Caregiving in schizophrenia: development, internal consistency and reliability of the Involvement Evaluation Questionnaire – European Version

EPSILON Study 4

BOB van WIJNGAARDEN, AART H. SCHENE, MAARTEN KOETER, JOSÉ LUIS VÁZQUEZ-BARQUERO, HELLE CHARLOTTE KNUDSEN, ANTONIO LASALVIA, PAUL McCRONE and the EPSILON STUDY GROUP

Background In international research on the consequences of psychiatric illnesses for relatives of patients, the need for an internationally standardised measure has been identified.

Aims To test the internal consistency and the test–retest reliability of the Involvement Evaluation Questionnaire (IEQ) in five European countries.

Method The IEQ was administered twice to a sample of relatives or friends of patients with an ICD–10 diagnosis of schizophrenia. Reliability was tested using Cronbach's α, intraclass correlation coefficients and standard error of measurement. Reliability estimates were tested between sites.

Results Test sample sizes ranged from 30 to 90 across sites, and retest sample sizes ranged from 21 to 77. Cronbach's α values of IEQ sub-scales and sumscore were substantial at most sites; but at two, α values were moderate. Intraclass correlation coefficients were substantial to high at all sites. The standard errors of measurement differed across sites, indicating differences in performance.

Conclusion The reliability of the IEQ in five languages varies across sites, but is sufficiently high in at least four out of five.

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

Severe mental illness such as schizophrenia often imposes a considerable burden on the patients who suffer from it, as well as on their families and the wider society (Hatfield & Lefley, 1987). Patients' symptoms and their often poor personal and social functioning have a far-reaching impact on their own quality of life, while the nature of schizophrenia and its early onset often impoverish the lives and lifestyles of those who care for them. This issue has become even more important because of ongoing changes in the organisation of mental health care services: in particular, the shift from hospital-based to community-based services has resulted in some of the caring for (mostly) adults falling again on their family (or others involved).

When this is the case, normal and reciprocal caregiving between two (or more) adults changes into caregiving where one adult is dependent on the care of the other(s). The recipient of the care is disabled by a mental disorder with a long-term course; and for the caregiver(s), their caregiving role is out of synchrony with the appropriate stage of their own lifecycle (Schene et al, 1996). The consequences resulting from such a non-synchronised caregiving situation have for a long time been described as family or caregiver 'burden'. However, this definition concentrates too much on the negative aspects of caring. Although mental disorders, particularly if they are long-term, disrupt family life, not all relatives experience their caring role as burdensome. Because of this we prefer the more neutral term 'caregiving consequences'.

Research on the consequences of mental illness for patients' relatives can be divided into three distinctive periods. First, starting in the 1950s, researchers described in detail all the different consequences for family members (Wing et al, 1959;

Mandelbrote & Folkard, 1961). They paid particular attention to negative aspects. In the second period, starting in the early 1970s, family burden became one of the outcome measures in mental health service evaluation (Fenton et al, 1979; Tessler et al, 1980). Instruments were developed and used in studies that compared community approaches with the more classical approaches (Schene et al, 1994). In the third period, beginning in the early 1980s, interventions or treatment programmes with a psycho-educational approach which aimed at a reduction of family burden, family stress or expressed emotion became the central point of interest (Kuipers & Bebbington, 1988). Recently, a fourth period has started, with the emphasis on relatives' needs, perceptions and attributions (Barrowclough et al, 1996; Scazufca & Kuipers, 1996), coping style (Budd et al, 1998; Magliano et al, 1998) and mental health (Schene et al, 1998a).

In 1994, Schene et al reviewed instruments measuring family or caregiver burden. They described 21 instruments, mostly from English-speaking countries, of which 15 were developed in the 5-year period prior to the review, an indication of the growing importance of caregiving consequences. Since 1994, more new instruments have been developed, such as the Experience of Caregiving Inventory (Szmukler et al, 1996), the Perceived Family Burden Scale (Levine et al, 1996), and a generic instrument to assess the experience of caregiving (Schofield et al, 1997). Most of these instruments are interviews that can only be administered by interviewers, which make them time-consuming and expensive; in addition, they differ considerably in the number of items and domains covered. As far as we know, none of the above instruments has been translated from English into other languages, which limits their application, especially in Europe with its variety in languages. Therefore the European Psychiatric Services: Inputs Linked to Outcome Domains and Needs (EPSILON) Study also included the translation and validation of an instrument to assess caregiving consequences (Becker et al, 1999).

The instrument chosen was the Involvement Evaluation Questionnaire (IEQ) (Schene & van Wijngaarden, 1992; Schene et al, 1996) (see Appendix). The IEQ was chosen because: (a) it is easy to administer; and (b) it is based on a variety of instruments developed in earlier years and covers a broad domain of caregiving consequences

(Schene *et al*, 1994). In this paper we shall describe the development of the IEQ and the results of earlier instrument testing; then we shall describe the translation procedure and reliability testing used in the EPSILON Study.

INVOLVEMENT EVALUATION QUESTIONNAIRE

Development of the IEQ: a brief history

The development of the IEQ started in 1987. In a randomised controlled trial, comparing in-patient and day patient treatment, an instrument was needed which could measure caregiving consequences (Schene *et al*, 1993). Since no such instrument existed in The Netherlands we developed one ourselves, starting with an extensive review of all empirical studies on family burden (Schene, 1986; Schene *et al*, 1994). We were then able to define the following desirable characteristics for such a research instrument: (a) it should be a questionnaire, covering all important domains; (b) it should be valid, reliable, easy to understand, and not time-consuming; and (c) it should cover a limited time frame and be sensitive to changes.

Items collected in the review of literature and existing instruments formed the basis of an IEQ item pool (Schene *et al*, 1996). This item pool was further extended with items emerging from interviews with professionals. A series of draft versions were piloted, and adapted if necessary. Since the principal aim in developing the IEQ was a reliable measure which would be sensitive to change, items relating to stigma, guilt, social network loss, suicide attempts by patients and other events that either happen rarely or are not sensitive to change were dropped.

The first version of the IEQ was used in four Dutch studies conducted between 1987 and 1990: (a) a comparative study of day treatment v. in-patient treatment ($n=80$); (b) a study of patients who had recently attempted suicide ($n=80$); (c) a study in the psychiatric department of a general hospital ($n=80$); and (d) a study of acute psychiatric patients in a community mental health centre ($n=30$). A psychometric analysis of these data and an updated literature review (in particular with regard to depression and families) resulted in the construction of a second Dutch version in 1992, which in the same year was translated into

English, followed by translations into Portuguese, Finnish and German. These translations, however, did not follow the procedures used in the EPSILON Study (see Knudsen *et al*, 2000, this supplement) and therefore have a different status from the translations used in the EPSILON Study.

Structure and item content of the IEQ

The IEQ is a 31-item questionnaire which is completed by the caregiver. The items relate to the encouragement and care that the caregiver has to give to the patient, to personal problems between patient and caregiver, and to the caregiver's worries, coping and subjective burden. All items are scored on 5-point Likert scales (never, sometimes, regularly, often, always). Carers who have had less than one hour's contact with the patient during the 4 weeks previous to the completion of the instrument skip items that refer to actual help and encouragement, because those items are considered to be not applicable.

A total of 27 items can be summarised in four distinct sub-scales: (a) *tension* (nine items), which refers to the strained interpersonal atmosphere between patient and relatives; (b) *supervision* (six items), which refers to the caregiver's tasks of guarding the patient's medicine intake, sleep and dangerous behaviour; (c) *worrying* (six items), which covers painful interpersonal cognitions, such as concern about the patient's safety and future, general health and health care; and (d) *urging* (eight items), which refers to activation and motivation; for instance, stimulating the patient to take care of her/himself, to eat enough and to undertake activities. In addition, a 27-item sumscore can be computed. The items and sub-scales of the IEQ are presented in the Appendix. In this paper the reliability testing of the sub-scales and total score will be described.

For research purposes, the IEQ is normally extended with extra modules. In the EPSILON Study the following modules were used: (a) 15 socio-demographic and contact variables, such as age, gender, household composition, and amount of contact between patient and respondent; (b) eight items on extra financial expenses incurred on behalf of the patient; (c) three items on the caregiver's use of professional help; (d) 11 items on the consequences for the patient's children; (e) one open question

for comments and additions; and (f) the 12-item General Health Questionnaire as a measure of caregiver's distress (Goldberg & Williams, 1988; Dutch translation: Koeter & Ormel, 1991; Spanish translation: Gaite, 1997; Danish translation: Nielsen, undated; Italian translation: Servizio di Psicologia Medica Verona, undated).

The entire set takes about 20–30 minutes to complete; the IEQ module itself, about 10 minutes. The IEQ can also be administered as a structured (telephone) interview. The IEQ can be used as both a research and a clinical instrument. Different ways of scoring are recommended, depending on the use of the IEQ: for research purposes the average scale scores based on the 5-point Likert scales are used for the computation of correlations with other instruments; in clinical use, where average scale scores are not easy to interpret, item scores are dichotomised to 0=('never' or 'sometimes') and 1=('regularly', 'often' or 'always'). In this case, the scale scores directly reflect the number of consequences that are experienced. Also, major changes in consequences can easily be detected when an item score changes from 0 to 1 or the reverse. Collected data can be interpreted on both sub-scales at item level.

Validity, reliability and applicability of the Dutch version of the IEQ

The 1992 version of the IEQ was tested in two Dutch studies, one among 680 members of an organisation of relatives of patients with psychotic disorders (Schene & van Wijngaarden, 1995), and one among 260 relatives of patients with affective disorders (van Wijngaarden *et al*, 1996). From the results of both studies it was concluded that the IEQ adequately covers all major domains of caregiving consequences. The four identified sub-scales were obtained by factor analysis and cover the caregiver consequences of both psychotic and affective disorders.

The reliability of the IEQ proved to be satisfactory in the Dutch samples. The internal consistency (Cronbach's α) ranges from 0.74–0.85 for the four sub-scales to 0.90 for the sumscore; test–retest effects were not found, and there were indications that the IEQ is sensitive to change (van Wijngaarden *et al*, 1996). Validity was also satisfactory. The construction process as described above secured the content validity of the IEQ. This validity was confirmed by a qualitative analysis of the open question no. 81, in which respondents were

asked to add any issue that bothered, stressed or satisfied them in their relationship with the patient that was not already covered by the IEQ. Analysis of the replies to this question in about 1000 questionnaires did not reveal any missing domains or variables (Schene & van Wijngaarden, 1993). In addition, separate analyses of the data regarding relatives of psychotic or depressed patients revealed factor structures that were very comparable with that of the combined sample. This consistency can also be considered as an indicator of content validity.

In an analysis of the relation between (a) caregiving consequences and the characteristics of the patient, the caregiver and their relationship, and (b) caregiving consequences and caregiver distress, it was found that caregiver consequences were related to the patient's symptomatology and the amount of time spent together. A path analysis revealed that caregiving consequences measured with the IEQ explained a substantial part of the relation between the caregiver's distress and the patient, caregiver and relationship characteristics (Schene *et al*, 1998*b*), emphasising the relevance of the concept.

Finally, the applicability of the IEQ proved to be good. Response rates were high, ranging from 70% to 81% (mailed survey with one reminder). The quality of the response was also high. Of 960 completed questionnaires, only 25 (2.6%) could not be used due to missing values.

RELIABILITY OF THE IEQ

Sample

The IEQ was completed by relatives (or other significant persons) of patients with an ICD–10 diagnosis of schizophrenia in Amsterdam, Copenhagen, London, Santander and Verona. For details on catchment areas and inclusion criteria see particular papers (Kastrup, 1998; Schene *et al*, 1998*a*; Tansella *et al*, 1998; Thornicroft & Goldberg, 1998; Vázquez-Barquero & García, 1999; Becker *et al*, 2000, this supplement).

All patients who entered the study were asked to name a relative or other significant person who could be asked to complete the IEQ. The number of patients in each site included in the study varied from 52 to 107, with a total of 404. However, not all of them were able to indicate someone who could complete the IEQ, either because they were not in close contact with others or

because they refused to cooperate. In Amsterdam and Copenhagen, where more patients live alone than in the other sites, the attrition was highest (about 40%). The number of respondents who completed the IEQ ranged from 30 (Copenhagen) to 78 (Santander), with a total of 285, which means an average response of about 70%. The number of retests ranged from 21 (Copenhagen) to 73 (London).

In the reliability protocol it was stated that at least 50 test–retest sets should be necessary for reliability testing. Once it became clear that the Amsterdam numbers would be lower, it was decided to do an extra sampling among 100 members of the Dutch organisation for relatives of schizophrenia patients. This sampling resulted in 52 test and 47 retest assessments, bringing the Amsterdam figures to a total of 90 tests and 77 retests.

Translation and cultural validation

The translation of the IEQ into the other languages largely followed the protocol described in this supplement by Knudsen *et al* (2000). This protocol included: (a) a translation into the four target languages by professional translators, who were informed on the content of the IEQ; (b) a discussion of this translation by the translator and the research group, leading to a revision and a list of disputed items; (c) a back-translation into Dutch by a native speaker, who also gave his/her comments on the first translation and the disputed items; (d) a comparison of the back-translation with the original IEQ, discussed by the first translator and the researchers, leading to a second revision and list of disputed items; (e) a discussion of this revision in focus groups; (f) a discussion of the focus group result by the researchers and one of the translators; (g) a third revision leading to the final version.

The focus group method is an arranged communication session among a selected group of persons who represent different parties involved. In case of the IEQ, they were representatives of patients, relatives, professionals and researchers. The conditions under which the focus groups took place were well defined (see Knudsen *et al*, 2000, this supplement).

In the IEQ focus groups the translation and content of the instrument was discussed, with special emphasis on linguistic problems, the applicability and relevance of items, redundancy, and missing items.

It was concluded that the instrument covers the domain of family burden. There were some problems with the response categories and items regarding education, type of professional help, income categories, and use of illegal drugs. The IEQ was adjusted in accordance with these comments (Knudsen *et al*, 2000, this supplement).

Methodology reliability study

For the reliability testing of the IEQ two measures were used: Cronbach's α for the internal consistency of the IEQ sub-scales, and the intraclass correlation coefficient (ICC) to estimate the test–retest reliability of the sub-scales. These estimates were computed for each site separately, and subsequently tested for inter-site differences. Since the reliability estimates are dependent on true score variance, inter-site comparisons will be affected if the variances are too different. In case of differences in score distribution, therefore, the standard errors of measurement, $(s.e.)_m$ were computed. These $(s.e.)_m$ scores, which are independent of the true score variance, were computed in two ways, using either Cronbach's α or the ICC in the formula (see Schene *et al*, 2000, this supplement). In addition, pooled reliability estimates were computed to assess overall reliability. As is pointed out by Schene *et al* (2000) in this supplement, these pooled estimates are influenced by differences in score distribution and differences in reliability between sites. Pooled estimates should be treated with caution, especially where reliabilities are generally not very high, or lower in one or more sites.

The following analysis scheme was used:

(a) test on inter-site differences in score distribution (mean and variance; ANOVA and Levene test);

(b) assessment of the site-specific reliability estimates (Cronbach's α, ICC, $(s.e.)_m$; the benchmark for substantial reliability was set to 0.70;

(c) test on inter-site differences in reliability estimates;

(d) reliability estimates from pooled data.

The reliability methodology of the EPSILON Study and the computer programs used are described in detail elsewhere in this supplement (Schene *et al*, 2000).

RESULTS

Score distributions

Table 1 presents means and standard deviations of the five samples, together with the

Table 1 Involvement Evaluation Questionnaire – European Version (IEQ–EU) sub-scales in the pooled sample and by site

Sub-scale	Pooled n=278[1]		Amsterdam n=65		Copenhagen n=25		London n=54		Santander n=77		Verona n=57		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.		
Tension	14.6	5.3	14.3	4.3	12.3	3.4	14.6	4.3	13.9	5.7	16.7	6.3	<0.01	0.10
Supervision	8.3	3.8	7.7	3.2	7.1	2.0	8.0	3.1	8.1	3.8	9.9	4.9	<0.01	<0.01
Worrying	15.6	6.3	14.3	5.8	11.8	4.1	14.2	4.8	18.8	7.5	17.0	6.2	<0.01	<0.01
Urging	15.4	6.4	14.1	5.6	12.7	3.6	16.5	7.6	15.6	6.0	16.7	6.8	<0.01	<0.01
Sumscore	50.6	16.3	46.7	14.7	41.3	9.4	49.9	15.1	53.0	16.5	56.6	18.6	<0.01	0.08

1. Samples quoted for sumscore; sample sizes varied between 278 (sumscore) and 335 (worrying).

Table 2 Internal consistency of the Involvement Evaluation Questionnaire – European Version (IEQ–EU): α coefficients (95% CI) and $(s.e.)_m$ in the pooled sample and by site

Sub-scale	Items	Pooled		Amsterdam		Copenhagen		London		Santander		Verona		Test of equality of α (P-value)
		n	α (CI) $(s.e.)_m$	n	α (CI) $(s.e.)_m$	n	α (CI) $(s.e.)_m$	n	α (CI) $(s.e.)_m$	n	α (CI) $(s.e.)_m$	n	α (CI) $(s.e.)_m$	
Tension	9	287	0.81 (0.77–0.84) 2.54	68	0.78 (0.68–0.84) 2.02	25	0.75 (0.54–0.86) 1.70	55	0.80 (0.69–0.86) 1.92	77	0.80 (0.71–0.85) 2.55	62	0.84 (0.77–0.89) 2.52	0.58
Supervision	6	285	0.77 (0.73–0.81) 1.82	69	0.80 (0.70–0.86) 1.43	25	0.73 (0.49–0.86) 1.04	54	0.68 (0.51–0.79) 1.75	77	0.75 (0.64–0.82) 1.90	60	0.82 (0.72–0.88) 2.08	0.47
Worries	6	335	0.84 (0.81–0.86) 2.52	88	0.86 (0.81–0.90) 2.17	30	0.84 (0.71–0.91) 1.64	75	0.77 (0.66–0.83) 2.30	78	0.83 (0.76–0.88) 3.09	64	0.82 (0.73–0.87) 2.63	0.55
Urging	8	291	0.79 (0.75–0.82) 2.93	70	0.82 (0.73–0.87) 2.38	25	0.71 (0.45–0.84) 1.94	55	0.86 (0.80–0.91) 2.84	77	0.68 (0.55–0.77) 3.39	64	0.81 (0.72–0.87) 2.96	0.03[1]
Sumscore	27	278	0.90 (0.88–0.92) 5.15	65	0.91 (0.87–0.94) 4.41	25	0.87 (0.77–0.93) 3.39	54	0.89 (0.84–0.93) 5.01	77	0.87 (0.82–0.91) 5.95	57	0.91 (0.87–0.94) 5.58	0.45

1. Differences were found between: Amsterdam – Santander $P<0.001$; London – Santander $P<0.001$; London – Copenhagen $P<0.05$; Santander – Verona $P<0.05$.

test on homogeneity of these variances. It shows that both means and variances differ between sites. Mean scale scores and variances are generally high in Verona and low in Copenhagen. In all cases means differ significantly, and in three cases (supervision, worrying and urging) this also holds for the variances. As the 95% confidence intervals show, the contrasts are mainly between Copenhagen on the one hand and Santander and Verona on the other.

Internal consistency

Cronbach's α values for the IEQ sub-scales are presented in Table 2. The α values range from 0.68 to 0.86 for the sub-scales and from 0.87 to 0.91 for the sumscore. In two cases the benchmark for substantial reliability was just not reached, the α for 'supervision' in London and that for 'urging' in Santander both having a value of 0.68.

Alpha testing between sites showed that on three sub-scales and the sumscore the α values do not differ. Only on the sub-scale 'urging' were differences significant. The α in Santander is lower than those in Amsterdam, London and Verona. Copenhagen also showed a lower α than London.

Differences in α values may be due to differences in score distribution. For that reason Table 2 also gives the $(s.e.)_m$ values.

The $(s.e.)_m$ is lowest for all scales in Copenhagen. On the other hand, Santander and Verona showed relatively high $(s.e.)_m$ scores. This means that the IEQ seems more precise in Copenhagen and less precise in Santander and Verona.

As the $(s.e.)_m$ scores suggest, the differences in α values between sites are caused as much by differences in sample variance as by differences in true reliability. For instance, in Copenhagen, α values are somewhat lower because of the low sample variance (Table 1), and in Verona it is the other way round. Where a lower α is combined with a higher $(s.e.)_m$, reliability seems a bit problematic. This is the case for the

s24

Table 3 Test–retest reliability of the Involvement Evaluation Questionnaire – European Version (IEQ–EU) in the pooled sample and by site

Sub-scale[1]	Pooled n=198[2]		Amsterdam n=51		Copenhagen n=16		London n=46		Santander n=48		Verona n=37		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.)$_m$	
Tension	0.89	1.71	0.92	1.21	0.95	0.71	0.97	0.72	0.82	2.51	0.88	2.09	<0.01
Supervision	0.83	1.54	0.87	1.05	0.98	0.50	0.97	0.51	0.70	2.15	0.82	2.14	<0.01
Worries	0.84	2.43	0.87	0.87	0.93	1.07	0.98	0.72	0.78	3.50	0.69	3.54	<0.01
Urging	0.89	2.03	0.93	1.39	0.80	1.55	0.98	1.12	0.73	3.10	0.90	2.18	<0.01
Sumscore	0.90	5.07	0.94	3.47	0.93	2.37	0.99	1.74	0.81	7.22	0.86	6.79	<0.01

1. ICC, intraclass correlation coefficient; (s.e.)$_m$, standard error of measurement (square root of error components of variance).
2. Samples quoted for sumscore; size of samples varied between 198 (sumscore) and 261 (worrying).

sub-scale 'urging' in the Santander sample. Here the lowest α value is combined with the highest (s.e.)$_m$. Such a combination is not found in any other case. In the case of pooled data, all reliability estimates are substantial, ranging between 0.77 and 0.90.

Test–retest reliability

Table 3 presents the results of the test–retest reliability analysis. In all but one case, reliability is substantial to high (at least 0.7). The ICC values prove to be highest in Amsterdam, Copenhagen and London, demonstrating good reliability. In Verona, one ICC value ('worrying') is only moderate. Although all reliabilities are substantial, in Santander the ICCs for 'supervision', 'worrying' and 'urging' are somewhat lower than the very high values in the other sites. The (s.e.)$_m$ indicate that the somewhat lower reliability of the IEQ in Verona and Santander can be attributed, at least partly, to higher measurement error. All differences between sites proved to be significant. The ICCs of the pooled data sets were all fairly high, ranging between 0.83 and 0.90.

The comparison of test and retest scores showed that retest scores in general were somewhat lower. In seven out of 25 cases these differences were significant, four times involving the sub-scale 'worrying', twice the sumscore, and once 'urging'.

The length of the interval between test and retest may influence reliability. The IEQ does not measure a stable characteristic, but consequences for relatives, which may change over time. Thus, the longer the interval between test and retest, the higher the probability that the situation has changed. However, as was stated in the general reliability paper (Schene *et al*, 2000, this supplement), a short time interval may also produce biased reliability estimates, due to the effect of memory. In the EPSILON Study the interval was set at 1–2 weeks,

but the actual average interval varied considerably between sites. In Santander the average was 6.3 days (s.d.=2.6), in London 10.0 days (s.d.=8.7), in Amsterdam 11.6 days (s.d.=8.2), in Copenhagen 14.9 days (s.d.=5.4) and in Verona 22.4 days (s.d.=14.6). The somewhat lower reliabilities were found in the sites where time intervals were longest or shortest, so that there appears to be no direct connection between interval and reliability.

DISCUSSION

The internal consistencies of the IEQ scales in general turned out to be satisfactory. In Amsterdam, Copenhagen and Santander all α values were substantial, while in both London and Verona, α was moderate in only one case. The comparison of α between sites was hindered by the differences in score distribution. As stated earlier, comparison of Cronbach's α and ICCs directly reflects measurement error, if it is assumed that the true score variances are comparable. The differences in score distribution, with Santander and Verona having the highest means and largest variances, as opposed to Copenhagen, have certainly influenced the test results. These differences seem to have caused the Santander α and the Verona α to be higher than the Copenhagen α. The standard errors of measurement show that when the effect of variance differences is cancelled out, at face value the Copenhagen data seem the most precise, and the Santander and Verona data less precise. A relatively low α combined with a high (s.e.)$_m$ is an indication of a somewhat lower performance of the instrument itself, rather than a characteristic of the sample on which the IEQ was tested. Considering this, the Santander α score on the sub-scale 'urging' seems to be a bit problematic. It is recommended that

the reasons why only a moderate reliability estimate was found here should be explored further.

The estimates from pooled data were all substantial to high. Only in the case of the sub-scale 'urging' should one be careful in the interpretation of these findings, due to the significant differences between sites. In all other cases, no differences were found and the pooled reliability estimates can be considered valid.

The test–retest reliability estimates were all substantial, but differed between sites. ICC values were highest in Amsterdam, Copenhagen and London. Although ICC values are affected by differences in sample variance, the lower reliabilities tended to be those with high (s.e.)$_m$ values, and a low ICC combined with high (s.e.)$_m$ may be an indication of lower performance. In Verona and Santander, the (s.e.)$_m$ values did indeed appear to be higher than at the other sites (although the differences have not been formally tested).

Whether the differences in reliabilities are caused by cultural differences, sampling, or test effects is not yet known. It was found that there was no direct connection between the test–retest interval and reliability. The way in which the IEQ was administered might yield a possible explanation. In Verona and Santander, the IEQ was administered regularly as an interview or completed under the supervision of a research assistant, while in the three other sites, most IEQs were completed by the respondents themselves. As the IEQ is designed to be a self-administered questionnaire, an interview might add some extra bias. Additional data on self-administered IEQs will be necessary to investigate this hypothesis. Finally, in Santander and Verona more patients live with other people (parents, relatives). Because of the more intense contact between patient and relative in

these situations, real changes – even in a short time – will be detected earlier than in the case in which the relative does not live with the patient.

Retest values were generally slightly lower than test values, indicating a certain test–retest effect. The absolute differences, however, were rather small, and were no higher in any one site than in any other. This systematic test–retest effect does not explain the lower ICC scores in Verona and Santander.

Pooled data analyses resulted in rather high reliability estimates. As was stated earlier in this paper, these pooled estimates should be treated with caution, because ICC values are somewhat lower in two sites. On the other hand, the pooled estimates are sufficiently high, and, combined with the fact that the lowest values are either moderate or (in one case) just below moderate, it is reasonable to conclude that overall reliability is good.

In summary, although the differences in sample variance make an exact test of the scales somewhat difficult, in general the IEQ scales have a substantial reliability in all sites. This conclusion is supported by the results of a simultaneous component analysis in which factor analyses of all five separate data sets were compared in one single analysis. As it turned out, in all sites very similar factors were found, indicating that the IEQ scales sufficiently cover all five samples.

The IEQ test–retest reliability analyses have been conducted on rather small samples, especially in Copenhagen ($n=21$). The number of samples in the other sites ranged from 47 to 77. In sites with larger samples, reliability issues could be studied in more detail. Bearing this in mind, one can conclude that despite some questions that still have to be answered, the reliability of the IEQ in the five EPSILON sites seems to be good enough for the moment to encourage the use of the instrument in European research. In doing so, larger datasets can be produced to study the validity and reliability of the IEQ in greater detail.

ACKNOWLEDGEMENTS

APPENDIX: IEQ ITEMS AND SUB-SCALES

Tension sub-scale (nine items)
How often during the past 4 weeks:

has your relative/friend disturbed your sleep?*

has the atmosphere been strained between you both, as a result of your relative/friend's behaviour?

has your relative/friend caused a quarrel?

have you been annoyed by your relative/friend's behaviour?

have you heard from others that they have been annoyed by your relative/friend's behaviour?

have you felt threatened by your relative/friend?

have you thought of moving out, as a result of your relative/friend's behaviour?

have you worried about your own future?

have your relative/friend's mental health problems been a burden to you?*

Supervision sub-scale (six items)
How often during the past 4 weeks:

have you guarded your relative/friend from committing dangerous acts?

have you guarded your relative/friend from self-inflicted harm?

have you ensured that your relative/friend received sufficient sleep?*

have you guarded your relative/friend from drinking too much alcohol?

---

*Items used in more than one sub-scale.

have you guarded your relative/friend from taking illegal drugs?

has your relative/friend disturbed your sleep?*

Worrying sub-scales (six items)
How often during the past 4 weeks:

have you worried about your relative/friend's safety?

have you worried about the kind of help/treatment your relative/friend is receiving?

have you worried about your relative/friend's general health?

have you worried about how your relative/friend would manage financially if you were no longer able to help?

have you worried about your relative/friend's future?

have your relative/friend's mental health problems been a burden to you?*

Urging sub-scale (eight items)
How often during the past 4 weeks:

have you encouraged your relative/friend to take proper care of her/himself?

have you helped your relative/friend take proper care of her/himself?

have you encouraged your relative/friend to eat enough?

have you encouraged your relative/friend to undertake some kind of activity?

have you accompanied your relative/friend on some kind of outside activity, because he/she did not dare to go alone?

have you ensured that your relative/friend has taken the required medicine?

have you carried out tasks normally done by your relative/friend?

have you encouraged your relative/friend to get up in the morning?

Items not included in a sub-scale (four items)

How often during the past 4 weeks have you been able to pursue your own activities and interests?

Have you got used to your relative/friend's mental problems?

How often have you felt able to cope with your relative/friend's mental health problems?

Has your relationship with your relative/friend changed since the onset of the mental health problems?

REFERENCES

Barrowclough, C., Tarrier, N. & Johnston, M. (1996) Distress, expressed emotion and attributions in relatives of schizophrenia patients. *Schizophrenia Bulletin*, **22**, 691–702.

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 1. *British Journal of Psychiatry*, **177** (suppl. 39), s1–s7.

Budd, R. J., Oles, G. & Hughes, I. C. T. (1998) The relationship between coping style and burden in the carers of relatives with schizophrenia. *Acta Psychiatrica Scandinavica*, **98**, 304–309.

Fenton, F. R., Tessier, L. & Struening, E. L. (1979) A comparative trial of home and hospital psychiatric care. One year follow-up. *Archives of General Psychiatry*, **36**, 1073–1079.

Gaite, L. (1997) Cuestionario de salud general (G.H.G.-12 Spanish translation). Santander: University of Cantabria, Clinical and Social Psychiatry Research Unit, Department of Psychiatry.

Goldberg, D. & Williams, P. (1988) *A User's Guide to the General Health Questionnaire.* Windsor: NFER–Nelson.

Hatfield, A. B. & Lefley, H. P. (1987) *Families of the Mentally Ill: Coping and Adaptation.* New York: Guilford Press.

Kastrup, M. (1998) Mental health in the city of Copenhagen, Denmark. In *Mental Health in our Future Cities* (eds D. Goldberg & G. Thornicroft), pp. 101–123. Hove: Psychology Press.

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.

Koeter, M. W. J. & Ormel, J. (1991) *G.H.Q.-12* (Dutch translation). Lisse: Swets en Zeitlinger.

Kuipers, L. & Bebbington, P. (1988) Expressed emotion research in schizophrenia: theoretical and clinical implications. *Psychological Medicine*, **18**, 893–809.

Levine, J. E., Lancee, W. J. & Seeman, M. V. (1996) The perceived family burden scale: measurement and validation. *Schizophrenia Research*, **22**, 151–157.

Magliano, L., Fadden, G., Madianos, M., *et al* (1998) Burden on the families of patients with schizophrenia: results of the BIOMED-1 study. *Social Psychiatry & Psychiatric Epidemiology*, **33**, 405–412.

Mandelbrote, B. & Folkard, S. (1961) Some factors related to outcome and social adjustment in schizophrenia. *Acta Psychiatrica Scandinavica*, **37**, 223–235.

Nielsen, H. (undated) *General Health Questionnaire* (G.H.Q.-12 Danish translation). Odense: University Hospital, Department of Neurology.

Scazufca, M. & Kuipers, E. (1996) Links between expressed emotion and burden of care in relatives of patients with schizophrenia. *British Journal of Psychiatry*, **168**, 580–587.

Schene, A. H. (1986) *Worried at Home; A Monograph on Burden on the Family in Psychiatry* (in Dutch). Utrecht: Nederlands Centrum Geestelijke Volksgezondheid (Netherlands Institute for Mental Health).

___ **& van Wijngaarden, B. (1992)** *The Involvement Evaluation Questionnaire.* Amsterdam: Department of Psychiatry, University of Amsterdam.

___ **& ___ (1993)** *Family Members of People with a Psychotic Disorder; A Study Among Members of Ypsilon* (in Dutch). Amsterdam: Department of Psychiatry, University of Amsterdam.

___, ___, **Poelijoe, N. W., *et al* (1993)** The Utrecht comparative study on psychiatric day treatment and inpatient treatment. *Acta Psychiatrica Scandinavica*, **87**, 427–436.

___, **Tessler, R. C. & Gamache, G. M. (1994)** Instruments measuring family or caregiver burden in severe mental illness. *Social Psychiatry and Psychiatric Epidemiology*, **29**, 228–240.

___ **& van Wijngaarden, B. (1995)** A survey of an organization for families of patients with serious mental illness in the Netherlands. *Psychiatric Services*, **46**, 807–813.

___, **Tessler, R. C. & Gamache, G. M. (1996)** Caregiving in severe mental illness: conceptualization and measurement. In *Mental Health Service Evaluation* (eds H. C. Knudsen & G. Thornicroft), pp. 296–316. Cambridge: Cambridge University Press.

___, **Hoffmann, E. & Goethals, A. L. J. (1998a)** Mental health in Amsterdam. In *Mental Health in our Future Cities* (eds D. Goldberg & G. Thornicroft), pp. 33–55. Hove: Psychology Press.

___, **van Wijngaarden, B. & Koeter, M. W. J. (1998b)** Family caregiving in schizophrenia: domains and distress. *Schizophrenia Bulletin*, **24**, 609–618.

___, **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.

Schofield, H. L., Murphy, B., Herrman, H. E., *et al* (1997) Family caregiving: measurement of emotional well-being and various aspects of the caregiving role. *Psychological Medicine*, **27**, 647–657.

Servizio di Psicologia Medica Verona (undated) *Questionario sulla salute G.H.Q.-12* (Italian translation). Verona: Università di Verona, Instituto di Psichiatria.

Szmukler, G. I., Burgess, P., Herrman, H., *et al* (1996) Caring for relatives with serious mental illness: the development of the Experience of Caregiving Inventory. *Social Psychiatry and Psychiatric Epidemiology*, **31**, 137–148.

Tansella, M., Amaddeo, F., Burti, L., *et al* (1998) Community-based mental health care in Verona, Italy. In *Mental Health in our Future Cities* (eds D. Goldberg & G. Thornicroft), pp. 239–262. Hove: Psychology Press.

Tessler, R. C., Killian, L. M., Tessler, M. A., *et al* (1980) Alternative to mental hospital treatment: III. Social cost. *Archives of General Psychiatry*, **37**, 409–412.

Thornicroft, G. & Goldberg, D. (1998) London's mental health services. In *Mental Health in our Future Cities* (eds D. Goldberg & G. Thornicroft), pp. 15–31. Hove: Psychology Press.

van Wijngaarden, B., Schene, A. H. & Koeter, M. W. J. (1996) *The Consequences of Depressive Disorders for Those Involved with the Patient: A Study on the Psychometric Qualities of the Involvement Evaluation Questionnaire* (in Dutch). Amsterdam: Department of Psychiatry, University of Amsterdam.

Vázquez-Barquero, J. L. & García, J. (1999) Deinstitutionalization and psychiatric reform in Spain. *European Archives of Psychiatry and Clinical Neuroscience*, **249**, 120–135.

Wing, J. K., Monck, E., Brown, G. W., *et al* (1959) Morbidity in the community of schizophrenic patients discharged from London mental hospitals in 1959. *British Journal of Psychiatry*, **110**, 10–21.

BOB VAN WIJNGAARDEN, MA, AART SCHENE, MD, MAARTEN KOETER, PhD, Department of Psychiatry, Academic Medical Centre, Amsterdam, The Netherlands; JOSÉ LUIS VÁZQUEZ-BARQUERO, FRCPsych, Clinical and Social Psychiatry Research Unit, University of Cantabria, Santander, Spain; HELLE CHARLOTTE KNUDSEN, MD, Institute of Preventive Medicine, Copenhagen University Hospital, Denmark; ANTONIO LASALVIA, MD, Department of Medicine and Public Health, University of Verona, Italy; PAUL McCRONE, Section of Community Psychiatry (PRiSM), Institute of Psychiatry, King's College London, UK

Correspondence: Professor Aart H. Schene, Academic Medical Centre, Rm. A3.254, PO Box 22700, 1100 DE Amsterdam, The Netherlands. Tel: +31 20 566 2088; fax: +31 20 697 1971