

Factors Associated With Agreement Between Experts in Evidence About Psychiatric Injury

Matthew M. Large, BSc (Med), MB, BS, and Olav Nielssen, MB, BS

Psychiatrists and psychologists acting as expert witnesses in court cases are often accused of bias or error. We examined the level of agreement and factors influencing agreement between expert reports admitted into evidence during adversarial civil proceedings. The inter-rater reliability of the psychiatric diagnosis was examined in 51 pairs of civil medicolegal reports written by experts engaged by the same side and 97 pairs of experts engaged by opposite sides. Reports written by experts engaged by the same adversarial side had good agreement about the presence of a mental disorder ($\kappa = .74$) but had only fair agreement about the specific psychiatric diagnosis (average $\kappa = .31$). Reports written by experts engaged by opposing adversarial sides had poor agreement about the presence of any mental disorder and also the specific psychiatric diagnosis. Experts were more likely to agree about the presence of a mental disorder if the plaintiff was involved in a fatal accident. The agreement of treating doctors and experts was similar to that of pairs of experts.

J Am Acad Psychiatry Law 36:515–21, 2008

The role of advocates in adversarial proceedings is to represent their clients to the best of their ability. Advocates sometimes engage the services of expert witnesses to provide evidence in areas that are beyond the usual knowledge of the court. Unlike the advocate, the principal duty of an expert witness is to provide an accurate and unbiased opinion. The expert witness is required to resist pressure arising from efforts by the competing advocates to obtain the best possible outcome for their clients.

The possible causes of bias in the opinions of experts have been widely discussed.^{1–3} Advocates naturally select experts whose previous opinions are known to support a client's case. Other possible sources include the understandable wish to please the hiring party, the financial inducement of the prospect of further work, and the nature of the instructions and the selection of documents given to the expert witness by the lawyer. The conclusions of medicolegal assessments may also be influenced by the interaction between expert and plaintiff. Obvious examples are the desire of treating doctors not to

damage a therapeutic relationship,⁴ the sympathy evoked by a bereaved plaintiff, and the countertransference evoked in the expert during the assessment.⁵ Consideration of the consequences of the expert's opinion for the patient and the defendant may also have an influence on the conclusions.

Even without the potential sources of bias, the formulation of psychiatric opinion regarding the effect of trauma is a complex task that includes some subjective assessment. Most psychiatric disorders have no biological markers or objective signs, and the expert must assess whether symptoms are present and are severe enough to meet the accepted criteria to make a diagnosis. In addition to making a diagnosis, experts are often asked to comment on the likely cause of symptoms, the patient's level of disability, and the prognosis. Despite longstanding concerns about the reliability of medicolegal assessments after trauma, the reliability of assessments of disorders such as post-traumatic stress disorder (PTSD) has received little scientific study. The Diagnostic and Statistical Manual of Mental Disorders (DSM-III) field trials reported only a very modest level of agreement between independent assessors in the diagnosis of minor mood disorders and anxiety disorders, did not report the reliability of PTSD, and may not even have included any patients with this diagnosis.⁶

Dr. Large is in Private Practice, Sydney, Australia. Dr. Nielssen is in Private Practice, Sydney, and is a Psychiatrist, Clinical Research Unit for Anxiety Disorders, School of Psychiatry, University of New South Wales, St. Vincent's Hospital, Sydney, Australia. Address correspondence to: Matthew M. Large, BSc (med), MB, BS, PO Box 110, Double Bay, 1360, NSW, Australia. E-mail: mmbml@bigpond.com

PTSD was not one of the conditions studied in the ICD-10 field trials,⁷ and no attempt was made to establish whether PTSD could be distinguished from other psychiatric conditions in the DSM-IV trials.⁸

There are many studies that demonstrate the reliability of structured and semistructured interviews used in diagnosing depressive and anxiety disorders, including PTSD. There are also studies that show adequate inter-rater reliability of agreement between clinical diagnosis and structured interviews in some specialized settings.⁹⁻¹² However, the suitability of semistructured and structured interviews in forensic settings has been challenged, and the results of structured interviews rarely form the basis of experts' opinions.^{13,14}

When the existing uncertainties about the reliability of clinical diagnosis are considered in a legal context in which the patient, the legal representatives, and even the expert witness may have a stake in a particular outcome, it is not surprising that doubts have been expressed about the reliability of psychiatric diagnoses generated in assessments for the courts.¹⁵ In this retrospective study of a complete series of concluded claims for psychological injury after motor vehicle accidents, we examined the extent of agreement between experts and factors that may influence agreement.

