Psychological Medicine (2013), **43**, 1929–1939. © Cambridge University Press 2013
doi:10.1017/S0033291712002693

ORIGINAL ARTICLE

Predicting onset of major depression in general practice attendees in Europe: extending the application of the predictD risk algorithm from 12 to 24 months

M. King[1]*, C. Bottomley[2], J. Bellón-Saameño[3], F. Torres-Gonzalez[4], I. Švab[5], D. Rotar[5], M. Xavier[6] and I. Nazareth[2,7]

[1] *Mental Health Sciences Unit, Faculty of Brain Sciences, University College London Medical School, London, UK*
[2] *MRC Tropical Epidemiology Group, London School of Hygiene and Tropical Medicine, London, UK*
[3] *El Palo Health Centre, Department of Preventive Medicine, Malaga, Spain*
[4] *Centro de Investigación Biomedica en Red de Salud Mental (CIBERSAM), Departmental Section of Psychiatry and Psychological Medicine, University of Granada, Granada, Spain*
[5] *Department of Family Medicine, University of Ljubljana, Ljubljana, Slovenia*
[6] *Faculdade Ciências Médicas, University of Lisbon, Lisbon, Portugal*
[7] *Medical Research Council General Practice Research Framework, London, UK*

Background. PredictD is a risk algorithm that was developed to predict risk of onset of major depression over 12 months in general practice attendees in Europe and validated in a similar population in Chile. It was the first risk algorithm to be developed in the field of mental disorders. Our objective was to extend predictD as an algorithm to detect people at risk of major depression over 24 months.

Method. Participants were 4190 adult attendees to general practices in the UK, Spain, Slovenia and Portugal, who were not depressed at baseline and were followed up for 24 months. The original predictD risk algorithm for onset of DSM-IV major depression had already been developed in data arising from the first 12 months of follow-up. In this analysis we fitted predictD to the longer period of follow-up, first by examining only the second year (12–24 months) and then the whole period of follow-up (0–24 months).

Results. The instrument performed well for prediction of major depression from 12 to 24 months [c-index 0.728, 95% confidence interval (CI) 0.675–0.781], or over the whole 24 months (c-index 0.783, 95% CI 0.757–0.809).

Conclusions. The predictD risk algorithm for major depression is accurate over 24 months, extending it current use of prediction over 12 months. This strengthens its use in prevention efforts in general medical settings.

Received 21 February 2012; Revised 16 August 2012; Accepted 16 October 2012; First published online 4 January 2013

Key words: Depression, family practice, prediction.

Introduction

Depression occurs in 15–20% of general practice attendees (Goldberg & Huxley, 1992), relapse after first episode is common (Thornicroft & Sartorius, 1993) and residual impairment and premature death are relatively common consequences (Cassano & Fava, 2002). Estimating overall risk of major depression across a range of key risk factors is a key first priority in planning prevention. We recently developed predictD, a risk algorithm for the onset of major depression over 12 months in general practice attendees. The algorithm was developed in six European countries and externally validated in general practice attendees in Chile (King *et al*. 2006, 2008). We modelled our approach on indices used to predict risk of cardiovascular disease (Anderson *et al*. 1991), which provide a percentage risk estimate over a given time period. PredictD contains 10 factors: age; sex; educational level achieved; lifetime screen for depression; family history of psychological difficulties; physical health and mental health subscale scores on the Short Form 12 (Jenkinson *et al*. 1997); unsupported difficulties in paid or unpaid work; and experiences of discrimination; country. The algorithm's precision (onset of major depression over 12 months) in terms of mean c-index for all six participating European

* Address for correspondence: M. King, M.D. Ph.D., Director of Mental Health Sciences, Faculty of Brain Sciences, University College London Medical School, Charles Bell House, 67–73 Riding House Street, London W1W 7EH, UK.
(Email: michael.king@ucl.ac.uk)

countries was 0.790 [95% confidence interval (CI) 0.767–0.813]. Effect size for difference in predicted log odds of depression between those who became depressed and those who did not was 1.28 (95% CI 1.17–1.40). The external validation of the algorithm in Chilean general practice attendees resulted in a c-index of 0.710 (95% CI 0.670–0.749).

Following our first study, resources were available to enable us to follow up participants in the UK, Spain, Slovenia and Portugal for a further 12 months. Although prediction over 12 months is a pragmatic objective in terms of the day-to-day work of general practitioners (GPs), it is important to know whether the algorithm is accurate over longer periods. Given that a number of the risk variables are long term in their effect, risk prediction (and therefore prevention efforts) might be accurate over longer time courses. Our aim was to assess the precision of predictD over a further 12 months, or 24 months in all.

Method

Study setting and design

The original predictD study was conducted in six European centres:

(1) *UK*: 25 general practices in the Medical Research Council's General Practice Research Framework (MRC GPRF), in the UK. There were 40 GPs in eight rural practices and 58 GPs in 17 urban practices.
(2) *Spain*: 57 GPs in nine large primary care centres in Andalucía, Spain, in which seven were urban and two rural.
(3) *Solvenia*: 74 general practices nationwide in Slovenia.
(4) *Estonia*: 23 general practices nationwide in Estonia. Of these, 15 were urban and eight rural and all had single-handed GPs.
(5) *The Netherlands*: seven large general practice centres near Utrecht, The Netherlands. Of these, two were urban and five rural and there were 38 GPs working across the seven centres.
(6) *Portugal*: two large primary care centres in the Lisbon area of Portugal; one was urban and contained 18 GPs and the other rural and contained seven GPs.

