Psychological Medicine (2011), **41**, 2075–2088. © Cambridge University Press 2011
doi:10.1017/S0033291711000468

ORIGINAL ARTICLE

Predicting the onset of major depression in primary care: international validation of a risk prediction algorithm from Spain

J. Á. Bellón[1]*, J. de Dios Luna[2], M. King[3], B. Moreno-Küstner[4], I. Nazareth[5], C. Montón-Franco[6], M. J. GildeGómez-Barragán[7], M. Sánchez-Celaya[8], M. Á. Díaz-Barreiros[9], C. Vicens[10], J. A. Cervilla[11], I. Švab[12], H.-I. Maaroos[13], M. Xavier[14], M. I. Geerlings[15], S. Saldivia[16], B. Gutiérrez[11], E. Motrico[17], M. T. Martínez-Cañavate[18], B. Oliván-Blázquez[19], M. S. Sánchez-Artiaga[20], S. March[21], M. del Mar Muñoz-García[22], A. Vázquez-Medrano[7], P. Moreno-Peral[23] and F. Torres-González[11]

[1] *Centro de Salud El Palo, Unidad de Investigación del Distrito de Atención Primaria de Málaga (redIAPP, grupo SAMSERAP); Departamento de Medicina Preventiva, Universidad de Málaga, Spain;* [2] *Departamento de Bioestadística (redIAPP, grupo SAMSERAP), Universidad de Granada, Spain;* [3] *Department of Mental Health Sciences, University College London, UK;* [4] *Fundación IMABIS; Unidad de Investigación del Distrito de Atención Primaria de Málaga (redIAPP, grupo SAMSERAP); Departamento de Personalidad, Evaluación y Tratamiento Psicológico, Universidad de Málaga, Spain;* [5] *Department of Primary care and Population Health, University College London and Medical Research Council General Practice Research Framework, UK;* [6] *Centro de Salud Casablanca. (redIAPP, grupo Aragón); Departamento de Medicina y Psiquiatría, Universidad de Zaragoza, Spain;* [7] *Unidad Docente de Medicina Familiar y Comunitaria de La Rioja, Servicio Riojano de la Salud, Logroño, La Rioja, Spain;* [8] *Unidad Docente de Medicina Familiar y Comunitaria, Área I de Atención Primaria, Madrid, Coordinadora de Investigación de la Sociedad Española de Medicina Familiar y Comunitaria, Spain;* [9] *Centro de Salud Vecindario, Gerencia de Atención Primaria de Gran Canaria, Servicio Canario de Salud, Las Palmas, Spain;* [10] *Centro de Salud son Serra-La Vileta, Unidad Docente de Medicina Familiar y Comunitaria de Mallorca, Instituto Balear de la Salud (redIAPP, grupo Baleares), Palma de Mallorca, Illes Balears, Spain;* [11] *CIBERSAM, Departamento de Psiquiatría y Medicina legal, Universidad de Granada, Spain;* [12] *Department of Family Medicine, University of Ljubljana, Slovenia;* [13] *Faculty of Medicine, University of Tartu, Estonia;* [14] *Faculdade Ciências Médicas, University of Lisbon, Portugal;* [15] *University Medical Centre, Utrecht, The Netherlands;* [16] *Departamento de Psiquiatría y Salud Mental, Universidad de Concepción, Chile;* [17] *Fundación IMABIS, Unidad de Investigación del Distrito de Atención Primaria de Málaga (redIAPP, grupo SAMSERAP); Departamento de Psicología Social, Universidad de Málaga, Spain;* [18] *Fundación IAVANTE, Granada, Spain;* [19] *Unidad de Investigación de Atención Primaria (redIAPP, grupo Aragón); Instituto Aragonés de Ciencias de la Salud, Zaragoza, Spain;* [20] *Centro de Salud Condes de Barcelona-Boadilla, Área 6 de Atención Primaria, Madrid, Spain;* [21] *Unidad de Investigación de Atención Primaria de Baleares (redIAPP, grupo Baleares), Mallorca, Spain;* [22] *Departamento de Psiquiatría y Medicina legal, Universidad de Granada, Spain;* [23] *Fundación IMABIS, Unidad de Investigación del Distrito de Atención Primaria de Málaga (redIAPP, grupo SAMSERAP), Málaga, Spain*

Background. The different incidence rates of, and risk factors for, depression in different countries argue for the need to have a specific risk algorithm for each country or a supranational risk algorithm. We aimed to develop and validate a predictD-Spain risk algorithm (PSRA) for the onset of major depression and to compare the performance of the PSRA with the predictD-Europe risk algorithm (PERA) in Spanish primary care.

Method. A prospective cohort study with evaluations at baseline, 6 and 12 months. We measured 39 known risk factors and used multi-level logistic regression and inverse probability weighting to build the PSRA. In Spain (4574), Chile (2133) and another five European countries (5184), 11 891 non-depressed adult primary care attendees formed our at-risk population. The main outcome was DSM-IV major depression (CIDI).

Results. Six variables were patient characteristics or past events (sex, age, sex × age interaction, education, physical child abuse, and lifetime depression) and six were current status [Short Form 12 (SF-12) physical score, SF-12 mental score, dissatisfaction with unpaid work, number of serious problems in very close persons, dissatisfaction with living together at home, and taking medication for stress, anxiety or depression]. The C-index of the PSRA was 0.82 [95% confidence interval (CI) 0.79–0.84]. The Integrated Discrimination Improvement (IDI) was 0.0558 [standard error (S.E.) = 0.0071, $Z_{exp} = 7.88$, $p < 0.0001$] mainly due to the increase in sensitivity. Both the IDI and calibration plots showed that the PSRA functioned better than the PERA in Spain.

Conclusions. The PSRA included new variables and afforded an improved performance over the PERA for predicting the onset of major depression in Spain. However, the PERA is still the best option in other European countries.

Received 15 July 2010; Revised 23 February 2011; Accepted 2 March 2011; First published online 5 April 2011

Key words: Depression, prediction, primary health care, risk factors.

* Address for correspondence: J. Á. Bellón, M.D., Ph.D., Departamento de Medicina Preventiva, Facultad de Medicina, Universidad de Málaga, Campus de Teatinos, 29071 Málaga, Spain.
(Email: JABELLON@terra.es)

Introduction

Effective strategies for preventing depression and reducing disease burden are hindered by lack of evidence about whether the risk for major depression can be quantified in the same way as other clinical disorders, such as cardiovascular disease (Conroy *et al*. 2003). The predictD study is a pioneering international study whose main objective was to develop a risk index for the onset of major depression in general practice attendees (King *et al*. 2006). From 39 potential risk factors for depression, a risk index of 10 risk factors was drawn up for Europe; it has excellent predictive power and good external validity (King *et al*. 2008*b*). However, risk models do not always apply well across countries. For example, the incidence of myocardial infarction is relatively low in Spain and other southern European countries and individual cardiovascular risk estimates based on the classic Framingham model have been shown to overestimate the actual individual risk in Spanish persons (Conroy *et al*. 2003; Marrugat *et al*. 2007). Likewise, evidence exists of different prevalence rates of depression across Europe, even after adjusting for likely confounding factors (King *et al*. 2008*a*). Moreover, important variations are found across European countries concerning the prevalence of factors associated with depression, including co-morbidity of mood and anxiety disorders, living arrangements, unemployment, perception of life events and social support, quality of life, impact of work on mental health, and seeking help from services for mental health problems or misuse of psychotropic drugs (European Commission, 2004; De Girolamo *et al*. 2006; König *et al*. 2009). Consequently, rather than simply calibrating existing risk prediction tools, new country-specific prediction risk scores for the onset of depression are required. We aimed to develop and validate a risk algorithm for the onset of major depression in Spanish primary care attendees, and to compare the performance of the European and Spanish risk algorithms in Spanish data.