For the purpose of the study, we assumed that any intrinsic unreliability in the specific diagnostic categories of psychiatric disorder is a source of random error. Thus if the experts were subject to this form of error, pairs of experts, irrespective of their roles, would be unlikely to agree about diagnosis. It was also assumed that the influences on the expert in favor of a particular outcome are a cause of systematic bias. If experts were subject to this form of bias then experts in the same role would be more likely to agree than experts from opposing sides.

The specific hypothesis was that experts from the same adversarial side would be more likely to agree about the psychiatric diagnosis than would experts from opposite sides. Our further hypotheses were that there would be higher levels of agreement about the presence of any mental disorder and the principal psychiatric diagnosis in pairs of reports by experts from the same profession (psychologist or psychiatrist), pairs of experts (rather than a pair made up of a treating doctor and an expert), and pairs of reports with respect to a plaintiff who was severely injured or bereaved.

Methods

The Sample of Reports

The reports used in the study have been described elsewhere¹⁶ and were provided to M.M.L. by the only law firm acting for the National Roads and Motorist's Association (NRMA), which at the time was a mutual society that provided insurance and other services to about half the motorists in the state of New South Wales, Australia. The files are the property of the NRMA (now called the Insurance Australia Group, or IAG), are held in secure storage by M.M.L., and will be returned for shredding in 2011. For the purpose of this study no identifying information was recorded, as the claims were identified with a number, and the experts were identified with a three-letter code. The electronic and paper records of the research are securely stored by M.M.L. The IAG has not sought any information from the study and has received copies of papers only after submission for publication to peer reviewed journals.

The reports were written about patients in a series of 559 consecutive third-party personal-injury claims. At the time, claims for personal injury after motor vehicle accidents were dealt with under common law, and most cases were settled by agreement between the plaintiff and defendant before a court hearing. In 67 claims, there were two or more reports written by a psychiatrist or a psychologist (excluding the reports of neuropsychologists). The reports consisted of all the psychiatric reports served on the defendant's lawyers by the plaintiffs and all the reports of assessments of the claimant by the defendant's experts.

Statistical Analysis

A κ statistic¹⁷ was used to measure inter-rater agreement about each specific psychiatric diagnosis, as it is an index of agreement about categorical variables that takes into account the differing prevalence of the conditions. The κ statistic is generally used to measure the level of agreement between two raters and generates a number between -1 and 1 , with 1 indicating agreement in all cases, 0 indicating the level of agreement that may be expected by chance, and -1 indicating disagreement in all cases. It was used to classify the level of agreement according to the following scores: 0 to $.2$, poor; $.2$ to $.4$, fair; $.4$ to $.6$, moderate; $.6$ to $.8$, good; and $.8$ to 1.0 , very good. There are other ways of measuring agreement be-

tween multiple raters,¹⁸ but these methods were unsuitable for this naturalistic study that had a varying number of raters per case. There is some controversy about the use of κ and the resultant scale of agreement (from poor to very good), and the statistic is also sensitive to the number of cases and the number of categories.^{19,20} However, the κ statistic was chosen to compare the inter-rater agreement between multiple sets of raters in the DSM-III,⁶ DSM-IV,⁸ and ICD-10⁷ field trials.

A contingency table ($df = 2$) was used for three group comparisons of categorical data. The injury severity (IS) score was compared by using the Kruskal-Wallis test, and the plaintiff's age was analyzed with a one-way ANOVA. All tests were in two-tailed form.