Because the study required the cooperation of the GPs in each centre, they could not be selected randomly or comprehensively except in Slovenia where almost all the primary care centres in the country participated. Nevertheless, the general practices and health centres covered urban and rural populations with considerable socio-economic variation. The 24 months' follow-up of patients was conducted in a cohort of 4190 general practice attendees in the UK, Spain, Slovenia and Portugal who were not depressed at baseline. The study was approved by ethical committees in each country.

Study participants

Consecutive attendees aged 18–75 years were recruited to the predictD study in Europe between April 2003 and September 2004. Exclusion criteria were not being resident in the country, an inability to understand one of the main languages involved, psychosis, dementia and incapacitating physical illness. Recruitment differed slightly in each country because of local service preferences. In the UK researchers spoke to patients waiting to see practice staff. In the other three European countries doctors introduced the study before contact with researchers. Participants gave informed consent and undertook a research evaluation within 2 weeks. We followed up patients at 6, 12 and 24 months.

Major depression

A Diagnostic and Statistical Manual of Mental Disorders, fourth edition (DSM-IV) diagnosis of major depression in the preceding 6 months at recruitment, and at 6 and 12 months' follow-ups was made using the Depression Section of the Composite International Diagnostic Interview (CIDI) (Robins *et al*. 1988; WHO, 1997). However, at the 24-month follow-up the CIDI covered the preceding 12 months.

Known risk factors

We selected risk factors in order to cover key areas identified in a systematic review of the literature (Weich, 2001). Where possible, we used standardized self-report measures. Each instrument or question not available in the relevant languages was translated from English and back-translated by professional translators. Questions adapted from standardized questionnaires or developed for the study were evaluated for test–retest reliability before the main study began. A total of 285 patients across the six European countries completed the battery of questions on two occasions, 2 weeks apart. The full range of risk factors tested is described elsewhere (King *et al*. 2006, 2008). Those factors that became a part of the predictD algorithm are as follows:

- Country of participation.
- Age, sex, educational level.
- Life-time depression was based on affirmative answers to the first two questions of the CIDI life-time depression section (Robins *et al*. 1988; WHO, 1997).

https://doi.org/10.1017/S0033291712002693 Published online by Cambridge University Press

- Questions on history of serious psychological problems in first-degree relatives were based on those contained in a family history questionnaire developed for genetic research (Qureshi *et al*. 2005). As described in our earlier paper, κ/intra-class correlations for test–retest reliability of these questions ranged from 0.70 to 1.0 (King *et al*. 2006). One combined question was included in the predictD algorithm (see Appendix, available online).
- Unsupported difficulties in paid and/or unpaid work in the preceding 6 months using questions from the job content instrument (Karasek & Theorell, 1990). This instrument focuses on paid work, so we adapted the questions to include unpaid work and subjected these to test–retest reliability (King *et al*. 2006). The κ statistics for the two questions on the final predictD algorithm (for unpaid work) were 0.48 and 0.67.
- Self-rated physical and mental health was assessed by the Short Form 12 (Jenkinson *et al*. 1997). This instrument has high validity and reliability and is used extensively in clinical and epidemiological research. The summary physical and mental health measures in the Short Form 12 have a high degree of correspondence with those in the parent instrument, the Short Form 36 (Gandek *et al*. 1998). The weights used to calculate scores are from version 1.
- Experiences of discrimination in the preceding 6 months on grounds of sex, age, ethnicity, appearance, disability or sexual orientation using questions from a European study (Janssen *et al*. 2003). Although these questions had been used in Europe previously, their reliability was unknown and thus they were included in our assessment of test–retest reliability. The κ statistics for these questions ranged from 0.53 to 1.0.

Statistical analysis

All analyses were performed using Stata release 11 (StataCorp LP, USA). We included only patients without major depression at baseline. Participants with missing depression diagnoses at any point were excluded, as these data were necessary for the validation. In our original analysis to derive the predictD algorithm (King *et al*. 2008), and in our analysis of its extended function here, we calculated the c-index (Harrell, 2001) in each country and all countries combined. The c-index is equivalent to the area under a receiver operating characteristic curve (plot of sensitivity *v*. 1 – specificity), in that it measures the model's ability to discriminate between individuals who become depressed and those who do not. For example, a c-index of 0.8 means that if you randomly selected one of the individuals who became depressed and one who did not, then on 80% of occasions the individual who became depressed will have the higher risk as predicted by the model. We assessed the goodness of fit of the final risk model by grouping individuals into deciles of risk and comparing the observed probability of major depression within these groups with the average predicted risk.

