Psychological Medicine, 2006, **36**, 69–80. © 2005 Cambridge University Press
doi:10.1017/S0033291705005969 First published online 29 September 2005. Printed in the United Kingdom

The diagnostic interview for psychoses (DIP): development, reliability and applications

D. J. CASTLE[1], A. JABLENSKY[2]*, J. J. McGRATH[3], V. CARR[4], V. MORGAN[2], A. WATERREUS[2], G. VALURI[2], H. STAIN[2], P. McGUFFIN[5] AND A. FARMER[5]

[1] *University of Melbourne and Mental Health Research Institute, Melbourne, Australia/School of Psychiatry and Clinical Neurosciences, The University of Western Australia, Perth, Australia*; [2] *The University of Western Australia, Perth, Australia*; [3] *Queensland Centre for Mental Health Research and University of Queensland, Brisbane, Australia*; [4] *Centre for Mental Health Studies and University of Newcastle, Newcastle, Australia*; [5] *MRC Social, Genetic and Developmental Psychiatry Research Centre, Institute of Psychiatry, King's College London, UK*

ABSTRACT

Background. We describe the development, reliability and applications of the Diagnostic Interview for Psychoses (DIP), a comprehensive interview schedule for psychotic disorders.

Method. The DIP is intended for use by interviewers with a clinical background and was designed to occupy the middle ground between fully structured, lay-administered schedules, and semi-structured, psychiatrist-administered interviews. It encompasses four main domains: (*a*) demographic data; (*b*) social functioning and disability; (*c*) a diagnostic module comprising symptoms, signs and past history ratings; and (*d*) patterns of service utilization and patient-perceived need for services. It generates diagnoses according to several sets of criteria using the OPCRIT computerized diagnostic algorithm and can be administered either on-screen or in a hard-copy format.

Results. The DIP proved easy to use and was well accepted in the field. For the diagnostic module, inter-rater reliability was assessed on 20 cases rated by 24 clinicians: good reliability was demonstrated for both ICD-10 and DSM-III-R diagnoses. Seven cases were interviewed 2–11 weeks apart to determine test–retest reliability, with pairwise agreement of 0·8–1·0 for most items. Diagnostic validity was assessed in 10 cases, interviewed with the DIP and using the SCAN as 'gold standard': in nine cases clinical diagnoses were in agreement.

Conclusions. The DIP is suitable for use in large-scale epidemiological studies of psychotic disorders, as well as in smaller studies where time is at a premium. While the diagnostic module stands on its own, the full DIP schedule, covering demography, social functioning and service utilization makes it a versatile multi-purpose tool.

INTRODUCTION

The adoption of explicit diagnostic criteria in psychiatry gave rise in the last two decades to a wave of new epidemiological research which aimed at a diagnostically differentiated and reliable characterization of the frequency, clinical profiles, course and treatment of mental disorders ascertained in representative population samples. Examples include the Epidemiological Catchment Area Study (Keith *et al.* 1991) and the National Comorbidity Survey (Kessler *et al.* 1994) in the USA; the National Psychiatric Morbidity Survey in the UK (Jenkins *et al.* 1997); and the Australian National Survey of Mental Health and Well-being (ANSMHW; Jablensky *et al.* 1999, 2000). In the course of such research, it became evident that, in addition to sampling and case-finding methods, the properties of the instruments used

* Address for correspondence: Professor Assen Jablensky, Level 3 Medical Research Foundation Building, 50 Murray Street, Perth WA 6000, Australia.
(Email: assen@cyllene.uwa.edu.au)

to obtain diagnostic assessment were critical to the subsequent interpretation of the epidemiological findings. The instruments developed for research in the field range from 'hard-wired', fully structured interview schedules designed for use by lay interviewers, such as the Diagnostic Interview Schedule (Robins *et al.* 1981) and the Composite International Diagnostic Interview (CIDI; Robins *et al.* 1988) to comprehensive semi-structured interviews, such as the Schedules for Clinical Assessment in Neuropsychiatry (SCAN; Wing *et al.* 1990) whose administration requires clinical experience and judgement. Comparisons of survey findings obtained by means of lay- and clinician-administered instruments have suggested that the two types of interview may produce discrepant diagnostic classification of cases, especially in the assessment of psychotic disorders (Brugha *et al.* 1999). In addition to variations in the sensitivity and rating thresholds of specific symptom items, the sources of discrepant diagnoses include inherent differences in the interviewing procedure – script-bound in the instance of lay-administered schedules, as contrasted to a more flexible approach allowing probing, cross-examination and judgement in the instance of clinician-administered interviews.

There is thus a need for an instrument that allows clinicians to establish accurate diagnoses using an interview schedule that has structure (ensuring uniformity of use) but which also allows clinical expertise and experience to be factored into the decision-making about reported symptoms. Such an instrument needs to be relatively brief (able to be completed in under 30 min) so that it can be used in large-scale epidemiological studies, yet sufficiently detailed to allow a fine-tuned analysis of symptoms, as well as the distinction of current and lifetime symptom profiles. Furthermore, given the fact that there is still a lack of consensus about the optimal classification of psychotic disorders (Kendell & Jablensky, 2003; Castle & Jablensky, 2005), the instrument should allow the user to establish diagnoses and subtype classification according to several sets of commonly used criteria. The packaging of this diagnostic instrument with schedules to establish disability and service utilization would allow a comprehensive assessment of individuals with psychotic illnesses, across the range of domains that are important to their lives, for researchers, clinicians, and service providers.

