Title

Development of a prognostic model for predicting depression severity in adult primary patients with depressive symptoms using the *diamond* longitudinal study

Authors

Patty Chondros, PhD[1]

Sandra Davidson, PhD[1]

Rory Wolfe, PhD[2]

Gail Gilchrist, PhD[3]

Christopher Dowrick, MD[1,4]

Frances Griffiths, PhD[1,5]

Kelsey Hegarty, PhD[1]

Helen Herrman, MD[6]

Jane Gunn, PhD[1]

Affiliation

1 Department of General Practice, The University of Melbourne

2. Department of Epidemiology and Preventive Medicine, Monash University

3. National Addiction Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College, London

4. Institute of Psychology, Health and Society, University of Liverpool

5. WMS - Social Science and Systems in Health, University of Warwick

6. Centre for Youth Mental Health, The University of Melbourne

Corresponding Author

Dr Patty Chondros

200 Berkeley Street, CARLTON, 3053, VICTORIA, AUSTRALIA

Email: p.chondros@unimelb.edu.au

Phone: +61 3 8344 4766

Highlights

- Model developed to predict future depression severity in primary care patients
- Prognostic model is brief and easily administered in a busy primary care setting
- Model using psychosocial items is embedded in a clinical prediction tool (CPT)
- CPT tailors type and intensity of treatment to predicted depression severity
- Is a systematic approach designed to support clinician treatment decision making

Abstract [Word count: 249]

Background

Depression trajectories among primary care patients are highly variable, making it difficult to identify patients that require intensive treatments or those that are likely to spontaneously remit. Currently, there are no easily implementable tools clinicians can use to stratify patients with depressive symptoms into different treatments according to their likely depression trajectory. We aimed to develop a prognostic tool to predict future depression severity among primary care patients with current depressive symptoms at three months.

Methods

Patient-reported data from the *diamond* study, a prospective cohort of 593 primary care patients with depressive symptoms attending 30 Australian general practices. Participants responded affirmatively to at least one of the first two PHQ-9 items. Twenty predictors were pre-selected by expert consensus based on reliability, ease of administration, likely patient acceptability, and international applicability. Multivariable mixed-effects linear regression was used to build the model.

Results

The prognostic model included eight baseline predictors: depressive symptoms, anxiety, history of depression, self-rated health, chronic physical illness, living alone, and perceived ability to manage on available income. Discrimination (*c*-statistic =0.74; 95% CI: 0.70-0.78) and calibration (agreement between predicted and observed symptom scores) were acceptable and comparable to other prognostic models in primary care.

Limitations

More complex model was not feasible because of modest sample size. Validation studies needed to confirm model performance in new primary care attendees.

Conclusion

A brief, easily administered algorithm predicting the severity of depressive symptoms has potential to assist clinicians to tailor treatment for adult primary care patients with current depressive symptoms.

Key words: prediction; prognostic; depression; depressive symptom severity; primary health care; mental health

List of Abbreviations

CES-D	Center of Epidemiological Studies Depression
CPT	Clinical prediction tool
GP	General Practitioner
MAR	Missing at random
MBS	Medicare Benefits Schedule
PBS	Pharmaceutical Benefits Scheme
PHQ-9	Patient Health Questionnaire (9 item version)
PHQ-2	Patient Health Questionnaire (2 item version)

Word count (exclusive of abstract, required statements, references and tables): 3130

Introduction

Mental health disorders account for 7.4% of the total disease burden with depression the main contributor.(Whiteford et al., 2013) Most people seeking help for depressive symptoms are treated in primary care,(Australian Bureau of Statistics, 2011; Australian Institute of Health and Welfare, 2015) and around one quarter of primary care attendees report current depressive symptoms.(Gunn et al., 2008; Herrman et al., 2002) Ten percent of attendees with subthreshold symptoms and no history of depression develop major depression over six months(Davidson et al., 2015) and 21% over two years.(Karsten et al., 2011) Nearly 60% of those with current major depression meet criteria for major depression at least once over the next three years.(Stegenga et al., 2012)

In a busy primary care practice, it can be difficult for clinicians to identify which patients with current depressive symptoms are likely to recover and which are likely to worsen, and to provide treatment appropriate for each trajectory. Primary care clinicians are often criticised for either over-treating patients with subthreshold depression(Davidson et al., 2015) or for not providing minimally adequate treatment for patients with major depression.(Wang et al., 2007) One systematic approach to informing clinician's treatment decisions is to use a clinical prediction tool.

A clinical prediction tool is built around a prognostic model that uses clinical and psychosocial information to predict future depression severity. The clinical prediction tool uses the information provided by the prognostic model to stratify patients into different depression severity groups. Type and intensity of treatment is tailored to each group to optimise clinical outcomes with the least intensive treatment.(Rubenstein et al., 2007) To

date, no such clinical prediction tool exists that can be used to stratify primary care patients with depressive symptoms into different treatment options based on their predicted depressive symptoms.

We also conducted a literature search to identify existing prognostic models that would be suitable for inclusion in a clinical prediction tool that predicts future depressive symptoms in primary care patients with depressive symptoms, ranging from sub-threshold to severe. The literature search identified nine different prognostic models for depression developed using data from five unique primary care studies. Only two of the models focussed on predicting future depression within samples experiencing current depressive symptoms.(Dowrick et al., 2011; Rubenstein et al., 2007) Of the remaining studies, three developed or validated prognostic models to predict the onset of depression (primary prevention)(Bellon et al., 2011; King et al., 2013; King et al., 2008), two studies developed a prediction rule to screen for the presence of current mood disorders(Vohringer et al., 2013; Zuithoff et al., 2009) and two studies developed algorithms to predict treatment response to antidepressants(Chekroud et al.; Perlis, 2013).