Recalibration of predictD (logit) risk score to predict incidence in year 2

Recalibration is a simple method of adjusting an existing model to predict risk in a new setting. It involves estimating only two new parameters and as such is expected to produce reasonable predictions beyond the dataset used for recalibration (i.e. the model does not over-fit the data) (Steyerberg *et al*. 2004; Janssen *et al*. 2008). Typically recalibration is used when a risk score is applied outside the population for which the original model was developed. Here we use the method to improve risk prediction when the model is applied over a different time period. The logit risk score, Z, is recalibrated to predict year 2 incidence by fitting a logistic model to data on incidence between month 12 and month 24 with Z as the predictor variable, i.e. the slope (a) and intercept (b) are estimated for the model logit (incidence year 2) $= a + bZ$. The logit risk score (Z) was calculated using imputed data that was created for the original predictD study (King *et al*. 2008).

Predicting incidence over 24 months

To predict incidence over the whole 24 months we used the following relationship between the 24-month incidence and the incidence in each year:

$$\text{Overall incidence} = 1 - [(1 - \text{incidence year 1}) \times (1 - \text{incidence year 2})],$$

where the predicted incidence during year 1 is simply the original 12-month risk score, and the predicted incidence in year 2 comes from the recalibrated score.

Finally we report the threshold values of risk score, and the associated sensitivity, for a range of specificity that would be practical (minimizing false positives) when using the instrument in a clinical setting.

Results

Numbers and missing data

Response to recruitment was high in Portugal (76%), Estonia (80%), Slovenia (80%) and Chile (97%) but lower in the UK (44%) and The Netherlands (45%). Ethical considerations prevented the collection of data

https://doi.org/10.1017/S0033291712002693 Published online by Cambridge University Press

Table 1. *Demographic characteristics*

	All European countries	UK	Spain	Slovenia	Portugal	The Netherlands	Estonia
n (% of European sample)	6190 (100.0)	1131 (18.3)	1006 (16.3)	1048 (16.9)	1005 (16.2)	1077 (17.4)	923 (14.9)
Mean age, years (S.D.)	48.9 (15.5)	52.2 (14.7)	50.8 (15.5)	48.8 (14.5)	50.2 (15.4)	48.9 (14.9)	41.6 (16.0)
Sex, *n* (%)							
Female	4081 (55.9)	750 (66.3)	689 (68.5)	660 (63.0)	649 (64.6)	668 (62.0)	665 (72.0)
Marital status, *n* (%)							
Married or living together	4491 (72.6)	844 (74.6)	708 (70.4)	732 (69.9)	750 (74.6)	827 (76.8)	630 (68.3)
Separated or divorced	421 (6.8)	100 (8.8)	49 (4.9)	56 (5.3)	69 (6.9)	64 (5.9)	83 (9.0)
Single	872 (14.1)	121 (10.7)	181 (18.0)	152 (14.5)	132 (13.1)	121 (11.2)	165 (17.9)
Widowed	383 (6.2)	65 (5.8)	67 (6.7)	105 (10.0)	53 (5.3)	48 (4.5)	45 (4.9)
Missing	23 (0.4)	1 (0.1)	1 (0.1)	3 (0.3)	1 (0.1)	17 (1.6)	0 (0)
Household status, *n* (%)							
Not living alone	5483 (88.6)	981 (86.7)	948 (94.2)	915 (87.3)	929 (92.4)	894 (83.0)	816 (88.4)
Living alone	707 (11.4)	150 (13.3)	58 (5.8)	133 (12.7)	76 (7.6)	183 (17.0)	107 (11.6)
Missing	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Education, *n* (%)							
Higher education	1879 (30.4)	448 (39.6)	135 (13.4)	181 (17.3)	129 (12.8)	458 (42.5)	528 (57.2)
Secondary	2038 (32.9)	465 (41.1)	215 (21.4)	385 (36.7)	182 (18.1)	508 (47.2)	283 (30.7)
Primary and no education	1767 (28.6)	25 (2.2)	656 (65.2)	235 (22.4)	662 (65.9)	78 (7.2)	111 (12.0)
Trade and other	451 (7.3)	171 (15.1)	0 (0)	247 (23.6)	32 (3.2)	0 (0)	1 (0.1)
Missing	55 (0.9)	22 (1.9)	0 (0)	0 (0)	0 (0)	33 (3.1)	0 (0)
Employment, *n* (%)							
Employed or full-time student	3256 (52.6)	574 (50.8)	349 (34.7)	563 (53.7)	486 (48.4)	602 (55.9)	682 (73.9)
Unemployed	300 (4.8)	21 (1.9)	62 (6.2)	53 (5.1)	108 (10.7)	35 (3.2)	21 (2.3)
Unable to work	322 (5.2)	86 (7.6)	101 (10.0)	16 (1.5)	38 (3.8)	48 (4.5)	33 (3.6)
Retired or looking after family	2269 (36.7)	450 (39.8)	493 (49.0)	409 (39.0)	372 (37.0)	358 (33.2)	187 (20.3)
Missing	43 (0.7)	0 (0)	1 (0.1)	7 (0.7)	1 (0.1)	34 (3.2)	0 (0)

S.D., Standard deviation.

on non-responders at baseline. A total of 4190 general practice attendees from the UK, Spain, Slovenia and Portugal were not depressed at baseline and were followed up for 24 months (Table 1). Within this cohort 3427 (81.7%) had data on their depression status at 12 months and 2670 (63.7%) had data at 12 months and 24 months. The pattern of missing depression data at each outcome is shown in Table 2. Attrition over 24 months was highest in Spain and lowest in Slovenia. It was also higher in 18- to 20-year-olds than in the other age groups.