Univariate measures that were found to be significant in the comparison of groups of reports that agreed about the presence of mental disorder or the principal psychiatric diagnosis were included in two generalized estimating equations (GEEs). For the GEE, the individual claim was used as the subject variable, and the particular report pair was the within-subject variable. A GEE was used in preference to logistic regression, as it assumes observations made within identified data clusters (in this case, report pairs from the same claim) lack statistical independence. We used agreement about the presence of a mental disorder and agreement about the principal diagnosis as two binary dependent variables, with a logit link function, a binomial probability distribution, and an unstructured correlation matrix.

All the statistical analyses were performed with SPSS for Windows, version 15.0.

The Sample of Report Pairs

Of the 67 claims with two or more psychiatric reports, there were 42 claims with two reports, 16 with three, 8 with four, and 1 with five, making a total of 169 reports. Psychiatrists wrote 119 of the reports and psychologists wrote 50. Fifty-six of the reports were from experts engaged by the plaintiff, 68 were from the defendant's experts, and the remaining 45 were written by treating psychologists or psychiatrists.

There were 148 possible pairings of the 169 reports. Eighty-four of the reports appeared in one report pair, 48 reports were used in two report pairs, 32 in three report pairs, and 5 in four report pairs.

Of these, there were 73 report pairs in which both reports were written by psychiatrists, 17 in which

both were written by psychologists, and a further 58 in which one report was written by a psychiatrist and the other by a psychologist. In 77 report pairs, both of the reports were written by experts who were not involved in the patient's treatment, in 63 one report was written by an expert and the other by a treating practitioner, and in 8 both reports were written by treating practitioners.

The sample presented a methodological dilemma of whether to use the findings of all possible report pairs in the analysis, because it would overestimate the number of degrees of freedom in claims in which all the possible pairs of three or more reports were used. For example, if two of three pairs of reports about a plaintiff agreed about a specific diagnosis, then the third pair must logically also be in agreement. This problem could have been approached by including only an incomplete set of report pairs per claim, or by excluding any claims with three or more reports. Either method results in an arbitrary omission of some report pairs and would alter the degree of agreement reported. Hence, we used an approach that included all the possible report pairs.

This approach prevented the use of the κ statistic to calculate an overall measure of agreement among all 148 report pairs, as the conclusion with respect to psychiatric diagnosis could be logically deduced in 46 of 148 report pairs from the results of the other 102 report pairs, and hence almost a third of the observations would have lacked statistical independence. However, within the group of 97 pairs of results from opposite sides, there were no instances when agreement between reports could be deduced from other report pairs. Within the group of 51 report pairs from the same side, there were six claims with three reports written by experts engaged by the same side in which the agreement between any one pair of raters could be predicted by the findings of the other two pairs. An analysis of the κ statistics after excluding these cases found minimally changed levels of agreement about the specific psychiatric diagnosis (see the footnotes to Table 1).

The GEE is an appropriate way to analyze all 148 pairs, as it assumes a lack of independence within cases.

Diagnostic Variables

Each diagnosis from each report was recorded. There is no ideal way of measuring inter-rater agreement between raters making multiple diagnoses.^{21,22} In the κ analysis, we restricted the consideration of

Expert Agreement About Psychiatric Injury

Table 1 Kappa Statistics for Common Psychiatric Diagnoses After Motor Vehicle Accidents

Specific Psychiatric Diagnosis	Diagnosis in Any of 148 Report Pairs <i>n</i> (%*)	κ for Opposing Adversarial Role <i>n</i> = 97 Pairs	κ for Same Adversarial Role <i>n</i> = 51 pairs†
Traumatic brain injury	36 (12)	.31	.61
Post-traumatic stress disorder	103 (35)	.27	.26
Depressive disorders‡	88 (30)	-.12	.29
Other anxiety disorders	61 (21)	-.10	.38
Pain-related disorder	24 (8)	.26	.43
Adjustment disorder	25 (8)	.18	.19
Pathological grief	11 (4)	.18	.00
Any mental disorder (includes others, <i>n</i> = 5)	250 (84)	.09	.74
Average κ for all disorders	-	.14	.31

*Total exceeds 100% due to multiple diagnoses.

