Citation: Transl Psychiatry (2013) 3, e214; doi:10.1038/tp.2012.133

www.nature.com/tp

Concordance of psychiatric symptom ratings between a subject and informant, relevancy to post-mortem research

PM Thompson[1], CG Bernardo[1], DA Cruz[1], NS Ketchum[2] and JE Michalek[2]

Investigators are interested in determining whether lifetime behavioral traits and specific mood states experienced close to death affect brain gene and protein expression as assessed in post-mortem human brains. Major obstacles to conducting this type of research are the uncertain reliability of the post-mortem psychiatric diagnoses and clinical information because of the retrospective nature of the information. In this study, we addressed the concordance of clinical information obtained through an informant compared with information obtained through a clinician interview of the subject. To test this, we measured both lifetime and within the week psychiatric symptoms of subjects ($n = 20$) and an informant, their next-of-kin ($n = 20$) who were asked identical questions. We found Diagnostic and Statistical Manual (DSM)-IV axis 1 diagnoses by Mini-International Neuropsychiatric Interview proportion of positive agreement for major depression was 0.97, bipolar disorder was 0.81, whereas proportion of negative agreement was 0.97 for schizophrenia. Symptom scale intra-class correlation coefficients and 95% confidence interval were: Bipolar Inventory of Signs and Symptoms = 0.59 (0.23, 0.81), Brief Psychiatric Rating Scale = 0.58 (0.19, 0.81), Hamilton Depression Rating Scale = 0.44 (0.03, 0.72), Montgomery Asberg Depression Rating Scale = 0.44 (0.03, 0.72), Young Mania Rating Scale = 0.61 (0.30, 0.82), Barratt Impulsiveness Score = 0.36 (− 0.11, 0.70) and Childhood Trauma Questionnaire = 0.48 (− 0.15, 0.83). We show that DSM-IV diagnoses; lifetime impulsivity severity, childhood trauma score and symptom scores were significantly consistent between the subjects and their informants. These data suggest, with some limitations, that both retrospective and informant obtained information can provide useful clinical information in post-mortem research.
Translational Psychiatry (2013) **3,** e214; doi:10.1038/tp.2012.133; published online 15 January 2013

Introduction

Neuropathological discoveries of the early 1900s[1,2] identified gross and cellular neuropathology changes with the classical degenerative diseases. Despite extensive research no single neuropathological signature has been found with the mental illnesses. Although there may be no gross or cellular neuropathology with mental illnesses, the issue of molecular neuropathology remains a question. With improvements in the level of investigative technology there has been renewed interest in identifying this pathology with some success. Both schizophrenia and bipolar disorder gene expression analyses have identified changes in genes encoding mitochondrial[3] and synaptic proteins.[4–6] However, evidence of molecular neuropathology that is consistently replicated in different cohorts is lacking in the field. One hindrance to achieving additional insights on molecular pathology is the limited post-mortem research being conducted on mental illness, most likely due to the scarcity of available tissue. An additional confound is the lack of reliable clinical information, which is important to interpreting the meaning of the biological results. In clinical psychiatric research, there are many well-validated clinical instruments to measure a wide variety of psychiatric symptoms. The same cannot be said of instruments used to collect retrospective information. It is difficult to address this issue because the nature of data collection relies, in part, on information obtained by an informant that is retrospective and subject to the vagaries of memory, as well as the closeness of relationship between subject and informant. Thus, obtaining accurate informant descriptions of lifetime psychiatric diagnoses, behavioral traits and clinical symptoms for donors of post-mortem brain tissue can be extremely difficult.

The goal of this project was to better understand if individuals who know a subject well and are often the informants for post-mortem brain research can accurately describe the mood state and identify lifetime psychiatric symptoms of their family member. To accomplish this goal, we interviewed a subject with a known axis 1 diagnosis and an informant who was their next-of-kin (NOK) using the same diagnostic and symptom severity scales. The goal was to determine the level of concordance of answers between subject and NOK pairs of well-established clinical instruments.