Incidence of major depression

During the first 12 months the incidence of major depression had been 8.8% (302/3427). The incidence between 12 and 24 months was lower – of the 2440 participants who did not have an episode of major depression in the first 12 months and who had 24-month data, only 78 (3.2%) became depressed during the second year of follow-up. Incidence over the 24 months as a whole was 11.5% (308/2670), with the highest rate in Spain (19.9%) and the lowest in Slovenia (5.3%) (Table 3).

Prediction from 12 to 24 months

The logit risk score for incidence in year 2 was estimated to be $-1.51 + (0.76 \times Z)$, where Z is the original logit risk score for depression in the first 12 months. The recalibrated risk score appears to predict incidence in year 2 well in the combined data from four countries (Fig. 1). The overall c-index was 0.728 (95% CI 0.675–0.781) (Table 3). When this new risk score was applied separately in each country, the predicted risks were in reasonable agreement with observed risk

https://doi.org/10.1017/S0033291712002693 Published online by Cambridge University Press

Table 2. *Percentage missing 12-month and 24-month depression data by baseline characteristics*

		% Missing		
	n	12 months	24 months	24 months: *p*
Age				
18–29 years	474	27.0	45.4	
30–49 years	1415	18.1	36.1	
50–69 years	1833	16.6	30.2	
70+ years	468	15.8	26.7	<0.001
Sex				
Male	2748	17.9	32.6	
Female	1442	18.9	35.3	0.075
Education				
Beyond secondary	893	16.3	32.4	
Secondary	1247	15.8	29.8	
Primary or no education	1578	21.9	37.6	
Trade and other	450	15.6	32.0	<0.001
Difficulties in paid and unpaid work				
No difficulties or often supported	3529	17.8	33.1	
Difficulties without support	623	20.5	36.1	0.141
Physical health, SF12 score[a]				
10–29.9	599	17.2	33.6	
30–49.9	2002	17.6	33.0	
50–69.9	1589	19.3	34.2	0.749
Mental health, SF12 score[a]				
10–29.9	273	19.0	35.5	
30–49.9	1678	20.3	36.0	
50–79.9	2239	16.6	31.4	0.008
First-degree relative with emotional problem				
No	2821	17.6	33.7	
Yes	1299	19.4	32.8	0.547
Discrimination				
None	3852	18.4	33.4	
Discrimination in one area	237	14.8	32.9	
Discrimination in more than one area	91	22.0	40.7	0.343
Lifetime depression				
No	2624	17.8	33.6	
Yes	1562	18.9	33.4	0.884
Country				
UK	1131	17.0	27.3	
Spain	1006	29.0	48.0	
Slovenia	1048	12.7	31.6	
Portugal	1005	14.5	28.0	<0.001

[a] Short Form 12 score (Jenkinson *et al*. 1997).

in all except Slovenia, where the model also performed poorly in terms of the c-index (i.e. the model was poor at discriminating between those who became depressed and those who did not).

Prediction over the full 24 months

We also predicted incidence of major depression over 24 months (Fig. 2). The model was a good

https://doi.org/10.1017/S0033291712002693 Published online by Cambridge University Press

Table 3. *Incidence of major depression over 24 months*

	UK	Spain	Slovenia	Portugal	Total
Year 1 incidence					
No	856 (91.2)	606 (84.9)	877 (95.8)	786 (91.5)	3125 (91.2)
Yes	83 (8.8)	108 (15.1)	38 (4.2)	73 (8.5)	302 (8.8)
Total	939 (100)	714 (100)	915 (100)	859 (100)	3427 (100)
Year 2 incidence					
No	684 (95.9)	374 (96.1)	666 (98.5)	638 (96.4)	2362 (96.8)
Yes	29 (4.1)	15 (3.9)	10 (1.5)	24 (3.6)	78 (3.2)
Total	713 (100)	389 (100)	676 (100)	662 (100)	2440 (100)
Overall incidence					
No	684 (87.7)	374 (80.1)	666 (94.7)	638 (88.6)	2362 (88.5)
Yes	96 (12.3)	93 (19.9)	37 (5.3)	82 (11.4)	308 (11.5)
Total	780 (100)	467 (100)	703 (100)	720 (100)	2670 (100)

Data are given as number of participants (percentage).

