Reliability of the Camberwell Assessment of Need – European Version

EPSILON Study 6

PAUL McCRONE, MORVEN LEESE, GRAHAM THORNICROFT,
AART H. SCHENE, HELLE CHARLOTTE KNUDSEN,
JOSÉ LUIS VÁZQUEZ-BARQUERO, ANTONIO LASALVIA, SARAH PADFIELD,
IAN R. WHITE, GWYN GRIFFITHS and the EPSILON STUDY GROUP

Background The five-country European Psychiatric Services: Inputs Linked to Outcome Domains and Needs (EPSILON) Study aimed to develop standardised and reliable outcome instruments for people with schizophrenia. This paper reports reliability findings for the Camberwell Assessment of Need – European Version (CAN–EU).

Method The CAN–EU was administered in each country, at two points in time to assess test–retest reliability, and was rated by two interviewers at the first administration. Cronbach's α, test–retest reliability and interrater reliability were compared between the five sites. Reliability coefficients and standard errors of measurement for summary scores were estimated.

Results Sites varied in levels and spread of needs. Alphas were 0.48, 0.58 and 0.64 for total, met and unmet needs respectively. Test–retest reliability estimates, pooled over sites, were 0.85 for the total needs, 0.69 for met needs and 0.78 for unmet needs. Pooled estimates for interrater reliability were higher, at 0.94, 0.85 and 0.79 for total, met and unmet needs respectively. There were statistically significant differences in interrater reliability between sites.

Conclusion The results confirm the feasibility of using CAN–EU across sites in Europe and its psychometric adequacy.

Declaration of interest No conflict of interest. Funding detailed in Acknowledgements.

This paper summarises the methods used in a five-nation European study to establish the reliability of the European Version of the Camberwell Assessment of Need (CAN–EU), presents the results, and discusses the implications of these findings. The paper should be read in close conjunction with other related papers in this series, which give more detailed accounts of key related aspects of the study (Becker *et al*, 1999, 2000, this supplement; Knudsen *et al*, 2000, this supplement; van Wijngaarden *et al*, 2000, this supplement).

Within the domain of the assessment of individual patient outcomes, increasing importance has recently been attached to the needs of those who suffer from mental illness. This emphasises the active role of the users of mental health services, and also raises a series of important questions: How can needs be defined, and by whom? How can they be measured and compared? What importance should be accorded to both met and unmet needs in the assessment of individual patients, and in the planning and evaluation of mental health services as a whole? How should the needs of those suffering from schizophrenia be prioritised in relation to the needs of other diagnostic groups?

CAMBERWELL ASSESSMENT OF NEED – EUROPEAN VERSION (CAN–EU)

The assessments of need in the European Psychiatric Services: Inputs Linked to Outcome Domains and Needs (EPSILON) Study were made using the Camberwell Assessment of Need – European Version (CAN–EU), based on the CAN Research Version 3.0 (Phelan *et al*, 1995). This is an interviewer-administered instrument comprising 22 individual domains of need: accommodation, food, looking after the home, self-care, physical health, psychological distress, psychotic symptoms, information about condition and treatment, daytime activities, company, safety to self, safety to others, alcohol, drugs, intimate relationships, sexual expression, basic education, child care, transport, using a telephone, money and welfare benefits. There is good agreement between staff and user ratings of the overall numbers of needs, although there may be substantial differences for individual items. It is important to note that in the present study the service user (patient) ratings are used.

Each item of the CAN–EU contains the same question structure. The first question asks whether a need exists, and, if it does, whether it is a met or an unmet need. If there is no need in any particular area, then the interviewer proceeds straight to the next item. If a met or unmet need does exist, then further questions relating to service receipt for that item are asked. The first of these finds out how much care is received from friends or relatives (0=no help, 1=low help, 2=moderate help, 3=high help). The same question is asked about care received from formal services and also how much care is *required* from formal services. Finally, the person being interviewed is asked whether overall they receive the right sort of help, and whether they receive the right amount of help. Both of these are rated as zero (no) or one (yes).

Summary scores of the total number of needs (the number of 1s or 2s), the met needs (the number of 1s) and the unmet needs (the number of 2s) are computed. If the number of valid items (i.e. excluding missing values) needs is 18 or more, a prorated total is computed from the valid items, otherwise the summary score is regarded as missing.

AIMS OF THE EPSILON STUDY

The aims of the EPSILON Study are:

(a) To produce standardised versions of five instruments in key areas of mental health service research in five European languages (Danish, Dutch, English, Italian and Spanish) by a rigorous conversion process from the original version into the other four languages. This involves (i) accurate and independent translation and back-translation from the original into the other four languages, (ii) checks of cross-cultural applicability using focus groups, and (iii) assessment of instrument reliability. Full details of these procedures are given elsewhere (Becker *et al*, 1999,

2000, this supplement; Knudsen *et al*, 2000, this supplement; van Wijngaarden *et al*, 2000, this supplement). This paper reports the reliability study on the CAN–EU.

(b) To obtain and compare data from five regions in different European countries, each with its particular system of health care, about social and clinical variables, characteristics of mental health care and its costs. The results of this study component are being prepared for publication.

(c) To test both instrument-specific and cross-instrument hypotheses. Full details of this stage of the study will be published in due course.

Although it is now relatively common for the authors of outcome scales to publish details of scale reliability in the original language, it is rare for the authors of translated versions to repeat the reliability exercise in the new languages, or indeed to do more than undertake a literal translation. This study therefore aims to undertake the conversion and cultural adaptation of each of the five main study scales into all the study languages in a comprehensive and scientifically rigorous manner.

METHOD

Study sites

The criteria used to identify study centres are given in full in Becker *et al* (2000, this supplement). The criteria were similar to those employed by Dowrick *et al* (1998). Six partners in five centres joined forces for this collaborative study, with the teams located in Amsterdam, Copenhagen, London (Centre for the Economics of Mental Health and Section of Community Psychiatry, Institute of Psychiatry), Santander and Verona. Full details of the general population characteristics of the study sites are given in Becker *et al* (2000, this supplement).

Case identification

Cases included in the study were adults aged 18–65, selected as representative of all people suffering from schizophrenia utilising mental health services in each of the five study sites. Study samples were identified either from psychiatric case registers (in Copenhagen and Verona) or case-loads of local special mental health services (in-patient, out-patient and community). Patients included had been in contact with

mental health services during the 3 months before the start of the study in 1997. Patients with a clinical diagnosis of any ICD–10 categories F20–F25 were eligible to enter screening, undertaken with the Item Group Checklist (IGC), which is part of the Schedule for Clinical Assessment in Neuropsychiatry (SCAN) developed by the World Health Organization (1992). Only patients with an ICD–10 F20 research diagnosis were finally included.

The exclusion criteria were: current residence in prison, secure residential services or hostels for long-term patients; co-existing mental retardation, primary dementia or other severe organic disorder; and extended in-patient treatment episodes longer than one year. The numbers of patients finally included in the study varied from 52 to 107 between the five sites, with a total of 404 for the study as a whole.

Outcome scales

The study included the conversion of five scales from their original language into the other four study languages. The scales are: the Camberwell Assessment of Need – European Version (CAN–EU), the Client Socio-Demographic and Service Receipt Inventory – European Version (CSSRI–EU), the Involvement Evaluation Questionnaire – European Version (IEQ–EU), the Lancashire Quality of Life Profile – European Version (LQoLP–EU), and the Verona Service Satisfaction Scale – European Version (VSSS–EU). These instruments were based on the following original versions: the Camberwell Assessment of Need (Phelan *et al*, 1995), the Client Service Receipt Inventory (Beecham & Knapp, 1992), the Involvement Evaluation Questionnaire (Schene *et al*, 1998), the Lancashire Quality of Life Profile (Oliver, 1991), and the Verona Service Satisfaction Scale (Ruggeri & Dall'Agnola, 1993). The CAN–EU reliability results are presented in this paper, and the results for the other scales appear in this supplement in the papers by Schene *et al* (2000), Gaite *et al* (2000) and Ruggeri *et al* (2000). Reliability tests were not appropriate for the CSSRI–EU, and its development is described by Chisholm *et al* (2000, this supplement).

Two other groups of questionnaires were used in the main study. The first group consisted of a number of instruments which had been developed previously by other authors. Local services were described using the European Service Map-

ping Schedule (ESMS) (Johnson *et al*, 1998). The Brief Psychiatric Rating Scale (Overall & Gorham, 1962) was used to measure symptomatology. Disability was measured by the Global Assessment of Functioning (American Psychiatric Association, 1987). These were not converted to different languages and were used or produced in English. Second, we also used instruments documenting the sampling process (Prevalence Cohort Data Sheet), area socio-demographic descriptors (Area Socio-Demographic Data Sheet) and patients' psychiatric history (Psychiatric History Data Sheet). These were developed for the purpose of the study in English, and all are available from the first author on request. Becker *et al* (1999) describe the study and the methodology employed.

Interviewing and data preparation

All interviewers received training at the Institute of Psychiatry, London, UK, in the use of SCAN and the other study instruments. There were regular contacts to ensure standard use of instruments and a series of study co-ordinating meetings. Data consistency and homogeneity were ensured by the co-ordinating centre (in London) preparing the SPSS templates used at all the participating sites. Consistent data structures were adhered to.

Reliability assessment procedures

Reliability testing was conducted on several levels depending on the nature of the responses involved and whether the instruments are administered as interviews or questionnaires. Three kinds of reliability test were used: (a) Cronbach's α statistic, to estimate the internal consistency of scales and sub-scales consisting of more than one item; (b) Cohen's κ statistic to estimate the interrater reliability and test–retest reliability of single items where these are expressed as binary variables; and (c) intraclass correlations, to estimate the interrater reliability and test–retest reliability of scales and sub-scales. These statistics are discussed in Streiner & Norman (1995). Each step in the analysis was described in an analysis protocol which was followed by all sites.

First, summary statistics were computed for each site, and differences in sample variances were explored using the Levene test (Levene, 1960). Cronbach's α was computed for each site, and for the pooled sample, and a test for differences

in α values between sites was performed (Feldt *et al*, 1987). Intraclass correlation coefficients (ICCs) were computed by maximum likelihood estimation of a variance components model, with patients entered as random effects, and (in the case of pooled estimates), site entered as a fixed effect. The data for each patient were either all time 1 ratings (for interrater reliability), or all ratings by the first rater (for test–retest reliability). All available data were used for these analyses, including cases where only one rating was present; however, values of *n* quoted in those tables which relate to reliability are the numbers of complete pairs. Only at one site (Verona) were there sufficient raters to estimate a specific interrater component of variance. However, for consistency in estimation between the sites, rater was not specifically included in the model. Interrater variance is thus reflected in the ICC by being incorporated in the error variance.

The ratio of the between-patient component of variance to total variance was used to estimate the ICC, and the delta technique (Dunn, 1989) was used to obtain standard errors for the ICC from the variance–covariance matrix for the components. Fisher's Z transformation was applied (Donner & Bull, 1983), and differences between sites were then tested for significance by the method of weighting (Armitage & Berry, 1994), before transforming back to the ICC scale. The standard error of measurement was obtained from the 'error' component of variance. Pooled reliability of the individual items was estimated, without between-site testing. Finally, a paired *t*-test on the test–retest data was carried out in order to assess systematic changes from time 1 to time 2.

For reasons of comparability, all sites used the same procedure and the same software for all instruments: SPSS 7.5 or higher, the Amsterdam α-testing program ALPHA.EXE based on Feldt *et al* (1987), and EXCEL for tests of the homogeneity of ICCs.

Test–retest reliability was conducted at intervals of between 1 and 2 weeks, although in a few cases up to 7 weeks elapsed, depending on the practicalities of contacting patients. The same rater interviewed at test and retest. For the CAN, patients' responses are rated by an interviewer, and therefore interrater reliability is an issue as well as test–retest. For interrater reliability, a second rater present at the interviews rated in parallel with the primary interviewer who asked the questions of the patient. The numbers of raters at time 1 and time 2 were as follows: Amsterdam: 4, 4; Copenhagen: 5, 5; London: 2, 5; Santander: 3, 3; Verona: 11, 13.

Answers to the service receipt part of the CAN–EU depend on answers to Part 1 (presence of a need), and therefore interrater reliability for the subsequent sections is hard to define, since the parallel rater has no control over the flow of questions. Furthermore, the service receipt sections are mainly useful in a clinical situation. For these reasons, and in common with reliability testing of the original CAN, these sections have not been analysed here for reliability purposes.

RESULTS

Table 1 shows summary statistics for the primary rater at time 1, including the results of a homogeneity of variance test. Mean total needs and unmet needs differed significantly between sites, with Amsterdam and London tending to show higher values than the other three sites on both measures. This will be discussed further in a paper on the substantive results of the study. Of more relevance to the reliability estimation is the lack of homogeneity in variance for met and total needs, as shown by the Levene test in Table 1.

The α coefficients, which reflect correlations between individual CAN items, were moderate to low, as shown in Table 2. For total needs, the pooled α was 0.64 (95% CI 0.58–0.70). Only for met needs (pooled mean 0.48, 95% CI 0.40–0.56) was there strong evidence for differences between sites, with Santander having the lowest value at 0.16. For unmet needs (pooled mean 0.58, 95% CI 0.51–0.64) the differences were less marked, but

Table 1 The Camberwell Assessment of Need – European Version (CAN–EU): total, met and unmet needs in the pooled sample and by site; results of primary rater at time 1

	Pooled n=400		Amsterdam n=59		Copenhagen n=51		London n=84		Santander n=100		Verona n=106		Test of equality of means (P-value)	Test of equality of s.d. (P-value)
	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.		
Total needs	5.35	3.07	6.31	2.96	5.19	3.39	5.95	2.78	4.81	2.52	4.93	3.28	<0.01	0.02
Met needs	3.56	2.26	3.79	2.38	3.86	2.46	3.77	1.88	3.19	1.64	3.46	2.80	0.27	<0.01
Unmet needs	1.79	1.98	2.52	2.02	1.33	1.96	2.18	2.09	1.61	1.68	1.48	2.00	<0.01	0.42

Table 2 Internal consistency of the Camberwell Assessment of Need – European Version (CAN–EU): α coefficients (95% CI) in the pooled sample and by site

	Pooled n=327	Amsterdam n=57	Copenhagen n=42	London n=69	Santander n=94	Verona n=65	Test of equality of α (P-value)
Total needs	0.64 (0.58–0.70)	0.64 (0.49–0.76)	0.73 (0.59–0.83)	0.55 (0.39–0.69)	0.61 (0.48–0.71)	0.67 (0.54–0.77)	0.42
Met needs	0.48 (0.40–0.56)	0.58 (0.41–0.72)	0.54 (0.31–0.72)	0.36 (0.11–0.56)	0.16 (−0.10 to 0.39)	0.62 (0.47–0.74)	0.03
Unmet needs	0.58 (0.51–0.64)	0.52 (0.32–0.68)	0.70 (0.56–0.82)	0.55 (0.38–0.69)	0.57 (0.43–0.69)	0.58 (0.42–0.72)	0.04

α estimated only for cases with complete data (22 items).

Table 3 Test–retest reliability of the Camberwell Assessment of Need – European Version (CAN–EU) summary scores in the pooled sample and by site

	Pooled n=247		Amsterdam n=45		Copenhagen n=33		London n=53		Santander n=48		Verona n=68		Test of equality of ICCs (P-value)
	ICC	(s.e.)$_m$	ICC	(s.e.)$_m$	ICC	(s.e.)$_m$	ICC	(s.e.)$_m$	ICC	(s.e.)$_m$	ICC	(s.e.)$_m$	
Total needs	0.85	1.12	0.81	1.23	0.90	1.02	0.82	1.13	0.89	0.82	0.85	1.26	0.33
Met needs	0.69	1.22	0.65	1.35	0.75	1.15	0.71	1.01	0.75	0.81	0.65	1.59	0.70
Unmet needs	0.78	0.90	0.72	1.08	0.84	0.84	0.86	0.76	0.83	0.69	0.69	1.06	0.03

ICC, Intraclass correlation coefficient; (s.e.)$_m$, standard error of measurement (square root of error component of variance).

Copenhagen showed a somewhat higher value than the other sites, at 0.70.

The ICCs between the two time points are given in Table 3, which shows that test–retest reliability is at an acceptable level. There were no significant differences between sites, except for unmet needs. Pooled values were 0.85 (95% CI 0.82–0.88) for total needs, 0.69 (95% CI 0.63–0.74) for met needs and 0.78 (95% CI 0.74–0.82) for unmet needs.

For estimating κ coefficients for the individual items, unmet, met and total needs were each expressed as binary variables in turn. Table 4 shows that κ coefficients for test–retest reliability were high for total needs (0.55–0.84), and moderately high for met needs (0.40–0.76, excluding κ for met needs for drugs, which was zero) and unmet needs (0.34–0.85). Standard errors for these κ estimates were typically 0.06, 0.08 and 0.09 respectively. Only one item had a κ coefficient below 0.4 (unmet needs for physical health).

There was evidence of site differences in interrater reliability (Table 5) for total, met and unmet needs. However, all the sites had coefficients for total and met needs above 0.8. In the case of unmet needs, coefficients for Amsterdam, Santander and Verona were under 0.8, although they were all still over 0.65. The pooled estimates were 0.93 (95% CI 0.92–0.95) for total needs, 0.85 (95% CI 0.81–0.87) for met needs and 0.79 (95% CI 0.75–0.83) for unmet needs.

Interrater reliability for individual items, pooled over sites (Table 6), was very good for total needs (0.75–0.99) and met needs (0.57–0.95), and moderately good

Table 4 Test–retest reliability of the Camberwell Assessment of Need – European Version (CAN–EU) items in the pooled sample

CAN no.	Area of need	% agreement	Total needs	Met needs	Unmet needs
1	Accommodation	92	0.84	0.76	0.60
2	Food	91	0.82	0.74	0.77
3	Looking after the home	85	0.66	0.60	0.75
4	Self-care	92	0.74	0.62	0.62
5	Daytime activities	82	0.74	0.60	0.73
6	Physical health	84	0.70	0.66	**0.34**
7	Psychotic symptoms	86	0.75	0.66	0.70
8	Information about condition and treatment	80	0.69	**0.49**	**0.50**
9	Psychological distress	72	0.64	**0.43**	**0.55**
10	Safety to self	93	0.76	**0.57**	0.68
11	Safety to others	95	0.60	**0.51**	**0.45**
12	Alcohol	96	0.75	0.70	0.62
13	Drugs	97	0.65	–¹	0.85
14	Company	73	**0.59**	**0.44**	**0.53**
15	Intimate relationships	82	**0.55**	**0.54**	**0.51**
16	Sexual expression	88	**0.59**	**0.40**	**0.58**
17	Child care	97	0.76	0.61	0.79
18	Basic education	92	0.70	0.67	**0.54**
19	Using a telephone	98	0.84	**0.59**	0.74
20	Travel	93	0.79	0.72	0.78
21	Money	86	0.74	0.60	0.62
22	Welfare benefits	91	0.68	**0.53**	0.72

Percentage agreement over all three needs and κ coefficients; $\kappa < 0.6$ in bold type.
1. Low base rate (4 and 6 cases of met need at times 1 and 2 respectively) with no agreement.

Table 5 Interrater reliability of the Camberwell Assessment of Need – European Version (CAN–EU) in the pooled sample and by site

	Pooled n=274		Amsterdam n=47		Copenhagen n=40		London n=79		Santander n=50		Verona n=58		Test of equality of ICCs (P-value)
	ICC	(s.e.)$_m$	ICC	(s.e.)$_m$	ICC	(s.e.)$_m$	ICC	(s.e.)$_m$	ICC	(s.e.)$_m$	ICC	(s.e.)$_m$	
Total needs	0.93	0.77	0.90	0.91	0.93	0.84	0.99	0.29	0.93	0.65	0.91	0.93	<0.01
Met needs	0.85	0.86	0.83	0.95	0.82	0.99	0.96	0.38	0.82	0.72	0.83	1.08	<0.01
Unmet needs	0.79	0.85	0.71	1.01	0.84	0.75	0.98	0.33	0.77	0.77	0.68	1.06	<0.01

ICC, intraclass correlation coefficient; (s.e.)$_m$, standard error of measurement (square root of error component of variance).

Table 6 Interrater reliability of the Camberwell Assessment of Need – European Version (CAN–EU) items in the pooled sample

CAN no.	Area of need	% agreement	Total needs	Met needs	Unmet needs
1	Accommodation	95	0.94	0.83	**0.41**
2	Food	95	0.88	0.86	0.72
3	Looking after the home	94	0.86	0.84	0.76
4	Self-care	98	0.95	0.94	0.66
5	Daytime activities	90	0.93	0.79	0.73
6	Physical health	95	0.92	0.90	0.73
7	Psychotic symptoms	86	0.83	0.66	0.64
8	Information about condition and treatment	87	0.92	0.69	**0.45**
9	Psychological distress	89	0.94	0.78	0.73
10	Safety to self	96	0.82	0.75	0.83
11	Safety to others	99	0.84	0.95	0.83
12	Alcohol	99	0.89	0.92	0.75
13	Drugs	98	0.94	**0.57**	0.80
14	Company	88	0.97	0.74	0.77
15	Intimate relationships	92	0.94	0.77	0.78
16	Sexual expression	91	0.84	0.72	**0.57**
17	Child care	96	0.79	0.61	0.69
18	Basic education	97	0.84	0.88	0.71
19	Using a telephone	98	0.75	0.83	0.66
20	Travel	98	0.99	0.90	0.83
21	Money	93	0.91	0.87	0.68
22	Welfare benefits	94	0.86	0.68	0.74

Percentage agreement over all three needs and κ coefficients; $\kappa < 0.6$ in bold type.

for unmet needs (0.41–0.83). Standard errors were typically 0.04, 0.07 and 0.10 respectively. All items had κ coefficients over 0.4.

Paired sample *t*-tests revealed a tendency for a decrease in the rating of total needs over time, pooled across sites, but this was significant only at a borderline level (P=0.053). At individual sites, there were no significant differences between mean scores at test and retest, with the exception of total needs in Verona, where the time 2 total values were rated lower: 4.39 at time 2 compared with 5.11 at time 1, difference 0.72 (95% CI 0.52–0.92), P=0.001. This is most likely to be a chance finding, given the large number of tests employed.

DISCUSSION

This paper has described how the reliability of the CAN–EU was tested in five different centres. Face validity was not specifically tested. However, focus groups were used in the translation process and these indicated that the CAN–EU was largely acceptable in its format and content. This confirmed the attainment of face validity for the original English version (Phelan *et al*, 1995). Very high internal consistency between the items for the CAN–EU is not expected or even necessarily desirable, and the moderate levels of α are quite acceptable in this context. Indeed, they are not surprising, given the diverse range of needs assessed with the instrument, which were deliberately selected to cover the entire

range of difficulties commonly encountered by people suffering from severe mental illnesses. In this context the α coefficients are not that informative, but have been reported so as to maintain consistency with the other papers in this supplement.

The very low value for α for met needs for Santander is interesting, and may be connected with the lower level and smaller degree of variation in met needs at that particular site, as shown in Table 1. Alternatively it may be connected with one particular item, "help with psychotic symptoms". When this item is removed, the α is doubled to 0.32, more in keeping with its value at other sites.

Overall, the test–retest reliability is at least moderately good, although usually lower than interrater reliability. Lack of reliability may, in some cases, have been due to changes in patient status that occurred between the two time points, in addition to lack of consistency in a patient's responses from one time point to the next. However, interviews were generally made within intervals of 1–2 weeks, so real changes in status were unlikely.