†Six of 51 report pairs lacked statistical independence. An analysis that excluded all the report pairs from claims with three reports from the same side found a minimally changed agreement for traumatic brain injury ($\kappa = .62$), post traumatic stress disorder ($\kappa = .31$), depression ($\kappa = .34$), other anxiety disorders ($\kappa = .34$), and any mental disorder ($\kappa = .68$).

‡Major depression, dysthymia and other depressive diagnoses.

the agreement to the specific psychiatric diagnosis, so that if one report contained a diagnosis of traumatic brain injury and an adjustment disorder and the second report in the pair had only a diagnosis of adjustment disorder, the result was coded as an agreement for adjustment disorder, but a disagreement for traumatic brain injury. Agreement between the earlier and later reports was measured when the reports were both from the same side.

In the GEE analysis, agreement was defined as agreement about the principal psychiatric diagnosis of the patient and agreement about the presence or absence of any mental disorder attributed to the accident. In these claims, we used a previously described diagnostic hierarchy based on an existing ranking of diagnoses of traumatic brain injury, PTSD, major depression, and other anxiety disorders and then adjustment disorders, bereavement, other disorders, and no diagnosis, to determine the principal diagnosis.¹⁶ PTSD was placed above major depression in this hierarchy, as PTSD was more frequently diagnosed than depression in this setting, and depression was often reported to be a complication of PTSD. Each report pair was rated 1 if the raters agreed about the principal psychiatric diagnosis (including no diagnosis), and/or 1 if they agreed about the presence or absence of mental disorder, and 0 in either category if they disagreed.

As a result, two measures of agreement were generated about the patient for use as dependent variables in the GEE: agreement about the presence of any mental disorder and agreement about the principal psychiatric diagnosis.

Expert Variables

The details of the expert were collected from each report. The profession of the author (psychiatrist or psychologist), the role (defendant's or plaintiff's expert; treating doctors were considered to be on the plaintiff's side), and the status as a court-recognized expert (expert versus treating practitioner) was recorded for each case. For the GEE, each report pair was scored 1 in each of these variables if they were the same and 0 if they differed. These pair-wise measures of agreement were within-subject variables in the GEE.

Plaintiff Variables

The patient's injuries were recorded using the injury severity (IS) score, an instrument for classifying the overall severity of multiple injuries. It classifies the injuries in bodily regions on a six-point scale from no injury to lethal injury and is scored by calculating the sum of the square of the three most severe injuries, making a maximum score of 75 for injuries that are not immediately lethal. In an acute medical setting, the IS score is correlated with mortality,²³ and, applied retrospectively to medical files, it predicts the most acute setting of medical care (in intensive care, a general ward, emergency department, or outpatient treatment) and the number of subsequent days in the hospital.²⁴ The age, sex, and whether another person was killed were also used as independent variables in the GEE.

The 67 plaintiffs had an average age of 36 (SD 15) years, 37 were male, and 13 had been involved in an accident in which there was a fatality. The median IS

Table 2 Variables in the Principal Diagnosis Groups

Independent Variables	No Agreement <i>n</i> = 30	Agree About Mental Disorder <i>n</i> = 65	Agree on Principal Diagnosis <i>n</i> = 53		<i>p</i>
Report writer factors					
Same expert status, <i>n</i> (%)	16 (53)	39 (69)	30 (57)	$\chi^2 = .40$.82
Same adversarial side, <i>n</i> (%)	3 (10)	24 (37)	24 (45)	$\chi^2 = 10.87$.004
Same profession, <i>n</i> (%)	21 (70)	37 (57)	32 (60)	$\chi^2 = 1.48$.48
Plaintiff factors					
Male, <i>n</i> (%)	18 (60)	38 (58)	27 (51)	$\chi^2 = 1.93$.75
Age, mean years (SD)	38.9 (7.3)	34.3 (7.0)	32.6 (8.1)	$F = 1.72$.18
IS score, median (IQR)	12 (16)	12 (21)	12 (21)	KW = .70	.70
Fatal accident, <i>n</i> (%)	0 (0)	16 (25)	14 (26)	$\chi^2 = 9.62$.008

score was 12, indicating that most plaintiffs sustained moderate or severe physical injuries. The characteristics of plaintiffs who were the subject of three or more psychiatric or psychological reports did not differ significantly from those who were the subject of only two reports.