To meet these aims, we developed the Diagnostic Interview for Psychoses (DIP), a comprehensive interview schedule that bridges the gap between highly structured lay interviews such as the CIDI and long, comprehensive schedules such as the SCAN. It was originally designed for the psychosis ('low-prevalence disorders') arm of the ANSMHW, which aimed to establish point (1-month) and 1-year prevalence rates for psychotic disorders in geographically defined catchment areas across Australia (Jablensky *et al.* 2000). The DIP is intended for use by interviewers with a clinical background (mental health nurses, clinical psychologists, and allied disciplines). The complete DIP interview requires 60–90 min to administer (30 min or less for the diagnostic module alone) and encompasses the following main domains (see Appendix 1): (*a*) demographic data; (*b*) assessment of social functioning and disability; (*c*) a profile of symptoms, signs and past history items required for the diagnosis of psychotic disorders; and (*d*) information on patterns of service utilization and patient-perceived unmet need for services. It generates diagnoses according to several sets of criteria using the OPCRIT (McGuffin *et al.* 1991) computerized diagnostic algorithm and can be administered either on-screen or in a hard-copy format. This paper describes the development, reliability and applications of the DIP.

METHOD

Structure and content of the DIP

The demography and social functioning module

Apart from standard demographic material (e.g. age, sex, marital status), the DIP includes items related to migrant status (e.g. country of birth, age at migration), family and household (e.g. number of children, carer role), education (e.g. age at leaving school, highest qualification), and accommodation (a detailed set of questions related to type of accommodation and frequency of change of setting, including homelessness). In relation to accommodation, respondents are also asked whether they felt safe in their current locality and whether they had

been victim of violence in the last 12 months. Aspects of social functioning and disability in key role domains are assessed by rating performance of household duties, general social contact (isolation and withdrawal), access to friends and family, and intimacy. A set of 14 items explores participation in the workforce and perceived capacity for work (including housework and studying). Items related to finances, activities of daily living, self-care and use of leisure time are also examined.

For the majority of these items, the period rated is either the past year or the past month. The experience of the interviewers is that respondents find these items relatively easy to answer, which allows time to build rapport prior to inquiring about symptoms and service utilization.

The diagnostic module

The structure of the diagnostic module of the DIP follows the Operational Criteria for Psychosis, OPCRIT, version 3.31 (McGuffin *et al*. 1991; Williams *et al*. 1996) 90-item checklist. The OPCRIT is essentially a phenomenological checklist that can be rated from practically any source and which, through the allied computerized algorithm, generates diagnoses according to the criteria of ICD-10 (WHO, 1993), DSM-III (APA, 1980) and DSM-III-R (APA, 1987), the Research Diagnostic Criteria (Spitzer *et al*. 1978); and the St Louis criteria (Feighner *et al*. 1972). It should be noted that at the time of the ANSMHW, OPCRIT had not been updated for DSM-IV, but that in future versions of the DIP, DSM-IV compatibility will be provided. OPCRIT also allows subtyping of psychotic disorders, according to the typologies proposed by Crow (1980), Tsuang & Winokur (1974), and Farmer and colleagues (1983). The OPCRIT checklist and algorithm have been used in a number of clinical, epidemiological and genetic studies (e.g. McGuffin *et al*. 1991; Farmer *et al*. 1992; van Os *et al*. 1996; Williams *et al*. 1996; Castle *et al*. 1998; McGrath *et al*. 2001, 2002; Rosenman *et al*. 2003).

The DIP-diagnostic module (DM) consists of a series of interview questions and probes either written *de novo* or, where relevant, using wording from the SCAN, version 2.0 (Wing *et al*. 1990, 1998), to elicit the OPCRIT checklist items. Responses to the probes are entered onto a computer database where the underlying OPCRIT algorithm generates diagnoses according to several diagnostic classification systems. The use of SCAN questions to elicit the OPCRIT items is justified on the basis that these have been developed and refined over many years, and are known to tap the specific symptom items accurately (Wing *et al*. 1990). Where SCAN wording is used, this is clearly identified by reference to the number of the SCAN item (Appendix 2). Since the underlying diagnostic algorithm used by the DIP-DM is the OPCRIT, the DIP-DM is best regarded as a clinical interview version of OPCRIT, and not as a 'mini-SCAN'.

The development of a clinical interview-based version of the OPCRIT allows the use of primary patient interview data in generating scores and achieves a better exploration of longitudinal course of illness, and ratings of present state and lifetime diagnosis. This overcomes the limitations of the original OPCRIT, notably the lack of a diagnostic interview to generate scores and its reliance on secondary sources of information, which can be of variable quality. Since the Present State Examination, PSE-9 (Wing *et al*. 1974), which was the precursor of SCAN, had been used in the original item definitions in OPCRIT, the use of SCAN questions and probes in the DIP-DM provides wording that is well established, developed by experts, and honed over years of use. In the development of the module, the first two authors independently decided and agreed on which SCAN questions and probes best reflected the OPCRIT items. The wording of some of the questions was modified to ensure they elicited items rateable under OPCRIT rules (for example, prompts were added to ensure better coverage of duration of symptoms, as required in OPCRIT). Questions were formulated in such a way as to allow the interviewer to ask about present state, past year, or lifetime occurrence of symptoms. The items were then ordered in a way allowing a natural progression to be followed, with symptoms being grouped into sections on depression, psychotic symptoms, and behaviour and affect. The DIP also includes a section of drug and alcohol use, to allow rating of these factors. The full list of items and their ordering is detailed in Appendix 1.

Whilst essentially interview-based, the DIP also encourages use of other sources of information, where available. For example, information on pre-morbid functioning and family history of psychiatric illness can be augmented by interview with a family member, although this is not mandatory. Signs, such as affect, psychomotor behaviour, or form and flow of speech, can be rated on the basis of observation during the interview, as well as using relevant information provided by informants or in clinical case-notes.

Being a semi-structured clinical interview, the DIP relies upon clinical judgement being used and presupposes certain clinical skills and experience. Once the interviewer is familiar with it, the instrument can be applied in a flexible manner, with clinical expedience deciding the order in which questions are asked. It is, however, recommended that each 'block' of questions (e.g. those on depression or mania), is kept intact, as questions have been grouped to enable the natural flow of the interview, with a number of built-in cut-offs and skips between and within sections to avoid redundancy when initial screening questions have indicated that psychopathology is unlikely to be present in that section.