Materials and methods

All research was approved by the University of Texas Health Science Center Institutional Review Board and was

[1]Department of Psychiatry, University of Texas Health Science Center San Antonio, San Antonio, TX, USA and [2]Department of Epidemiology and Biostatistics, University of Texas Health Science Center San Antonio, San Antonio, TX, USA
Correspondence: Dr PM Thompson, Department of Psychiatry, Psychiatry-Southwest Brain Bank, UTHSCSA-University of Texas Health Science Center San Antonio, 7703 Floyd Curl Drive, San Antonio, TX, 78229, USA.
E-mail: thompsonp@uthscsa.edu
Keywords: BISS; HAM-D; informant interview; MADRS; next-of-kin interview; post-mortem; psychological autopsy
Received 13 September 2012; accepted 10 November 2012

performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki. The interview with the subject occurred in person in a University office and the NOK interview was conducted either by telephone or in person. Subject recruitment was from the patient Mood Disorders Clinic at the University of Texas Health Science Center San Antonio or via advertisement. The inclusion criteria for subject participation were (1) a psychiatric diagnosis of bipolar disorder 1, major depression or schizophrenia and (2) a NOK who had regular contact with them and was willing to participate in the research. In the first set of 10 subject–NOK pairs, PMT interviewed the subject and CGB interviewed the NOK. For the second 10 sets, the interviewers were reversed.

All subjects and informants were administered the following instruments; Mini International Neuropsychiatric Interview (MINI),[7] Barratt Impulsiveness Scale,[8] Childhood Trauma Questionnaire (CTQ),[9] Montgomery Asberg Depression Rating Scale,[10] Hamilton Depression Rating Scale, 31 question (Ham-D31),[11,12] Brief Psychiatric Rating Scale (BPRS)[13] and the Bipolar Inventory of Symptoms Scale (BISS).[14] With the seven NOK who were not married or engaged to the subjects, we did not ask sex-related questions. One subject did not complete the BPRS, and two subjects did not complete the Barratt Impulsiveness Scale. Nine NOK did not have knowledge of childhood events in the subjects and did not complete the CTQ.

Statistics. Based on the MINI, we categorized subjects and informants according to psychiatric diagnosis and we assessed diagnostic agreement. We reported results as proportions of either positive or negative agreement[15,16] with 95% confidence intervals.[17] That is, with a 95% level of confidence the range of values contained the 'true' proportion. As a result of the limited sample, we combined the alcohol abuse and alcohol-dependent results into alcohol use disorder, and the drug abuse and drug dependence into drug use disorder. We analyzed summative scores for the seven symptom severity scales administered to both subject and informant. We assessed agreement between subject and informant for the BPRS, BISS, BISS subscales, Montgomery Asberg Depression Rating Scale, Ham-D, Young Mania Rating Scale, CTQ and Barratt Impulsiveness Scale using intra-class correlation coefficients (ICC) and 95% confidence intervals.[18,19] We additionally assessed

Table 1 Demographic information on subjects and NOK

	Subject	*NOK*	*Total*
Race, N *(%)*			
Anglo	14 (70%)	11 (55%)	25 (62.5%)
Hispanic	6 (30%)	7 (35%)	13 (32.5%)
Black	0 (0)	2 (10%)	2 (5%)
Total	20	20	40
Sex, N *(%)*			
Female	12 (60%)	10 (50%)	22 (55%)
Male	8 (40%)	10 (50%)	18 (45%)
Total	20	20	40
Subject relation to NOK, N *(%)*			
Parent	3 (15%)		
Sibling	3 (15%)		
Spouse	12 (60%)		
Ex-husband	1 (5%)		
Caregiver	1 (5%)		
Total	20 (100%)		

Abbreviation: NOK, next-of-kin.