Of the two studies that developed prognostic models to predict future depression among people with current depressive symptoms, neither was suitable for inclusion in a clinical prediction tool.(Dowrick et al., 2011; Rubenstein et al., 2007) In the first study, the prognostic model developed using trial data from 220 participants in the THREAD study was insufficiently robust to use in the clinical prediction tool because it had low prognostic accuracy.(Dowrick et al., 2011) Furthermore, the development sample only included participants with mild to moderate depression, thus could not be generalised to new primary

care patients who present with severe depression. The second study described the development of the Diagnostic Prognostic Index, which was derived using data from 1471 primary care attendees with current major depression participating in one of four randomised trials.(Rubenstein et al., 2007) The Diagnostic Prognostic Index was also unsuitable because the development sample excluded patients with subthreshold depression. Given that in primary care subthreshold depression makes up the largest group of patients presenting with depressive symptoms, the prognostic model would not be generalisable to this population. Additionally, the Diagnostic Prognostic Index, consisting of over 60 items, would be too lengthy to administer in a primary care waiting room or during a consultation, limiting its usability and usefulness in routine clinical practice.(Toll et al., 2008)

This study aimed to develop a prognostic model for future depression severity among adult primary care attendees with current depressive symptoms, ranging from sub-threshold to severe depression. To increase the utility and uptake of the clinical prediction tool we aimed to develop a model with relatively few items that were easy to collect in routine practice.(Toll et al., 2008)

Methods

Source of data

We developed a prognostic model using data from the *diamond* (Diagnosis, Management and Outcomes of Depression) cohort study. *Diamond* is a 10-year prospective study of adult primary care patients with depressive symptoms.(Gunn et al., 2008)

Cohort participants were recruited from 30 general practitioners (GPs) working at 30 different urban, regional and rural practices in Victoria, Australia between January and December, 2005. Details of recruitment are published elsewhere.(Gunn et al., 2008) Briefly, 17,780 randomly selected patients of study GPs (approximately 600 patients per GP) were posted a survey containing the Center of Epidemiological Studies Depression (CES-D) scale.(Radloff, 1977) Patients were eligible if they were: aged between 18-75 years; able to read English; not terminally ill; and did not live in residential care. Forty-two percent (7509/17780) returned the survey with a completed CES-D. Twenty-four percent (1793/7509) scored ≥16 on the CES-D scale, of which 1007 were interested in hearing more about the study and provided contact details. Seventy-eight percent (789/1007) of eligible patients consented and formed the *diamond* cohort. Participants in the *diamond* cohort completed self-report surveys at baseline, at three monthly intervals for the first year, and annually thereafter until 2016. Computer assisted telephone interviews were conducted annually.

Depression severity was measured using the Patient Health Questionnaire (PHQ-9).(Kroenke et al., 2003; Spitzer et al., 1999) The PHQ-9 is a reliable and valid measure to assess and monitor severity of depressive symptoms over time in primary care.(Kroenke et al., 2001; Spitzer et al., 1999) To reflect usual practice, where only individuals likely to have current depressive symptoms would be administered the clinical prediction tool, the model was derived on 593 (75.2% of 789) cohort participants who scored ≥2 on the first two items on the Patient Health Questionnaire (PHQ-2).(Kroenke et al., 2003; Spitzer et al., 1999) That is, they reported that in the last two weeks they were bothered by ‘little interest or pleasure in doing things’ and ‘feeling down, depressed, or hopeless’ on ‘several days’ or reported having one or both problems on ‘more than half the days’ or ‘nearly every day’. This threshold

ensured that most cases of major depression were not missed (sensitivity=92.7%), but less likely to exclude individuals that did not satisfy formal criteria for major depression (specificity=73.7%).(Kroenke et al., 2003)

Ethics approval and consent to participate

The *diamond* study was approved by the Human Research Ethics Committee at the University of Melbourne (ID: 030613X). The Australian Government Department of Human Services Information Services Branch has approved the collection of Medicare Benefits Schedule (MBS) and Pharmaceutical Benefits Scheme (PBS) data (ID: MI3794). All participants provided informed consent to participate in the study. Separate informed consent was obtained to collect data on participant's health services use provided under the MBS and medicines prescribed under the PBS. The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

Outcome

The outcome was depressive symptoms at three months, calculated by summing the nine items of the PHQ-9,(Spitzer et al., 1999) with scores ranging from 0 to 27.(Kroenke et al., 2001) If two or fewer items on the PHQ-9 were missing the missing values were substituted with the mean response of the completed items, otherwise they were coded as missing.(Kroenke et al., 2010) Depressive symptom severity was stratified according the cut-points nominated by Kroenke *et al*;(Kroenke et al., 2001) specifically, minimal (0-4), mild (5-9), moderate (10-14), moderately severe (15-19), and severe (20-27).

Predictor selection

Over forty candidate predictors were considered for the model development. We identified potential predictors from the literature and through consensus opinion by a multi-disciplinary expert group. Variables included were patient demographics, history of depression or anxiety, health service use, antidepressant use, physical health, social support, social functioning and life events.(Gunn et al., 2008) We also considered potentially sensitive questions that, although they are associated with depression, may be distressing for some respondents (e.g. childhood abuse, intimate partner violence, obsessive, or compulsive behaviours).

Twenty predictors were selected for model development (See Supplementary Table). Five were potentially sensitive questions. Inclusion criteria for predictors for the model building were: easily administered (e.g. not time-consuming), reliable, internationally applicable, publicly available and measured in the *diamond* study. We also imposed criteria for data quality: less than 10% missing values; distribution not severely skewed and range of values not too narrow.

Sample size

There are no generally accepted methods for sample size calculations for building prognostic models.(Moons et al., 2009b; Steyerberg, 2009) Data from 593 participants were available for the development sample. Restricting the candidate predictors to 20 for the model development ensured there were approximately 30 observations per predictor which is sufficient to build a reliable model.(Steyerberg, 2009)

Statistical analysis and missing data assumption

Analyses were conducted using Stata version 13.1. Mixed effects linear regression, treating general practice as a random effect and predictors as fixed effects, was used to build the model. Under this model, data were assumed missing at random (MAR) conditional on the variables included in the model.(White et al., 2012)

Model development

Initially, all candidate predictors were included in a full model. Variables were dropped if the regression coefficient was close to zero (< 0.5 mean change per unit change on the variable) and the p-value was greater than 0.5. Adjacent categories for ordinal variables "managing on available income" and "self-rated health" were collapsed because the numbers of individuals in some of the categories were small.(Steyerberg, 2009) Adjusted R^2 measure was used to assess model fit with a penalty for model complexity. Model adequacy and robustness to outliers and influential values were examined with overall goodness of fit statistics and regression diagnostics, including residuals and identifying influential values.