Results

Reports written by experts from the same adversarial side had good agreement about the presence of any mental disorder ($\kappa = .74$) but had only fair agreement about the specific diagnosis (average $\kappa = .31$). The agreement between experts who were employed by the same side was good for traumatic brain injury, but was in the poor or fair range for the other common diagnoses. There was poor agreement about the presence of any mental disorder ($\kappa = .09$) and also the specific psychiatric diagnosis (average $\kappa = .14$) in report pairs written by experts from opposing adversarial sides (Table 1).

There were only six instances in which there were two experts for the defendant for a single claim, although all the reports written at the request of the defendants' lawyers were available to the researchers. In these six cases, one pair agreed that no diagnosis was warranted, five pairs agreed about the presence of mental disorder, but no pair agreed on the principal psychiatric diagnosis. It is likely that the sample of treating psychiatrist reports was also quite complete, as the insurance company generally paid for treatment before litigation only if the treating practitioner provided a report. All eight pairs of reports by treating practitioners agreed that the plaintiff had a mental disorder, but only four pairs agreed about the principal diagnosis.

In 53 report pairs, the report writers agreed about the principal psychiatric diagnosis, and in 30, they

did not agree about the presence of a mental disorder. In the remaining 65 pairs, the report writers agreed that the plaintiff had a mental disorder but did not agree about the principal psychiatric diagnosis.

Report pairs that did not agree about the presence of a mental disorder were more likely to be written by writers from opposite adversarial sides about a plaintiff who had not been in a fatal accident (Table 2).

The GEE suggested that pairs of reports with writers engaged by the same side and those written about a plaintiff involved in a fatal accident were more likely to agree that the plaintiff had a mental disorder and pairs of reports from the same adversarial side were also more likely to agree about the principal psychiatric diagnosis (Table 3).

Conclusions

The results of this study should be interpreted with caution because of the relatively small sample size and the likelihood that some plaintiff's reports that did not support the claim that the plaintiff sustained a psychiatric injury were not tendered. Hence, if a plaintiff's lawyer received an opinion that the plaintiff did not have a psychiatric disorder, the lawyer was free to seek a further opinion and was not required to disclose the contents of the earlier report. If a second report found that the plaintiff had a psychiatric disorder, it may not have agreed with the report of the defendant's expert, whereas the report that was not served on the defendant may well have agreed with the report of the defendant's expert. We were unable to find out how many plaintiff's reports were not served. The withholding of reports that did not support the plaintiff's case may have decreased the level of agreement between experts from opposite adversarial sides, but should have increased the level of agreement between pairs of experts engaged by the

Expert Agreement About Psychiatric Injury

Table 3 GEE Examining Factors Associated With Agreement Between Experts

Variable	B	SE	95% Wald CI	Wald χ^2	df	p
Agreement about the presence of mental disorder						
Intercept	-4.807	1.262	-7.281 to -2.332	14.494	1	.000
Same adversarial side	2.157	.862	.468 to 3.846	6.268	1	.012
Fatal accident*	1.835	.926	0.20 to 3.650	3.927	1	.048
Agreement about the principal psychiatric diagnosis						
Intercept	-.253	.4167	-1.070 to .564	.368	1	.544
Same adversarial side	.989	.447	.112 to 1.865	4.882	1	.027
Fatal accident	-.168	.424	-1.00 to .663	.158	1	.691

*A conservative estimate of fatal accidents was calculated by the inclusion of a false positive in the group with no agreement about mental disorder, to prevent having a zero in the denominator.

plaintiff. However, we found a low level of agreement about the actual diagnosis by experts from the same adversarial side.