Table 2 Proportion of positive and negative agreement between subjects and NOK with 95% confidence interval using the MINI semistructured exam for DSM-IV diagnoses

			Subject		*Positive agreement (95% CI)*	*Negative agreement (95% CI)*
			Yes	*No*		
Major depression	NOK	Yes	19	0	0.97 (0.89, 1)	0
		No	1	0		
Bipolar disorder	NOK	Yes	13	1	0.81 (0.65, 0.92)	0.25 (0.06, 0.59)
		No	5	1		
Schizophrenia	NOK	Yes	0	1	0	0.97 (0.89, 1)
		No	0	19		
Panic Disorder	NOK	Yes	2	0	0.67 (0.29, 0.92)	0.94 (0.82, 0.99)
		No	2	16		
Panic disorder with agoraphobia	NOK	Yes	1	4	0.18 (0.04, 0.47)	0.69 (0.51,0.83)
		No	5	10		
Generalized anxiety disorder	NOK	Yes	0	3	0	0.82 (0.67, 0.92)
		No	3	14		
OCD	NOK	Yes	0	1	0	0.97 (0.89, 1)
		No	0	19		
PTSD	NOK	Yes	0	1	0	0.92 (0.8, 0.98)
		No	2	17		
Alcohol use disorder	NOK	Yes	1	2	0.25 (0.06, 0.59)	0.81 (0.65, 0.92)
		No	4	13		
Drug use disorder	NOK	Yes	2	0	0.8 (0.37, 0.98)	0.97 (0.87, 1)
		No	1	17		
Adult ADHD	NOK	Yes	0	0	0	0.95 (0.84, 0.99)
		No	2	18		

Abbreviations: ADHD, attention defect, hyperactivity disorder; CI, confidence interval; DSM, Diagnostic and Statistical Manual; MIMI, Mini-International Neuropsychiatric Interview; NOK, next-of-kin (informant), OCD, obsessive compulsive disorder; PTSD, post-traumatic stress disorder.

agreement between mean scale scores for subjects and informants using two one-sided equivalence testing.[20]

Results

Table 1 shows the demographic information. The subjects were 30% Hispanic and 70% Anglo with 60% female and 40% male. Although in the NOK, 35% were Hispanic and 55% Anglo. Table 2 shows the proportion of positive and negative agreement between subject and informant based on responses to the MINI. The diagnostic positive agreement values range from 0.25 for alcohol use disorder to 0.97 with major depression. Diagnostic negative agreement ranged from 0 with major depression to 0.97 with obsessive compulsive disorder.

Agreement between subject and informant symptom severity scale scores is shown in Table 3. Concordance rates were as follows: BISS = 0.59 (0.23, 0.81), Ham-D = 0.44 (0.03, 0.72), Young Mania Rating Scale = 0.0.61 (0.26, 0.82), Montgomery Asberg Depression Rating Scale = 0.0.44 (0.03, 0.72), BPRS = 0.58 (0.19, 0.81), CTQ = 0.48 (− 0.15, 0.83) and Barratt Impulsiveness Scale = 0.36 (− 0.11, 0.70). Subdividing the BISS into its factor components[21] showed moderate concordance, with the mania, irritability and anxiety factors showing the greatest concordance and the depression factor the least (Table 3). We found the Barratt Impulsiveness Scale mean scores and the BPRS mean scores for subject and informant statistically equivalent ($P = 0.005$ and $P = 0.02$, respectively).

Discussion

We report that informant-gathered information on individuals with a major mental illness can identify most severe lifetime Diagnostic and Statistical Manual (DSM) diagnosis. The notable diagnostic exceptions are generalized anxiety disorder, agoraphobia without panic disorder, alcohol abuse, alcohol dependence and drug dependence, for which there was a moderate level of disagreement between the subject and NOK. Psychiatric symptoms experienced in the last week and childhood trauma scores were concordant between the subject and his or her NOK. Although the ICC for the Barratt Impulsiveness Scale was low, the mean scores were found to be statistically equivalent.