Two models were derived using this model building process. The first full model included fifteen predictors that were considered easy to collect and likely to be acceptable to respondents (minimising missing data). The predictors included: current depression severity, current anxiety, past depression and anxiety, chronic illness, self-rated health, antidepressant use, ability to manage on available income, social support, negative life events, living alone and exercise. The items "Have you ever been bothered by feeling down, depressed, or hopeless for longer than two weeks" and "Have you ever been bothered by little interest or pleasure in doing things for longer than two weeks" were combined into a single variable "Ever had depression/little interest" if both items were scored 'Yes'. The item "Depression in the past 12 months" met criteria for inclusion in Model 1, but was excluded because it was highly correlated with "Ever had depression/little interest" and baseline "PHQ-9 depression symptom severity" and contributed minimal independent information. The two depression items retained reflected both current (past two weeks) and lifetime depression. The second model included five additional potentially sensitive predictors. We repeated the model building process using all 20 predictors to determine whether including potentially sensitive predictors further improved the model performance.

To adjust for over-optimism, a uniform shrinkage factor can be multiplied with the model coefficients. A shrinkage factors was calculated using a heuristic formula that accounted for the number of predictors considered in the model building process.(Steyerberg, 2009)

Model performance

Model discrimination was assessed with the concordance (*c*) statistic, where the 95% confidence interval was calculated using bootstrap resampling for clusters and individuals.

Calibration was assessed with the calibration plot where the observed depressive symptoms (*y*-axis) were plotted against the predicted scores (*x*-axis). Perfect predictions would lie on the line of identity.

Stratification by predicted depressive symptoms

Participants were stratified into three groups based on their predicted depressive symptoms: minimal/mild (<=10), moderate (>10 and <13) and severe (>=13). The cut-points were based on the 50th and 75th percentile values of the predicted values. These percentiles reflect the percentage of participants stratified into three severity groups: minimal/mild (0 to 9), moderate (10 to14) and moderately severe/severe (15 to 27), based on the observed depressive symptoms at three months.(Kroenke et al., 2001)

Results

Participants

Distribution of participant characteristics in the development sample are shown in Table 1. The fraction of missing responses for each predictor variable for the development sample was small, ranging from zero for gender to 2.2 percent (13/593) for ever being afraid of a partner. Most participants (91%, 538/593) had complete data for the 20 candidate predictor variables, including the sensitive questions. Fourteen percent (82/593) had missing values for the outcome.

Model development

Table 2 shows the estimated coefficients of the prognostic models developed without (Model 1) and with (Model 2) the potentially sensitive questions. For Model 1, eight of 15 predictor variables were retained. When all 20 predictors were included in the model development, an additional four potentially sensitive questions were retained (Model 2). The model coefficients were not adjusted for over-fitting because the heuristic shrinkage factor was close to 1 for both models (0.96 for model 1 and 0.94 for model 2).

No outliers or influential values were identified (results not shown). In explanatory analyses, inclusion of splines to accommodate an unspecified non-linear relationship between depression severity at baseline and outcome did not improve the fit of the models (results not shown). There was no evidence for an interaction between sex and baseline depressive symptoms (results not shown).

Model 1 explained 39.2% of the variation in three-month depression severity (Table 2). Depression symptom severity scores were predicted using the coefficients for Model 1 and Model 2 (Table 2) respectively and the predictor values at baseline. Predicted depression severity for the 593 participants using Model 1 was similar to the mean observed depression severity at three months (n=511, mean=10.7, SD=6.2, range 0 to 27). Figure 1 shows that agreement between the observed and predicted values for the depressive symptoms was acceptable. Model performance did not improve with the inclusion of additional sensitive items (Model 2).

Stratification of depressive symptoms

Table 3 shows the distribution of the observed depression severity groups at three months across the three predicted severity groups. When predicted and observed depression severity scores were stratified into the three groups the *c*-statistic was 0.74 (95% CI: 0.70 to 0.78).

Discussion

We developed a brief, easily administered prognostic model to predict depression severity at three months in adult primary care patients with current depressive symptoms. The eight predictors were depressive symptoms, current anxiety, history of depression, self-rated health, chronic physical illness, living alone, and perceived ability to manage available on income. The final model consists of 17 questions, nine of which are from the PHQ-9. Including potentially sensitive or distressing questions did not improve the model performance, probably due to correlations with other indicators in the model. The simpler, user-friendly model could be administered in the waiting room or during a consultation and has the potential to be incorporated into routine clinical practice.(Toll et al., 2008)

The overall performance of the model measured by the R^2 was 39.2% in the development sample, well above 20% that is commonly found in prognostic research.(Steyerberg, 2009) Compared to the Diagnostic Prognostic Index, where the R^2 was 33% in the development sample,(Rubenstein et al., 2007) our model performed slightly better with substantially fewer items. The R^2 indicates the predictability of the outcome, and models that explain more the 20% of the variability have the potential to be clinically useful and warrant further evaluation and development.(Rubenstein et al., 2007; Steyerberg, 2009)

The prognostic model we developed has acceptable discrimination and calibration. A *c*-statistic of 0.74 was comfortably within the typical range of 0.60 and 0.85 for prognostic models predicting depression onset,(Bellon et al., 2011; King et al., 2008) current major depression,(Vohringer et al., 2013; Zuithoff et al., 2009) or, treatment depression outcome(Chekroud et al., 2016; Perlis, 2013) in primary care and in other health settings.(Royston et al., 2009) (Toll et al., 2008)

