Development of the EURO-D scale - a European Union initiative to compare symptoms of depression in 14 European centres

M. J. PRINCE, F. REISCHIES, A. T. F. BEEKMAN, R. FUHRER, C. JONKER, S.-L. KIVELA, B. A. LAWLOR, A. LOBO, H. MAGNUSSON, M. FICHTER, H. VAN OYEN, M. ROELANDS, I. SKOOG, C. TURRINA and J. R. M. COPELAND

Background In an 11-country European collaboration, 14 population-based surveys included 21 724 subjects aged ≥ 65 years. Most participating centres used the Geriatric Mental State (GMS), but other measures were also used.

Aims To derive from these instruments a common depression symptoms scale, the EURO-D, to allow comparison of risk factor profiles between centres.

Method Common items were identified from the instruments. Algorithms for fitting items to GMS were derived by observation of item correspondence or expert opinion. The resulting 12-item scale was checked for internal consistency, criterion validity and uniformity of factor-analytic profile.

Results The EURO-D is internally consistent, capturing the essence of its parent instrument. A two-factor solution seemed appropriate: depression, tearfulness and wishing to die loaded on the first factor (affective suffering), and loss of interest, poor concentration and lack of enjoyment on the second (motivation).

Conclusions The EURO-D scale should permit valid comparison of risk-factor associations between centres, even if between-centre variation remains difficult to attribute.

Declaration of interest The European Commission BIOMED I initiative funded this Concerted Action Programme.

The EURODEP Concerted Action Programme is a consortium of 14 research groups from 11 European countries all engaged in population-based research into the epidemiology of late-life depression. The aims of the consortium were three-fold: to compare the prevalence of late-life depression in different European countries and cultures; to assess the homogeneity or heterogeneity of aetiological associations with late-life depression between centres; and (in so far as patterns of association were homogeneous between centres) to pool data in order to increase the power and precision of these analyses.

The collaboration was initially intended to include nine European centres, all of which had used either the Geriatric Mental State–AGECAT (GMS–AGECAT; Copeland *et al*, 1986) or the very similar SHORT-CARE (Gurland *et al*, 1984) as their index of clinical case-level depression, allowing direct comparison of prevalence between centres (Table 1). Data from these centres are reported elsewhere (Copeland *et al*, 1999*a,b*, this issue). However, five large European studies had used other measures of depression (Kivela *et al*, 1988; Barberger Gateau *et al*, 1992; Skoog *et al*, 1993; Roelands *et al*, 1994; Beekman *et al*, 1995) (see also Table 1). Three centres had used the Centre for Epidemiological Studies Depression scale (CES–D) (Radloff, 1977), one the Zung Self-Rating Depression Scale (ZSDS) (Zung, 1965) and one the Comprehensive Psychopathological Rating Scale (CPRS) (Asberg *et al*, 1978). Incorporation of data from these centres into the analyses would include four countries (France, Belgium, Sweden and Finland) not otherwise represented, and would increase the overall sample by 8427, to 21 724 subjects. Our task was, therefore, to find a way of harmonising data from these five depression measures in such a way as to allow meaningful comparison, while retaining core validity.

METHOD

Measures

GMS–AGECAT is a semi-structured clinical interview schedule which can be administered by medically qualified or trained lay interviewers. Identification of symptoms requires some judgement to be exercised by the interviewer, who rates the severity of symptoms and also makes observations regarding the subject's demeanour and behaviour. Diagnoses are allocated by the computer program AGECAT, using a hierarchical typological algorithm.

SHORT-CARE is a very similar instrument to GMS, with the majority of items identically or nearly identically worded. It is, however, more structured; interviewers are not required to make judgements regarding symptom severity and there are no observational items. The diagnosis of pervasive depression is made on the basis of a simple algorithm requiring the presence of six or more of 17 key symptoms of depression. Like GMS–AGECAT depression, it is a broad diagnostic category signifying depression of clinical case-level severity.

CES–D is a 20-item self-report depression symptom rating scale. Subjects are asked to rate each symptom as being present over the previous week (a) rarely or none of the time, (b) some or a little of the time, (c) occasionally, or (d) most or all of the time. A cut-off point of 15/16 is commonly used to identify a group with a high probability of suffering from depression of clinical significance. This cut-off point has been validated against a diagnosis of clinical case-level depression in both younger and older subjects (Breslau, 1985; Beekman *et al*, 1994).

The modified ZSDS is a 20-item self-report depression symptom rating scale. Subjects are asked to rate each symptom as being present over the previous two weeks (a) never, (b) now and then, (c) quite often, or (d) most of the time.

CPRS is a wide-ranging semi-structured clinical psychiatric interview with operationalised symptom definitions and symptom ratings. Sufficient information can be generated to make a variety of ICD–10 (World Health Organization, 1993) or DSM–IV (American Psychiatric Association, 1994) diagnoses.