The reliability of psychiatric diagnoses in living individuals generated by a variety of instruments has been demonstrated by the SCID-I (Structured Clinical Interview for Axis-1),[22] SCID-II (Structured Clinical Interview for Axis-2),[23] MINI[7] and Diagnostic Interview for Genetic Studies.[24] However, there is very limited information regarding the reliability of retrospective diagnoses, especially as they apply to post-mortem research. The general approach to establish post-mortem psychiatric diagnoses includes a review of medical records and conducing a psychological autopsy about the decedent with the NOK (Table 4). Sundqvist *et al.*[25] reported a kappa coefficient of agreement for diagnoses solely from chart review between the ante and post-mortem diagnoses ranging from 0.35 for schizoaffective disorder to 0.95 with major depression. The inclusion of an interview with the NOK, in addition to the review of medical records, increases the information reliability across diagnostic classifications.

Table 3 Intra-class correlation coefficients with 95% confidence intervals between the subject and NOK for BISS, BPRS, MADRS, Ham-D, YMRS, CTQ, BISS illness subscale and Barratt Impulsivity Scale

Instrument	*Subject*	*NOK*	*Intra-class correlation coefficient (95% confidence interval)*
Ham-D			
N	20	20	
Mean (s.d.)	17.2 (12.5)	12.1 (11.6)	0.44 (0.03, 0.72)
Min, max	0, 50	0, 47	
MADRS			
N	20	20	
Mean (s.d.)	11.4 (9.4)	9 (10.2)	0.44 (0.03, 0.72)
Min, max	0, 33	0, 45	
YMRS			
N	20	20	
Mean (s.d.)	7.5 (7.3)	6 (8.9)	0.61 (0.26, 0.82)
Min, max	0, 26	0, 40	
BPRS			
N	19	19	
Mean (s.d.)	24.5 (8.7)	24.8 (8.6)	0.58 (0.19, 0.81)
Min, max	9, 48	18, 56	
CTQ			
N	14	11	
Mean (s.d.)	56.6 (18.2)	48.7 (13.8)	0.48 (− 0.15, 0.83)
Min, max	30, 102	21, 76	
BISS total score			
N	20	20	
Mean (s.d.)	33.8 (21.4)	24.4 (23.9)	0.59 (0.23, 0.81)
Min, max	6, 87	2, 112	
BISS score factors			
Mania			
N	20	20	
Mean (s.d.)	8.7 (7.9)	7 (8.7)	0.66 (0.34,0.85)
Min, max	0, 31	0, 39	
Depression			
N	20	20	
Mean (s.d.)	9.5 (7)	7 (7.2)	0.28 (− 0.16, 0.62)
Min, max	0, 26	0, 22	
Irritability			
N	20	20	
Mean (s.d.)	4.5 (3.9)	3.5 (3.7)	0.57 (0.2, 0.8)
Min, max	0, 14	0, 15	
Anxiety			
N	20	20	
Mean (s.d.)	3 (3.1)	2.1 (3.6)	0.46 (0.05, 0.74)
Min, max	0, 10	0, 15	
Barratt Impulsiveness Scale			
N	18	19	
Mean (s.d.)	73.9 (15.7)	76.2 (15.8)	0.36 (− 0.11, 0.70)
Min, max	47, 96	43, 112	

Abbreviations: BISS, Bipolar Disorder Inventory of Symptoms Scale; BPRS, Brief Psychiatric Rating Scale; CTQ, Childhood Trauma Questionnaire; Ham-D, Hamilton Depression Rating Scale; MADRS- Montgomery Asberg Depression Rating Scale, NOK, next-of-kin (informant); YMRS, Young Mania Rating Scale.