Strengths

Using the *diamond* cohort provided a strong study design to develop the prognostic model.(Moons et al., 2009a) We were able to prospectively map the natural course of depressive symptoms over time,(Gunn et al., 2013) and use predictors that were well defined and reproducible which increases the generalisability of the model.(Moons et al., 2009a) Unlike many prognostic models which are developed in secondary care and then applied to primary care,(Moons et al., 2009a) our prognostic model was specifically developed for use in general practice using data collected from general practice attendees. The model development using linear regression to predict the depression symptom scores maximised the available information, increased the statistical efficiency(Steyerberg, 2009) and enabled us to consider categorisation of the predictions as a final step rather than an initial one. Our approach to selecting the candidate predictors using information from several sources and limiting their number to 20 ensured that over-fitting was not a threat to the internal validity and generalisability of the model.(Altman et al., 2009; Steyerberg, 2009) The development sample included patients with depressive symptoms ranging from sub-threshold to severe that reflects the wider patient base seen in primary care.(Herrman et al., 2002)

Limitations

A prognostic model of excessive complexity was not feasible with the modest development sample size.(Steyerberg, 2009) If important predictors of depression course were not captured in the model building process, the model may not perform as well in new data.(Altman et al., 2009) Validation studies with samples drawn from primary care settings within Australia and other countries are required to assess the performance of the model in new patients.

Further research

We have incorporated the prognostic model into a web-based clinical prediction tool that stratifies primary care attendees into three groups (i.e., mild, moderate, or severe) based on their predicted depression severity at three months and then provides a matched treatment recommendation. Pilot work showed the clinical prediction tool was acceptable and feasible to use with primary care clinicians and patients. A randomised controlled trial is currently testing the clinical and cost effectiveness of the tool to reduce depressive symptoms, a key step in evaluating the impact of the clinical prediction tool in routine primary care.(Moons et al., 2009a) The trial data also provide the opportunity to externally validate and if required update the prognostic model.(Moons et al., 2009a)

Conclusion

We developed a brief, easily administered prognostic model for use in primary care across the depressive symptom range to predict depression severity at three months. A clinical prediction tool utilising this model has the potential to assist clinicians manage the large burden of mental health symptoms presenting to primary care. Widespread implementation of

tools like this offers the best chance of ensuring that limited resources are allocated based on need.

References

Altman, D.G., Vergouwe, Y., Royston, P., Moons, K.G.M., 2009. Prognosis and prognostic research: validating a prognostic model. BMJ 3381, 1432-1435.

Australian Bureau of Statistics, 2011. 4329.0 - Characteristics of people using mental health services and prescription medication, in: Table 1 Persons accessing MBS subsidised mental health-related services in 2011.

Australian Institute of Health and Welfare, 2015. Australian Institute of Health and Welfare Mental health services—in brief, Canberra.

Bellon, J.A., de Dios Luna, J., King, M., Moreno-Kustner, B., Nazareth, I., Monton-Franco, C., GildeGomez-Barragan, M.J., Sanchez-Celaya, M., Diaz-Barreiros, M.A., Vicens, C., Cervilla, J.A., Svab, I., Maaroos, H.I., Xavier, M., Geerlings, M.I., Saldivia, S., Gutierrez, B., Motrico, E., Martinez-Canavate, M.T., Olivan-Blazquez, B., Sanchez-Artiaga, M.S., March, S., del Mar Munoz-Garcia, M., Vazquez-Medrano, A., Moreno-Peral, P., Torres-Gonzalez, F., 2011. Predicting the onset of major depression in primary care: international validation of a risk prediction algorithm from Spain. Psychol. Med. 41, 2075-2088.

Chekroud, A.M., Zotti, R.J., Shehzad, Z., Gueorguieva, R., Johnson, M.K., Trivedi, M.H., Cannon, T.D., Krystal, J.H., Corlett, P.R., 2016. Cross-trial prediction of treatment outcome in depression: a machine learning approach. Lancet Psychiatry published online Jan 20. http://dx.doi.org/10.1016/S2215-0366(15)00471-X.

Davidson, S.K., Harris, M.G., Dowrick, C.F., Wachtler, C.A., Pirkis, J., Gunn, J.M., 2015. Mental health interventions and future major depression among primary care patients with subthreshold depression. J. Affect. Disord. 177, 65-73.

Dowrick, C., Flach, C., Leese, M., Chatwin, J., Morriss, R., Peveler, R., Gabbay, M., Byng, R., Moore, M., Tylee, A., Kendrick, T., 2011. Estimating probability of sustained recovery from mild to moderate depression in primary care: evidence from the THREAD study. Psychol. Med. 41, 141-150.

Gunn, J., Elliott, P., Densley, K., Middleton, A., Ambresin, G., Dowrick, C., Herrman, H., Hegarty, K., Gilchrist, G., Griffiths, F., 2013. A trajectory-based approach to understand the factors associated with persistent depressive symptoms in primary care. J. Affect. Disord. 148, 338-346.

Gunn, J.M., Gilchrist, G.P., Chondros, P., Ramp, M., Hegarty, K.L., Blashki, G.A., Pond, D.C., Kyrios, M., Herrman, H.E., 2008. Who is identified when screening for depression is undertaken in general practice? Baseline findings from the Diagnosis, Management and Outcomes of Depression in Primary Care (diamond) longitudinal study. Med. J. Aust. 188, S119-125.

Herrman, H., Patrick, D., Diehr, P., Martin, M., Fleck, M., Simon, G., Buesching, D., 2002. Longitudinal investigation of depression outcomes in primary care in six countries: the LIDO study. Functional status, health service use and treatment of people with depressive symptoms. Psychol. Med. 32, 889-902.

Hodgson, R., Alwyn, T., John, B., Thom, B., Smith, A., 2002. The FAST Alcohol Screening Test. Alcohol Alcohol. 37, 61-66.

Karsten, J., Hartman, C.A., Smit, J.H., Zitman, F.G., Beekman, A.T.F., Cuijpers, P., Van der Does, A.J.W., Ormal, J., Nolen, W.A., Penninx, B.W., 2011. Psychiatric history and subthreshold symptoms as predictors of the occurrence of depressive or anxiety disorder within 2 years. Br. J. Psychiatry 198, 206.