Harmonisation of scales
Overall strategy
GMS–AGECAT, SHORT-CARE and CPRS are interviewer-administered and generate

Table I The EURODEP consortium – studies and subjects

Centre	Number of subjects						Percentage of subjects aged 80+
	Age 65–69	Age 70–74	Age 75–79	Age 80–84	Age 85+	Total	
GMS centres							
Liverpool	926	967	894	910	1525	5222	47%
Berlin	–	83	84	78	243	488	66%
Amsterdam	1052	929	1023	983	–	3987	25%
Dublin	285	249	172	132	75	913	23%
Iceland	–	–	–	11	760	771	100%
Munich	–	–	–	–	293	293	100%
Verona, Italy	81	37	38	28	18	202	24%
Zaragoza, Spain	278	298	225	146	90	1037	23%
London	176	164	103	120	74	637	30%
CES–D centres							
LASA,[1] Netherlands	471	438	528	437	70	1944	26%
PAQUID,[2] France	1021	792	893	529	369	3604	25%
Antwerp, Belgium	201	293	245	224	167	1130	35%
Other instruments							
Gothenburg, Sweden (CPRS)	–	–	–	–	449	449	100%
Ahtari, Finland (ZSDS)	335	279	251	120	62	1047	17%
Total	4826	4529	4456	3718	4195	21 724	36%

GMS, Geriatric Mental State; CES–D, Centre for Epidemiological Studies Depression Scale; CPRS, Comprehensive Psychopathological Rating Scale; ZSDS, Zung Self-Rating Depression Scale.
1. LASA, Longitudinal Ageing Study Amsterdam
2. PAQUID, Personnes Agées QUID.

clinical diagnoses, whereas CES–D and ZSDS are simple self-report instruments assessing the presence and pervasiveness of depression symptoms. We decided, therefore, to develop a depression symptom scale by selecting those symptoms from the clinical instruments which were also common to the CES–D and the ZSDS, rather than attempting the unrealistic goal of a common categorical criterion.

Item selection

We reviewed each instrument to identify items that were common to all or most of the five, and had good face validity for late-life depression. We selected 12 items: depression, pessimism, wishing death, guilt, sleep, interest, irritability, appetite, fatigue, concentration, enjoyment, and tearfulness (Table 2). All were present in GMS–AGE-CAT (although some centres used a version in which enjoyment was not assessed) and in ZSDS, 11 were present in SHORT-CARE, and 10 in CES–D and CPRS (Table 2). In some cases, we judged that more than one item from CES–D related to a single GMS item.

Fitting CES–D, ZSDS and CPRS items to GMS

We next attempted to fit the CES–D, ZSDS and CPRS items to the GMS items, with the 0, 1, 2 GMS item scores (symptom not present, present to a moderate degree, present to a marked degree) collapsed into 0 for not present and 1 for present. Two methods were used.

Method 1 – expert opinion. Six old-age psychiatrists with experience of epidemiological research, familiar with the use of the five instruments, but independent of the EURODEP consortium, were asked to advise. For each item on each questionnaire they had three judgements to make. In the case of the CES–D, for example, they had to judge (a) in cases where the GMS item could be addressed by two items on CES–D, whether the GMS symptom be considered to be present if the subject scored above a threshold (i) on both CES–D items, (ii) on either item, or (iii) on only one of the two (the other adjudged to be superfluous); (b) at what level of frequency or pervasiveness of the CES–D symptom would the symptom be considered to be present (to match with the dichotomous coding of the GMS item); (c) overall, on a scale of 1 to

5, how good was the fit between the CES–D item or items and the GMS item.

Method 2 – probabilistic modelling. One centre, Berlin, had administered both GMS–AGECAT and CES–D to the same 488 subjects. Thus, it was possible to examine directly the correspondence between items from the two instruments. The EURO-D item score was calculated as the positive predictive value (PPV) for a GMS item score of 1 at each possible level of CES–D (0, 1, 2 or 3). Where the fit with GMS could be improved by combining information from two CES–D items, PPVs were calculated for all 16 possible combinations of CES–D score from the two items.

Imputation of missing data

Subjects with missing data for more than one relevant GMS, SHORT-CARE, CES–D, CPRS or ZSDS item were excluded from the analysis. We used statistical imputation to estimate missing data in two situations: where individual subjects were missing only one item, and where all subjects in one centre were missing only one item, and where all subjects in one centre were missing (by design) one or two items. For

Table 2 EURO–D scale items

Symptom	GMS-item	CES–D item	Fit[1]
EURO1 Depression	Have you been sad (depressed, miserable, in low spirits, blue) recently?	I felt depressed; and I felt that I could not shake off the blues even with the help of my family or friends	3.2
EURO2 Pessimism	How do you see your future? (Pessimistic, empty expectations or bleak future)	I feel hopeful about my future	2.3
EURO3 Wishing death	Have you ever felt that you would rather be dead? (Has ever felt suicidal or wished to be dead)	Imputed	
EURO4 Guilt	Do you tend to blame yourself or feel guilty about anything? (Obvious guilt or self blame)	I felt as good as other people; and, I thought my life had been a failure	1.5
EURO5 Sleep	Have you had trouble sleeping recently? (Trouble with sleep or recent change in pattern)	My sleep was restless	3.0
EURO6 Interest	What is your interest in things? (Less interest than is usual)	Imputed	
EURO7 Irritability	Have you been irritable recently?	I was bothered by things which don't usually bother me	2.3
EURO8 Appetite	What has your appetite been like? (Diminution in the desire for food)	I did not like eating; my appetite was poor	3.5
EURO9 Fatigue	Have you had too little energy (to do the things you want to do)? (Listlessness or subjective energy restriction)	I felt that everything I did was an effort	3.7
EURO10 Concentration	How is your concentration? (Difficulty in concentrating on entertainment or reading)	I had trouble keeping my mind on what I was doing	3.7
EURO11 Enjoyment	What have you enjoyed doing recently? (Almost nothing enjoyed)	I enjoyed life	3.7
EURO12 Tearfulness	Have you cried at all?	I had crying spells	3.3