Most research of this type relies on using a semistructured information gathering process to organize medical and psychological autopsy material. The two common ones are the Diagnostic Interview After Death,[26,27] Diagnostic Instrument for Brain Studies[28] and their variants.[29,30]

Deep-Scoboslay *et al.*,[31] Kelly and Mann[32] and Lehrmann *et al.*[33] used SCID-P (axis 1) and the SCID-II with either DSM-

Table 4 Previously published reliability assessments for post-mortem diagnoses

Authors	*Sundqvist* et al.[25]	*Roberts* et al.[29]	*Hill* et al.[30]	*Keilp* et al.[27]	*Deep-Soboslay* et al.[31]	*Lehrmann* et al.[33]	*Kelly and Mann*[32]	*Schneider* et al.[34]	*Zhang* et al.[35]
Comparison groups	Medical record diagnosis to post-mortem structured case review	Police reports of diagnoses to DEAD plus	ME diagnoses to DEAD plus	Medical record diagnoses to consensus post-mortem diagnoses	Psychological autopsy to last available psychiatric record	Case history (ME records and NOK interview) to toxicology results	Psychological autopsy to medical record	Subject to informant interviews	Subject to informant interview
Study instrument	Item group checklist for the schedules for clinical assessment in neuropsychiatry DIBS	DEAD for DSM-IV	DEAD for DSM-III-R	DEAD for DSM-III-R	SCID for DSM-IV	NIDA-PSAAQ, SCID for DSM-IV	SCID-P and SCID-II for DSM-III-R	SCID-I, SCID-II for DSM-IV	SCID-1 for DSM-III-R
Statistic	Kappa coefficient	% Agreement	Kappa coefficient	Kappa coefficient	Kappa coefficient	% Agreement	% Agreement	Kappa coefficient	Kappa coefficient
Psychiatric diagnoses									
Schizophrenia	0.61	69.6%, (Kappa between DSM-III-R and DSM-IV = 0.57)	0.59	0.74	Combined		100	NA	
Schizoaffective disorder	0.35			Combined other psychotic disorders, 0.40	0.94		67		
Delusional disorder							100 ($n = 1$)		
Major depression	0.95			Combined mood disorders,0.79	0.63		88	Combined mood disorders (excluding dysthymia), 0.79 (includes adjustment disorders), 0.79	1
Bipolar disorder	0.70				0.58		50		
Anxiety disorders									0.66
Cocaine dependence						42			
Cannabis dependence						76	Combined, substance disorder 100 ($n = 1$)		
Phencyclidine						20			
Opioids						44			
Alcohol dependence or abuse				0.77			100		1
All axis 1				0.57			Kappa = 0.84	0.65	
All axis 2				0.32			Kappa = 0.65		

Abbreviations: DIBS, Diagnostic Instrument for Brain Studies; DEAD, Diagnostic Evaluation After Death; DSM, Diagnostic and Statistical Manuel of Mental Disorders; ME, medical examiner; NA, not applicable; NIDA, PSAAQ-National Institute on Drug Abuse-Post-Mortem Substance Abuse Assessment Questionnaire; NOK, next-of-kin; SCID, Structured Clinical Interview for DSM-IV.

III-R and DSM-IV criteria. They combined this information with antemortem data organized through the Diagnostic Interview After Death and found the instruments demonstrate good reliability when compared with medical records. This study also shows good reliability of informant information for a majority of diagnoses. Because our sample was limited to primary diagnoses of mood disorders, the reliability determination of the other diagnoses such as schizophrenia was incomplete. For example, three subjects endorsed generalized anxiety disorder symptoms and two post-traumatic disorder symptoms but these symptom sets were not observed by the NOK. The subject–NOK interview provided the greatest discordance in the alcohol use disorders with four subjects reporting misuse but not by the NOK. This is consistent with the clinical experience of patients often under reporting their drinking. There was higher concordance with drug use, but the frequency of any positive response was low with only three NOK or subject reported misuse. Lehrmann *et al.*[33] looked at substance misuse in a post-mortem sample identified by medical examiner records, NOK interviews and toxicology. They showed that when medical records and toxicology data are combined, the detection rate drastically increases. Clearly, increasing the number of sources of information allows for greater reliability for all diagnoses. Two other studies looked at the concordance between psychiatric diagnosis generated by an informant compared with that of a subject and found high concordance.[34,35] Schneider *et al.*[34] found kappa correlation coefficients for mood disorders at 0.79, anxiety at 0.79 and any personality disorder at 0.92, which are comparable to our findings. Zhang *et al.*[35] also found high concordance with SCID-based diagnosis and also conducted a Ham-D with a subject and two informants. They found that the results were significantly correlated (Spearman's rho = 0.57). Their results had substantially higher concordance than we report and this is most likely because they used two informants for each subject.