The service utilization module

Since people with psychotic disorders are likely to use a large number of helping agencies, the service utilization module of the DIP aims to capture, as comprehensively as possible, a variety of services and to quantify the extent of their use in the year prior to the interview. Hospitalization, both psychiatric and non-psychiatric, public and private, is recorded in terms of number of voluntary and involuntary admissions and length of stay. In addition, visits to general accident and emergency departments and the reasons for those visits are recorded, as well as the number of contacts with psychiatric emergency or crisis services, both hospital and community based. The number and type of continuing care visits in the community or clinics are also recorded, as well as the type of health professional assigned as the patient's case manager. Other health professionals seen and services received, both psychiatric and non-psychiatric are noted, including whether the services met the patient's perceived needs. Involvement in rehabilitation or day programmes is recorded in terms of weeks of participation and frequency of attendances. Access to, and use of, government and non-government health and welfare agencies is recorded, as well as information about the availability of carers and legal guardianship. The module contains a detailed checklist of medication and the respondent is asked to identify the drugs prescribed and/or used in the past month; to make a judgement as to their perceived 'helpfulness'; and to describe any experienced side-effects, using prompts from a checklist (see Castle *et al.* 2002). The concluding part of the interview aims to elicit subjective quality of life judgements (satisfaction with own independence and satisfaction 'with life as a whole' in the past year), as well as an open-ended account of perceived need for services that were unavailable.

The DIP-DM software, reliability and validity

The DIP-DM software was written to allow data entry of information elicited with the diagnostic module directly into a Microsoft Access database. Once the data have been entered, the DIP-DM generates diagnoses according to the various operational definitions functional in the OPCRIT diagnostic algorithm. The software contains in-built validation rules to ensure the internal consistency of the data being entered. The database stores both the diagnostic data and the raw data, ready for export into other software for manipulation or analysis. These data may be printed for individual cases, using pre-formatted reports, one for the diagnostic data, the other for raw data. Thus, the clinician entering the data can have an immediate diagnostic printout, or the data can be stored for later analysis. The diagnostic data can also be viewed directly on screen.

All prospective interviewers ($n=33$) were trained to ensure inter-rater reliability. In an inter-centre workshop, 14 interviewers jointly rated a series of videotaped DIP interviews, using as 'gold standard' a set of ratings being independently agreed by a panel of experienced clinicians. Training was conducted at each site (Brisbane, Canberra, Melbourne and Perth), with both 'live' interviews and pre-recorded videotaped interviews that had been independently rated by an experienced interviewer.

Particular attention was paid to certain key items that have been shown to have an important impact on OPCRIT diagnoses, notably the item on the relationship between psychotic and affective symptoms (see Farmer *et al*. 1992).

A total of 108 interviews were used to assess instrument reliability. The inter-rater reliability of the diagnostic module was assessed using jointly rated interviews involving 20 cases, conducted by 24 interviewers who had received prior training in the use of the DIP from an experienced clinician. Of the 20 cases, 13 were rated by pairs of raters in joint live interviews and seven were multiple-rated by groups of 8–15 raters from live or videotaped interviews where the interviewer was an experienced clinician. Test–retest reliability was examined by using seven cases, each interviewed by two different raters with intervals of 2–11 weeks between the interviews. Reliability was determined for ICD-10 and DSM III-R diagnoses, as well as for individual OPCRIT items. Inter-rater reliability was assessed by calculating *overall pairwise agreement* (PAR), the ratio of the number of agreements between observers/raters to the total number of comparisons made, the *kappa statistic*, to measure the degree of agreement between two raters for each observation, taking into account agreement due to chance, and *generalized kappa*, which was used to measure the reliability of multiple raters (Fleiss, 1981; Haas, 1991).

RESULTS

Use of DIP in the ANSMHW (study of low-prevalence disorders)

Prior to the ANSMHW, the DIP was pilot-tested in a range of clinical settings, including some 30 interviews in Perth and over 100 interviews in Brisbane. Refinements were made to ensure ease of use and seamless 'flow' when used with patients. Largely these refinements related to wording of specific items, and ordering of items such that they were compatible with a free-flowing clinical interview style of administration.

In the course of the ANSMHW, the DIP was administered to a stratified sample of 980 individuals (586 men, 394 women), identified by screening service contacts in the census month. The majority (70·1%) were recruited through mainstream mental health services (in- or outpatients), with a further 17·7% being identified through their general practitioner or private psychiatrist. A special subsample of 120 people of 'no fixed abode', or living in marginal accommodation, were recruited through shelters or other services provided for homeless people. In addition, the DIP was administered to 146 individuals with history of a psychotic disorder, identified from register records as having been in contact with the mental health services within 3 years of the census, but without a service contact during the census month itself. All interviewed patients gave written informed consent to participation in the study.

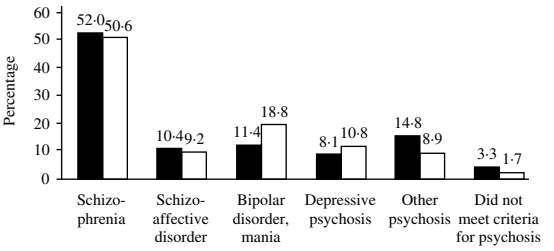

FIG. 1. Diagnostic distribution of 980 interviewed cases according to the DIP diagnostic algorithm, by diagnostic classification system. ■, ICD-10; □, DSM-III-R.