1. Expert panel's mean rating of overall match with GMS item, scored on a scale of 1 to 5.

single subjects we imputed the missing item score by applying to that subject's other item scores the relevant coefficients from the logistic regression (GMS and SHORT-CARE) or multiple regression (CES-D, ZSDS and CPRS) model predicting the missing item score, derived from all subjects in that centre with complete data. Similarly, where an item was missing by design from an entire centre, we imputed the item scores using regression models derived from all other centres with complete data. In the case of the CES-D centres, where EURO-D item scores were non-integer probabilities, Berlin CES-D item scores were used to predict, using logistic regression, GMS EURO3 and EURO6 scores on the same subjects. We then used the coefficients from the resulting logistic model to predict for each subject from the Longitudinal Ageing Study Amsterdam (LASA) (Netherlands), PAQUID (Personnes Agées QUID, France) and Antwerp centres a probability of an item score of 1 for the missing EURO3 and EURO6 items.

Scale construction

For the GMS centres, we constructed the EURO-D scale by summing the 0 or 1 scores for the 12 contributing items. For the CES-D centres we calculated the scale according to two methods: for the expert opinion method, by summing the 0,1 scores derived from the panel's recommendations; and, for the probabilistic modelling method, by summing the PPVs calculated from algorithms derived from the Berlin data set.

Reliability

We assessed the internal consistency of the resulting scales from each centre by calculating the inter-item correlations, the item-total correlations and the standardised alpha value. For these analyses, we excluded imputed items so as not to bias upwards the estimates of internal consistency. We tested whether internal consistency (judged by the standardised alpha) could be improved by omitting items with low item-total correlations.

Validity

The criterion validity of the EURO-D scales could be assessed directly in Berlin, where the EURO-D derived from the GMS could be compared with the CES-D, and in the Netherlands LASA study where the EURO-D derived from CES-D could be

compared with Composite International Diagnostic Interview (CIDI; World Health Organization, 1990) diagnoses of major depression. In other centres, comparison of EURO-D with GMS-AGECAT depression diagnoses or CES-D scale scores involved an element of circularity, in that the EURO-D scale had been derived from items from the instrument which provided the criterion measure. Agreement with continuous measures was assessed by Spearman non-parametric correlations, and for dichotomous measures by the area under the receiver operating characteristic curve.

We assessed whether a common factor structure existed across all the centres by carrying out a principal-components analysis of the EURO-D scale items (with varimax rotation) separately for each centre.

RESULTS

Harmonisation – the expert opinion and the probabilistic model approaches

The expert panel's recommendations were remarkably consistent. For nearly all of the items on each of the three scales there was a clear majority for one cut-off point

and, where appropriate, for a single approach in combining two items. There was also good agreement on the extent to which the non-GMS items could be matched to those from the GMS. The CPRS items were all felt to match adequately. For the CES–D, three items – numbers 2 (pessimism), 4 (guilt) and 7 (irritability) – were considered to give poor matches; for the ZSDS, there were five such items – numbers 2 (pessimism), 3 (wishing death), 6 (interest), 8 (appetite) and 10 (concentration). The accuracy of the panel's judgement on matching could be checked for CES–D using data from Berlin, where the correlation between the CES–D score and the GMS item score was measured directly. The Spearman correlation between the panel's rating and the observed correlation for 10 items was only 0.14 ($P=0.7$).

Table 3 compares the Berlin CES–D EURO–D calculated using the experts' opinion and the probabilistic model with the GMS EURO–D from the same centre. The panel's opinion led to the CES–D EURO–D underestimating the prevalence of the depression, wishing death, sleep and tearfulness items (with reference to GMS EURO–D from the same centre), and overestimating the prevalence of the pessimism and enjoyment items. The net result was a similar mean and standard deviation for the total scale score. The probabilist model inevitably resulted in

item 'prevalences' which were almost identical to those observed for the GMS EURO–D. The scale mean was also very similar to that for the GMS EURO–D. The probabilistic model, when applied to CES–D data from the LASA study, also seemed to result in a more plausible pattern of item prevalences (cf. Table 4) than did the expert panel's algorithm.