King, M., Bottomley, C., Bellon-Saameno, J., Torres-Gonzalez, F., Svab, I., Rotar, D., Xavier, M., Nazareth, I., 2013. Predicting onset of major depression in general practice attendees in Europe: extending the application of the predictD risk algorithm from 12 to 24 months. Psychol. Med. 43, 1929-1939.

King, M., Walker, C., Levy, G., Bottomley, C., Royston, P., Weich, S., Bellon-Saameno, J.A., Moreno, B., Svab, I., Rotar, D., Rifel, J., Maaroos, H.I., Aluoja, A., Kalda, R., Neeleman, J., Geerlings, M.I., Xavier, M., Carraca, I., Goncalves-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. Arch. Gen. Psychiatry 65, 1368-1376.

Kroenke, K., Spitzer, R.L., Janet, B.W.W., 2003. The Patient Health Questionnaire-2: Validity of a Two-Item Depression Screener. Med. Care 41, 1284-1292.
Kroenke, K., Spitzer, R.L., Williams, J.B.W., 2001. The PHQ-9. J. Gen. Intern. Med. 16, 606-613.
Kroenke, K., Spitzer, R.L., Williams, J.B.W., Löwe, B., 2010. Psychiatry and Primary Care: The Patient Health Questionnaire Somatic, Anxiety, and Depressive Symptom Scales: a systematic review. Gen. Hosp. Psychiatry 32, 345-359.
Moons, K.G.M., Altman, D.G., Vergouwe, Y., Royston, P., 2009a. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. BMJ 338, 1487-1490.
Moons, K.G.M., Royston, P., Vergouwe, Y., Grobbee, D.E., Altman, D.G., 2009b. Prognosis and prognostic research: what, why and how? BMJ 338, 1317-1320.
Perlis, R.H., 2013. A clinical risk stratification tool for predicting treatment resistance in major depressive disorder. Biol. Psychiatry 74, 7-14.
Radloff, L.S., 1977. The CES-D scale: A self-report depression scale for research in the general population. Appl. Psychol. Meas. 1, 385-401.
Royston, P., Moons, K.G.M., Altman, D.G., Vergouwe, Y., 2009. Prognosis and prognostic research: Developing a prognostic model. BMJ 338, 1373-1377.
Rubenstein, L.V., Rayburn, N.R., Keeler, E.B., Ford, D.E., Rost, K.M., Sherbourne, C.D., 2007. Predicting outcomes of primary care patients with major depression: Development of a depression prognosis index. Psychiatr. Serv. 58, 1049-1056.
Spitzer, R.L., Kroenke, K., W. Williams, J.B., 1999. Validation and Utility of a Self-report Version of PRIME-MD: The PHQ Primary Care Study. J. Am. Med. Assoc. 282, 1737-1744.
Stegenga, B., Kamphuis, M., King, M., Nazareth, I., Geerlings, M., 2012. The natural course and outcome of major depressive disorder in primary care: the PREDICT-NL study. Soc. Psychiatry Psychiatr. Epidemiol. 47, 87-95.
Steyerberg, E.W., 2009. Clinical prediction models : a practical approach to development, validation, and updating. Springer, New York.
Toll, D.B., Janssen, K.J.M., Vergouwe, Y., Moons, K.G.M., 2008. Validation, updating and impact of clinical prediction rules: A review. J. Clin. Epidemiol. 61, 1085-1094.
Vohringer, P.A., Jimenez, M.I., Igor, M.A., Fores, G.A., Correa, M.O., Sullivan, M.C., Holtzman, N.S., Whitham, E.A., Barroilhet, S.A., Alvear, K., Logvinenko, T., Kent, D.M., Ghaemi, S.N., 2013. A clinical predictive score for mood disorder risk in low-income primary care settings. J. Affect. Disord. 151, 1125-1131.
Wang, P.S., Aguilar-Gaxiola, S., Alonso, J., Angermeyer, M.C., Borges, G., Bromet, E.J., Bruffaerts, R., de Girolamo, G., de Graaf, R., Gureje, O., Haro, J.M., Karam, E.G., Kessler, R.C., Kovess, V., Lane, M.C., Lee, S., Levinson, D., Ono, Y., Petukhova, M., Posada-Villa, J., Seedat, S., Wells, J.E., 2007. Use of mental health services for anxiety, mood, and substance disorders in 17 countries in the WHO world mental health surveys. Lancet 370, 841-850.
White, I.R., Carpenter, J., Horton, N.J., 2012. Including all individuals is not enough: lessons for intention-to-treat analysis. Clin. Trials J. 9, **396-407**.
Whiteford, H.A., Degenhardt, L., Rehm, J., Baxter, A.J., Ferrari, A.J., Erskine, H.E., Charlson, F.J., Norman, R.E., Flaxman, A.D., Johns, N., Burstein, R., Murray, C.J.L., Vos, T., 2013. Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. Lancet 382, 1575-1586.
Zuithoff, N.P.A., Vergouwe, Y., King, M., Nazareth, I., Hak, E., Moons, K.G.M., Geerlings, M.I., 2009. A clinical prediction rule for detecting major depressive disorder in primary care: the PREDICT-NL study. Fam. Pract. 26, 241-250.