Method

Design

We undertook a prospective cohort study to develop and validate a risk prediction algorithm for the onset of major depression at 12 months in Spanish primary care attendees. The method has been described in detail elsewhere (King *et al*. 2006, 2008*b*; Bellón *et al*. 2008). The predictD-Spain study was approved by ethics committees in each Spanish province.

Setting

Seven provinces participated with 41 health centres and 231 physicians distributed throughout Spain: Malaga and Granada in southern Spain; Saragossa and La Rioja in northern Spain; Madrid, capital of Spain, situated in the centre; Las Palmas in the Canary Islands; and Majorca in the Balearic Islands. Each health centre covers a population of 15 000–30 000 inhabitants from a geographically defined area. The physicians in each health centre work as a group, with extensive primary care teams. The Spanish National Health Service provides free medical cover to 100% of the population. The health centres taking part cover urban and rural settings in each province.

The external validation study used data collected in the original predictD-International study in five other European countries: 25 health centres in the Medical Research Council General Practice Research Framework in the UK; 74 health centres nationwide in Slovenia; 23 health centres nationwide in Estonia; seven large health centres near Utrecht, The Netherlands; two large health centres in Portugal, one in Lisbon and the other in Alentejo; and 78 health centres in Concepción and Talcahuano in the Eighth Region of Chile.

Participants

In the six Spanish provinces, systematic random samples from physician appointment lists were taken at regular intervals of between four and six attendees with random starting points for each day. The study population, aged 18 to 75 years, was recruited between October 2005 and February 2006. The seventh province, Malaga, recruited between October 2003 and February 2004 as it was already participating in the predictD-International study. The external validation data were collected in consecutive attendees aged 18–75 years who had been recruited in Europe between April 2003 and September 2004 and in Chile between October 2003 and February 2005. Exclusion criteria for all participant countries were an inability to understand one of the main languages involved, psychosis, dementia, and incapacitating physical illness. In the UK and The Netherlands, patients were recruited in health centre waiting rooms whereas in the other countries recruitment was conducted in discussion with the family physician. In Chile, attendees were randomly selected stratified by age and sex in each centre. Participants who gave informed consent undertook a research interview within 2 weeks.

Variables

A DSM-IV diagnosis of major depression in the preceding 6 months was made using the depression

section of the Composite International Diagnostic Interview (CIDI) at baseline, 6 and 12 months (Robins *et al*. 1988; Rubio-Stipec *et al*. 1991; WHO, 1997). The risk factors selected cover all important areas identified in a systematic review of the literature. For the test–retest analysis, we selected in Spain a random sample of 401 patients stratified by province; 251 completed researcher-administered questionnaires and 150 self-administered questionnaires before the main study began (Bellón *et al*. 2008). Test–retest reliability of questions used in the predictD-International study has been reported previously (King *et al*. 2006). All potential risk factors for depression were measured at baseline:

- Sociodemographic factors: age, sex, marital status, occupation, employment status, ethnicity, nationality, country of birth, educational level, income, owner-occupier of an accommodation, living alone or with others.
- Controls, demands and rewards for unpaid and paid work, using an adapted version of the job content instrument (Karasek & Theorell, 1990).
- Debt and financial strain (Weich & Lewis, 1998).
- Physical and mental well-being, assessed by the 12-item Short Form (SF-12; Jenkinson *et al*. 1997; Gandek *et al*. 1998) and a question on the presence of long-standing illness, disability or infirmity.
- Alcohol misuse, assessed by the Alcohol Use Disorders Identification Test (AUDIT; Barbor *et al*. 1989; Rubio Valladolid *et al*. 1998; Pérula-de Torres *et al*. 2005).
- A lifetime screen for depression based on the first two questions of the CIDI (Arroll *et al*. 2003).
- Lifetime use of recreational drugs (WHO, 1997).
- Brief questions on the quality of sexual and emotional relationships with a partner, adapted from a standardized questionnaire (Reynolds *et al*. 1988).
- Anxiety symptoms using the anxiety section of the Primary Care Evaluation of Mental Disorders (PRIME-MD; Baca *et al*. 1999; Spitzer *et al*. 1999).
- Childhood experiences of physical, emotional or sexual abuse (Fink *et al*. 1995).
- Nature and strength of spiritual beliefs (King *et al*. 1995).
- Presence of serious physical, psychological or substance misuse problems, or any serious disability, in persons who were close friends or relations of participants; and difficulty getting on with people and maintaining close relationships, assessed using questions from a social functioning scale (Tyrer, 1990).
- Family psychiatric history in first-degree family members, and suicide in first-degree relatives (Qureshi *et al*. 2005).
- The living environment, including satisfaction with neighbourhood and perception of safety inside and outside the home using questions from the Health Surveys for England (Sproston & Primatesta, 2003).
- Recent life-threatening events, using a brief validated checklist (Brugha *et al*. 1985).
- Experiences of discrimination on the grounds of sex, age, ethnicity, appearance, disability or sexual orientation using questions from a European study (Janssen *et al*. 2003).
- Adequacy, availability and sources of social support from family and friends (Blaxter, 1990).

Statistical analyses

To develop the predictD-Spain risk algorithm (PSRA), we included only patients without major depression at baseline. Participants with missing depression diagnoses at both follow-up points (at 6 and 12 months) were excluded. We also excluded those with missing CIDI data at one follow-up who were not depressed at the other, as we could not conclude whether or not major depression had occurred over the follow-up period. However, we could include patients who were depressed at one follow-up point and missing at the other (at 6 or 12 months), as they met the outcome criterion of depression at some point over the 12 months. We conducted all analyses using Stata, release 10 (StataCorp, 2007).

Data imputation

Missing data in candidate risk factors were imputed using the method of chained equations, implemented in the Stata ICE program (Royston, 2005). We imputed 10 datasets (Schafer, 1999) and obtained combined estimates (Little & Rubin, 2002).

Model building

We performed multi-level logistic regressions to test the hierarchical data structure with the cumulative incidence of depression at 12 months as the dependent variable. The likelihood-ratio test of the null model with health centre as a random factor *versus* usual logistic regression was significant ($\chi^2 = 15.20$, $p < 0.001$). Nevertheless, the likelihood-ratio test of the null model with health centre and doctor as a random factor *versus* the null model with only health centre was not significant ($\chi^2 = 1.48$, $p = 0.11$). The intraclass correlation coefficients for incidence of depression at 12 months were 0.07 and 0.03 for health centre and

doctor respectively. Hence, we used multi-level logistic regression with two levels, patients and health centre. We built the risk model at 12 months using all the risk factors described earlier and the province of each participant. We developed these models in the imputed data using a threshold for inclusion of $p \leqslant 0.20$ to ensure that information lost as a result of exclusion of a variable from the equation was minimal (Greenland, 1989). We retained age and sex in all regression models because of their well-known associations with the onset of depression (Piccinelli & Wilkinson, 2000). We also retained province because of an *a priori* assumption of clustering within province, although it had few categories ($n = 7$) that could be considered as random factors (Snijders & Bosker, 1999). The usefulness of including first-degree interactions was considered. We considered especially the age × sex interaction because it has been found previously (Bebbington *et al*. 2003). Multi-variable fractional polynomial analysis was used to assess possible nonlinear effects of continuous predictors. From the model thus obtained, those variables with $p \geqslant 0.05$ were extracted step by step to obtain a more parsimonious model. The variables that modified coefficients by more than 10%, irrespective of the p value, remained in the model. For each patient the probability of remaining in the follow-up at 12 months was obtained (Bellón *et al*. 2010) and then inverse probability weighting was applied to the final model to adjust for a possible selection bias due to participants lost during the follow-up (Hernán *et al*. 2004), implemented through the Stata GLLAMM program (Rabe-Hesketh & Skrondal, 2008). We repeated the analyses in participants with complete data as a sensitivity analysis.