A further limitation of the study is that more than half the experts who performed the assessments may have been aware of an earlier psychiatric opinion. Awareness of a previous diagnosis usually increases inter-rater reliability, but in an adversarial legal setting, it may have the opposite effect and may have contributed to the lack of agreement between the experts in this study, including the lack of agreement between experts engaged by the same side.

We found poor agreement about both the presence of any mental disorder and all the specific psychiatric diagnoses in reports by experts engaged by opposite sides. Experts from the same adversarial side usually agreed that a mental disorder was present, and being on the same side was the most important predictor of agreement between experts about the presence of a specific psychiatric diagnosis. This finding suggests that even if the experts were not biased, the reports that were eventually relied on in litigation contained predictable opinions about the presence or absence of psychiatric disorder.

We also found evidence of a significant level of error in making psychiatric diagnoses, as there was only modestly greater agreement about the diagnosis of common outpatient psychiatric disorders in reports from the same adversarial side when compared with those from opposite sides. The inter-rater reliability of the most common diagnosis, PTSD, in reports written by experts on the same adversarial side was only fair.

The diagnosis of PTSD presents particular difficulties in medicolegal settings. The disorder is defined as a consequence of trauma, even though the

causal relationship between the traumatic event and subsequent symptoms may be the main matter before the court. Moreover, PTSD has few objective features and is a relatively new diagnostic category, for which the diagnostic criteria have been revised several times. We observed that some experts elicited a similar history but differed in their diagnoses because they disagreed about whether the plaintiff's experiences were sufficiently traumatic. Another reason for disagreement may have been the use of idiosyncratic diagnostic criteria in an attempt to avoid prompting the patient to report the symptoms or to exercise caution in relying on the DSM-III and DSM-IV criteria alone in medicolegal settings. However, the low level of agreement between experts reported in this study may not have been entirely due to the medicolegal context, as the κ statistics for the diagnosis of anxiety and depressive disorders were similar to the modest levels of agreement reported in the DSM-III field trials.⁶

The hypothesis that report writers would be more likely to agree about the presence of a mental disorder in those involved in fatal accidents was confirmed, but the seriousness of the plaintiff's overall injuries was not associated with increased agreement between experts. The level of agreement with respect to the psychiatric effects of fatal accidents suggests that feelings of sympathy evoked in the report writers may be a source of bias in expert opinions. Notably, all the experts who examined plaintiffs who were involved in fatal accidents diagnosed psychiatric disorders. However, the diagnoses varied considerably among both plaintiff's and defendant's experts, and the range of diagnoses applied to bereaved patients was similar to that of the diagnoses of nonbereaved plaintiffs.

The hypothesis that treating practitioners and experts would be less likely to agree than two experts or two treating practitioners was not confirmed. This result suggests that the theoretical concerns about the tendency of treating doctors to provide reports that favor the patients under their care can be overstated.

The study does provide some evidence of both error and bias in written evidence about psychiatric diagnosis. In general, if both error and bias are known to be present, minimization of bias should be the priority, as a biased but accurate measure is always wrong, whereas an objective but inaccurate opinion is sometimes correct. Measures to reduce bias obviously include the use of experts appointed and instructed by the court, disclosure of all reports including reports that are unhelpful to the plaintiff, and codes of conduct reinforcing the expert's primary duty to the court. However, the findings of this study do not support the reliance on court-appointed experts, as all experts may still be subject to bias arising from their perception of the wishes of the court and their subjective views of the plaintiff. Moreover, the degree of error in diagnosis suggests that it would be unwise to replace the current system with one in which psychiatric opinion is not tested by comparison with those of other experts.