The findings of the study have been reported in detail elsewhere (Jablensky *et al*. 1999, 2000). As an illustration of the 'polydiagnostic' applications of the DIP, Fig. 1 shows the diagnostic distribution, in terms of OPCRIT-generated ICD-10 and DSM-III-R diagnoses, for the same set of 980 cases, interviewed during the census month. The overall agreement between the two diagnostic systems (overlapping assignment of cases to the same diagnostic category) was high (80%), but considerably higher for schizophrenia and schizoaffective disorder, than for bipolar disorder and 'other' psychoses. The discrepant assignment of a proportion of the patients to the latter two categories may be due to the wider scope of 'other' psychoses in ICD-10, which encompasses the acute transient psychoses. Thus, a single acute episode with hypermotility, emotional turmoil, intense feelings of happiness and fleeting psychotic symptoms may be classified according to ICD-10 as acute polymorphic psychotic disorder without symptoms of schizophrenia (rubric F23.0). In the absence of a comparable diagnostic category

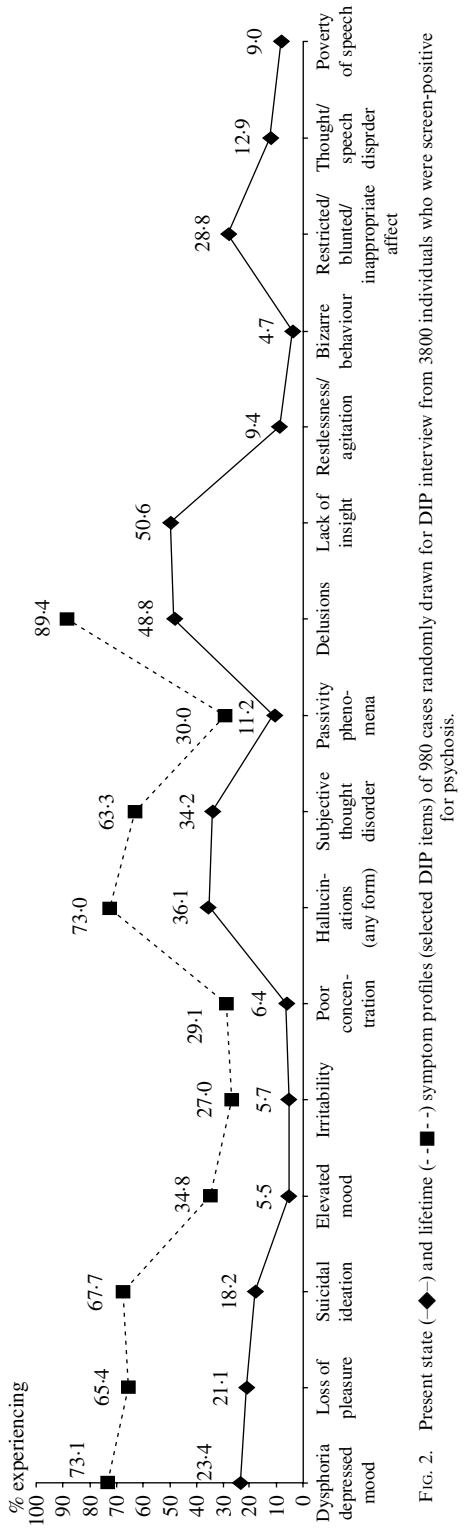

FIG. 2. Present state (—◆—) and lifetime (- - ■ - -) symptom profiles (selected DIP items) of 980 cases randomly drawn for DIP interview from 3800 individuals who were screen-positive for psychosis.

in DSM-III-R, such cases are more likely to be classified according to its criteria as single episodes of mania or hypomania.

Another application of the DIP-OPCRIT diagnostic module is the generation of individual and group symptom profile plots. Fig. 2 shows such plots for lifetime and present state (including the 4 weeks prior to interview) frequency of symptoms in the interviewed sample of 980 patients. When combined with a 'polydiagnostic' classification, such plots provide a convenient visualization of both the similarities and differences between alternative sets of diagnostic criteria in terms of actual symptomatology.

Reliability and validity of the diagnostic module

Inter-rater reliability

Table 1 shows the overall PAR and kappa ranges for a selection from the 90 OPCRIT items in the diagnostic module. A full list of the items with their individual kappa/PAR reliability coefficients is available on request. In terms of PAR, 65 items had a rate of 0·8–1·0. Using the kappa statistic, half of the items achieved a kappa value of $\geqslant 0{\cdot}6$, i.e. good to excellent concordance, with 20% (18 items) in the $>0{\cdot}8$ range. A kappa of $<0{\cdot}4$ was obtained for 21% (19 items), of which six items resulted in a kappa of zero (these items, and some of the items with a low kappa, actually had attained a high PAR). The majority of these items contained dichotomous response categories (yes/no) where almost all of the responses fell into one category, causing the data to be skewed. Thus, the zero or low kappa in these instances could be attributed to the instability of kappa when the data distributions are skewed and small variations can cause large fluctuations in kappa values. When these items were assessed for agreement on absence or presence of the condition, the agreement was excellent.

Table 2 summarizes inter-rater agreement on ICD-10 and DSM-III-R diagnoses when two alternative groupings of the diagnostic rubrics were used: (*a*) the narrow OPCRIT diagnostic subtypes; and (*b*) the broad 3-symbol (or 3-digit) diagnostic categories of the two classifications. At the level of detailed, narrow diagnostic breakdown, the inter-rater reliability results for both ICD-10 and DSM-III-R

Table 1. *DIP inter-rater and test–retest reliability – selected items*[a]