Internal validation

The ability to distinguish those who would develop major depression from those who would not was assessed using the C-index (Harrell, 2001). We used a calculation proposed by Copas (1983) to adjust for overfitting of our prediction models. To deal with the overfitting that arises through variable selection, we computed the shrinkage factor based on the initial model including all variables. We calculated effect sizes using Hedges' g (Cooper & Hedges, 1994) for the difference in log odds of predicted probability between patients who were later found to be depressed and those who were not. To obtain more information on the level of overoptimism of the Spanish C-index and Hedges' g, we recalculated them deriving the PSRA from a random sample of 75% of the Spanish data and testing it on the remaining 25%. 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 risk (calibration plots).

External validation

We used the C-index, Hedges' g and calibration plots to evaluate the performance of the PSRA (without province) in data from Chile and the other European countries. When the PSRA was tested in data from all of the European countries, we excluded Spanish patients from the European sample. We estimated the same parameters applying the predictD-Europe risk algorithm (PERA; King *et al*. 2008*b*) in Spanish data to compare both models, the European and the Spanish. In this comparison the sample from Malaga was excluded because it was used to develop the PERA. A test for the difference between two correlated C-index (PSRA and PERA), estimating standard error (S.E.) by bootstrap, was used (Pepe, 2003). Furthermore, we calculated the Integrated Discrimination Improvement (IDI) and the asymptotic test for the null hypothesis of IDI = 0 (Pencina *et al*. 2008). The IDI can be viewed as a difference between improvement in average sensitivity and any potential increase in average '1 – specificity'. Because the calculation of the IDI can be affected by the different incidence rates of depression in Europe and Spain, for each individual we multiplied the predicted probabilities by a 'calibration factor', defined as the ratio of the observed depression rate to the mean predicted probability (Pencina *et al*. 2008).

Results

A total of 6526 people in the seven Spanish provinces were asked to take part in the study. The response to recruitment was 83.4%; 5442 were interviewed and 1084 refused to participate at baseline (Fig. 1). Of those who refused to participate, 780 gave their consent for their age and sex data to be used in our analysis. A higher proportion of the 780 were male [360 (46.1%) *versus* 1756 of the 5442 (32.3%) patients who provided baseline information, $\chi^2 = 18.06$ and $p < 0.001$] and those who refused had a lower mean age, 46.9 [95% confidence interval (CI) 45.7–48.0] *versus* 48.5 years (95% CI 48.1–48.9), $p = 0.018$.

At recruitment, 5360 participants had full CIDI data to allow a depression diagnosis; of these, 4574 were not depressed. The response to follow-up was 70% at 6 months and 66% at 12 months. The analysis of variables associated with non-response has been described elsewhere (Bellón *et al*. 2010). In brief, province and sociodemographic factors were strong predictors of loss to follow-up: those who did not respond were younger, had lower levels of education and income,

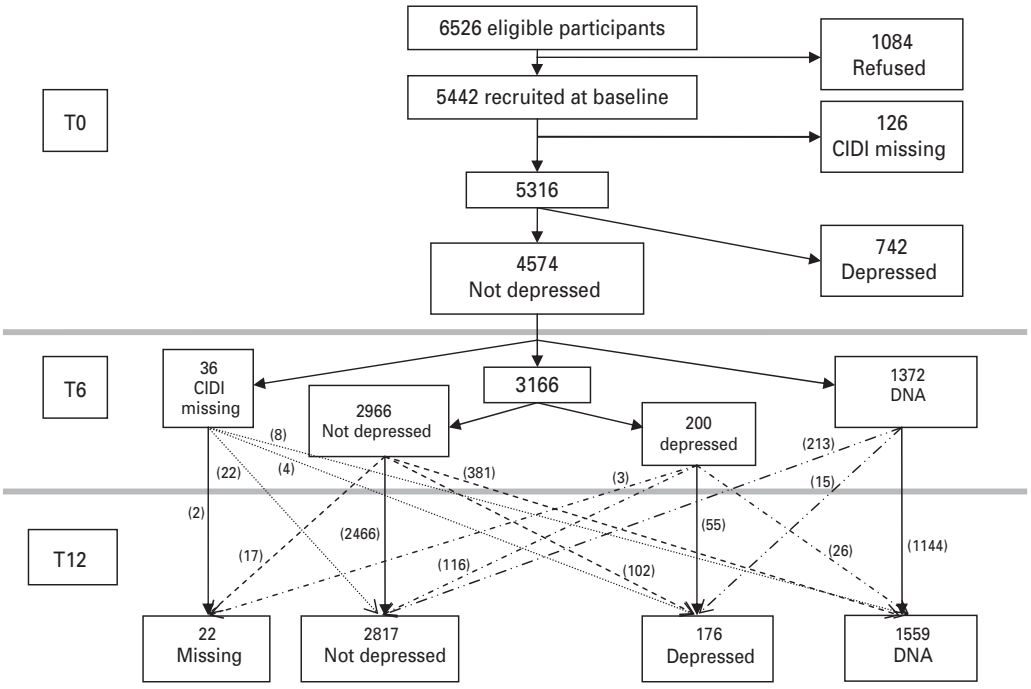

Fig 1. Flow of patients through the predictD-Spain study and numbers becoming depressed. CIDI, Composite International Diagnostic Interview; DNA, did not attend; T0, T6 and T12, baseline, 6 and 12 months interview.

and were more often male, single, born outside Spain, and less often students than those who responded. Major depression and anxiety had no effect but other psychosocial factors predicted attrition (Bellón *et al*. 2010).

In the six countries (apart from Spain) that had participated in predictD, 8567 people took part and their responses to recruitment were, in decreasing order: Chile (97%), Estonia (80%), Slovenia (80%), Portugal (76%), The Netherlands (45%) and the UK (44%). In Chile, 2133 participants who were not depressed started the follow-up and 5184 in the five other European countries; the respective responses to follow-up at 6 and 12 months were 89% and 82%, and 91% and 88% (King *et al*. 2006). The sociodemographic differences between Spanish provinces are shown in Table 1; differences between countries have been described elsewhere (King *et al*. 2008*b*).

The cumulative 12-months incidence of DSM-IV major depression was 11.5% in Spain, varying between provinces: Las Palmas 17.5%, Malaga 15.2%, Granada 14.9%, Majorca 14%, Saragossa 7.9%, Madrid 7.2% and La Rioja 5.6%. The incidence in the other countries was: the UK 8.8%, Slovenia 4.2%, Portugal 8.5%, The Netherlands 5.4%, Estonia 5.9% and Chile 11.6% (King *et al*. 2008*b*).

Missing information in Spanish data was less than 1% for most risk factors; exceptions were ethnicity (2.9%), suicide in brothers or sisters (3.30%), and sexual or emotional relationship with a spouse or partner (18.3%).

The results of reliability analyses have been reported elsewhere (Bellón *et al*. 2008). They were good or excellent for almost all the questionnaires and items. However, in Spain the questions on the use of any recreational drugs over the previous 6 months were removed due to poor reliability (Bellón *et al*. 2008).