Scale distribution

The EURO–D item prevalences, and scale means and standard deviations for all centres, are given in Table 4. In each centre the scale distribution was positively skewed. The median value, 1 for most centres, 2 for London and Berlin and 3 for Munich, was always less than the mean, which ranged from 1.34 to 3.58. When the three outlying centres were excluded (Dublin, mean 1.34; Ahtari, Finland 3.17; Munich 3.58), the mean EURO–D scores for the other nine centres ranged only from 1.79 to 2.54. The standard deviations were much lower for the CES–D EURO–D (range 0.94–1.13) than for the GMS EURO–D (1.83–2.66). Within each instrument category, however, the standard deviations were remarkably constant.

Reliability

In each centre, the EURO–D seemed to be adequately internally consistent, although

the inter-item and item-total correlations and the standardised alpha value were higher for the CES–D EURO–D than for the GMS EURO–D (Table 5). In most centres the internal consistency of the scale could not be improved upon; however, in Verona, Italy the omission of EURO7 (irritability), and in Zaragoza, Spain EURO4 (guilt) marginally increased the standardised alpha.

Validity

The strong associations between the EURO–D and its parent instruments, measured as correlations for continuous measures and as areas under the receiver operating characteristic curves for the dichotomous measures, suggest that the EURO–D has captured the essence of the instruments from which it was derived (Table 6). Evidence for the validity of comparing EURO–D results derived from different parent instruments comes from Berlin, where the GMS EURO–D had correlations of 0.72 with the CES–D EURO–D and 0.70 with CES–D. Our only evidence for true criterion validity of EURO–D comes from LASA, where CES–D EURO–D gave an area under the curve of 0.93 for the prediction of DSM major depression (Diagnostic Interview Schedule (DIS); Robins et al, 1981). The optimal cut-off point on the EURO–D scale for prediction of GMS depression (DN or DP 3 and above) and SHORT–CARE pervasive depression was generally 3/4; the predictive characteristics for these diagnoses at this cut-off point are also given in Table 6. In Dublin a 2/3 cut-off point may have been more appropriate, with a large gain in sensitivity in return for small falls in specificity and positive predictive value. For the CES–D centres, a score of 2.5 or above on the EURO–D seemed to correspond best to the conventional CES–D cut-off point of 15/16; in the LASA study this cut-off point was associated with 94% sensitivity, 90% specificity and a PPV of 64%.

Principal components analysis (Table 7) generated two or more factors with eigenvalues over one from each centre. Inspection of the items loading on these factors suggested that two factors were common to nearly every centre. The depression and tearfulness items, and less consistently the pessimism and wishing death items, tended to load on the first factor (which we shall call depressed affect), while the interest, concentration and enjoyment items tended

Table 3 EURO–D item prevalences and scale scores from Berlin (GMS and CES–D) and LASA (Longitudinal Ageing Study Amsterdam; CES–D), derived according to three different methods

Symptom	Berlin[1]	Berlin[2]	Berlin[3]	LASA[2]	LASA[3]
1 Depression	35	8	36	3	28
2 Pessimism	33	73	28	51	24
3 Suicidality	24	5	32	1	22
4 Guilt	5	8	5	2	3
5 Sleep	48	21	48	17	47
6 Interest	10	17	10	5	5
7 Irritability	10	17	10	12	9
8 Appetite	18	15	18	5	11
9 Fatigue	31	21	31	10	26
10 Concentration	8	11	10	7	8
11 Enjoyment	8	68	9	25	4
12 Tearfulness	17	5	17	4	18
Total score (mean)	2.48	2.67	2.52	1.41	2.06
s.d.	2.19	2.10	1.13	1.73	0.94

Item prevalences in **bold** signify data from imputed items.
1. Integer item scores derived from the GMS.
2. Integer item scores derived from the CES–D according to expert panel's algorithm.
3. Non-integer item scores calculated from the CES–D according to positive predictive values for GMS items derived from Berlin data set. Item 'prevalences' are in fact the mean item scores for the study population.

Table 4 Item prevalences and scale score according to centre

Symptom	LASA[1,2]	PAQUID[1,2]	Antwerp[1]	Amsterdam	London	Liverpool	Berlin	Dublin	Munich	Iceland	Zaragoza[4]	Verona[5]	Gothenburg[6]	Ahtari[7]
1 Depression	28	34	27	40	34	29	35	32	58	15	27	32	27	9
2 Pessimism	24	20	21	10	29	30	33	17	28	10	11	24	*	10
3 Suicidality	*	*	*	3	10	6	24	6	30	5	5	12	2	23
4 Guilt	3	4	3	4	9	4	5	3	9	8	3	3	8	*
5 Sleep	47	47	39	*	45	27	48	15	66	44	34	40	14	72
6 Interest	*	*	*	6	19	6	10	4	23	27	6	11	20	42
7 Irritability	9	8	8	23	11	13	10	16	13	10	14	6	29	7
8 Appetite	11	14	10	11	18	8	18	6	21	11	7	5	20	48
9 Fatigue	26	30	27	30	46	23	31	10	36	35	22	8	32	23
10 Concentration	8	10	8	14	*	8	8	2	20	21	10	16	5	24
11 Enjoyment	4	5	4	*	18	*	8	4	19	*	*	*	*	13
12 Tearfulness	18	19	18	28	23	18	17	21	35	6	19	17	41	45
Total score (mean)	2.06	2.23	1.93	1.98	2.54	1.79	2.48	1.34	3.48	2.03	1.61	1.84	2.11	3.17
s.d.	0.94	1.01	0.95	2.11	2.14	1.95	2.19	1.83	2.66	2.00	2.04	2.15	2.01	2.11