Although this was a relatively small study that considered only subjects involved in motor vehicle accidents in a single jurisdiction, the similarities in civil litigation and psychiatric practice in English-speaking countries make it relevant to other jurisdictions. To the authors' knowledge, it is the only study of its kind, despite the very large number of psychiatric reports prepared for courts each year around the world. The findings suggest the need for more research, not only to improve the reliability of psychiatric evidence, but also to establish the reliability of the diagnosis of minor psychiatric disorders in clinical settings.

Acknowledgments

The authors acknowledge Mr. Victor Kelly, formerly of Abbott Tout Lawyers, Sydney, for making the files available for study; and Professor Daniel Shuman, Southern Methodist University, Dallas, TX; Professor Chris Tennant, University of Sydney; and Dr. Tim Slade, University of New South Wales, for their assistance.

References

1. Beck M: The hired gun expert witness. *Mo Med* 91:179–82, 1994
2. Shea P. *Psychiatry in Court* (ed 2). Sydney, Australia: Hawkins Press, 1996

3. Gutheil TG, Commons M, Miller PM: Withholding, seducing, and threatening: a pilot study of further attorney pressures on expert witnesses. *J Am Acad Psychiatry Law* 29:336–9, 2001
4. Greenburg SA, Shuman DW: Irreconcilable conflict between therapeutic and forensic roles. *Prof Psychol Res Pract* 28:50–7, 1997
5. Groves JE: Taking care of the hateful patient. *N Engl J Med* 298:883–7, 1978
6. Spitzer RL, Forman JBW, Nee J: DSM-III field trials I. Initial interater diagnostic reliability. *Am J Psychiatry* 136:815–7, 1979
7. 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
8. Kilpatrick DG, Resnick HS, Freedy JR, *et al*: The posttraumatic stress disorder field trial: 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.
9. Andrews G, Peters L: The psychometric properties of the Composite International Diagnostic Interview. *Soc Psychiatry Psychiatr Epidemiol* 33:80–8, 1998
10. Peters L, Andrews G: Procedural validity of the computerised version of the Composite International Diagnostic Interview (CIDI-Auto) in the anxiety disorders. *Psychol Med* 25:1269–80, 1995
11. Komiti A, Jackson HJ, Judd FK, *et al*: A comparison of the Composite International Diagnostic Interview (CIDI-Auto) with clinical assessment in diagnosing mood and anxiety disorders. *Aust NZ J Psychiatry* 35: 224–30, 2001
12. 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
13. Lees-Haley PR, Dunn JT: The ability of naive subjects to report symptoms of mild brain injury, post-traumatic stress disorder, major depression, and generalized anxiety disorder. *J Clin Psychol* 50:252–6, 1994
14. Niessen O, Misrachi S: Prevalence of psychoses on reception to male prisons in New South Wales. *Aust NZ J Psychiatry* 39: 453–9, 2005
15. Campbell T: Challenging the evidentiary reliability of DSM-IV. *Am J Forensic Psychol* 17:47–68, 1999
16. Large M, Niessen O: An audit of medico-legal reports prepared for claims of psychiatric injury following motor vehicle accidents. *Aust NZ J Psychiatry* 35:535–40, 2001
17. Landis JR, Koch GG: The measurement of observer agreement for categorical data. *Biometrics* 33:159–74, 1977
18. Fleiss JL: Measuring nominal scale agreement among many raters. *Psychol Bull* 76:378–82, 1971
19. Sim J, Wright CC: The κ 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. Mezzich JE, Kraemer HC, Worthington DR, *et al*: Assessment of agreement among several raters formulating multiple diagnoses. *J Psychiatr Res* 16:29–9, 1981
22. Spitzer RL, Cohen J, Fleiss JL, *et al*: Quantification of agreement in psychiatric diagnosis: a new approach. *Arch Gen Psychiatry* 17:83–7, 1967
23. Baker SP, O'Neill B, Haddon W, *et al*: The Injury Severity Score: a method for describing patients with multiple injuries and evaluating emergency care. *J Trauma* 14:187–96, 1974
24. Large MM: Relationship between compensation claims for psychiatric injury and severity of physical injuries from motor vehicle accidents. *Med J Aust* 175:129–32, 2001