Development of the Spanish model

Eight variables were retained at $p<0.05$ (Tables 2 and 3), and one more (physical child abuse) was included at $p=0.085$ because the coefficients changed by more than 10% when this was removed. The age × sex interaction was also included in the final equation because the likelihood-ratio test was significant ($\chi^2=4.03$, 1 df, $p=0.0447$). Interaction between sex and each of the remaining risk factors in the model was not significant at $p\leqslant 0.10$. Nor were interactions between age and the other variables in the model significant. Nonlinear transformation of continuous variables did not significantly improve the model fit. Six variables were patient characteristics or past events (sex, age, sex × age interaction, education, physical child abuse and lifetime depression) and six were current status (SF-12 physical health subscale score, SF-12 mental health subscale score, dissatisfaction with unpaid

Table 1. *Demographic characteristics and response to follow-up of not depressed Spanish participants at baseline*

Characteristic	Spain	Malaga	Granada	Saragossa	Madrid	La Rioja	Balearic Isles	Las Palmas
Not depressed participants, *n* (%)	4574 (100)	1030 (22.5)	609 (13.3)	660 (14.4)	647 (14.1)	680 (14.9)	633 (13.8)	315 (6.9)
Age (years), mean (S.D.)	49.2 (15.8)	49.9 (15.5)	50.3 (16.7)	47.3 (15.4)	50.6 (15.8)	49.5 (15.8)	49.9 (15.7)	43.5 (14.5)
Female, *n* (%)	3020 (66)	704 (68.3)	424 (69.6)	399 (69.6)	432 (66.8)	431 (63.4)	413 (65.2)	217 (68.9)
Marital status, *n* (%)								
Married or living together	3030 (66.2)	725 (70.4)	408 (67)	443 (67.1)	440 (68)	447 (65.7)	397 (62.7)	170 (54.0)
Separated or divorced	279 (6.1)	51 (4.9)	29 (4.8)	27 (4.1)	42 (6.5)	36 (5.3)	54 (8.5)	40 (12.7)
Single	955 (20.9)	186 (18.1)	113 (18.6)	160 (24.2)	126 (19.5)	45 (6.6)	129 (20.4)	89 (28.3)
Widowed	308 (6.7)	68 (6.6)	57 (9.4)	30 (4.5)	39 (6)	152 (22.4)	53 (8.4)	16 (5.1)
Missing	2 (0.1)	0	2 (0.3)	0	0	0	0	0
Household status, *n* (%)								
Not living alone	4208 (92.0)	971 (94.3)	555 (91.1)	615 (93.2)	594 (91.8)	616 (90.6)	560 (88.5)	297 (94.3)
Living alone	366 (8.0)	59 (5.7)	54 (8.9)	45 (6.8)	53 (8.2)	64 (9.4)	73 (11.5)	18 (5.7)
Missing	0	0	0	0	0	0	0	0
Education, *n* (%)								
Higher education	552 (12.1)	141 (13.7)	75 (12.3)	93 (14.1)	59 (9.1)	112 (16.5)	33 (5.2)	39 (12.4)
Secondary	964 (21.1)	220 (21.4)	103 (16.9)	173 (26.1)	149 (23.0)	140 (20.6)	103 (16.3)	76 (24.1)
Primary/no education	2145 (46.9)	424 (41.2)	227 (37.3)	319 (48.3)	265 (41.0)	382 (56.2)	385 (60.8)	143 (45.4)
Trade/other	910 (19.9)	245 (23.8)	202 (33.2)	75 (11.4)	174 (26.9)	45 (6.6)	112 (17.7)	57 (18.1)
Missing	3 (0.1)	0	2 (0.3)	0	0	1 (0.1)	0	0
Employment, *n* (%)								
Employed/full-time student	2045 (44.8)	359 (34.9)	230 (37.8)	363 (55)	308 (47.6)	353 (51.9)	246 (38.9)	186 (59)
Unemployed	291 (6.4)	64 (6.2)	47 (7.7)	36 (5.5)	26 (4)	43 (6.3)	41 (6.5)	34 (10.8)
Unable to work	315 (6.9)	102 (9.9)	42 (6.9)	15 (2.3)	29 (4.5)	4 (0.6)	110 (17.4)	13 (4.1)
Retired/looking after family	1906 (41.7)	500 (48.5)	288 (47.3)	244 (37)	281 (43.4)	279 (41.1)	235 (37.1)	79 (25.1)
Missing	17 (0.2)	5 (0.5)	2 (0.3)	2 (0.2)	3 (0.5)	1 (0.1)	1 (0.2)	3 (1)
Born in country of residence, *n* (%)								
Yes	4327 (94.6)	977 (94.9)	594 (97.5)	632 (95.8)	607 (93.8)	648 (95.3)	586 (92.6)	283 (89.8)
Missing	20 (0.4)	0	2 (0.3)	1 (0.2)	5 (0.8)	0	7 (1.1)	5 (1.6)
Ethnicity, *n* (%)								
White European	4371 (95.6)	1017 (98.7)	595 (97.7)	551 (83.5)	631 (97.5)	647 (95.1)	617 (97.5)	313 (99.4)
Missing	131 (2.9)	1 (0.1)	6 (1.0)	104 (15.8)	1 (0.2)	14 (2.1)	5 (0.8)	0
6 months response, *n* (%)	3202 (70.0)	810 (78.6)	465 (76.4)	520 (78.8)	403 (62.3)	468 (68.8)	334 (52.8)	202 (64.1)
12 months response, *n* (%)	3014 (66.0)	743 (72.1)	434 (71.3)	448 (67.9)	403 (62.3)	507 (74.6)	295 (46.6)	184 (58.4)

S.D., Standard deviation.

Table 2. *Spanish predictD model with province[a] for predicting the onset of major depression at 12 months*

		Model with province			
Prognostic factors	Levels in factor	Coefficient	Coefficient[b]	S.E.	*p*
Constant		1.948	1.421	0.764	0.011
Province	Malaga				
	Granada	0.276	0.241	0.299	0.356
	Saragossa	0.166	0.145	0.321	0.606
	Madrid	−0.377	−0.330	0.362	0.297
	La Rioja	−0.329	−0.287	0.398	0.409
	Balearic Isles	0.337	0.294	0.263	0.200
	Las Palmas	0.308	0.269	0.191	0.106
Age	Each year	−0.032	−0.027	0.008	<0.001
Sex	Female				
	Male	−1.128	−0.985	0.752	0.134
Sex × age interaction	Each year	0.022	0.019	0.014	0.118
Education	Beyond secondary education				
	Secondary education	0.580	0.507	0.304	0.057
	Primary education	0.839	0.733	0.285	0.003
	Incomplete primary education or illiterate	1.490	1.302	0.369	<0.001
Physical childhood abuse	Never				
	Seldom	0.470	0.411	0.465	0.312
	Sometimes	0.320	0.279	0.300	0.286
	Often	0.818	0.715	0.474	0.085
	Frequently	0.106	0.093	0.459	0.816
Lifetime depression	No				0.001
	Yes	0.682	0.596	0.199	
Taking medication for anxiety, depression or stress	No				0.002
	Yes	0.480	0.419	0.151	
Dissatisfaction with unpaid work scale (possible range 3–22)	Satisfied (range 3–7)				
	Dissatisfied (range 8–12)	−0.029	−0.025	0.237	0.902
	Very dissatisfied (range 13–22)	0.472	0.412	0.235	0.045
Number of serious problems in very close persons (alcohol–drugs, psychological, physical, or disability)	None				
	One	0.049	0.042	0.179	0.786
	Two	0.451	0.394	0.166	0.007
	Three	0.824	0.719	0.288	0.004
	Four	0.766	0.669	0.639	0.231
Satisfied with living together at home	Very satisfied				
	Fairly satisfied	0.091	0.080	0.213	0.668
	Neither satisfied nor dissatisfied	−0.256	−0.224	0.262	0.328
	Fairly dissatisfied	0.892	0.780	0.333	0.008
	Very dissatisfied	0.479	0.418	0.545	0.380
Physical health (SF-12), possible range 0–100	Each point on SF-12 subscale score	−0.034	−0.030	0.009	<0.001
Mental health (SF-12), possible range 0–100	Each point on SF-12 subscale score	−0.055	−0.048	0.005	<0.001

S.E., Standard error; SF-12, 12-item Short Form Health Survey.

[a] Model derived in the 10 imputed datasets weighting for the inverse probability of remaining in the follow-up to 12 months.

[b] Coefficient after Copas shrinkage.

work, number of serious problems in very close persons, dissatisfaction with living together at home, and taking medication for stress, anxiety or depression); and one concerned Spanish province. The random component (health centre) was also significant even after including all variables of the fixed component in the regression models; these coefficients were 0.390 (S.E. = 0.085, $p < 0.0001$) and 0.469 (S.E. = 0.101,

Table 3. *Spanish predictD model without province[a] for predicting the onset of major depression at 12 months*

Prognostic factors	Levels in factor	Model without province			
		Coefficient	Coefficient[b]	S.E.	*p*
Constant		2.027	1.482	0.731	0.006
Age	Each year	−0.032	−0.028	0.008	<0.001
Sex	Female				
	Male	−1.111	0.968	0.751	0.139
Sex × age interaction	Each year	0.022	0.019	0.014	0.120
Education	Beyond secondary education				
	Secondary education	0.578	0.504	0.305	0.058
	Primary education	0.840	0.732	0.288	0.004
	Incomplete primary education or illiterate	1.500	1.307	0.370	<0.001
Physical childhood abuse	Never				
	Seldom	0.453	0.395	0.476	0.341
	Sometimes	0.345	0.300	0.295	0.243
	Often	0.820	0.715	0.478	0.087
	Frequently	0.119	0.104	0.465	0.798
Lifetime depression	No				
	Yes	0.689	0.601	0.198	0.001
Taking medication for anxiety, depression or stress	No				
	Yes	0.497	0.433	0.154	0.001
Dissatisfaction with unpaid work scale (possible range 3–22)	Satisfied (range 3–7)				
	Dissatisfied (range 8–12)	−0.038	−0.033	0.240	0.873
	Very dissatisfied (range 13–22)	0.480	0.419	0.234	0.040
Number of serious problems in very close persons (alcohol–drugs, psychological, physical, or disability)	None				
	One	0.041	0.036	0.179	0.820
	Two	0.435	0.380	0.161	0.007
	Three	0.824	0.718	0.293	0.005
	Four	0.769	0.670	0.644	0.233
Satisfied with living together at home	Very satisfied				
	Fairly satisfied	0.077	0.067	0.213	0.716
	Neither satisfied nor dissatisfied	−0.270	−0.235	0.267	0.312
	Fairly dissatisfied	0.926	0.807	0.332	0.005
	Very dissatisfied	0.473	0.412	0.540	0.381
Physical health (SF-12), possible range 0–100	Each point on SF-12 subscale score	−0.035	−0.030	0.009	<0.001
Mental health (SF-12), possible range 0–100	Each point on SF-12 subscale score	−0.055	−0.048	0.005	<0.001

S.E., Standard error; SF-12, 12-item Short Form Health Survey.

[a] Model derived in the 10 imputed datasets weighting for the inverse probability of remaining in the follow-up to 12 months.

[b] Coefficient after Copas shrinkage.

$p<0.0001$) for the models with and without province respectively.

The model derived in participants with complete data ($n=2544$) and the model derived in the 10 imputed datasets ($n=2787$) were very similar, except for the variables set 'age, sex, and age × sex interaction', which was more significant in the model with complete data (see Appendix 1, available online); nevertheless, there were more differences between the model derived in the 10 imputed datasets and the same weighted for the inverse probability of remaining in the follow-up to 12 months (Appendix 1).

Internal validation

The average C-index and the effect size (Hedges' *g*) in data sets were 0.817 (95% CI 0.790–0.843) and 1.35 (95% CI 1.21–1.48) respectively; and 0.816 (95% CI 0.755–0.878) and 1.34 (95% CI 1.16–1.53) when deriving the PSRA from a random sample of 75% of the

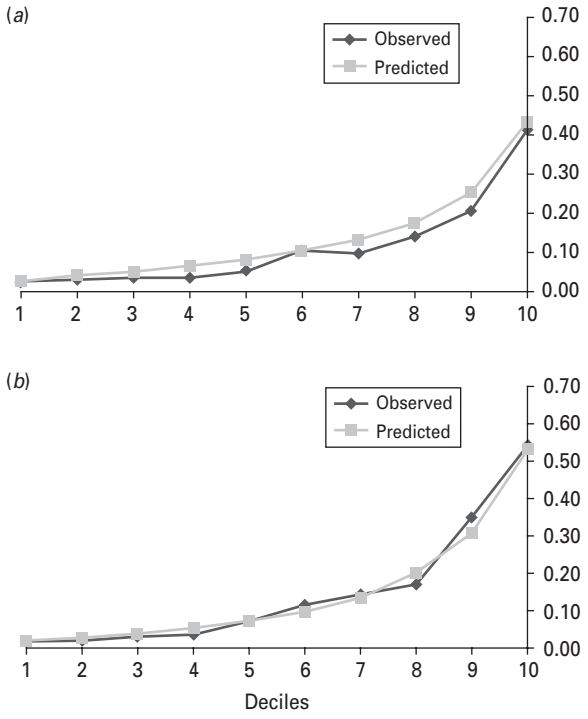

Fig. 2. Calibration plots (mean predicted probability against observed probability of depression within deciles of predicted risk) of the (*a*) predictD-Europe risk algorithm (PERA) and (*b*) the predictD-Spain risk algorithm (PSRA) in Spain.

Spanish data and testing it on the remaining 25%. The calibration plot of the PSRA in Spain is shown in Fig. 2. The predicted probability of depression at 0.113 was associated with estimates of sensitivity, specificity and likelihood ratio (+) of 72.8%, 72.6% and 2.67 respectively. Examples of the kinds of participants scoring at increasing levels of predicted probability of depression are shown in Table 4. The predicted probability of major depression over 12 months can be calculated through the PSRA at www.rediapp.org/predict.php.

Table 4. *Examples of a range of predicted probabilities of depression at baseline*

Case 1: Risk score 8.5% (3.1%)
A man of 55 years living in Saragossa
Secondary education
No personal history of depression
Never suffered physical childhood abuse
Does not take medication for anxiety, depression or stress
SF-12 mental score 30
SF-12 physical scale score 56
One serious problem in very close persons
Satisfied with unpaid work
Fairly satisfied with living together at home
Case 2: Risk score 39.3% (11.4%)[a]
A woman of 62 years living in Balearic Isles
Primary education
Sometimes suffered physical childhood abuse
No personal history of depression
Taking medication for anxiety, depression or stress
SF-12 mental score 45.6
SF-12 physical scale score 52.8
Two serious problems in very close persons
Very dissatisfied with unpaid work
Fairly dissatisfied with living together at home
Case 3: Risk score 93.5% (18.4%)[a,b]
A woman of 45 years living in Malaga
Incomplete primary education
Often suffered physical childhood abuse
Personal history of depression
Taking medication for anxiety, depression or stress
SF-12 mental score 29.2
SF-12 physical scale score 25.7
Three serious problems in very close persons
Very dissatisfied with unpaid work
Very dissatisfied with living together at home

Mean (standard deviation) Short Form 12 (SF-12) mental and physical subscale scores for Spain were 47.1 (12.4) and 43.8 (11.4) respectively. High scores indicate good health/well-being. Scores in parentheses correspond to eliminating dissatisfaction with unpaid work and living together at home, perception of serious problems in close persons and correcting SF-12 physical and mental health scores to the Spanish mean.

[a] Perception of serious problems in close persons did not change.

[b] Very dissatisfied with living together at home changed to neither satisfied nor dissatisfied and stopped taking medication for anxiety, depression or stress.

External validation

The Copas shrinkage factor for the Spanish model was 0.873 including the province and 0.872 without the province. The shrunk regression coefficients are shown in Tables 2 and 3. The C-index ranged from 0.70 in Chile to 0.83 in The Netherlands and Hedges' g from 0.77 in Chile to 1.50 in The Netherlands (Table 5). Calibration plots of the PSRA in Chile and the other European countries are shown in Appendix 2 (available online).

When we applied the PERA to the Spanish data (excluding the sample recruited from Malaga): the C-index was 0.78 (95% CI 0.73–0.83) and Hedges' g was 1.14 (95% CI 0.98–1.31) (Table 5); the test for the C-index difference between the PSRA and PERA was significant (difference = 0.0316, 95% CI 0.0121–0.0530, $Z_{exp} = 3.10$, $p < 0.0022$). The IDI was 0.0558 (S.E. = 0.0071, $Z_{exp} = 7.88$, $p < 0.0001$) because of the increase in average sensitivity (0.2744 *v.* 0.2256, $Z_{exp} = 7.00$,

Table 5. *C-Index statistic and effect sizes computed using Hedges'* g

	PSRA[a]		PERA[b]	
Country (*n*)	C-Index (95% CI)[c]	Hedges' *g* (95% CI)[c]	C-Index (95% CI)[c]	Hedges' *g* (95% CI)[c]
Spain (2787)	0.82 (0.79–0.84)	1.35 (1.21–1.48)	0.78 (0.75–0.81)	1.15 (1.02–1.29)
Spain without Malaga (2045)	0.82 (0.77–0.86)	1.36 (1.19–1.54)	0.78 (0.73–0.83)	1.14 (0.98–1.31)
Chile (1844)	0.70 (0.66–0.74)	0.77 (0.61–0.93)	0.71 (0.67–0.74)	0.85 (0.68–1.02)
UK (811)	0.76 (0.71–0.82)	1.02 (0.76–1.28)	0.76 (0.70–0.81)	1.02 (0.78–1.27)
Slovenia (866)	0.82 (0.75–0.88)	1.38 (0.98–1.78)	0.83 (0.77–0.89)	1.40 (1.06–1.75)
Portugal (844)	0.71 (0.65–0.77)	0.78 (0.54–1.03)	0.75 (0.69–0.80)	0.99 (0.73–1.25)
The Netherlands (731)	0.83 (0.77–0.90)	1.50 (1.09–1.92)	0.85 (0.80–0.90)	1.55 (1.25–1.85)
Estonia (823)	0.73 (0.66–0.80)	0.91 (0.58–1.23)	0.76 (0.69–0.83)	1.09 (0.76–1.42)
Europe without Spain[d] (4075)	0.76 (0.74–0.79)	1.04 (0.90–1.17)	0.79 (0.77–0.82)	1.19 (1.05–1.32)

PSRA, PredictD-Spain risk algorithm; PERA, predictD-Europe risk algorithm; CI, confidence interval.

[a] The risk score was computed using unshrunk estimates in Spain and shrunk estimates in Chile and other European countries.

[b] The risk score was computed using shrunk estimates in Spain, Chile and other European countries.

[c] Average C-Index and Hedges' *g* over 10 imputed data sets.

[d] The UK + Slovenia + Portugal + The Netherlands + Estonia.

$p < 0.0001$) and a small decrease in 'average 1 – specificity' (0.0942 *v*. 0.1012, $Z_{exp} = 5.55$, $p < 0.0001$). Calibration plots showed that the PSRA functioned better in Spain than the PERA (Fig. 2).

Discussion

We have developed and validated a risk score for the development of major depression over 12 months in 2787 general practice attendees in Spain. The PSRA included new variables and afforded an improved performance over the PERA for predicting the onset of major depression in Spain. To our knowledge, Spain is the first country to have developed its own risk score for predicting new episodes of major depression in primary care. However, the PERA is still the best option for predicting the onset of major depression in other European countries.

The PSRA worked better than the PERA in Spain; however, this conclusion cannot be generalized to other countries because it has only been studied in Spain. In general, it is expected that countries with similar incidence rates of depression and a similar distribution of their risk factors can probably share the same risk algorithm.

Studies are needed to provide data about whether the improvement in 5.6 IDI points translates into improvements for the health of patients (depressions avoided) and/or decreased costs, but so far no study has been published about the primary prevention of depression using the PSRA. Meanwhile, we can again use the analogy with cardiovascular disease, where an increase of 1 IDI point or more has been suggested to represent a meaningful improvement (Pencina *et al*. 2008). From this viewpoint, the improvement in 5.6 IDI points could lead to substantial clinical differences and important public health implications.

The C-indexes were very similar (differing in one thousandth) when we derived and applied the PSRA on the whole sample or when we derived it from 75% and applied it to the remaining 25%. This supports the hypothesis that the differences between the PSRA and PERA cannot be explained by overoptimism.

We recruited a systematic random sample of primary care attendees and we used a criterion of stratification to include urban and rural health centres in each province and included provinces from different geographical areas in both mainland Spain (north, central and south) and the Spanish islands. Although we did not select health centres randomly and our sample could under-represent patients who attend very infrequently (Lee *et al*. 2002), the study population is likely to be fairly representative of primary care attendees in Spain. Further studies to develop risk algorithms in other countries will have to consider the external validity of the sample chosen, especially if there are data to suggest that within the same country there are different incidence rates of depression and different risk factors. In the case of Spain, we found that unadjusted incidence rates of depression were very different between provinces; for example, 17.5% in Las Palmas and 5.6% in La Rioja, with an ascending gradient from the north to the south. However, after adjusting for risk factors, these differences were

largely dissipated (Table 2), though not so with the PERA, where differences between countries remained (King *et al*. 2008*b*). If the PSRA is applied in a different country, or a province other than one of the seven participating provinces, we recommend using the shrunk coefficients of the model without province (Table 3).

We used multi-level regression because of the hierarchical structure of the data. In these cases, this approach improves the accuracy of estimates of coefficients and standard errors (Snijders & Bosker, 1999). Our large sample size and the number of events (major depression) per variable included in the model (>29) contributed to reducing the risk of selecting unimportant variables and failing to include important ones (Altman & Royston, 2000). The multiple imputation strategy allowed us to gain statistical power and to avoid potentially biased estimates obtained from a reduced complete-case dataset (Little & Rubin, 2002). We have lack of certainty about the reasons for missing data, but we do know that, at baseline, the outcome variable (depression) was not associated with loss during the follow-up (Bellón *et al*. 2010) and no major discrepancy was found between imputed data and complete-case analyses (Appendix 1). From this point of view, we would be more inclined to think they are 'at random'. There were important differences between the Spanish models with and without inverse probability weighting, indicating that loss to follow-up might lead to selection bias and suggesting that this strategy could provide unbiased estimates of coefficients, even in the presence of selection bias (Hernán *et al*. 2004). We consider that follow-up at 12 months is appropriate for the prediction of the onset of depression in primary care because this is sufficient time to develop major depression (11.5% of incident cases). Furthermore, doctors and patients may be more motivated to undertake interventions and behavioural changes when depression is likely to happen sooner rather than later.