*, Imputed items.
1. Non-integer item scores calculated from CES–D according to positive predictive values for GMS items derived from Berlin data set. Item 'prevalences' are in fact the mean item scores for that centre.
2. Longitudinal Ageing Study Amsterdam (LASA), Netherlands.
3. PAQUID (Personnes Agées QUID), France.
4. Zaragoza, Spain.
5. Verona, Italy.
6. Gothenberg, Sweden.
7. Ahtari, Finland.

Table 5 Reliability

	Inter-item correlations		Item-total correlation – range	Alpha value	Alpha value if item deleted
	Mean	Range			
CES–D centres					
LASA,[1] Netherlands	0.23	0.08–0.41	0.31–0.57	0.74	0.62–0.68
PAQUID,[2] France	0.23	0.12–0.40	0.31–0.55	0.75	0.61–0.67
Antwerp, Belgium	0.28	0.15–0.49	0.31–0.55	0.80	0.70–0.74
CPRS centre					
Gothenburg, Sweden	0.12	−0.06–0.39	0.13–0.53	0.58	0.48–0.60
ZSDS centre					
Ahtari, Finland	0.15	−0.03–0.51	0.19–0.48	0.66	0.58–0.65
GMS–SHORT–CARE centres					
London	0.14	0.04–0.33	0.20–0.44	0.65	0.59–0.64
Liverpool	0.15	0.04–0.46	0.18–0.50	0.66	0.61–0.65
Dublin	0.20	−0.01–0.55	0.25–0.55	0.75	0.68–0.73
Berlin	0.16	−0.01–0.59	0.17–0.57	0.69	0.63–0.70
Munich	0.20	0.03–0.59	0.26–0.52	0.75	0.72–0.75
Zaragoza, Spain	0.20	0.01–0.58	0.11–0.60	0.73	0.67–0.74
Verona, Italy	0.18	−0.06–0.55	0.08–0.57	0.71	0.66–0.73
Amsterdam	0.13	0.02–0.45	0.15–0.42	0.61	0.54–0.60
Iceland	0.14	−0.02–0.46	0.06–0.47	0.64	0.56–0.64

For all these analyses, imputed items have been excluded. CES–D, Centre for Epidemiological Studies Depression Scale; CPRS, Comprehensive Psychopathological Rating Scale; ZSDS, Zung Self-Rating Depression Scale; GMS, Geriatric Mental State.
1. LASA, Longitudinal Ageing Study Amsterdam.
2. PAQUID, Personnes Agées QUID.

to load on the second factor (motivation). Other, relatively distinct factors emerged in a minority of centres: a somatic factor (sleep, appetite and fatigue) in Amsterdam, Iceland, Gothenburg and Ahtari; an irritability factor in Berlin, London, Iceland and Verona; and a guilt factor in London, Zaragoza and Verona. When a two-factor solution was forced, a recognisable depressed effect and motivation factor was generated in each centre (Table 7). When data were pooled from all centres, a scree plot again favoured a two-factor solution, the depressed affect factor accounting for 24.6% of the variance in the items, and the motivation factor for 12.1%. The depression, tearfulness, sleep and wishing death items loaded on the depressed effect factor, with smaller contributions (0.4–0.5) from pessimism, appetite and fatigue. The motivation factor was clearly characterised, with large contributions from the interest, concentration and enjoyment items but negligible loadings from other items.

DISCUSSION

The principal aim of this exercise was to harmonise data derived from a variety of

Table 6 Validity

Centre	Spearman correlation		Area under ROC curve	Values at 3/4 cut-off point		
				Sensitivity	Specificity	PPV
LASA, Netherlands	0.92[a]		0.97[1]	0.93[2]		
PAQUID, France	0.93[a]		0.96[1]			
Antwerp, Belgium	0.92[a]		0.98[1]			
Ahtari, Finland	0.84[d]		0.83[3]			
London	0.79[b]		0.93[4]	0.86[5]	76	83	49
Liverpool				0.89[5]	70	88	37
Berlin (CES–D)	0.92[a]			0.84[5]		
Berlin (GMS)	0.70[a]	0.72[c]	0.85[1]	0.92[5]	83	84	52
Dublin				0.93[5]	63	95	62
Munich				0.79[5]	79	66	42
Zaragoza, Spain				0.94[5]	79	92	53
Verona, Italy				0.95[5]	76	95	76
Amsterdam				0.88[5]	72	86	41
Iceland				0.85[5]	76	49	32

Spearman correlations:
a. CES–D total score.
b. SHORT-CARE depression diagnostic scale score.
c. EURO–D derived from CES–D.
d. Zung total score.
Area under ROC curves:
1. CES–D ≥16.
2. Major depression (CIDI).
3. Major depression, dysthymia and atypical depression (DSM–III clinical diagnoses).
4. SHORT-CARE pervasive depression.
5. GMS–AGECAT depression (DN3+ or DP3+).
ROC, receiver operating characteristic; PPV, positive predictive values; LASA, Longitudinal Ageing Study Amsterdam; PAQUID, Personnes Agées QUID.

depression measures for the purposes of a multi-centre collaborative analysis of putative risk factors for late-life depression. For this aim, we can claim a qualified success. The EURO–D, whether derived from GMS, SHORT-CARE, CES–D, CPRS or ZSDS, is an internally consistent scale, which seems to capture the essence of its parent instruments, and which has a common factor structure whatever its origins.