Genetic and family studies using the family history method also collect and utilize informant-based information. Rougemont-Buecking *et al.*[36] in a large well-designed study showed fair to good agreement between a family member and direct interview for panic disorder and obsessive compulsive disorder, whereas poor agreement was seen with overall anxiety disorder and generalized anxiety disorder.[36] Mendlewicz *et al.*[37] reported an agreement kappa of 0.5 to identify affective disorders between a direct psychiatric interview and probands recollection.[37] Gershon and Guroff[38] reported kappa's for bipolar disorder 1 = 0.63, major depression = 0.42, whereas for bipolar disorder 2 and schizoaffective disorder the kappa = 0.[38] One possibility that our values showed greater agreement than the genetic studies is that all of our subjects were long-term psychiatric patients with family that were knowledgeable of their medical history. In this report, we show that the MINI can also provide an accurate psychiatric diagnosis, and can be completed in a shorter amount of time in comparison with the SCID. This is important as it limits the intrusiveness of the NOK interview.

In this work, we attempted to simulate a typical NOK post-mortem interview with clinic patients and their NOK to see if the NOK were aware of the severity of psychiatric symptoms and mood state in the subjects. The ICC of all the scales ranged from 0.66 to 0.44 with the exception of the BISS depression subscale (0.28) and Barratt Impulsiveness Scale (0.36). Using Landi and Koch[39] interpretation of the similar kappa, these scores showed at least a 'moderate' level of agreement. Although Barratt Impulsiveness Score had a poor level of agreement by ICC, the mean scores were found to be statistically equivalent.

This study gathered retrospective information obtained by an informant. The reliability of this type of data has several areas of potential confounds. Most NOK under report symptoms and when interpreting the results, care must be given to who the informant is. For example, parents may not be aware of their children's alcohol/drug use nor of their sexual drive and many spouses may not have detailed history of the other spouse's childhood abuse. This study uses a small sample size of non-randomly selected subjects; even so, our results are similar to other reports using a variety of instruments.[34,35,40] Because our focus was on mood disorders not all DSM axis 1 diagnoses were encountered and we were unable to report positive agreement data for the diagnoses of: obsessive compulsive disorder, post-traumatic stress disorder, schizophrenia and adult attention deficit hyperactivity disorder. Additional work is needed to study the concordance of these disorders and understand how long after death a NOK can provide reliable mood symptom ratings.

Overall, we show in a group with severe mental illness that an informant interview of the NOK can provide useful information, which can be used to better analyze post-mortem biological information. There are significant caveats to reliable post-mortem data collection: (1) the interviewer must have extensive experience conducting clinical interviews, (2) the informant must have regular contact with the subject, (3) multiple informant interviews should be conducted if available and (4) the psychometric instruments used must be geared toward clinically obvious symptom levels.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements. This work was supported in part by grants: DOD W81XWH-07-1-0244 and R24MH07603901 to PMT.

1. Holdorff B. Friedrich Heinrich Lewy (1885-1950) and his work. *J Hist Neurosci* 2002; **11**: 19–28.
2. Alzheimer A. The Early Story of Alzheimer's Disease. *Translation of the Historical Papers by Alois Alzheimer*. Raven Press:: New York, 1987.
3. Iwamoto K, Bundo M, Kato T. Altered expression of mitochondria-related genes in postmortem brains of patients with bipolar disorder or schizophrenia, as revealed by large-scale DNA microarray analysis. *Hum Mol Genet* 2005; **14**: 241–253.
4. Mirnics K, Middleton FA, Lewis DA, Levitt P. Analysis of complex brain disorders with gene expression microarrays: schizophrenia as a disease of the synapse. *Trends Neurosci* 2001; **24**: 479–486.
5. Scarr E, Gray L, Keriakous D, Robinson PJ, Dean B. Increased levels of SNAP-25 and synaptophysin in the dorsolateral prefrontal cortex in bipolar I disorder. *Bipolar Disord* 2006; **8**: 133–143.
6. Sequeira A, Klempan T, Canetti L, ffrench-Mullen J, Benkelfat C, Rouleau GA *et al.* Patterns of gene expression in the limbic system of suicides with and without major depression. *Mol Psychiatry* 2007; **12**: 640–655.
7. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E *et al.* The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a

structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J Clin Psychiatry* 1998; **59**: 22–33.
8. Patton JH, Stanford MS, Barratt ES. Factor structure of the Barratt impulsiveness scale. *J Clin Psychol* 1995; **51**: 768–774.
9. Bernstein DP, Fink L, Handelsman L, Foote J, Lovejoy M, Wenzel K *et al.* Initial reliability and validity of a new retrospective measure of child abuse and neglect. *Am J Psychiatry* 1994; **151**: 1132–1136.
10. Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. *Br J Psychiatry* 1979; **134**: 382–389.
11. Hamilton M. Development of a rating scale for primary depressive illness. *Br J Soc Clin Psychol* 1967; **6**: 278–296.
12. Williams JB. Standardizing the Hamilton Depression Rating Scale: past, present, and future. *Eur Arch Psychiatry Clin Neurosci* 2001; **251**(Suppl 2): II6–12.
13. Overall JE, Gorham DR. The brief psychiatric rating scale. *Psychol Rep* 1962; **10**: 799–812.
14. Bowden CL, Singh V, Thompson P, Gonzalez JM, Katz MM, Dahl M *et al.* Development of the bipolar inventory of symptoms scale. *Acta Psychiatr Scand* 2007; **116**: 189–194.
15. Cicchetti DV, Feinstein AR. High agreement but low kappa: II. Resolving the paradoxes. *J Clin Epidemiol* 1990; **43**: 551–558.
16. Spitzer RL, Fleiss JL. A re-analysis of the reliability of psychiatric diagnosis. *Br J Psychiatry* 1974; **125**: 341–347.
17. Brown LD, Cai T, DasGupta A. Interval estimation for a binomial proportion. *Stat Sci* 2001; **16**: 101–133.
18. Fleiss JL. Block design: application to an interexaminer reliability study. *The Design and Analysis of Clinical Experiments*. John Wiley and Sons: New York, 1986 291–305.
19. Lu L, Shara N. *Reliability Analysis: Calculate and Compare Intra-class Correlation Coefficients (ICC) in SAS*. NorthEast SAS Users Group, 2007.
20. Schuirmann DJ. A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *J Pharmacokinet Biopharm* 1987; **15**: 657–680.
21. Thompson PM, Gonzalez JM, Singh V, Schoolfield JD, Katz MM, Bowden CL. Principal domains of behavioral psychopathology identified by the Bipolar Inventory of Signs and Symptoms Scale (BISS). *Psychiatry Res* 2010; **175**: 221–226.
22. Skre I, Onstad S, Torgersen S, Kringlen E. High interrater reliability for the Structured Clinical Interview for DSM-III-R Axis I (SCID-I). *Acta Psychiatr Scand* 1991; **84**: 167–173.
23. Zanarini MC, Skodol AE, Bender D, Dolan R, Sanislow C, Schaefer E *et al.* The Collaborative Longitudinal Personality Disorders Study: reliability of axis I and II diagnoses. *J Pers Disord* 2000; **14**: 291–299.
24. Nurnberger JIJ, Blehar MC, Kaufmann CA, York-Cooler C, Simpson SG, Harkavy-Friedman J *et al.* Diagnostic interview for genetic studies. Rationale, unique features, and training. NIMH Genetics Initiative. *Arch Gen Psychiatry* 1994; **51**: 849–859.
25. Sundqvist N, Garrick T, Bishop I, Harper C. Reliability of post-mortem psychiatric diagnosis for neuroscience research. *Aust N Z J Psychiatry* 2008; **42**: 221–227.
26. Zalcman S, Endicott J, Clayton PJ, Winokur G. *Diagnostic Evaluation After Death (DEAD)*. National Institure of Mental Health, Neuroscience Research Branch: Rockville MD, 1983.
27. Keilp JG, Waniek C, Goldman RG, Zemishlany Z, Alexander GE, Gibbon M *et al.* Reliability of post-mortem chart diagnoses of schizophrenia and dementia. *Schizophr Res* 1995; **17**: 221–228.
28. Keks NA, Hill C, Opeskin KO, Copolov DL, Dean B. Psychiatric diagnosis after death: the problems of accurate diagnosis from case hisotry review and relative interviews. In: Dean B, Kleinman JE, Hyde TM eds. *Using CNS Tissue In Psychiatric Research*. Harwood Academic Publishers: Amsterdam, 1999, pp 19–37.
29. Roberts SB, Hill CA, Dean B, Keks NA, Opeskin K, Copolov DL. Confirmation of the diagnosis of schizophrenia after death using DSM-IV: a Victorian experience. *Aust N Z J Psychiatry* 1998; **32**: 73–76.
30. Hill C, Keks N, Roberts S, Opeskin K, Dean B, MacKinnon A *et al.* Problem of diagnosis in postmortem brain studies of schizophrenia. *Am J Psychiatry* 1996; **153**: 533–537.
31. Deep-Soboslay A, Akil M, Martin CE, Bigelow LB, Herman MM, Hyde TM *et al.* Reliability of psychiatric diagnosis in postmortem research. *Biol Psychiatry* 2005; **57**: 96–101.
32. Kelly TM, Mann JJ. Validity of DSM-III-R diagnosis by psychological autopsy: a comparison with clinician ante-mortem diagnosis. *Acta Psychiatr Scand* 1996; **94**: 337–343.
33. Lehrmann E, Afanador ZR, Deep-Soboslay A, Gallegos G, Darwin WD, Lowe RH *et al.* Postmortem diagnosis and toxicological validation of illicit substance use. *Addict Biol* 2008; **13**: 105–117.
34. Schneider B, Maurer K, Sargk D, Heiskel H, Weber B, Frölich L *et al.* Concordance of DSM-IV Axis I and II diagnoses by personal and informant's interview. *Psychiatry Res* 2004; **127**: 121–136.
35. Zhang J, Conwell Y, Wieczorek WF, Jiang C, Jia S, Zhou L. Studying Chinese suicide with proxy-based data: reliability and validity of the methodology and instruments in China. *J Nerv Ment Dis* 2003; **191**: 450–457.
36. Rougemont-Buecking A, Rothen S, Jeanpretre N, Lustenberger Y, Vandeleur CL, Ferrero F *et al.* Inter-informant agreement on diagnoses and prevalence estimates of anxiety disorders: direct interview versus family history method. *Psychiatry Res* 2008; **157**: 211–223.
37. Mendlewicz J, Fleiss JL, Cataldo M, Rainer JD. Accuracy of the family history method in affective illness. Comparison with direct interviews in family studies. *Arch Gen Psychiatry* 1975; **32**: 309–314.
38. Gershon ES, Guroff JJ. Information from relatives. Diagnosis of affective disorders. *Arch Gen Psychiatry* 1984; **41**: 173–180.
39. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; **33**: 159–174.
40. Conner KR, Duberstein PR, Conwell Y. The validity of proxy-based data in suicide research: a study of patients 50 years of age and older who attempted suicide. I. Psychiatric diagnoses. *Acta Psychiatr Scand* 2001; **104**: 204–209.