Table 1: Participant characteristics at baseline for the development sample

Participant characteristics	Development sample (N=593)	Missing
Age in years – Mean (standard deviation)	47·7 (12·9)	
Female	422 (71·2)	
Current marital status		5 (0.8)
Never married/single	138 (23·5)	
Widowed/divorced/separated	178 (30·3)	
Married	272 (46·3)	
Live alone*	132 (22·3)	2 (0.3)
Born in Australia	494 (83·6)	2 (0.3)
English first language	567 (95·8)	1 (0.2)
Highest education level reached		1 (0.2)
Less than high school education	237 (40·1)	
Completed high school	93 (15.7)	
Certificate/Diploma	144 (24·3)	
Bachelor Degree or higher	118 (19·9)	
Health Care Card	117 (20·3)	17 (2·9)
Employment status		2 (0·3)
Employed/Student	348 (58·9)	
Not employed/Not in paid employment	145 (24·5)	
Unable to work	98 (16·6)	
Hazardous drinking in past 12 months†	134 (22·8)	4 (0·7)
Current smoker	210 (35·6)	3 (0·5)
Ever depressed and/or ever had little interest in doing things for greater than 2 weeks	464 (78·5)	2 (0·3)
Depression in past 12 months	352 (59·4)	
Anxiety on the past 12 months	279 (47·0)	
PHQ current anxiety		6 (1·0)
Not at all	30 (5·1)	
Several days	311 (53·0)	
More than half these days	246 (41·9)	
Long term illness*	324 (55·5)	9 (1·5)
Self-rated health*		
Excellent	16 (2·7)	
Very Good	101 (17·0)	
Good	207 (34·9)	
Fair	194 (32·7)	
Poor	75 (12·6)	
Depression medication in past 12 months	305 (51·8)	4 (0·7)
Social support/confidant past 4 weeks		3 (0·5)
Not bothered	182 (30·8)	
Bothered a little	220 (37·3)	
Bothered a lot	188 (31·9)	

Participant characteristics		Development sample (N=593)	Missing
Negative life event past 4 weeks			7 (1·2)
	Not bothered	277 (47·3)	
	Bothered a little	155 (26·5)	
	Bothered a lot	154 (26·3)	
Managing on available income*			2 (0·3)
	Easily	61 (10·3)	
	Not too bad	171 (28·9)	
	Difficult some of the time	223 (37·7)	
	Difficult all of the time	113 (19·1)	
	Impossible	23 (3·9)	
Compulsions*†			9 (1·5)
	Not at all	250 (42·8)	
	Some days	237 (40·6)	
	More than half the days	97 (16·6)	
Obsession*†			1 (0·2)
	Not at all	110 (18·6)	
	Some days	312 (52·7)	
	More than half the days	170 (28·7)	
Ever afraid of partner*‡§		211 (36·4)	13 (2·2)
Child sexual abuse*‡		182 (31·2)	9 (1·5)
Child physical abuse*‡		296 (50·6)	8 (1·4)
Depression symptom severity score (PHQ-9)*‡			
	Mild (0-9)	215 (36·3)	
	Moderate/moderately severe (10-14)	190 (32·0)	
	Severe (15-27)	188 (31·7)	

Count (percentage) presented unless otherwise stated
Development sample: 30 GPs, 593 patients, mean cluster size=20, range 6-32
* Candidate predictor variables selected for the model building process
† Hazardous drinking in the past 12 months measured using the "The Fast Alcohol Screening Test"(Hodgson et al., 2002)
‡ Potentially sensitive questions
§ No partner coded as "No"

Table 2: Prognostic models without sensitive questions (Model 1) and with sensitive questions (Model 2)

Predictive factors	Levels of the factor	Coefficient (SE) † Model 1 (N=497)	Coefficient (SE) † Model 2 (N=473)
Sex	Male	0	0
	Female	-0·69 (0·50)	-0·80 (0·52)
Depression symptom severity score (PHQ-9)	Each point on depression symptom severity scale (Range 0-27)	0·50 (0·05)	0·46 (0·05)
PHQ current anxiety	Not at all	0	0
	Several days	0·75 (0·99)	0·81 (1·04)
	More than half these days	1·56 (1·03)	1·66 (1·09)
Ever depressed and/or ever had little interest in doing things for greater than 2 weeks	Yes	1·59 (0·54)	1·72 (0·56)
	No	0	0
Self-rated health	Excellent/very good/good	0	0
	Fair	0·75 (0·52)	0·80 (0·53)
	Poor	2·19 (0·73)	2·58 (0·78)
Long term illness	Yes	1·16 (0·50)	1·19 (0·52)
	No	0	0
Live alone	Yes	0·86 (0·51)	0·89 (0·53)
	No	0	0
Managing on available income	Easily/not to bad/difficult some of the time	0	0
	Difficult all of the time/impossible	1·16 (0·55)	0·42 (0·24)
Compulsion*	Not at all		0
	Some days		0·84 (0·51)
	More than half the days		1·49 (0·69)
Obsession*	Not at all		0
	Some days		-0·14 (0·61)
	More than half the days		-1·18 (0·74)
Ever afraid of partner*	Yes		0·39 (0·48)
	No/No partner		0
Childhood sexual abuse*	Each point increase for total number (Range 0-4)		0·31 (0·18)
Constant		1·05 (1·16)	-0·025 (1·27)
Adjusted R^2		39·2%	39·5%
c-statistic (95% CI)		0·71 (0·68 to 0·74)	0·71 (0·68 to 0·74)
Predicted depressive symptom scores			
N		593	574
Mean (SD)		10·9 (3·8)	11.3 (3·7)
Range		(3 to 22)	(3 to 22)

SE = Standard Error

***** Considered potentially sensitive or distressing questions

† Variance of random effects for general practice was truncated to zero in both models

Table 3: Distribution of the stratified observed depressive symptoms by the stratified predicted values for the development sample