The expert panel gave us a consistent opinion on the fitting of CES–D, ZSDS and CPRS items to the GMS items on which the EURO–D was based. However, application of the experts' algorithm gave rise to some unusual item prevalences; for example, only 3% of LASA subjects said that they felt depressed or felt that they could not shake off the blues for at least 3–4 days per week. The item prevalence for the analogous GMS item "have you been sad (depressed, miserable, in low spirits, blue) recently?" ranged between 15% and 58%. Using the alternative method of assigning to each subject, on the basis of their CES–D scores, the PPV observed in Berlin for a GMS item score of 1 resulted in

an estimated item prevalence of 27%. The item prevalences for sleep, fatigue and tearfulness calculated from the panel's algorithms were similarly aberrant, while again those generated using the probabilistic approach were more consistent with the pattern observed in other centres. The EURO–D was not developed to allow meaningful comparison of item prevalences between centres using different instruments. Also, the aberrant item prevalences would tend to even themselves out in the total EURO–D scale score. However, the Berlin data suggested that it was the vagaries of the panel's algorithm for calibrating CES–D to GMS, rather than any genuine between-centre difference in GMS item prevalence, which were responsible for the discrepancies. We felt that such wide discrepancies were undesirable and therefore preferred the probabilistic approach.

There were two disadvantages associated with this decision. First, the method could only be applied to CES–D, as no centres had used GMS with ZSDS or CPRS. The ZSDS and CPRS EURO–D scales may

be of questionable validity if the panel's judgement on calibrating items to GMS was as flawed as it proved to be for CES–D. There may be a particular problem with the ZSDS, as the panel rated five of the 12 items as matching only weakly with their GMS analogues; however, their judgement on the adequacy of the match turned out to be inaccurate for the CES–D. Also, the standard deviation of the CES–D EURO–D calculated by this method was much lower than for the GMS EURO–D. This was almost certainly an artefact of the item-scoring method for the two scales. CES–D EURO–D used non-integer PPVs ranging between 0 and 1, but for most items scores of 0 and 1 were impossible. Thus, the range of possible total scores was 1.09–7.89 instead of 0–12 for the GMS EURO–D.

There are three potential uses for the EURO–D in the EURODEP collaboration:

(a) comparison of EURO–D item prevalence between centres;

(b) comparison of EURO–D scale distribution between centres; and

(c) comparison of effect sizes for associations between risk factors and EURO–D score between centres.

Use (a) is likely to be valid for GMS centres and CES–D centres, to the extent to which the pattern of prediction observed in Berlin pertains elsewhere. Use (b) is again likely to be valid for GMS centres, and possibly for CES–D centres. However, even though the central tendency of the CES–D EURO–D may accurately reflect that which would have been measured with the GMS EURO–D, the artefactual constraint of the variance of the former will be problematic when making statistical inferences about observed differences. One solution might be to standardise the variance of the scales in each centre by dividing by the standard deviation adding the mean and subtracting mean/standard deviation. This transformation would leave the means (and the between-centres differences in mean) intact, but each scale would have a standard deviation of one. Use (c) may be valid for all versions of the EURO–D, particularly if we focus away from between-centre main effects by z-scoring all scales (subtracting the mean and dividing by the standard deviation). Each centre's scale would then have a mean of zero and a standard deviation of one. Effect sizes – for example, the difference between the mean EURO–D scores for men and women – could then be compared directly between centres, and

Table 7 Factor analysis

	Two-factor solution		Number of factors with eigenvalue > 1	Content of factors not captured in two-factor solution
	Factor 1	Factor 2		
CES–D centres				
LASA,[1] Netherlands			2	
Variance (%)	33.8	9.2		
Items loading	Depression	(Suicidality)		
	(Suicidality)	(Interest)		
	Appetite	Enjoyment		
	Fatigue			
	Concentration			
PAQUID,[2] France			2	
Variance (%)	33.0	8.3		
Items loading	Tearfulness	Concentration		
	Irritability	Fatigue		
	Depression			
	(Interest)			
	(Suicidality)			
Berlin (CES–D)			2	
Variance (%)	33.7	9.9		
Items loading	(Suicidality)	Pessimism		
	Depression	Enjoyment		
	Concentration			
	Fatigue			
	Irritability			
	Tearfulness			
Antwerp, Belgium			2	
Variance (%)	38.3	8.7		
Items loading	(Suicidality)	Enjoyment		
	Sleep	Pessimism		
	Depression	(Interest)		
	Appetite			
	Irritability			
CPRS centre Gothenburg, Sweden			4	Sleep, appetite, fatigue 3/4 Tearfulness 4/4
Variance (%)	25.7	12.3		
Items loading	(Pessimism)	Concentration		
	Depression	(Enjoyment)		
	Guilt			
ZSDS centre Ahtari, Finland			3	Interest, concentration, enjoyment 1/3
Variance (%)	22.4	10.4		
Items loading	Pessimism	Sleep		
	Depression	Tearfulness		
		Appetite		
GMS centres				
Berlin			4	Irritability 3/4 Concentration 4/4
Variance (%)	24.0	12.4		
Items loading	Depression	Interest		
	Tearfulness	Enjoyment		
	Sleep			