	Inter-rater reliability (No. of cases: 20) (No. of raters: 24)			Test–retest reliability (No. of cases: 7) (No. of raters: 10)		
	Overall pairwise agreement	Generalized kappa	Level of agreement[b]	Overall pairwise agreement	Generalized kappa	Level of agreement[b]
Age of onset	0·71	0·68	Good	0·57	0·51	Moderate
Mode of onset	0·78	0·71	Good	0·71	0·53	Moderate
Psychosocial stressor prior to first episode	0·74	0·43	Moderate	0·89	0·69	Good
Pre-morbid personality disorder	0·94	0·39	Fair	1·00	1·00	—[c]
Dysphoria	0·74	0·51	Moderate	0·71	0·45	Moderate
Suicidal ideation	0·90	0·68	Good	0·71	0·42	Moderate
Loss of energy	0·80	0·50	Moderate	0·60	0·00	—[c]
Diminished libido	0·95	0·89	Excellent	0·80	0·52	Moderate
Early morning waking	0·58	0·17	Poor	0·80	0·57	Moderate
Excessive sleep	0·78	0·59	Moderate	0·80	0·60	Good
Delusions of guilt	0·83	0·52	Moderate	1·00	1·00	Excellent
Elevated mood	0·79	0·68	Good	0·86	0·74	Good
Thoughts racing	0·86	0·73	Good	0·67	0·00	—[c]
Excessive activity	0·75	0·35	Fair	1·00	1·00	Excellent
Increased sociability	0·79	0·68	Good	0·67	0·33	Fair
Non-affective hallucinations in any modality	0·90	0·65	Good	1·00	1·00	Excellent
Third person auditory hallucination	0·94	0·87	Excellent	0·75	0·47	Moderate
Thought insertion	0·85	0·58	Moderate	0·86	0·69	Good
Thought broadcast	0·89	0·76	Good	0·57	0·14	Poor
Thought withdrawal	0·92	0·77	Good	0·86	0·58	Moderate
Thought echo	0·96	0·79	Good	0·86	0·69	Good
Delusions of passivity	0·90	0·78	Good	0·57	0·07	Poor
Persecutory delusions	0·94	0·86	Excellent	1·00	1·00	Excellent
Grandiose delusions	0·66	0·47	Moderate	0·57	0·29	Fair
Bizarre delusions	0·81	0·61	Good	0·71	0·30	Fair
Lack of insight	0·59	0·17	Poor	0·86	0·58	Moderate
Persecutory delusions and hallucinations	0·80	0·58	Moderate	1·00	1·00	Excellent
Lifetime diagnosis of alcohol abuse/dependence	0·87	0·65	Good	0·86	0·58	Moderate
Lifetime diagnosis of cannabis abuse/dependence	0·96	0·87	Excellent	1·00	1·00	Excellent
Course of the disorder	0·54	0·37	Fair	0·43	0·14	Poor
Bizarre behaviour	0·97	0·68	Good	1·00	1·00	Excellent
Blunt affect	0·99	0·42	Moderate	0·86	0·00	—[c]

[a] Based on lifetime ratings, where applicable.
[b] See Landis & Koch (1977).
[c] Kappa reflects skewed data due to a dichotomous response category with almost all of the responses in the one category.

diagnoses showed good agreement beyond chance according to PAR (0·81 for ICD-10 and 0·67 for DSM-III-R). Using kappa, agreement was high for ICD-10 diagnoses (kappa = 0·73) but only moderate (kappa = 0·49) for DSM-III-R diagnoses. Analysis of the discrepancy between the ICD-10 and DSM-III-R results on a case by case basis indicated that the main reason for the discrepancy was that, for DSM-III-R, the OPCRIT algorithm was sensitive to small differences between raters on the coding of items used to allocate a case to a specific diagnostic category (for example, in one case of disagreement with two raters, the DSM-III-R OPCRIT diagnosis for rater A was 'mania with psychosis' and for rater B 'bipolar with psychosis'). In light of such findings, additional analysis was conducted based on aggregating the subtype diagnoses of the two classifications into broader diagnostic groupings, identical for both ICD-10 and DSM-III-R: schizophrenia, schizoaffective disorder, bipolar disorder, depressive disorder with/without psychotic features, and other psychosis. At this level of aggregated diagnoses, the reliability for ICD-10 was only slightly higher (kappa = 0·74) compared to the level of detailed diagnostic breakdown, but for DSM-III-R it was considerably improved, attaining a kappa of 0·65 (good agreement).

Test–retest reliability

The results of the test–retest reliability analysis (seven cases, two different raters, interviews

Table 2. *DIP inter-rater and test–retest reliability – diagnosis*[a]

	Inter-rater reliability Narrow categories (No. of cases: 20) (No. of raters: 24)			Inter-rater reliability Broad categories (No. of cases: 20) (No. of raters: 24)			Test–retest reliability Narrow categories (No. of cases: 7) (No. of raters: 10)			Test–retest reliability Broad categories (No. of cases: 7) (No. of raters: 10)		
	Overall pairwise agreement	Generalized kappa	Level of agreement[b]	Overall pairwise agreement	Generalized kappa	Level of agreement[b]	Overall pairwise agreement	Generalized kappa	Level of agreement[b]	Overall pairwise agreement	Generalized kappa	Level of agreement[b]
ICD-10	0·81	0·73	Good	0·82	0·74	Good	0·57	0·51	Moderate	0·71	0·65	Good
DSM III-R	0·67	0·49	Moderate	0·78	0·62	Good	0·57	0·48	Moderate	0·86	0·81	Excellent

[a] Narrow categories: OPCRIT generates up to 17 DSM-III-R diagnostic categories and 20 ICD-10 diagnostic categories. Broad categories: the OPCRIT-generated diagnostic categories were aggregated into five broad categories: schizophrenia; schizoaffective; bipolar, mania; depressive psychosis; and other psychosis.

[b] The Level of agreement given for the generalized kappa is based on Landis & Koch (1977).

2–11 weeks apart) are summarized in Table 1. At the level of individual OPCRIT items, kappa of $\geqslant 0{\cdot}60$ was obtained for 42% of the items. However, over a third of the items showed an agreement lower than 0·4. In over half of these items, the low kappa was due to the small number of cases. When analysed for presence/absence of the item, using PAR, agreement was high (0·80–1·00) for the majority of items.

Agreement was moderate at the level of narrow ICD-10 and DSM-III-R diagnoses, but good or excellent when broad disorder categories were used (Table 2). Considering that the design, involving different raters and intervals of 2–11 weeks, is a stringent test of stability of diagnostic assessment, these results can be regarded as satisfactory.

Diagnostic validity

An assessment of the validity of the DIP-generated diagnosis was possible by comparing diagnoses for 10 cases that had been assessed using both the DIP interview and a comprehensive SCAN interview, with the SCAN interview as the 'gold standard'. The level of agreement between ICD-10 SCAN-generated diagnoses and ICD-10 DIP-generated diagnoses was good, with nine out of the 10 DIP diagnoses matching the SCAN diagnosis at the 3-digit level.

DISCUSSION

Methodological issues

There are inevitable methodological constraints inherent in the development of a complex diagnostic instrument such as the DIP. The fact that the instrument is derived from existing instruments (i.e. the OPCRIT and SCAN) is both a strength and a weakness. In its favour, this approach allowed us to use well-established diagnostic questions and to tailor them to the OPCRIT checklist. However, we were constrained with respect to which items we included and which we excluded: indeed, some of the DIP items are fairly rare and specific, whilst we did not include other symptoms which are of interest because of either their singularity (e.g. some of Schneider's 'first rank' symptoms), or their common occurrence in clinical practice (e.g. religious delusions).

The reliability and validity data presented here are based on relatively small samples

(notably for the test–retest analyses), but do give an indication of the robustness of the diagnostic module of the DIP. Also, we present data from its use only in an Australian setting (albeit across four geographically discrete areas). Whilst the diagnostic module is expected to be widely applicable across different settings, given the similarities in the presentation of psychotic illness across different cultures (Jablensky *et al.* 1992), the service utilization module will not be so readily applicable. There are already a number of translated versions of the DIP (see below), and we await reliability and validity data using these versions.

COMMENT

The DIP is a semi-structured interview for use in epidemiological and clinical settings, designed for the diagnostic assessment of persons with a psychotic illness, description of their symptom profiles, as well as evaluation of their social functioning, disablement, and service utilization.

The diagnostic module of the DIP is an interview version of the OPCRIT, which allows present state, past year and lifetime diagnoses to be made. The module is linked to the OPCRIT computer algorithm, allowing diagnoses according to a wide range of diagnostic criteria for psychotic illness. It is available in a computerized format (DIP-DM), allowing direct data entry and determination of symptom profile and diagnosis. The DIP is a relatively brief and 'user-friendly' instrument, allowing completion in 60–90 min; where appropriate, the diagnostic module, which takes 20–30 min to complete, can be used alone. The interview was well accepted by respondents, even when acutely unwell, and interviewers reported satisfaction with its structure and use.

A standard, comprehensive DIP training kit and a 2-day training programme have been developed and tested at multiple clinical and research settings across Australia. Based on user feedback, the training package has been further fine-tuned and is currently being used for training by correspondence at clinical and research sites outside Australia (including the UK, Bulgaria, France, Italy, Indonesia and Ghana). The training kit includes a DIP manual complete with glossary and software instructions, training videotapes supplied with 'gold standard' reference ratings by experienced clinicians, as well as a user feedback questionnaire. User feedback has highlighted satisfaction with the ease and brevity of the administration of the instrument, as well as its capacity to provide an almost immediate diagnosis. In addition to its use in research projects, many practising clinicians are now using the DIP to monitor changes in symptomatology over time, underscoring its potential utility as a tool for routine clinical practice. The DIP is currently being updated to incorporate the OPCRIT DSM-IV algorithm.

The robust performance of the DIP as a diagnostic instrument for epidemiological studies was demonstrated in the ANSMHW (Jablensky *et al.* 1999). In addition, the DIP has proved useful in case-control studies where unaffected control subjects need to be screened in order to exclude the presence of a psychotic disorder (McGrath *et al.* 2001, 2002).

The information recorded in the social functioning and service utilization modules has enabled detailed analyses of levels of disability associated with psychotic illness (Gureje *et al.* 2001, 2002), co-morbid substance use disorders (Kavanagh *et al.* 2004), as well as an evaluation of the extent and patterns of service use and their relationship to other variables, including the economic costs of psychosis (Carr *et al.* 2003, 2004). Inevitably the service utilization items will require some modification for use in service settings which are structured differently from those in Australia, but it provides a template for a comprehensive assessment of all contacts, both planned and unplanned, with mental health and other services.

In addition to being relatively easy to administer, the DIP is an efficient instrument, allowing an immediate diagnosis to be made according to several alternative sets of operational diagnostic criteria. This makes the DIP suitable for use in large-scale epidemiological studies that are essential to mapping the incidence, prevalence and course of psychoses across populations and over time, as well as in smaller studies where time is at a premium. While the diagnostic module stands on its own, the full DIP schedule, covering demography, social functioning and service utilization permits a breadth of data capture that provides researchers and clinicians with a versatile, multi-purpose tool.

ACKNOWLEDGEMENTS

The present report is based on data collected in the Collaborative Study on Low-Prevalence (Psychotic) Disorders, an epidemiological and clinical investigation which is part of the National Survey of Mental Health and Well-being, Australia 1997–1998. The study was funded by the Australian Commonwealth Department of Health and Aged Care (for fieldwork in Brisbane, Melbourne and Perth); and by the Australian Capital Territory Department of Health and Community Care, and the Psychiatric Epidemiology Research Centre, Australian National University (for fieldwork in Canberra). The Chief Investigators of the study are: Assen Jablensky (Project Director), Vaughan Carr, David Castle, Mandy Evans, Oye Gureje, Carol Harvey, Helen Herrman, Ailsa Korten, John McGrath and Vera Morgan. A complete list of the investigators is available in: Jablensky *et al.* (1999, 2000). The DIP interviewers at the four study sites were: Helen Stoeckel, David Kirk, David Keys, Johann Napier (Australian Capital Territory); Susette Cardy, Chris Young, Dianne Briffa, Ben Chapple, Suzanne Drake, Madonna Hirning, Troy Winterflood (Queensland); Helen Evert, Rosemary Thomas, Monique Decortis, Naomi Harris, Maggie McIntosh, Andrew Green, Susan Roberts (Victoria); Anna Waterreus, Jenny Griffith, Frances Waru, Elizabeth Trainer, Colin Sansom, Michelle Lusty (Western Australia). Assistance in the development of the DIP training package was provided by John Dean, Helen Stain and Rachael Williams. The DIP data entry interface was developed by Max Maller, The University of Western Australia.