Interrater reliability is excellent, with only a slight fall-off for unmet needs. Although there were significant differences between sites, all values of interrater reliability coefficients were over 0.65. The two slightly lower reliabilities for unmet needs (Amsterdam and Verona) are due to higher standard errors of measurement rather than the differences in variances between the samples shown in Table 1. It should be noted that Verona had a larger pool of primary raters and, in this respect, the data from that site may more realistically reflect the range of raters who might use the instrument in practice. The very small standard error of measurement in London might reflect a longer history of CAN training and use.

For individual items, both for test–retest and interrater, the items with the lowest κ values tend to be those where there are low base rates for the need: for example, drugs. Two items showed both low κ and low percentage agreement in the test–retest comparison: 'psychological distress' (item 9) and 'company' (item 14). These two items are not of this character, and there seems to be no obvious pattern in the inconsistent responses over time. It may be that these two items are hard to rate because they are very much related to mood and reflect relatively transient situations.

A point which applies generally, both over time and also between raters, is that there are greater levels of agreement for total needs than for the component items. However, the very skewed nature of the data relating to individual items (i.e. the low base-rates in many cases) makes reliability tests problematic. Indeed it reduces the feasibility of analysing these variables individually, except in very large samples.

Mean scores did not differ significantly between test and retest, except for one score in one site. The pooled ratings for total needs did decrease slightly over time (at a borderline level of significance) but in general there is little evidence for substantial increase or decrease over time, a problem which might occur if patients tended to reflect on and modify their ideas following an interview. In these respects the CAN can be seen to be stable over time.

This analysis has concentrated on the three total needs scores, rather than individual items. This is because the 22 CAN items, while clearly important in considering the needs of individual patients, are of limited use for analytical purposes when treated in isolation, since most of them are encountered infrequently in individual cases. Similarly, the sections of the CAN relating to levels of formal and informal care received, and formal care required, are most relevant for clinical rather than research purposes. With large samples, the data on particular needs and on care required or received could be analysed, but such samples have hitherto been scarce.

Bearing in mind these caveats, we suggest that the summary scores for the CAN–EU (total, met and unmet needs) are generally reliable over time and between raters. Despite some evidence for differences in levels of reliability between sites for unmet needs at test–retest, and between raters for all three total scores, the results are good at each site, and encouraging for the use of this instrument in its five translations.

ACKNOWLEDGEMENTS

REFERENCES

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

Armitage, P. & Berry, G. (1994) *Statistical Methods in Medical Research* (3rd edn). Oxford: Blackwell Scientific.

Becker, T., Knapp, M., Knudsen, H. C., et al (1999) The EPSILON study of schizophrenia in five European countries: design and methodology for standardising outcome measures and comparing patterns of care and service costs. *British Journal of Psychiatry*, **175**, 514–521.

__, __, __, et al (2000) Aims, outcome measures, study sites and patient sample. Epsilon Study I. *British Journal of Psychiatry*, **177** (suppl. 39), s1–s7.

Beecham, J. & Knapp, M. (1992) Costing psychiatric interventions. In *Measuring Mental Health Needs* (eds G. Thornicroft, C. R. Brewin & J. Wing). London: Gaskell.

Chisholm, D., Knapp, M. R. J., Knudsen, H. C. et al (2000) The Client Socio-Demographic and Service Receipt Inventory: development of an instrument for international research. EPSILON Study 5. *British Journal of Psychiatry*, **177** (suppl. 39), s28–s33.

Donner, A. & Bull, S. (1983) Inferences concerning a common intraclass correlation coefficient. *Biometrics*, **39**, 771–775.

Dowrick, C., Casey, P., Dalgard, O., et al (1998) Outcomes of Depression International Network (ODIN). Background, methods and field trials. *British Journal of Psychiatry*, **171**, 359–363.

Dunn, G. (1989) *Design and Analysis of Reliability Studies.* London: Edward Arnold.

Feldt, L. S., Woodruff, D. J. & Salih, F. A. (1987) Statistical inference for coefficient alpha. *Applied Psychological Measurement*, **11**, 93–103.

Gaite, L., Vázquez-Barquero, J. L., Arriaga Arrizabalaga, A., et al (2000) Quality of life in schizophrenia: development, reliability and internal consistency of the Lancashire Quality of Life Profile – European Version. EPSILON Study 8. *British Journal of Psychiatry*, **177** (suppl. 39), s49–s54.

Johnson, S., Salvador-Carulla, L. & the EPCAT group (1998) Description and classification of mental health services: a European perspective. *European Psychiatry*, **13**, 333–341.

Knudsen, H. C., Vázquez-Barquero, J. L., Welcher, B., et al (2000) Translation and cross-cultural adaptation of outcome measurements for schizophrenia. EPSILON Study 2. *British Journal of Psychiatry*, **177** (suppl. 39), s8–s14.

Levene, H. (1960) Tests for equality of variances. In *Contributions to probability and statistics: Essays in honor of Harold Hotelling* (eds I. Olkin, S. G. Ghurye, W. Hoeffding, *et al*), pp. 278–292. Stanford, CA: Stanford University Press.

Oliver, J. (1991) The social care directive: development of a quality of life profile for use in the community services for the mentally ill. *Social Work and Social Sciences Review*, **3**, 5–45.

Overall, J. & Gorham, D. (1962) Brief Psychiatric Rating Scale. *Psychological Reports*, **10**, 799–812.

Phelan, M., Slade, M., Thornicroft, G., et al (1995) The Camberwell Assessment of Need: the validity and reliability of an instrument to assess the needs of people with severe mental illness. *British Journal of Psychiatry*, **167**, 589–595.

Ruggeri, M. & Dall'Agnola, R. (1993) The development and use of the Verona Expectations for Care Scale (VECS) and the Verona Service Satisfaction Scale (VSSS) for measuring expectations and satisfaction with community-based psychiatric services in patients, relatives and professionals. *Psychological Medicine*, **23**, 511–523.

—— , Lasalvia, A. Dall'Agnola, R., *et al* **(2000)** Development, internal consistency and reliability of the Verona Service Satisfaction Scale — European Version. EPSILON Study 7. *British Journal of Psychiatry*, **177** (suppl. 39), s41–s48.

Schene, A. H., van Wijngaarden, B & Koeter, M. W. J. (1998) Family caregiving in schizophrenia: domains and distress. *Schizophrenia Bulletin*, **24**, 618–618.

—— , Koeter, M., van Wijngaarden, B., *et al* **(2000)** Methodology of a multi-site reliability study. EPSILON Study 3. *British Journal of Psychiatry*, **177** (suppl. 39), s15–s20.

Streiner, D. & Norman, G. (1995) *Health Measurement Scales: A Practical Guide to their Development and Use.* Oxford: Oxford University Press.

van Wijngaarden, B., Schene, A. H., Koeter, M., et al (2000) Caregiving in schizophrenia: development, internal consistency and reliability of the Involvement Evaluation Questionnaire — European Version. EPSILON Study 4. *British Journal of Psychiatry*, **177** (suppl. 39), s21–s27.

World Health Organization (1992) *Schedules for Clinical Assessment in Neuropsychiatry* (ed.-in-chief J. K. Wing). Geneva: WHO.

PAUL McCRONE, MORVEN LEESE, PhD, GRAHAM THORNICROFT, MRCPsych, GWYN GRIFFITHS, SARAH PADFIELD, Section of Community Psychiatry (PRiSM), Institute of Psychiatry, King's College London, UK; AART SCHENE, MD, Department of Psychiatry, Academic Medical Centre, Amsterdam, The Netherlands; HELLE CHARLOTTE KNUDSEN, MD, Institute of Preventive Medicine, Copenhagen University Hospital, Denmark; JOSÉ LUIS VÁZQUEZ-BARQUERO, FRCPsych, Clinical and Social Psychiatry Research Unit, University of Cantabria, Santander, Spain; ANTONIO LASALVIA, MD, Department of Medicine and Public Health, University of Verona, Italy; IAN R. WHITE, Medical Statistics Unit, London School of Hygiene and Tropical Medicine, London, UK

Correspondence: Paul McCrone, Section of Community Psychiatry (PRiSM), Institute of Psychiatry, King's College, London, De Crespigny Park, Denmark Hill, London SE5 8AF. Tel: 020 7848 0711; fax: 020 7277 1462