The PERA and PSRA share many risk factors; however, the new risk factors included in the Spanish equation improved its results for prediction in Spain. 'Dissatisfaction with living together at home' and 'number of serious problems in very close persons' are risk factors consistent with the geography of family systems; the central and northern parts of Europe, together with North American society, have been characterized by relatively weak family links, whereas the Mediterranean region has strong family ties (Reher, 1998). Spain belongs to the regions where the family group has traditionally had priority over the individual. Moreover, the association between marital discord, family dysfunction and depression is well known (Whisman & Uebelacker, 2009). The inclusion of 'dissatisfaction with unpaid work' instead of 'difficulties in paid and unpaid work' may be due to different ways of measuring these variables. We used two scales in Spain, one for unpaid work and another for paid work, with seven items each that were valid and reliable (Bellón *et al*. 2010), whereas in the predictD-Europe study two items were used to summarize both dimensions together. When we included these two items in the predictD-Spain model instead of work scales, it was not significant. However, we cannot rule out the influence of other factors, such as a higher participation of Spanish women in domestic work as compared with other European regions (Drew *et al*. 1998). Although relationships between 'physical childhood abuse' and depression are well documented (Arnow, 2004), they are complex, vary between countries, and have cross-cultural differences (Sebre *et al*. 2004). Finally, the variable 'taking medication for anxiety, depression or stress' might be associated with patients who have suffered previous depressive episodes and were still taking antidepressants. However, the question is phrased in such a way that it might also include those taking anxiolytics, often in an inadequate way, for anxiety, comorbidity or even just taking medicines (vitamins, placebos, etc.) for other minor emotional problems. A hypothesis might be that these patients share coping styles, such as 'external health locus', with a tendency to ask their doctors for more psychotropic drugs for emotional problems encountered in everyday life (Demyttenaere *et al*. 2008). Spain is also among those European countries that have a higher use of psychotropic drugs (Alonso *et al*. 2004). We might expect the PSRA to work better than the PERA in Chile or Portugal because the Spanish equation includes specific risk factors that may be shared with Mediterranean or southern Europe countries and Latin America, that is those related to family ('dissatisfaction with living at home', 'serious problems in families and close persons'). However, this was not the case.

Although the PSRA adds two items on top of those already included in the PERA, both algorithms need a computer for risk calculation. Nowadays, it would be easy to incorporate our algorithm into a computerized medical records system. As the questionnaire can be checked online (www.rediapp.org/predict.php), completing it just takes about three minutes. One of the uses of our PSRA could be to select relevant patients for studies of the primary prevention of depression, although the main use of any risk score is to help physicians with complex decisions (Moons *et al*. 2009). Our PSRA could help physicians with decisions by providing more objective estimates of the likelihood of risk of major depression, as a supplement to

other relevant clinical information; perhaps in a similar manner to the way cardiovascular risk scores are used to determine the indication for lowering cholesterol. However, trials are needed using our PSRA to test different strategies of primary prevention of major depression. Impact studies are also needed to quantify the effect of using the PSRA on physicians' behaviour, patient outcome or cost-effectiveness of care. When such evidence is available, the PSRA could also be used by any patient for self-assessment using the web-based calculator. Meanwhile, we have taken the first step; we have an accurate, valid and reliable tool that provides an objective and individualized measure of the likelihood of risk of the onset of major depression in primary care.

Note

Supplementary material accompanies this paper on the Journal's website (http://journals.cambridge.org/psm).

Acknowledgements

We thank the Primary Care District of Malaga and, in particular, Dr J.M. Morales and Dr M. Vilaseca for their support. We also thank P. Royston and C. Bottomley (both UK) for their contribution to the analytical strategy in predictD and predictD-Spain, and P. Zuithoff (The Netherlands) for his statistical comments on the article. Finally, we thank all the interviewers, physicians and patients for their participation in the predictD study.

This work was supported in Spain by grants from the Spanish Ministry of Health (grant FIS references: PI041980, PI041771, PI042450 and PI06/1442); the Andalusian Council of Health (grant references: 05/403 and 06/278) and the Spanish Ministry of Education and Science (grant reference SAF 2006/07192); the Spanish Network of Primary Care Research 'redIAPP' (RD06/0018), the 'Aragón group' (RD06/0018/0020), the 'Baleares group' (RD07/0018/0033), and the 'SAMSERAP group' (RD06/0018/0039). The Malaga sample, as part of the predictD-International study, was also co-funded by a grant from the European Commission (reference QL4-CT2002-00683). The research in Europe was funded by a grant from the European Commission (reference PREDICT-QL4-CT2002-00683). Funding in Chile was provided by project FONDEF DO2I-1140. Partial support in Europe was from the Estonian Scientific Foundation (grant 5696) and the Slovenian Ministry for Research (grant 4369-1027). The UK National Health Service Research and Development Office provided 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.

Declaration of Interest

None.

References

Alonso J, Angermeyer MC, Bernert S, Bruffaerts R, Brugha TS, Bryson H, de Girolamo G, Graaf R, Demyttenaere K, Gasquet I, Haro JM, Katz SJ, Kessler RC, Kovess V, Lépine JP, Ormel J, Polidori G, Russo LJ, Vilagut G, Almansa J, Arbabzadeh-Bouchez S, Autonell J, Bernal M, Buist-Bouwman MA, Codony M, Domingo-Salvany A, Ferrer M, Joo SS, Martínez-Alonso M, Matschinger H, Mazzi F, Morgan Z, Morosini P, Palacín C, Romera B, Taub N, Vollebergh WA; ESEMeD/MHEDEA 2000 Investigators (2004). Psychotropic drug utilization in Europe: results from the European Study of the Epidemiology of Mental Disorders (ESEMeD) project. *Acta Psychiatrica Scandinavica. Supplementum* **420**, 55–64.

Altman DG, Royston P (2000). What do we mean by validating a prognostic model? *Statistics in Medicine* **19**, 453–473.

Arnow BA (2004). Relationships between childhood maltreatment, adult health and psychiatric outcomes, and medical utilization. *Journal of Clinical Psychiatry* **65** (Suppl. 12), 10–15.

Arroll B, Khin N, Kerse N (2003). Screening for depression in primary care with two verbally asked questions: cross sectional study. *British Medical Journal* **327**, 1144–1146.

Baca E, Saiz J, Agüera L, Caballero L, Fernández-Liria A, Ramos J, Gil A, Madrigal M, Porras A (1999). Validation of the Spanish version of PRIME-MD: a procedure for diagnosing mental disorders in primary care [in Spanish]. *Actas Españolas de Psiquiatría* **27**, 375–383.

Barbor TF, de la Fuente JR, Saunders J, Grant M (1989). *The Alcohol Use Disorders Identification Test: Guidelines for Use in Primary Health Care*. World Health Organization: Geneva.

Bebbington P, Dunn G, Jenkins R, Lewis G, Brugha T, Farrell M, Meltzer H (2003). The influence of age and sex on the prevalence of depressive conditions: report from the National Survey of Psychiatric Morbidity. *International Review of Psychiatry* **15**, 74–83.

Bellón JA, Luna JD, Moreno B, Montón-Franco C, Gildegómez-Barragán MJ, Sánchez-Celaya M, Díaz-Barreiros MA, Vicens C, Motrico E, Martínez-Cañavate MT, Olivan-Blázquez B, Vázquez-Medrano A, Sánchez-Artiaga MS, March S, Muñoz-García MD, Moreno-Peral P, Nazareth I, King M, Torres-González F (2010). Psychosocial and sociodemographic predictors of attrition in a longitudinal study of major depression in primary care: the predictD-Spain study. *Journal of Epidemiology and Community Health* **64**, 874–884.

Bellón JA, Moreno B, Torres-González F, Montón-Franco C, GildeGómez-Barragán MJ, Sánchez-Celaya M, Díaz-Barreiros MA, Vicens C, de Dios Luna J, Cervilla JA, Gutierrez B, Martínez-Cañavate MT, Oliván-Blázquez B, Vázquez-Medrano A, Sánchez-Artiaga MS, March S, Motrico E, Ruiz-García VM, Brangier-Wainberg PR, Del Mar Muñoz-García M, Nazareth I, King M; predictD group (2008). Predicting the onset and persistence of episodes of depression in primary health care. The predictD-Spain study: methodology. *BMC Public Health* **8**, 256.

Blaxter M (1990). *Health and Lifestyles*. Routledge: London.

Brugha T, Bebbington P, Tennant C, Hurry J (1985). The List of Threatening Experiences: a subset of 12 life event categories with considerable long-term contextual threat. *Psychological Medicine* **15**, 189–194.

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

Cooper H, Hedges LV (1994). *The Handbook of Research Synthesis*. Russell Sage Foundation: New York.

Copas JB (1983). Regression, prediction and shrinkage. *Journal of the Royal Statistical Society* (*Series B*) **45**, 311–354.

De Girolamo G, Alonso J, Vilagut G (2006). The ESEMeD-WMH project: strengthening epidemiological research in Europe through the study of variation in prevalence estimates. *Epidemiologia e Psichiatria Sociale* **15**, 167–173.

Demyttenaere K, Bonnewyn A, Bruffaerts R, De Girolamo G, Gasquet I, Kovess V, Haro JM, Alonso J (2008). Clinical factors influencing the prescription of antidepressants and benzodiazepines: results from the European Study of the Epidemiology of Mental Disorders (ESEMeD) project. *Journal of Affective Disorders* **110**, 84–93.

Drew E, Emerek R, Mahon E (1998). *Women, Work, and the Family in Europe*. Routledge: London.

European Commission. Directorate General for Health and Consumer Protection (2004). *The State of Mental Health in the European Union* (http://ec.europa.eu/health/archive/ph_projects/2001/monitoring/fp_monitoring_2001_frep_06_en.pdf). Accessed 1 July 2010.

Fink LA, Bernstein D, Handelsman L, Foote J, Lovejoy M (1995). Initial reliability and validity of the childhood trauma interview: a new multidimensional measure of childhood interpersonal trauma. *American Journal of Psychiatry* **152**, 1329–1335.

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. International Quality of Life Assessment. *Journal of Clinical Epidemiology* **51**, 1171–1178.

Greenland S (1989). Modeling variables selection in epidemiologic analysis. *American Journal of Public Health* **79**, 340–349.

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

Hernán MA, Hernández-Díaz S, Robins JM (2004). A structural approach to selection bias. *Epidemiology* **15**, 615–625.

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 Psychiatry* **182**, 71–76.

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, Nazareth I, Levy G, Walker C, Morris R, Weich S, Bellón-Saameño JA, Moreno B, Svab I, Rotar D, Rifel J, Maaroos HI, Aluoja A, Kalda R, Neeleman J, Geerlings MI, Xavier M, de Almeida MC, Correa B, Torres-Gonzalez F (2008*a*). Prevalence of common mental disorders in general practice attendees across Europe. *British Journal of Psichiatry* **192**, 362–367.

King M, Speck P, Thomas A (1995). The Royal Free interview for religious and spiritual beliefs: development and standardization. *Psychological Medicine* **25**, 1125–1134.

King M, Walker C, Levy G, Bottomley C, Royston P, Weich S, Bellon-Saameno JA, Moreno B, Svab I, Rotar D, Rifel J, Maaroos HI, Aluoja A, Kalda R, Neeleman J, Geerlings MI, Xavier M, Carraca I, Goncalves-Pereira M, Vicente B, Saldivia S, Melipillan R, Torres-Gonzalez F, Nazareth I (2008*b*). 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.

König HH, Bernert S, Angermeyer MC, Matschinger H, Martinez M, Vilagut G, Haro JM, de Girolamo G, de Graaf R, Kovess V, Alonso J; ESEMeD/MHEDEA 2000 Investigators (2009). Comparison of population health status in six European countries: results of a representative survey using the EQ-5D questionnaire. *Medical Care* **47**, 255–261.

Lee ML, Yano EM, Wang M, Simon BF, Rubenstein LV (2002). What patient population does visit-based sampling in primary care setting represent? *Medical Care* **40**, 761–770.

Little RJA, Rubin DB (2002). *Statistical Analysis with Missing Data*, 2nd edn. Wiley: New York.

Marrugat J, Subirana I, Comin E, Cabezas C, Vila J, Elosua R, Nam BH, Ramos R, Sala J, Solanas P, Cordón F,

Gené-Badia J, D'Agostino RB; VERIFICA Investigators (2007). Validity of an adaptation of the Framingham cardiovascular risk function: the VERIFICA study. *Journal of Epidemiology and Community Health* **61**, 40–47.

Moons KGM, Royston P, Vergouwe Y, Grobbee DE, Altman DG (2009). Prognosis and prognostic research: what, why, and how? *British Medical Journal* 338, b375.

Pencina MJ, D'Agostino RB, D'Agostino RB, Vasan RS (2008). Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Statistics in Medicine* **27**, 157–172.

Pepe MS (2003). *The Statistical Evaluation of Medical Tests for Classification and Prediction*. Oxford University Press: New York.

Pérula-de Torres LA, Fernández-García JA, Arias-Vega R, Muriel-Palomino M, Marquez-Rebollo E, Ruiz-Moral R (2005). Validity of AUDIT test for detection of disorders related with alcohol consumption in women [in Spanish]. *Medicina Clinica* **125**, 727–730.

Piccinelli M, Wilkinson G (2000). Gender differences in depression. *Critical review. British Journal of Psychiatry* **177**, 486–492.

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.

Rabe-Hesketh S, Skrondal A (2008). *Multilevel and Longitudinal Modelling Using STATA*, 2nd edn. STATA Press: College Station, TX.

Reher DS (1998). Family ties in Western Europe: persistent contrasts. *Population and Development Review* **24**, 203–234.

Reynolds CF, Frank E, Thase ME, Houck PR, Jennings JR, Howell JR, Lilienfeld SO, Kupfer DJ (1998). Assessment of sexual function in depressed, impotent, and healthy men: factor analysis of a Brief Sexual Function Questionnaire for men. *Psychiatry Research* **24**, 231–250.

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.

Royston P (2005). Multiple imputation of missing values: update of ice. *Stata Journal* **5**, 527–536.

Rubio Valladolid G, Bermejo Vicedo J, Caballero Sánchez-Serrano MC, Santo-Domingo Carrasco J (1998). Validation of the alcohol use disorders identification test (AUDIT) in primary care [in Spanish]. *Revista Clinica Española* **198**, 11–14.

Rubio-Stipec M, Bravo M, Canino G (1991). The Composite International Diagnostic Interview (CIDI): an epidemiologic instrument suitable for using in conjunction with different diagnostic systems in different cultures [in Spanish]. *Acta psiquiátrica y psicológica de América latina* **37**, 191–204.

Schafer JL (1999). Multiple imputation: a primer. *Statistical Methods in Medical Research* **8**, 3–15.

Sebre S, Sprugevica I, Novotni A, Bonevski D, Pakalniskiene V, Popescu D, Turchina T, Friedrich W, Lewis O (2004). Cross-cultural comparisons of child-reported emotional and physical abuse: rates, risk factors and psychosocial symptoms. *Child Abuse and Neglect* **28**, 113–127.

Snijders TAB, Bosker RJ (1999). *Multilevel Analysis. An Introduction to Basic and Advanced Multilevel Modelling*. Sage Publications: London.

Spitzer RL, Kroenke K, Williams JB (1999). Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. *Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. Journal of the American Medical Association* **282**, 1737–1744.

Sproston K, Primatesta P (2003). *Health Survey for England 2002: A Survey carried out on behalf of the Department of Health. Volume 1: The health of children and young people*. The Stationery Office: London.

StataCorp (2007). *Stata Statistical Software: Release 10*. Stata Corporation: College Station, TX.

Tyrer P (1990). Personality disorder and social functioning. In *Measuring Human Problems: A Practical Guide* (ed. D. F. Peck and C. M. Shapiro), pp. 119–142. Wiley & Sons: Chichester.

Weich S, Lewis G (1998). Poverty, unemployment, and common mental disorders: population based cohort study. *British Medical Journal* **317**, 115–119.

Whisman MA, Uebelacker LA (2009). Prospective associations between marital discord and depressive symptoms in middle-aged and older adults. *Psychology and Aging* **24**, 184–189.

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