnosis of other disorders casts further doubt on the reliability of the diagnosis of PTSD in medicolegal settings.^{23,24}

The first hypothesis, that experts from the same side are more likely to agree than experts from opposite sides, was not confirmed. The kappa statistics for the diagnosis of the principal Axis I diagnosis and schizophrenia-spectrum disorders were slightly higher in the pairs of reports from the same side than in those from opposite sides, but this finding was not the case for any other disorder, and same expert role

was not associated with agreement on the principal diagnosis in the GEEs.

The second hypothesis, that experts from the same profession are more likely to agree about the diagnosis than experts from differing professions, was not confirmed. Report pairs written by two psychiatrists were no more likely to agree on the diagnosis than were report pairs written by a psychiatrist and a psychologist.

There was some evidence that demographic factors and aspects of the charge itself were associated with agreement on the principal diagnosis. In the univariate analysis, there was less agreement on the principal diagnosis of married defendants, and in the multivariate analysis, there was less agreement on the psychiatric diagnosis of males. These findings should be interpreted cautiously because there were only 16 female defendants and because the association between being married and a lower level of agreement was not independent of other factors. Both univariate and multivariate test results showed a lower probability of agreement on the diagnosis in homicides. The apparent explanation for this finding is that patients with an established diagnosis of schizophrenia were more likely to face nonhomicide charges, whereas defendants in their first episode of psychosis, with no history of treatment for psychosis, committed 21 of the 30 homicide offenses. This finding

concur with that of Bourget and associates,²⁵ who noted the difficulty in arriving at a reliable diagnosis in cases of homicide in first-episode psychosis. However, associations of agreement on diagnosis with age, marital status, and the presence of homicide charges should be regarded with caution and may represent type I errors, because we were not examining a specific hypothesis and did not apply a statistical correction for multiple comparisons of demographic and criminologic variables.

The higher level of reliability of psychiatric diagnoses made by expert witnesses in criminal matters is in contrast with findings in our earlier study of expert reports in civil proceedings and may be due to the nature of disorders found among those charged with serious criminal offenses. Many of the defendants had severe and disabling forms of schizophrenia, a condition that is more reliably diagnosed than are nonpsychotic disorders.^{8,10,26} Another reason could be the way in which the experts were engaged by the lawyers in criminal matters. The prosecution experts were selected from a panel of experienced forensic psychiatrists, most of whom were still in clinical practice and who had also prepared reports for the defense in other cases. The experts chosen by the defense on the basis of their experience and reputation in court were also often on the prosecution panel of experts. By contrast, the experts in our study of civil matters were often retired from clinical practice and had become full-time expert witnesses, usually providing reports for just one side or even for just one of the insurance companies or firms specializing in personal injury. A further difference between criminal and civil cases is that experts in civil matters are less likely to face cross-examination.¹⁵

Another limitation of the finding of a satisfactory level of agreement between experts on major psychiatric disorders in this study is that levels of agreement on the absence of a mental disorder were higher than levels of agreement on the presence of a disorder. Although agreement on uncommon disorders relies on concurrence on their absence, this difference was also found in relation to the most common diagnosis, schizophrenia-spectrum psychosis.

Perhaps the most significant limitation of our assessment of agreement on psychiatric diagnosis resulted from the use of observational methods in the study. We were unable to establish whether any reports had been withheld by the defense, either because they were of no assistance or because they in-

cluded history that was prejudicial to the defendant's case. Moreover, although we were able to control for a loss of statistical independence in cases with three or more reports with the use of GEEs, even the diagnoses in cases in which there were just two reports might not always have been entirely independent, because the writers of the prosecution reports would frequently have been aware of the diagnoses made in the reports by the defense experts, and both report writers might have been aware of the earlier diagnoses made by other treating doctors. Furthermore, the reason for requesting more reports in some cases could have been the disagreement about the diagnosis between the initial reports. Additional reports would be expected to lower agreement in most cases.

Agreement could also have been reduced by some experts who omitted diagnoses that they knew to be present but did not consider relevant. For example, the diagnosis of a substance abuse disorder was often not made despite a history of harmful substance use in the body of the report. This omission could also have contributed to the lower levels of reliability in the diagnoses of some disorders, including PTSD.

The role of expert witnesses in Australia is similar to that of expert witnesses in other jurisdictions with adversarial legal systems, and the question they are usually asked to address is not the defendant's psychiatric diagnosis, but the level of criminal responsibility or competence to stand trial. Hence, although all of the reports in this study included an opinion about psychiatric diagnosis, agreement on the presence or absence of a disorder did not necessarily mean that there was agreement on competence to stand trial or the availability of a mental illness defense. However, in a related study of the reliability of expert opinion about legal issues in the same case sample with the same statistical methods, we found moderate levels of agreement between defense and prosecution experts on the availability of the defense of mental illness ($\kappa = 0.508$; 95% CI, 0.295–0.720) and fair levels of agreement regarding competence to stand trial ($\kappa = 0.293$, 95% CI, 0.134–0.451), with little evidence of bias arising from the expert's role.²⁷

Finally, although we examined a large consecutive sample of cases, the sample size did not provide sufficient statistical power to exclude the possibility that bias can sometimes influence the psychiatric diagnosis provided by experts. We showed that bias arising from whether the expert was engaged by the defense or prosecution in criminal proceedings did not have

a great influence on the diagnosis. The good level of agreement, particularly about the diagnosis of psychotic disorders, is also likely to exist in other jurisdictions that use similar methods for training psychiatrists and qualifying experts. However, the standard of written opinions may have improved in recent years in NSW because of the introduction of codes of conduct for expert witnesses, which require experts to state that any report is complete, truthful, and objective and that they are aware that their first duty is to assist the court.²⁸

Conclusions

The results of this study show that the psychiatric diagnoses made by expert witnesses in criminal matters are generally reliable. An exception appears to be the diagnosis of PTSD, either because it was not relevant to criminal proceedings, or because it is difficult, with the current diagnostic criteria, to make the diagnosis in a reliable way in legal settings.

Acknowledgments

We thank Nicholas Cowdrey, QC, New South Wales Director of Public Prosecutions, for permission to conduct the study; Craig Hyland, Office of the Director of Public Prosecutions, for assistance in collating the reports; and Peter Arnold for invaluable assistance in preparing the manuscript.

References

1. Beck M: The hired gun expert witness. *Mo Med* 91:179–82, 1994
2. Mossman D: "Hired guns," "whores," and "prostitutes": case law references to clinicians of ill repute. *J Am Acad Psychiatry Law* 27:414–25, 1999
3. Faust D, Ziskin J: The expert witness in psychology and psychiatry. *Science* 241:31–5, 1988
4. *Daubert v. Merrell Dow Pharmaceuticals*, 509 U.S. 579 (1993)
5. *Kumho Tire Company v. Carmichael*, 526 U.S. 137 (1999)
6. Beck AT: Reliability of psychiatric diagnoses: 1, a critique of systematic studies. *Am J Psychiatry* 119:210–16, 1962
7. Spitzer RL, Fleiss JL: A re-analysis of the reliability of psychiatric diagnosis. *Br J Psychiatry* 125:341–7, 1974
8. Spitzer RL, Forman JBW, Nee J: DSM-III field trials, I: initial inter-rater diagnostic reliability. *Am J Psychiatry* 136:815–7, 1979
9. American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders, Third Edition, Revised*. Washington, DC: American Psychiatric Association, 1987
10. Sartorius N, Ustun TB, Korten A, *et al*: Progress toward achieving a common language in psychiatry, II: results from the international field trials of the ICD-10 diagnostic criteria for research for mental and behavioral disorders. *Am J Psychiatry* 152:1427–37, 1995
11. Kilpatrick DG, Resnick HS, Freedy JR: The posttraumatic stress disorder field trial, in *Emphasis on Criterion A and Overall PTSD Diagnosis*. DSM-IV Sourcebook (vol 4). Edited by Widiger T, Frances A, Pincus H, *et al*. Washington, DC: American Psychiatric Press, 1998, pp 303–44
12. Miller PR: Inpatient diagnostic assessments, 2: interrater reliability and outcomes of structured vs. unstructured interviews. *Psychiatry Res* 105:265–71, 2001
13. Lees-Haley PR: Malingering mental disorder on the Impact of Event Scale (IES): toxic exposure and cancerphobia. *J Trauma Stress* 3:315–21, 1990
14. Nielssen O, Misrachi S: Prevalence of psychoses on reception to male prisons in New South Wales. *Aust N Z J Psychiatry* 39: 453–9, 2005
15. Large M, Nielssen O: An audit of medico-legal reports prepared for claims of psychiatric injury following motor vehicle accidents. *Aust N Z J Psychiatry* 35:535–40, 2001
16. Large M, Nielssen O: Factors associated with agreement between experts in evidence about psychiatric injury. *J Am Acad Psychiatry Law* 36:515–21, 2008
17. Mezzich JE, Kraemer HC, Worthington DR, *et al*: Assessment of agreement among several raters formulating multiple diagnoses. *J Psychiatr Res* 16:29–39, 1981
18. Fleiss JL: Measuring nominal scale agreement among many raters. *Psychol Bull* 76:378–82, 1971
19. Sim J, Wright CC: The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys Ther* 85:257–68, 2005
20. Shrout PE: Measurement reliability and agreement in psychiatry. *Stat Methods Med Res* 7:301–17, 1998
21. Landis JR, Koch GG: The measurement of observer agreement for categorical data. *Biometrics* 33:159–74, 1977
22. Fleiss JL: *Statistical Methods for Rates and Proportions* (ed 2). New York: John Wiley & Sons, 1981
23. Rosen GM, Spitzer RL, McHugh PR: Problems with the post-traumatic stress disorder diagnosis and its future in DSM V. *Br J Psychiatry* 192:3–4, 2008
24. Nielssen O, Large M: The future of post-traumatic stress disorder. *Br J Psychiatry* 192:394, 2008
25. Bourget D, Labelle A, Gagné P: First episode psychosis and homicide: a diagnostic challenge. *Can Psychiatr Assoc Bull* 36:6–9, 2004
26. Rosenman SJ, Korten AE, Levings CT: Computerised diagnosis in acute psychiatry: validity of CIDI-Auto against routine clinical diagnosis. *J Psychiatr Res* 31:581–92, 1997
27. Large M, Nielssen O, Elliott G: Reliability of evidence in serious criminal matters: fitness to stand trial and the defence of mental illness. *Aust N Z J Psychiatry* 43:446–52, 2009
28. Expert witness code of conduct, Supreme Court of NSW. Available at: http://www.accessmq.com.au/site_files/s1001/files/Expert_advice/Supreme%20Court%20Code%20of%20Conduct.pdf. Accessed October 26, 2010