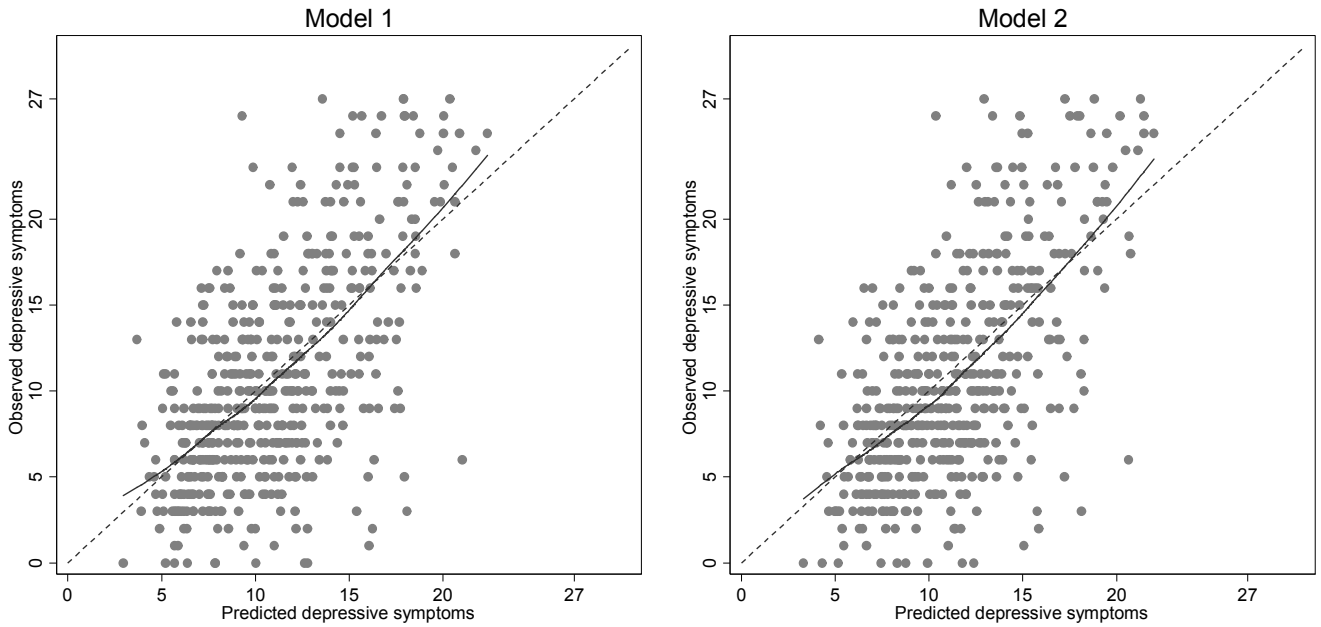
			Stratification for observed depressive symptoms at 3 months*					
Stratification for predicted depressive symptoms at 3 months†			Minimal/Mild (0-9)		Moderate (10-14)		Moderately severe/Severe (15-27)	
Development sample (n=511)								
Minimal/Mild	234	(46)	169	(65)	49	(41)	16	(12)
Moderate	138	(27)	68	(26)	41	(34)	29	(22)
Severe	139	(27)	24	(9)	29	(24)	86	(66)

Count (percentage) presented

* Cut-points used for the observed depressive symptoms were nominated by Kreonke et al, 2001

† Predicted scores estimated using Model 1; The cut-points used to stratify participants as mild, moderate and severe using the predicted depressive symptom scores were: less than 10; 10 to 13; 13 or more

Figure 1: Calibration plot between the observed and predicted values for the depression symptom severity in the development sample (N=511)



Grey dashed line represents the line of identity (45^{o} line), that is perfect model calibration; Grey solid line represents is a smooth line though the scatter plot created using *lowess* smoothing

Conflict of interest

Authors declare no conflict of interest relevant to this study.

Contributors

PC, SD and JG drafted the manuscript. PC conducted the analysis and produced the tables and figures. SD, GG, CD, FG, KH, HH, JG, and PC formed the multi-disciplinary expert group to identify and select candidate predictor variables. RW provided statistical expertise on the development of the prognostic model. All the authors contributed to development and drafting of the manuscript

Acknowledgements

We acknowledge the 30 dedicated general practitioners, their patients, and practice staff for making this research possible. We thank the cohort participants for their ongoing involvement in the study. We also thank the diamond project team and associate investigators involved in the study: A/Prof. Lena Sanci, Ms Maria Potiriadis, Ms Konstancja Densley, Ms Aves Middleton, and the casual research staff.

The authors submit this manuscript on behalf of the diamond study investigators: Prof. Jane M. Gunn, Prof. Helen Herrman, Prof. Mike Kyrios, Prof. Kelsey Hegarty, Prof. Christopher Dowrick, Dr Gail Gilchrist, A/Prof Grant Blashki, Prof. Dimity Pond, Dr Patty Chondros, A/Prof. Renata Kokanovic, and Dr Victoria Palmer.

Funding sources

The *diamond* study was initiated with pilot funding from the Victoria Centre of Excellence in Depression and Related disorders, an initiative between *beyondblue* and the State Government of Victoria, Australia. The main *diamond* cohort was supported with project grant funding from the National Health and Medical Research Council (grant numbers: 299869 (2004), 454463 (2007), 566511 (2009), and 1002908 (2011)). The one year Computer Assisted Telephone Interview was funded by a Stream 3 grant from the Australian Primary Health Care Research Institute (APHCRI). Refinement and testing of the *diamond* Clinical Prediction Tool was supported by National Health and Medical Research Council (grant number 1059863 (2014)). Dr Davidson received funding from the National Health and Medical Research Council Early Career Fellowship scheme and from a National Health and Medical Research Council Centre of Research Excellence Grant during the conduct of the study.

Role of Funding Source

No funding body had a role in the study design, the collection, analysis or interpretation of the data; the preparation, review or approval of the manuscript; or the decision to submit the manuscript for publication. The corresponding author had full access to all the data in the study. All authors had final responsibility for the decision to submit for publication.