DECLARATION OF INTEREST

None.

APPENDIX 1

DIP Part 1: Demography and social functioning module
Contains 49 items under the following item headings:

- General information including sociodemographic data
- Children, carer role
- Education
- Accommodation
- Household and participation in household activities
- Socializing; social withdrawal
- Confiding relationships, intimacy, sex life
- Work, housework, studying
- Finances
- Activities of daily living and self-care
- Interests

DIP Part 2: Diagnostic module
Contains 94 items under the following item headings:

- General items
- Pre-morbid characteristics and onset
- Family history
- Depression
- Mania
- Hallucinations
- Subjective thought disorder
- Delusions
- Insight
- Response to medication
- General ratings on psychotic symptoms
- Substance use: alcohol; non-medical use of drugs; tobacco and caffeine
- Alcohol and drug abuse and dependence
- Duration and course
- Behaviour
- Affect
- Speech

DIP Part 3: Service utilization module
Contains 40 items under the following item headings:

- In-patient treatment
- Care received from emergency/casualty department
- Treatment in the community (out-patient clinic/community mental health clinic)
- Other health professionals seen and services received
- Rehabilitation or day programme
- Health and welfare and voluntary agencies
- Guardianship, carers
- Medication and the perceived benefits
- Impairment due to side-effects of medication
- Self-harm
- Offending behaviour
- Satisfaction with life
- Unmet need for services
- Social and Occupational Functioning Assessment Scale (SOFAS)

APPENDIX 2

Extract from the DIP interview schedule

Present state, past year and lifetime coding

2.48 Running commentary (OPCRIT 74)
Voice(s) commenting on thoughts or actions (SCAN 17.008)

☐ ☐ ☐
PS PY LT

- *Does a voice comment on what you are thinking or doing?*
- *Do you hear a voice saying what you are reading, or describing what you are seeing on television as you see it?*
- *Do you hear them in your head, or through your ears, as though comming from outside?*
- *How often does it happen?*

0 = Not present
1 = Either internal voices ('pseudo' hallucinations) or external voices ('true' hallucinations)

Further information on the DIP, including contact details, is available at the following website: http://dip.ccrn.uwa.edu.au

REFERENCES

APA (1980). *Diagnostic and Statistical Manual of Mental Disorders* (3rd edn). American Psychiatric Association: Washington, DC.

APA (1987). *Diagnostic and Statistical Manual of Mental Disorders* (3rd edn, revised). American Psychiatric Association: Washington, DC.

Brugha, T. S., Nienhuis, F., Bagchi, D., Smith, J. & Meltzer, H. (1999). The survey form of SCAN: the feasibility of using experienced lay survey interviewers to administer a semi-structured systematic clinical assessment of psychotic and non-psychotic disorders. *Psychological Medicine* **29**, 703–711.

Carr, V. J., Johnston, P. J., Lewin, T. J., Rajkumar, S., Carter, G. L. & Issakidis, C. (2003). Patterns of service use among persons with schizophrenia and other psychotic disorders. *Psychiatric Services* **54**, 226–235.

Carr, V., Lewin, T., Neil, A., Halpin, S. & Holmes, S. (2004). Premorbid, psychosocial and clinical predictors of the costs of schizophrenia and other psychoses. *British Journal of Psychiatry* **184**, 517–525.

Castle, D., Morgan, V. & Jablensky, A. (2002). Antipsychotic use in Australia: the patients' perspective. *Australian and New Zealand Journal of Psychiatry* **36**, 633–641.

Castle, D. J. & Jablensky, A. (2005). Diagnosis and classification in psychiatry. In *Core Psychiatry* (2nd edn) (ed. P. Wright, M. Phelan and J. Stern), pp. 507–515. W. B. Saunders: London.

Castle, D. J., Wessley, S., Van Os, J. & Murray, R. M. (1998). *Psychosis in the Inner City: The Camberwell First Episode Study*. Psychology Press: East Sussex.

Crow, T. (1980). The molecular pathology of schizophrenia: more than one disease process? *British Medical Journal* **280**, 66–68.

Farmer, A., Wessely, S., Castle, D. & McGuffin, P. (1992). Methodological issues in using a polydiagnostic approach to define psychotic illness. *British Journal of Psychiatry* **161**, 824–830.

Farmer, A. E., McGuffin, P. & Spitznagel, L. (1983). Heterogeneity in schizophrenia: a cluster analytic approach. *Psychiatry Research* **8**, 1–12.

Feighner, J. P., Robins, E., Guze, S. B., Woodruff, R. & Munoz, R. (1972). Diagnostic criteria for use in psychiatric research. *Archives of General Psychiatry* **26**, 57–61.

Fleiss, J. L. (1981). *Statistical Methods for Rates and Proportions*, pp. 212–236. John Wiley & Sons: New York.

Gureje, O., Herrman, H., Harvey, C., Morgan, V. & Jablensky, A. (2002). The Australian National Survey of Psychotic Disorders: profile of psychosocial disability and its risk factors. *Psychological Medicine* **32**, 639–647.

Gureje, O., Herrman, H., Harvey, C., Trauer, T. & Jablensky, A. (2001). Defining disability in psychosis: performance of the diagnostic interview for psychosis – disability module (DIP-DIS) in the Australian National Survey of Psychotic Disorders. *Australian and New Zealand Journal of Psychiatry* **35**, 846–851.