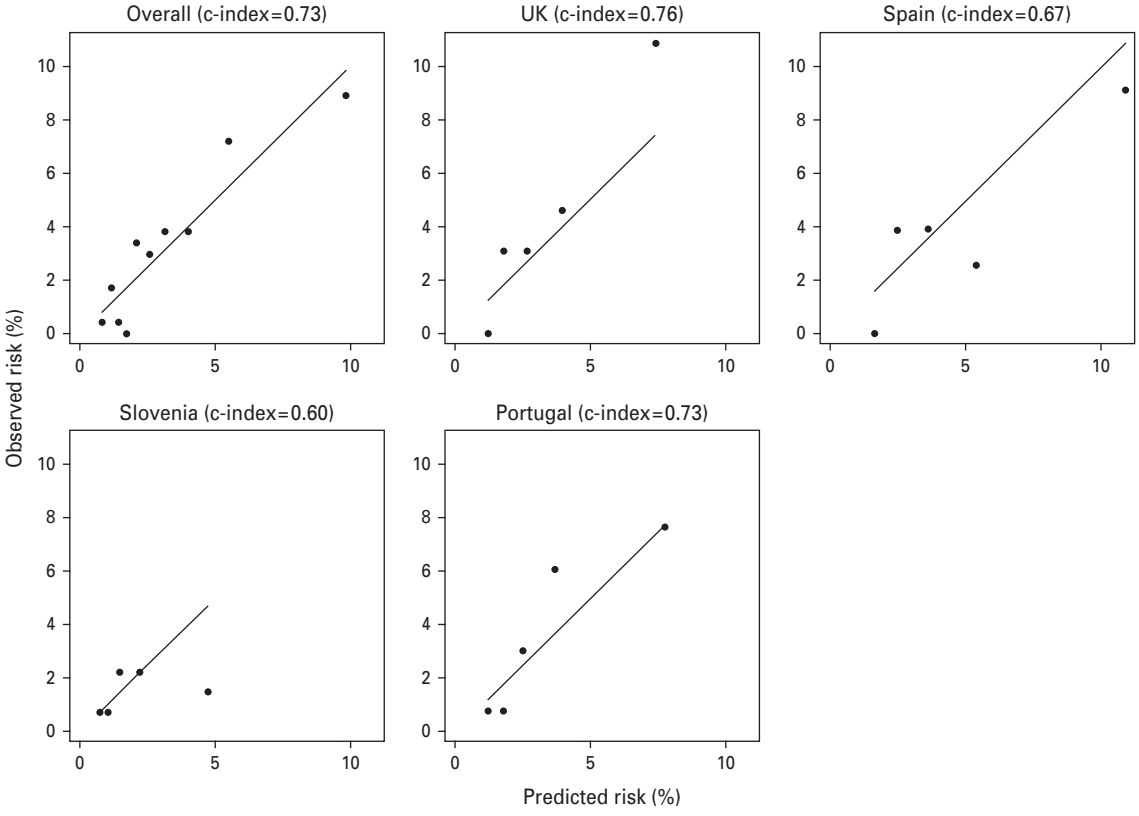

Fig. 1. Recalibrated risk score used to predict incidence between 12 and 24 months. Mean predicted risk *versus* observed risk for each decile (overall) or quintile (within countries) of predicted risk. Overall (c-index = 0.73); UK (c-index = 0.76); Spain (c-index = 0.67); Slovenia (c-index = 0.60); Portugal (c-index = 0.73). The c-index is equivalent to the area under a receiver operating characteristic curve (plot of sensitivity *v.* 1 – specificity), in that it measures the model's ability to discriminate between individuals who become depressed and those who do not.

predictor of incidence over this period but this is partly because most events occurred within the first year and these data were used to build the original model. The overall c-index is 0.783 (95% CI 0.757–0.809) (Table 4). Sensitivity and specificities for different thresholds of risk over 24 months are given

https://doi.org/10.1017/S0033291712002693 Published online by Cambridge University Press

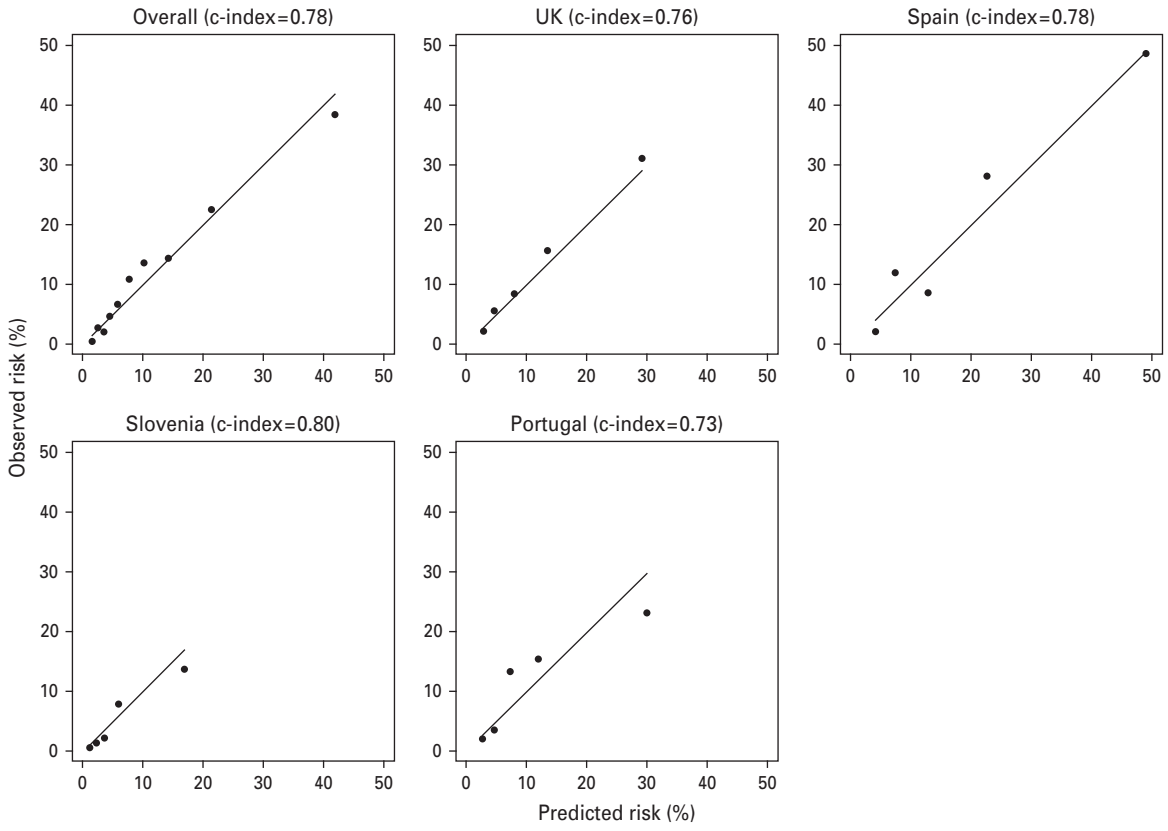

Fig. 2. Recalibrated risk score used to predict incidence over 24 months. Mean predicted risk *versus* observed risk for each decile (overall) or quintile (within countries) of predicted risk. Overall (c-index = 0.78); UK (c-index = 0.76); Spain (c-index = 0.78); Slovenia (c-index = 0.80); Portugal (c-index = 0.73). The c-index is equivalent to the area under a receiver operating characteristic curve (plot of sensitivity *v.* 1 – specificity), in that it measures the model's ability to discriminate between individuals who become depressed and those who do not.

Table 4. *c-Indices*[a] *by country*

Country	12- to 24-month incidence	24-month incidence
Overall	0.728 (0.675–0.781)	0.783 (0.757–0.809)
UK	0.756 (0.678–0.835)	0.757 (0.706–0.807)
Spain	0.675 (0.544–0.805)	0.782 (0.728–0.836)
Slovenia	0.602 (0.433–0.771)	0.795 (0.724–0.865)
Portugal	0.734 (0.650–0.818)	0.733 (0.682–0.784)

Data are given as c-index (95% confidence interval).

[a] The c-index is equivalent to the area under a receiver operating characteristic curve (plot of sensitivity *v.* 1 – specificity), in that it measures the model's ability to discriminate between individuals who become depressed and those who do not.

in Table 5. Risk factors and their coefficients are given in Table 6 and a worked example of how to calculate a risk score in an individual is shown in Table 7.

Table 5. *Sensitivity and specificity at different score cut-offs for use of predictD over 24 months*

Country	Specificity	Sensitivity	Score cut-off
Overall	0.8	0.58	0.144
	0.85	0.52	0.177
	0.9	0.41	0.223
UK	0.8	0.53	0.161
	0.85	0.46	0.185
	0.9	0.38	0.222
Spain	0.8	0.65	0.233
	0.85	0.57	0.265
	0.9	0.45	0.343
Slovenia	0.8	0.53	0.078
	0.85	0.50	0.093
	0.9	0.47	0.122
Portugal	0.8	0.43	0.140
	0.85	0.38	0.181
	0.9	0.27	0.239

https://doi.org/10.1017/S0033291712002693 Published online by Cambridge University Press

Table 6. *Risk factors in the predictD algorithm and their regression coefficients*

Prognostic factors	Levels in factor	Coefficient[a]
Constant		1.155
Age	Each year	−0.005
Sex	Female	
	Male	−0.212
Education	Beyond secondary education	
	Secondary education	0.089
	Primary or no education	0.409
	Trade and other	0.566
Difficulties in paid and unpaid work	No difficulties or often supported	
	Difficulties without support	0.366
Physical health	Each point on SF12 subscale score[b]	−0.03
Mental health	Each point on SF12 subscale score[b]	−0.055
First-degree relative with emotional problem	No	
	Yes	0.395
Discrimination	None	
	In one area	0.161
	In more than one area	0.736
Lifetime depression	No	
	Yes	0.489
Country[c]	UK	
	Spain	0.23
	Slovenia	−0.729
	Estonia	−0.467
	The Netherlands	−0.115
	Portugal	−0.169

[a] Regression coefficients after shrinkage for over-fitting. See King *et al.* (2008).
[b] Short Form 12 score (Jenkinson *et al*. 1997).
[c] We include all countries in the original predictD study here because the algorithm for 24 months can in principle be used in Estonia and The Netherlands even though it has not been validated over 24 months in these countries.

Discussion

Our aim in developing the predictD algorithm was to select key risk factors from the many reported in the literature in a simple and yet valid risk equation for prediction of major depression over 12 months. We have now demonstrated that predictD also predicts risk over 24 months but at slightly lower levels of precision. The c-index provides a standardized way of comparing the discriminative power of tests that use different measurement units in different settings (Pepe *et al*. 2004). The 24-month predictD risk score (c-index = 0.783) compares favourably with a risk index for cardiovascular events developed in 12 European cohorts (Conroy *et al*. 2003), which reported c-indices between 0.71 and 0.82. Although developed in general practice attendees, it is likely to have validity in general populations, as most people in the UK and other European countries attend their GP at least once per year.