Supplementary Table: Candidate predictor variables selected for the model building process (N=593 participants)

#	Variable	Question/item	Coding	Scoring of items	% missing
1	Depression symptom severity score (PHQ-9)*†	Over the last 2 weeks, how often have you been bothered by any of the following problems? 1. Little interest or pleasure in doing things 2. Feeling down, depressed or hopeless 3. Trouble falling or staying asleep, or sleeping too much 4. Feeling tired or having little energy 5. Poor appetite or overeating 6. Feeling bad about yourself, or that you are a failure, or have let yourself or your family down 7. Trouble concentrating on things, such as reading the newspaper or watching television 8. Moving or speaking so slowly that other people could have noticed. Or the opposite-being so fidgety or restless that you have been moving around a lot more than usual 9. Thoughts that you would be better off dead, or of hurting yourself in some way	For each item the responses are: 0 = Not at all 1 = Several days 2 = More than half the days 3 = Nearly every day	Sum of the 9 items. The score range is between 0 and 27. **Dealing with missing responses** If > 2 of the 9 items had missing responses, the total score was coded as missing. If 1 or 2 items had a missing response, the missing responses were substituted with the mean of the responses to the other items.	0
2	Ever depressed and/or ever had little interest in doing things for greater than 2 weeks*	Have you ever been bothered by feeling down, depressed or hopeless for longer than 2 weeks? Have you ever been bothered by little interest or pleasure in doing things for longer than 2 weeks?	0 = No 1 = Yes 0 = No 1 = Yes	Combined responses of the two items to create a new binary variable: 1 if responded yes to both items and 0 (no) otherwise	2 (0.4%)
3		Depression in past 12 months	0 = No 1 = Yes		0
4		Anxiety in the past 12 months	0 = No 1 = Yes		0
5	PHQ current anxiety*	Over the last 4 weeks, how often have you been bothered by feeling nervous, anxious, on edge or worrying a lot about different things?	0 = Not at all 1 = Several days 2 = More than half these days		6 (1.0%)
6	Long term illness*	Do you have any long-term illness, health problem, which limits your	0 = No 1 = Yes		9 (1.5%)

#	Variable	Question/item	Coding	Scoring of items	% missing
		daily activities or the work you can do (including problems that are due to old age)?			
7	Self-rated health*	In general, would you say your health is...	1 = Excellent 2 = Very Good 3 = Good 4 = Fair 5 = Poor	Collapsed the first 3 categories: 1 = Excellent/Very Good/Good 2 = Fair 3 = Poor	0
8	Depression medication in past 12 months	Have you tried depression medication in past 12 months?	0 = No 1 = Yes		4 (<1%)
9	Social support/confidant	In the past 4 weeks, how much have you been bothered by having no one to turn to when you have a problem.	0 = Not bothered 1 = Bothered a little 2 = Bothered a lot		3 (0.5%)
10	Negative life events	In the past 4 weeks, how much have you been bothered by something bad that happened recently.	0 = Not bothered 1 = Bothered a little 2 = Bothered a lot		7 (1.2%)
11	Live alone*	Do you live alone?	0 = No 1 = Yes		2 (0.3%)
12	Managing on available income*	How do you manage on your available income?	1 = Easily 2 = Not to bad 3 = Difficult some of the time 4 = Difficult all of the time 5 = Impossible	Collapsed to a **binary** variable: 0 = Easily/Not to bad/Difficult some of the time 1 = Difficult all of the time or impossible	2 (0.3%)
13	Exercise	In a NORMAL week, how many times do you engage in VIGOROUS exercise lasting for 20 minutes or more? *(exercise which makes you breathe harder or puff and pant, such as netball, squash, jogging, aerobics, vigorous swimming, etc.)* In a NORMAL week, how many times do you engage in LESS VIGOROUS exercise which lasts for 20 minutes or more? *(exercise which does not make you breathe harder or puff and pant, like walking, gardening, swimming and lawn bowls)*	0=Never 1=Once a week 2=Two or three times a week 3=Four, five or six times a week 4=Once every day 5=More than once every day	A "recreational physical activity measure" was derived based on the algorithm as specified in pages 3-7 in the "Data Technical Report of the ALSWH" referenced below‡	0
14	Sex*	Are you male or female?	0 = Male 1 = Female		0
15	Age	Age in years	Range: 18 to 76		0
Considered potentially sensitive or distressing questions					
16	Compulsions	Over the past 4 weeks, how often have you been bothered by repetitive thoughts, ideas, doubts, images or impulses that distress you and that you regard as unwanted and senseless?	0 = Not at all 1 = Some days 2 = More than half the days		9 (1.5%)

#	Variable	Question/item	Coding	Scoring of items	% missing
17	Obsession	Over the past 4 weeks, how often have you been compelled to do or think certain things repeatedly, excessively or according to strict rules in order to prevent something bad from happening or to make sure things are "just right"?	0 = Not at all 1 = Some days 2 = More than half the days		1 (0.2%)
18	Ever afraid of partner	Ever afraid of partner	0 = No 1 = Yes 2 = No intimate relationship	Collapsed "No intimate relationship with "No"	13 (2.2%)
19	Child sexual abuse	When you were growing up, did any adult do any of these things against your will? Exposed themselves to you more than once? Threatened to have sex with you? Touched the sex parts of your body? Tried to have sex with you or sexually attacked you?	0 = No 1 = Yes	Coded as yes if reported that they had experienced at least one sexual abuse item	9 (1.5%)
20	Child physical abuse	When you were growing up, how often did any adult do any of the things to you: Pushed, grabbed or shoved you Threw something at you Hit you with something Kicked, bit or punched you Choked, burned or scalded you Physically attacked you in some other way	0 = Never 2 = Rarely 3 = Sometimes 4 = Often	The responses to the first three abuse items were collapsed to a binary response: 1 = Often/Sometimes 0 = Never/Rarely The responses to the last three abuse items were collapsed to a binary response: 1 = Often/Sometimes/Rarely 0 = Never Collapsed response across the six items: 1 responded Created a new binary variable that combined responses across the six items: Coded as "yes" if responded as "Often/Sometimes" on any of the first three items or coded as "1" for any of the six items they experienced at least one of the sexual abuse items	8 (1.3%)

* Variables retained in the multivariable prognostic model

† Spitzer, R.L., Kroenke, K., W. Williams, J.B., 1999. Validation and Utility of a Self-report Version of PRIME-MD: The PHQ Primary Care Study. J. Am. Med. Assoc. 282, 1737-1744.

‡ Russell A. Data Technical Report of the ALSWH: Derived variables not included in datasets. December 2006; Accessed online January 2011: http://www.alswh.org.au/images/content/pdf/InfoData/data_technical_reports/DataTechRep_VariablesNOTinDatasets_Dec2006.pdf