Haas, M. (1991). Statistical methodology for reliability studies. *Journal of Manipulative and Physiological Therapeutics* **14**, 119–132.

Jablensky, A., McGrath, J., Herrman, H., Castle, D., Gureje, O., Evans, M., Carr, V., Morgan, V., Korten, A. & Harvey, C. (2000). Psychotic disorders in urban areas: an overview of the Study on Low Prevalence Disorders. *Australian and New Zealand Journal of Psychiatry* **34**, 221–236.

Jablensky, A., McGrath, J., Herrman, H., Castle, D., Gureje, O., Morgan, V. & Korten, A. (1999). People Living with Psychotic Illness: An Australian Study 1997–98. National Survey of Mental Health and Wellbeing Report 4. Commonwealth Department of Health and Aged Care, Canberra.

Jablensky, A., Sartorius, N., Ernberg, G., Anker, M., Korten, A., Cooper, J. E., Day, R. & Bertelsen, A. (1992). *Schizophrenia: Manifestations, Incidence and Course in Different Cultures*. Psychological Medicine Monograph Supplement 20. Cambridge University Press.

Jenkins, R., Bebbington, P., Brugha, T., Farrell, M., Gill, B., Lewis, G., Meltzer, H. & Petticrew, M. (1997). The National Psychiatric Morbidity Survey of Great Britain – strategy and methods. *Psychological Medicine* **27**, 765–774.

Kavanagh, D. J., Waghorn, G., Jenner, L., Chant, D. C., Carr, V., Evans, M., Herrman, H., Jablensky, A. & McGrath, J. J. (2004).

Demographic and clinical correlates of comorbid substance use disorders in psychosis: multivariate analyses from an epidemiological sample. *Schizophrenia Research* **66**, 115–124.

Keith, S. J., Regier, D. A. & Rae, D. S. (1991). Schizophrenic disorders. In: *Psychiatric Disorders in America: The Epidemiologic Catchment Area Study* (ed. L. N. Robins and D. A. Regier), pp. 33–52. Free Press: New York.

Kendell, R. & Jablensky, A. (2003). Distinguishing between the validity and utility of psychiatric diagnoses. *American Journal of Psychiatry* **160**, 4–12.

Kessler, R. C., McDonagle, K. A., Zhao, S., Nelson, C. B., Hughes, M., Eshleman, S., Wittchen, H. U. & Kendler, K. S. (1994). Lifetime and 12-month prevalence of DSM-IIIR psychiatric disorders in the United States: results from the National Comorbidity Study. *Archives of General Psychiatry* **51**, 8–19.

Landis, J. & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics* **33**, 159–174.

McGrath, J., El-Saadi, O., Cardy, S., Chapple, B., Chant, D. & Mowry, B. (2001). Urban birth and migrant status as risk factors for psychosis: an Australian case-control study. *Social Psychiatry & Psychiatric Epidemiology* **36**, 533–536.

McGrath, J., El-Saadi, O., Grim, V., Cardy, S., Chapple, B., Chant, D., Lieberman, D. & Mowry, B. (2002). Minor physical anomalies and quantitative measures of the head and face in patients with psychosis. *Archives of General Psychiatry* **59**, 458–464.

McGuffin, P., Farmer, A. & Harvey, I. (1991). A polydiagnostic application of operational criteria in studies of psychotic illness. Development and reliability of the OPCRIT system. *Archives of General Psychiatry* **48**, 764–770.

Robins, N. L., Helzer, J. E., Croughan, J. & Ratcliff, K. S. (1981). National Institute of Mental Health Diagnostic Interview Schedule: its history, characteristics and validity. *Archives of General Psychiatry* **38**, 381–389.

Robins, L. N., Wing, J., Wittchen, H. U., Helzer, J. E., Babor, T. F., Burke, J., Farmer, A., Jablensky, A., Pickens, R., Regier, D. A., Sartorius, N. & Towle, L. H. (1988). The Composite International Diagnostic Interview. An epidemiologic instrument suitable for use in conjunction with different diagnostic systems in different cultures. *Archives of General Psychiatry* **45**, 1069–1077.

Rosenman, S., Korten, A., Medway, J. & Evans, M. (2003). Dimensional vs. categorical diagnosis in psychosis. *Acta Psychiatrica Scandinavica* **107**, 378–384.

Spitzer, R. L., Endicott, J. & Robins, E. (1978). Research diagnostic criteria: rationale and reliability. *Archives of General Psychiatry* **35**, 773–782.

Tsuang, M. T. & Winokur, G. (1974). Criteria for subtyping schizophrenia. *Archives of General Psychiatry* **31**, 43–47.

Van Os, J., Castle, D., Takei, N., Der, G. & Murray, R. (1996). Psychotic illness in ethnic minorities: clarification from the 1991 census. *Psychological Medicine* **26**, 203–208.

WHO (1993). *The ICD-10 Classification of Mental and Behavioural Disorders. Diagnostic Criteria for Research*. World Health Organization: Geneva.

Williams, J., Farmer, A. E., Ackenheil, M., Kaufmann, C. A., McGuffin, P. and the OPCRIT Reliability Research Group (1996). A multicentre inter-rater reliability study using the OPCRIT computerized diagnostic system. *Psychological Medicine* **26**, 775–783.

Wing, J. K., Babor, T., Brugha, T., Burke, J., Cooper, J. E., Giel, R., Jablensky, A., Regier, D. & Sartorius, N. (1990). SCAN: Schedules for Clinical Assessment in Neuropsychiatry. *Archives of General Psychiatry* **47**, 589–593.

Wing, J. K., Cooper, J. E. & Sartorius, N. (1974). *The Measurement and Classification of Psychiatric Symptoms*. Cambridge University Press: Cambridge.

Wing, J. K., Sartorius, N. & Üstün, T. B. (eds) (1998). Diagnosis and Clinical Measurement in Psychiatry. A Reference Manual for SCAN/PSE-10. Cambridge University Press: Cambridge.