Strengths and limitations

The main strength is that use of predictD was extended to 24 months in a large population in four European countries. Lower recruitment rates occurred in the UK and The Netherlands, possibly because the study was not so obviously endorsed by doctors. Although this was compensated for by high follow-up rates at 6 and 12 months, the rate was lower at 24 months for the people in this four-country cohort who were not depressed at baseline. There were differences in the geographical distribution of general practices in each country, which reflected the varying networks available to the centres. Although our risk factors are based on self-report, in the main we used standardized instruments and where unstandardized questions were used, all were tested for reliability (King *et al*. 2006). Assessing the occurrence of major depression on the CIDI over 12 months at the 24 months follow-up (instead of over 6 months as

https://doi.org/10.1017/S0033291712002693 Published online by Cambridge University Press

Table 7. *Worked example of use of predictD*

	Example and workings
Factor	
Age	23 years
Sex	Female
Education	Secondary
Difficulties in paid and unpaid work	No difficulties or often supported
Physical health	SF12 subscale score[a] = 55.1
Mental health	SF12 subscale score[a] = 26.2
First-degree relative with emotional problem	Yes
Discrimination	No
Lifetime depression	Yes
Country	UK
Calculations	
Using the coefficients in Table 4, the score (on the logit scale) for this individual is −1.081, hence her predicted risk for the following 12 months is	$\exp(-1.081)/[1+\exp(-1.081)]=0.253$
The risk of depression in the second 12 months (assuming she did not become depressed in the first 12 months) is obtained by recalibrating the score. On the logit scale the recalibrated score is	$-1.51+(0.76\times-1.081)=-2.332$
Giving a risk of	$\exp(-2.332)/[1+\exp(-2.332)]=0.089$
The predicted 24-month risk can be obtained from the 0- to 12-month risk and the 12- to 24-month risk and is	$1-(1-0.253)\times(1-0.089)=0.319$

[a] Short Form 12 score (Jenkinson *et al*. 1997).

at baseline, 6 and 12 months outcomes) potentially introduces a degree of imprecision in that recall may have been less accurate over this longer period.

Implementation

Our study does not address how predictD might best be used in general practice. However, the questions making up the algorithm are short and easy to complete and thus it has potential as a clinical tool for prediction of future episodes of depression in this setting. In reporting a range of thresholds for sensitivity and specificity (Table 5), we would maximize specificity at the cost of reduced sensitivity in order to limit the workload for family doctors engaging with false positives. For example, if a primary care doctor were to apply a European threshold for risk of 22.3% (i.e. specificity of 0.9 and sensitivity of 0.41) they could be sure that the numbers of patients wrongly identified as people at risk of major depression (false positives) would be kept to a minimum. Although this would be at the cost of missing some of those who would go on to develop major depression over 24 months, use of a high cut-off ensures that prevention efforts are less likely to be wasted on those not at risk. On the other hand, if the prevention interventions require little input by way of physician time and effort (e.g. a web-based depression self-help prevention package), a lower cut-off, say 14.4% (i.e. specificity of 0.8 and sensitivity of 0.58), might be considered, as the larger number of positives caught in the net could be offered the intervention without substantially increasing costs to the health service. These threshold values at the same specificity (i.e. 0.9 and 0.8) levels using the 12-month algorithm were 16.5% and 10.6%, respectively (King *et al*. 2008). It appears that in general a higher cut-off can be considered if an overall 24-month perspective were to be considered rather than 12 months. This would result in delivery of the preventive intervention to fewer patients.

Application in the general population

PredictD was developed in general practice attendees, as that setting was considered to be appropriate both for discussing risk with people and for prevention efforts. We cannot be sure that it will function similarly in general populations. However, in the countries in which predictD was developed and validated there are national health services that are free at the point of delivery. Thus, in these countries we suspect the algorithm will function just as it does in general

https://doi.org/10.1017/S0033291712002693 Published online by Cambridge University Press

practice attendees. However, this needs confirmation in further research.

Prevention

Even though a number of predictD risk factors are immutable, advising patients on the nature of depression, or on brief cognitive behaviour strategies they might undertake to reduce their risk, might be helpful. The same might be true for starting or restarting antidepressant medication. We are currently undertaking two randomized clinical trials in which we are evaluating predictD as part of brief prevention strategies in general practice for people at high risk. One is an ongoing evaluation of the effectiveness of a brief GP intervention in a large number of general practices in Spain, while the second is a feasibility trial in the UK of the effectiveness of a web-based package of education and information for patients at risk.

Conclusions

The predictD risk algorithm is an accurate indicator of the risk of major depression over 24 months but performs slightly less well over the second 12 months. It may be useful as a strategy to identify those at risk in prevention efforts in general medical settings.

Supplementary material

For supplementary material accompanying this paper visit http://dx.doi.org/10.1017/S0033291712002693.

Acknowledgements

The research in Europe was funded by a grant from The European Commission (grant no. PREDICT-QL4-CT2002-00683). We are also grateful for part support in Europe from: the Slovenian Ministry for Research (grant no. 4369-1027); the Spanish Ministry of Health (grant FIS references no. PI041980, PI041771 and PI042450) and the Spanish Network of Primary Care Research, redIAPP (ISCIII-RETIC RD06/0018) and SAMSERAP group; and the UK National Health Service (NHS) Research and Development office for providing service support costs in the UK. The funders had no direct role in the design or conduct of the study, interpretation of the data or review of the manuscript.

M.K. had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. We thank all patients and general practice staff who took part; the European Office at University College London for their administrative assistance at the coordinating centre; Mr Kevin McCarthy, the project's scientific officer in the European Commission, Brussels, for his helpful support and guidance; the UK MRC GPRF; Ms Louise Letley, from the MRC GPRF; and the Camden and Islington NHS Foundation Trust, London, UK.

Declaration of Interest

None.

References

Anderson KM, Wilson PW, Odell PM, Kannel WB (1991). An updated coronary risk profile. A statement for health professionals. *Circulation* **83**, 356–362.

Cassano P, Fava M (2002). Depression and public health: an overview. *Journal of Psychosomatic Research* **53**, 849–857.

Conroy RM, Pyorala K, Fitzgerald AP, Sans S, Menotti A, De Backer G, De Bacquer D, Ducimetiere P, Jousilahti P, Keil U, Njolstad I, Oganov RG, Thomsen T, Tunstall-Pedoe H, Tverdal A, Wedel H, Whincup P, Wilhelmsen L, Graham IM, and on behalf of the SCORE Project Group (2003). Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. *European Heart Journal* **24**, 987–1003.

Gandek B, Ware JE, Aaronson NK, Apolone G, Bjorner JB, Brazier JE, Bullinger M, Kaasa S, Leplege A, Prieto L, Sullivan M (1998). Cross-validation of item selection and scoring for the SF-12 Health survey in nine countries: results from the IQOLA Project. *Journal of Clinical Epidemiology* **11**, 1171–1178.

Goldberg DP, Huxley P (1992). *Common Mental Disorders: A Bio-Social Model*. Routledge: London/New York.

Harrell FE (2001). *Regression Modelling Strategies*. Springer: New York.

Janssen I, Hanssen M, Bak M, Bijl RV, de Graaf R, Vollebergh W, McKenzie K, van Os J (2003). Discrimination and delusional ideation. *British Journal of Psychiatry* **182**, 71–76.

Janssen KJ, Moons KG, Kalkman CJ, Grobbee DE, Vergouwe Y (2008). Updating methods improved the performance of a clinical prediction model in new patients. *Journal of Clinical Epidemiology* **61**, 76–86.

Jenkinson C, Layte R, Jenkinson D, Lawrence K, Petersen S, Paice C, Stradling J (1997). A shorter form health survey: can the SF-12 replicate results from the SF-36 in longitudinal studies? *Journal of Public Health Medicine* **19**, 179–186.

Karasek RA, Theorell T (1990). *Healthy Work: Stress, Productivity, and the Reconstruction of Working Life*. Basic Books: New York.

King M, Walker C, Levy G, Bottomley C, Royston P, Weich S, Bellón-Saameño J, Moreno B, Svab I, Rotar D, Rifel J, Maaroos H, Aluoja A, Kalda R, Neeleman J, Geerlings MI, Xavier M, Carraça I, Gonçalves-Pereira M, Vicente B, Saldivia S, Melipillan R, Torres-Gonzalez F, Nazareth I (2008). Development and validation of an

international risk prediction algorithm for episodes of major depression in general practice attendees: the PredictD study. *Archives of General Psychiatry* **65**, 1368–1376.

King M, Weich S, Torres F, Svab I, Maaroos H, Neeleman J, Xavier M, Morris R, Walker C, Bellon JA, Moreno B, Rotar D, Rifel J, Aluoja A, Kalda R, Geerlings MI, Carraca I, Caldas de Almeida M, Vicente B, Saldivia S, Rioseco P, Nazareth I (2006). Prediction of depression in European general practice attendees: the PREDICT study. *BMC Public Health* **6**, 6.

Pepe MS, Janes H, Longton G, Leisenring W, Newcomb P (2004). Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. *American Journal of Epidemiology* **159**, 882–890.

Qureshi N, Bethea J, Modell B, Brennan P, Papageorgiou A, Raeburn S, Hapgood R, Modell M (2005). Collecting genetic information in primary care: evaluating a new family history tool. *Family Practice* **22**, 663–669.

Robins LN, Wing J, Wittchen HU, Helzer JE, Babor TF, Burke J, Farmer A, Jablenski A, Pickens R, Regier DA (1988). The Composite International Diagnostic Interview. An epidemiologic instrument suitable for use in conjunction with different diagnostic systems and in different cultures. *Archives of General Psychiatry* **45**, 1069–1077.

Steyerberg EW, Borsboom GJJM, van Houwelingen HC, Eijkemans MJC, Habbema JD (2004). Validation and updating of predictive logistic regression models: a study on sample size and shrinkage. *Statistics in Medicine* **23**, 2567–2586.

Thornicroft G, Sartorius N (1993). The course and outcome of depression in different cultures: 10-year follow-up of the WHO Collaborative Study on the Assessment of Depressive Disorders. *Psychological Medicine* **23**, 1023–1032.

Weich S (2001). Risk factors for the common mental disorders in primary care (Ph.D. thesis). University of Cambridge: Cambridge.

WHO (1997). *Composite International Diagnostic Interview* (*CIDI*). *Version 2.1*. World Health Organization: Geneva.