(continued)

Table 7 *(continued)*

	Two-factor solution		Number of factors with eigenvalue > 1	Content of factors not captured in two-factor solution
	Factor 1	Factor 2		
GMS centres (continued)				
London			4	Irritability 3/4
Variance (%)	23.0	10.7		Guilt 4/4
Items loading	Depression	Interest		
	Sleep	Enjoyment		
Liverpool			2	
Variance (%)	24.1	14.4		
Items loading	Depression	Concentration		
	Tearfulness	Enjoyment		
		Interest		
Amsterdam			4	(Sleep), fatigue, appetite 2/4
Variance (%)	26.9	11.7		Guilt, suicidality 4/4
Items loading	(Enjoyment)	Depression		
	Interest	Tearfulness		
	Concentration	(Sleep)		
Dublin			3	Guilt, concentration 3/3
Variance (%)	27.3	11.2		
Items loading	Interest	Depression		
	Enjoyment	Tearfulness		
	Fatigue	Pessimism		
		Suicidality		
Iceland			4	Fatigue, appetite, sleep 3/4
Variance (%)	24.6	12.6		Guilt, irritability 4/4
Items loading	Depression	(Enjoyment)		
	Pessimism	Interest		
	Suicidality	Concentration		
	Tearfulness			
Munich			3	Suicidality, pessimism, guilt 2/3
Variance (%)	27.5	13.3		
Items loading	Depression	Enjoyment		
	Tearfulness	Interest		
	Suicidality	Concentration		
	Pessimism			
Zaragoza, Spain			3	Guilt 3/3
Variance (%)	29.5	13.1		
Items loading	Depression	Enjoyment		
	Tearfulness	Interest		
	Suicidality	Concentration		
	Pessimism			
Verona, Italy			5	Sleep, pessimism 2/5
Variance (%)	29.2	13.5		Guilt, fatigue 4/5
Items loading	(Enjoyment)	Suicidality		Irritability 5/5
	Interest	Depression		
	Concentration	Sleep		
		Fatigue		

CES–D, Centre for Epidemiological Studies Depression Scale; CPRS, Comprehensive Psychopathological Rating Scale; ZSDS, Zung Self-Rating Depression Scale; GMS, Geriatric Mental State.
Roman type: Items loading at ⩾0.60.
Italics: Items loading at 0.55–0.59.
(Brackets): Imputed variables.
1. LASA, Longitudinal Ageing Study Amsterdam.
2. PAQUID, Personnes Agées QUID.

heterogeneity tested for by fitting centre-by-main-effect interaction terms.

Discussion of the observed differences in EURO–D scores between centres is reserved for the accompanying paper on the effect of age, gender and marital status (Prince *et al*, 1999, this issue).

ACKNOWLEDGEMENTS

REFERENCES

American Psychiatric Association (1994) *Diagnostic and Statistical Manual of Mental Disorders* (4th edn) (DSM–IV). Washington, DC: APA.

Asberg, M., Perris, C., Schalling, D., et al (1978) CPRS: Development and applications of a psychiatric rating scale. *Acta Psychiatrica Scandinavica*, **271** (suppl.), 1–69.

Barberger Gateau, P., Chaslerie, A., Dartigues, J. F., et al (1992) Health measures correlates in a French elderly community population: the PAQUID study. *Journal of Gerontology*, **47**, S88–S95.

Beekman, A. T. F., van Limbeek, J., Deeg, D. J. H., et al (1994) Een screeningsinstrument voor depressie bij ouderen in de bevolking: de bruikbaarheid van de Center for Epidemiologic Studies Depression scale (CES–D). *Tijdschrift voor Gerontologie en Geriatrie*, **25**, 95–103.

__ , **Deeg, D. J. H., van Tilburg, T., et al (1995)** Major and minor depression in later life: a study of prevalence and associated factors. *Journal of Affective Disorders*, **36**, 65–75.

Breslau, N. (1985) Depressive symptoms, major depression, and generalized anxiety: a comparison of self-reports on CES–D and results from diagnostic interviews. *Psychiatry Research*, **15**, 219–229.

Copeland, J. R. M., Dewey, M. E. & Griffiths-Jones, H. M. (1986) Computerised psychiatric diagnostic system and case nomenclature for elderly subjects: GMS and AGECAT. *Psychological Medicine*, **16**, 89–99.

__ , **Beekman, A. T. F., Dewey, M. E., et al (1999a)** Depression in Europe. Geographical distribution among older people. *British Journal of Psychiatry*, **174**, 312–321.

__ , __ , __ , **et al (1999b)** Cross-cultural comparison of depressive symptoms in Europe does not support stereotypes of ageing. *British Journal of Psychiatry*, **174**, 322–329.

Gurland, B., Golden, R. R., Teresi, J. A., et al (1984) The SHORT-CARE: an efficient instrument for the assessment of depression, dementia and disability. *Journal of Gerontology*, **39**, 166–169.

Kivela, S.-L., Pahkala, K. & Laipala, P. (1988) Prevalence of depression in an elderly Finnish population. *Acta Psychiatrica Scandinavica*, **78**, 401–413.

Prince, M. J., Beekman, A. T. F., Deeg, D. J. H., et al (1999) Depression symptoms in late-life assessed using the EURO–D scale. Effect of age, gender and marital

status in 14 European centres. *British Journal of Psychiatry*, **174**, 339–345.

Radloff, L. S. (1977) The CES–D scale: a self report depression scale for research in the general population. *Applied Psychological Measurement*, **1**, 385–401.

Robins, L., Melzer, J. E., Croughan, J., et al (1981) National Institute of Mental Health Diagnostic Interview Schedule: its history, characteristics and validity. *Archives of General Psychiatry*, **38**, 381–389.

Roelands, M., Wostyn, P., Dom, H., et al (1994) The prevalence of dementia in Belgium: a population-based door-to-door survey in a rural community. *Neuroepidemiology*, **13**, 155–161.

Skoog, I., Nilsson, L., Landahl, S., et al (1993) Mental disorders and the use of psychotropic drugs in an 85-year-old urban population. *International Psychogeriatrics*, **5**, 33–48.

World Health Organization (1990) *Composite International Diagnostic Interview* (CIDI, version 1.0). Geneva: WHO.

__ **(1993)** *The Tenth Revision of the International Classification of Diseases and Related Health Problems (ICD–10): Diagnostic Criteria for Research.* Geneva: WHO.

Zung, W. W. K. (1965) A self-rating depression scale. *Archives of General Psychiatry*, **12**, 62–70.

CLINICAL IMPLICATIONS

■ The subset of GMS items used in the EURO–D scale has strong conceptual validity, and there is now an abundance of population-based normative data from different European centres.

■ The main application of the EURO–D may be, as here, as a scaleable measure of depression severity in comparative epidemiological studies.

■ The EURO–D's brevity and simplicity may lead to its later development as a screening instrument, and as an outcomes measure for health services research.

LIMITATIONS

■ The EURO–D seems to be capable of identifying cases of depression, defined according to a variety of criteria, with adequate sensitivity and specificity. However, its validity needs to be checked against a truly independent external criterion.

■ The EURO–D's psychometric properties are not yet as well established as those of existing depression symptom scales validated for use in late-life, such as CES–D and GDS.

M. J. PRINCE, MD, Epidemiology Unit, London School of Hygiene and Tropical Medicine, London; F. REISCHIES, Priv. Doz. Dr med., Freie Universität Berlin, Department of Psychiatry, Universität Klinikum Eschenalle 3, Berlin, Germany; A. T. F. BEEKMAN, MD, Department of Psychiatry, Vrije Universiteit Amsterdam, The Netherlands; R. FUHRER, PhD, INSERM U360, Hôpital de la Salpetriere, 75652 Paris, Cedex 13, France; C. JONKER, MD, Faculty SCW/LASA, Vrije Universiteit, De Boelelaan 1081C, 1081 Amsterdam, The Netherlands; S.-L. KIVELA, MD, Oulu University Hospital, Unit of General Practice PB5000, 90401 Oulu, Finland; B. A. LAWLOR, MRCPsych, Department of Psychiatry of the Elderly, St James's Hospital, Dublin, Republic of Ireland; A. LOBO, Prof. Dr, Department of Psychiatry, Hospital Clinico Universitario, Zaragoza, Spain; H. MAGNUSSON, Prof. Dr, Heilsugaeslustod Grundarfjardar, Iceland; M. FICHTER, Prof. Dr, Department of Psychiatry, University of Munich, Munich, Germany; H. VAN OYEN, MD DrPH, Department of Epidemiology, Scientific Institute of Public Health, J. Wytsmanstraat 4, 1050 Brussels, Belgium; M. ROELANDS, MD, Department of Behaviour Therapy and Counselling, University of Gent, B-9000 Gent, Belgium; I. SKOOG, MD, Psykiatriska Kliniken, Sahlgrenska sjukhuset, 413 45 Goteborg, Sweden; C. TURRINA, MD, Clinica Psichiatrica, Ospendale Civile, Breschia, Italy; J. R. M. COPELAND, FRCPsych, Department of Psychiatry, Institute of Human Ageing, University of Liverpool, Liverpool

Correspondence: M. J. Prince, Epidemiology Unit, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT