Journal of Affective Disorders 151 (2013) 39–45

Contents lists available at ScienceDirect

Journal of Affective Disorders

journal homepage: www.elsevier.com/locate/jad

Research report

Development and validation of prediction algorithms for major depressive episode in the general population ☆

Jian Li Wang [a,*], Douglas Manuel [b], Jeanne Williams [c], Norbert Schmitz [d], Heather Gilmour [e], Scott Patten [f], Glenda MacQueen [g], Arden Birney [c]

[a] Departments of Psychiatry and of Community Health Sciences, Faculty of Medicine, University of Calgary, Calgary, AB, Canada T2N 4Z6
[b] Department of Clinical Epidemiology, Faculty of Medicine, University of Ottawa. Ottawa, ON, Canada K1N 6N5
[c] Department of Community Health Sciences, Faculty of Medicine, University of Calgary, University of Calgary, Calgary, AB, Canada T2N 1N4
[d] Department of Psychiatry, Faculty of Medicine, McGill University, Montreal, QC, Canada H3A 0G4
[e] Health Analysis Division, Statistics Canada. Ottawa, ON, Canada K1A 0T6
[f] Departments of Community Health Sciences and of Psychiatry, Faculty of Medicine, University of Calgary, Calgary, AB, Canada T2N 1N4
[g] Department of Psychiatry, Faculty of Medicine, University of Calgary, Calgary, AB, Canada T2N 1N4

ARTICLE INFO

Article history:
Received 22 April 2013
Accepted 17 May 2013
Available online 19 June 2013

Keywords:
Prediction algorithms
Major depression
Sex-specific
Population-based

ABSTRACT

Background: To develop and validate sex specific prediction algorithms for 4-year risk of major depressive episode (MDE) using data from a population-based longitudinal cohort.
Methods: Household residents from 10 provinces were randomly recruited and interviewed by Statistics Canada. 10,601 participants who were aged 18 years and older and who did not meet the criteria for MDE in the 12 months prior to a baseline interview in 2000/01 were included in algorithm development; data from 7902 participants who were aged 18 and older and who were free of MDE in 2004/05 were used for validation. Validation was also conducted in sub-populations that are of practice and policy importance. MDE was assessed using the World Health Organization's Composite International Diagnostic Interview (CIDI)—Short Form for Major Depression (CIDI-SFMD).
Results: In the training data, the C statistics for algorithms in men was 0.7953 and was 0.7667 for algorithm in women. The algorithms had good predictive power and calibrated well in the development and validation data.
Limitations: The data relied on self-report. MDE was assessed with CIDI-SFMD. It was not feasible to validate the algorithms in different populations from different countries.
Conclusions: More studies are needed to further validate and refine these algorithms. However, the ability of a small number of easily assessed variables to predict MDE risk indicates that algorithms are a promising strategy for identifying individuals in need of enhanced monitoring and preventive interventions. Ultimately, application of algorithms may lead to increased personalization of treatment, and better clinical outcomes.

© 2013 The Authors. Published by Elsevier B.V. All rights reserved.

1. Introduction

Major depression is prevalent and imposes considerable burden on society (Murray and Lopez, 1996). Major depressive episode (MDE) is the building block for mood disorders (American Psychiatric Association, 1994). In the past decades, research has identified risk factors for MDE. The most consistent demographic factors associated with the risk of MDE have been female sex, younger age and low socioeconomic status (Skapinakis et al., 2006; Weich and Lewis, 1998a, 1998b; Weich et al., 1998; Wang et al., 2010). Psychosocial risk factors include negative life events, experience of traumatic events, work stress, financial strain, poor marital or interpersonal relationships, lack of social support and low self-esteem and mastery (Blazer and Hybels, 2005; Goldberg, 2006; Hope et al., 1992; Janzing et al., 2009; Kendler et al., 2003; Kessing et al., 2003; Libby et al., 2005; Patel et al., 1999; Patten et al., 2005; Patton et al., 2003; Wang, 2005; Wang et al., 2012). Past occurrence of MDE and family history are strongly associated with the risk of MDE. These results have laid a strong foundation for early identification and intervention. However, the combined effects of a key set of risk factors that are the most predictive of future risk are unknown. To more accurately identify individuals who are at high risk of having MDE, prediction algorithms that include a key set of risk factors are needed.

☆This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial-No Derivative Works License, which permits non-commercial use, distribution, and reproduction in any medium, provided the original author and source are credited.

* Corresponding author. Tel.: +403 210 8653; fax: +403 210 9182.
E-mail address: jlwang@ucalgary.ca (J.L. Wang).

0165-0327/$ - see front matter © 2013 The Authors. Published by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jad.2013.05.045

Prediction algorithms are tools that aid physicians, individuals and population health policy makers to make informed decisions when it comes to the health problems. In the clinical setting, predictive risk algorithms are embedded in clinicians' daily practice as the primary tool to estimate individual risk of future disease. Thousands of risk algorithms are used to guide clinical decisions about disease prevention and treatment. Prediction algorithms are not about identifying new risk factors. Rather the algorithms are clinical and public health decision aids that are developed using known risk factors. The practical application of risk prediction algorithms is that they can be used to assess the risk or probability of developing certain health conditions among those who are free of the health conditions at the time of assessment, based on the exposure to risk factors included in the model (Anderson et al., 1991; Chen et al., 2006; Decarli et al., 2006; Park et al., 2009). Well-known examples include the Framingham risk prediction algorithms for cardiovascular disease (Anderson et al., 1991) and prediction algorithms for cancer risk (Chen et al., 2006; Decarli et al., 2006; Park et al., 2009). Risk prediction models can also be used for individual and population applications including self-assessment, prediction of the number of new cases of disease in populations and estimating the potential benefit of preventive interventions implemented community wide (Manuel and Rosella, 2010).

There is a paucity of research in risk prediction for mental disorders. To our knowledge, only the PredictD Study that was conducted in European primary care settings attempted to develop a risk prediction algorithm for MDE (King et al., 2008; Bellon et al., 2011). General population studies employing more feasible measurement strategies are needed so that the developed algorithms cannot only be used in clinical practice, but also be feasibly incorporated in general population health surveys for health planning purpose. Like the Framingham Risk Function (Anderson et al., 1991) and the Gal Score (Decarli et al., 2006), we believe that prediction algorithms should be developed for men and women separately due to their biological and social role differences and sex difference in the prevalence and incidence of MDE. Therefore, the objective of the current study was to develop and validate sex specific prediction algorithms for MDE over 4 years in the Canadian general population.

2. Methods

2.1. Study setting

This analysis is based on longitudinal data from the biennial National Population Health Survey (NPHS) (from cycle 1 to cycle 8). The household component of the survey covers the population living in private dwellings in the 10 provinces in 1994/1995. It excludes people on Indian reserves, in the territories, on Canadian Forces bases, and in some remote areas. Of 20,095 individuals selected for the longitudinal panel in 1994/1995, 17,276 agreed to participate. More detailed descriptions of the NPHS can be found in published reports (Catlin and Will, 1992; Tambay and Catlin, 1995).

2.2. Study participants

For this study, we chose Cycle 4 in the year of 2000/01 as baseline because comprehensive information about risk factors is available in this cycle. Participants who were aged 18 years and older and who were free of MDE in Cycle 4 were eligible to be included in model development ($n=10,601$). Prediction algorithms for 4-year risk of MDE were developed for men ($n=4737$) and women ($n=5864$), separately. Data from Cycle 5 (2002/03) and Cycle 6 (2004/05) were used to identify MDE occurring over the 4-year follow-up period (from 2002/03 to 2004/05).

2.3. MDE and risk factors

In the NPHS, MDE in the past 12 months was assessed by the Composite International Diagnostic Interview Short Form for Major Depression (CIDI-SFMD) (Kessler et al., 1998). The CIDI-SFMD was designed to provide an operationalization of the DSM-III-R diagnostic criteria (which have remained essentially unchanged in DSM-IV) for MDE. The instrument detects symptoms indicative of MDE, and identification of five such symptoms out of nine indicates a high probability that the subject fulfilled DSM-IV criteria for MDE in the past 12-months. The sensitivity and specificity of the CIDI-SFMD were 89.6% and 93.9% in relation to the full version of CIDI (Kessler et al., 1998). The CIDI-SFMD was used in all cycles of the NPHS.

We initially selected and examined the associations between the following potential risk factors and the 4-year risk of MDE: demographic and socioeconomic characteristics, self-rated general health and self-rated stress, activity restrictions, chronic medical conditions diagnosed by physicians, difficulty in mobility, cognitive function, levels of pain, monthly frequency of physical activity last 15 minutes or longer, ongoing chronic stress (18 items), recent negative life events (10 items), childhood traumatic events (7 items), work stress based on the brief version of Job Content Questionnaire(Karasek et al., 1998) (13 items), self esteem (Pearlin and Schooler, 1981) (6 items), mastery (Pearlin and Schooler, 1981) (7 items), anti-depressants and sleeping pill use in the past month, smoking, problematic alcohol consumption, K-6 psychological distress in the past month (Kessler et al., 2002) (6 items), having had depressed mood or loss of interests in activity in the past year, having talked or seen a health professionals for emotional or mental health issues in the past year, past MDE, and family history of depression. We used MDE occurring from Cycle 1 to Cycle 3 as an indicator for past MDE. It was decided to include this variable as it is likely to be one of the strongest predictors of MDE, and it was the goal of the study to predict new-onset episodes as opposed to new-onset disorders. The coding of the predictors included in the algorithms can be found in Appendix I.

2.4. Statistical analysis

All analyses were conducted using STATA 11 (StataCorp, 2010) We first examined the bi-variate relations between selected variables and the risk of MDE. Variables that were strongly associated with the outcome at the level of $p=0.002$, were kept as candidates for model development. Missing data in candidate risk factors were imputed using STATA's "ice" program with the outcome variable included in the imputation equation (Royston, 2005). We created 5 imputed data sets and obtained combined estimates using Rubin's rules (Rubin, 1987).

2.5. Development of prediction algorithms

Prediction algorithms for 4-year risk of MDE were developed using logistic regression modelling for men and women separately. We first included age (continuous variable), past MDE and family history of MDE in the model. We then examined whether adding an additional variable improved the model's discriminative power and fit with data. Once all the predictors from a specific domain were assessed, we re-evaluated the model to determine whether specific factors could be excluded without affecting the performance of the model. The observed effect (such as odds ratio) of specific predictor assumes that the effect is constant regardless of the levels of other predictors in the model. This is not always

true in medical research. Therefore, two-way interactions between age, past MDE, family history of MDE and selected predictors were also investigated. To be consistent to current medical literature (D'Agostino et al., 2001; Lloyd-Jones, 2010; Manuel et al., 2012; Steyerberg et al., 2010), the model's performance was assessed by discrimination and calibration.

Discrimination is the ability of a prediction model to separate those who experienced the outcome events from those who did not. We quantified this by calculating the C statistic. The C statistics is a rank order statistic for prediction against the true outcome and is analogous to the area under a receiver operating characteristic curve for a binary outcome variable. An uninformative model, such as coin flip, will have a C statistics of 0.5; a perfect prediction model will have a maximum C statistics of 1. Calibration measures how closely predicted outcomes agree with actual outcomes. For this we used D'Agostino's version of the Hosmer–Lemeshow (H–L) Chi 2 statistic (D'Agostino et al., 1997). A Chi 2 statistic was calculated to compare the differences between mean predicted and actual event rates; small values indicated good calibration. Once the combined algorithm coefficients were obtained, we applied the algorithms in each of the 5 imputed data sets. The C statistics and H–L chi 2 statistics were calculated for each imputed data set. Additionally, we visually compared the mean predicted risk of MDE versus observed risk of MDE by decile risk groups to exam calibration in specific risk groups.

2.6. Validation of the prediction algorithms

We validated the performance of the sex specific algorithms in two ways. First, we generated another imputed data set as for the training data. We applied the algorithms in sub-populations that are of practice and policy importance: immigration status (immigrants, non-immigrants), race (white, non-white), rural and urban areas and geographic regions (Atlantic Canada region, Ontario and Quebec, and Western Canada). We calculated the C statistics and H–L statistics of the algorithms by the sub-populations. Second, we applied the prediction algorithms in the Cycle 6 (2004/05) participants who were aged 18 years and older and who were free of MDE in this cycle to predict the 4-year risk of MDE from 2006/07 to 2008/09 based on data of Cycle 7 and Cycle 8. As described previously, some participants who were under 18 years old from Cycle 1 to Cycle 4 were excluded from the analysis using the training data. They became 18 years or older in Cycle 6 and were eligible for being included in the validation data (Women: $n=252$, 4.85%; men: $n=245$, 5.79%). From Cycle 4 (2000/01) to Cycle 6 (2004/05), some participants died or lost to follow-up ($n=476$, 4.5%). As such, we were able to validate the prediction algorithms in a slightly different population in a different time period. This study was approved the Conjoint Health Research Ethics Board of University of Calgary.

3. Results

The demographic and socioeconomic characteristics of the participants who were free of MDE and were aged 18 years or older in Cycle 4 (2000/01) are in Table 1. The developed prediction algorithms for men and women are in Table 2 and Table 3. The mean predicted 4-year risk of MDE in women based on the prediction algorithm was 8.31%; the observed risk was 8.27%. The mean predicted 4-year risk of MDE in men based on the algorithm was 5.25%; the observed risk was 5.15%. The mean predicted vs. observed risk of MDE by decile groups in men and in women are depicted in Fig. 1.

We applied the sex specific prediction algorithms in the 5 imputed data sets separately. The sex specific algorithms predicted the risk of MDE very well. The C statistics based on 5 data sets ranged from 0.7645 to 0.7684 in women, with an average C statistics of 0.7667. In men, the C statistics ranged from 0.7920 to 0.7969, with an average C statistics of 0.7953. There was no evidence against good calibration of the algorithm. The H–L Chi statistics ranged from 7.41 ($p=0.4934$) to 13.40 ($p=0.0989$).

Table 1
The demographic and socioeconomic characteristics of male and female participants who were free of major depressive episode in the development data.

Variables	Men (*n*, %)	Women (*n*, %)
	***N*=4737**	***N*=5864**
Age (mean, SE) year old	46.3 (0.24)	48.9 (0.24)
Married/common-low	3059 (64.6)	3283 (56.0)
Single/never married	1124 (23.7)	1041 (17.8)
Divorced/separated/widowed	554 (11.7)	1540 (26.2)
Immigrants	588 (12.4)	737 (12.6)
Non-immigrants	4146 (87.6)	5123 (87.4)
White	4391 (93.5)	5496 (94.4)
Non-white	307 (6.5)	329 (5.6)
Annual personal income*		
≤$29,999	1790 (40.5)	3869 (72.0)
$30,000–$59,999	1757 (39.8)	1256 (23.4)
$60,000+	869 (19.7)	249 (4.6)
Educational levels		
< Secondary school	1106 (23.5)	1353 (23.1)
Secondary school graduation	639 (13.5)	828 (14.1)
Some post-secondary	1265 (26.7)	1610 (27.6)
College/university degree	1720 (36.3)	2061 (35.2)

SE: standard error.

* In Canadian dollar.

Table 2
Prediction algorithm for women derived in imputed data sets.

Predictors	Coefficients	*P* value
Constant	−3.1127	< 0.0005
Age (continuous)	−0.0219	< 0.0005
Past major depressive episode	0.9747	< 0.0005
Family history of MDE	0.5844	< 0.0005
Annual personal income		
≤$29,999	0.5410	0.102
$30,000–$59,999	0.6332	0.061
Self-rated health		
Fair	0.3746	0.037
Poor	0.1130	0.748
Activity restrictions	0.4292	0.002
Satisfaction with oneself	1.0680	< 0.0005
Self-rated chronic stress	0.5278	< 0.0005
Childhood trauma		
Being traumatized for years	0.4314	0.004
Being sent away from home	1.4243	< 0.0005
Recent life events		
Major financial crisis	0.1714	0.266
Changed job for a worse one	0.1005	0.155
Chronic stress		
Other expect too much	0.1818	0.131
Family member in bad health	0.3105	0.102
Daily smoking	0.2780	0.028
Having depressed mood/lost of	0.5018	0.074
Interests for 2 weeks		
Talked to health professionals for mental heath issues	0.4534	0.009
Interaction terms		
Past MDE *x* changed for a worse job	−0.6840	0.064
Family history *x* being sent away from home	−1.8535	0.004
Family history *x* being traumatized for years	−0.4659	0.094

MDE: major depressive episode.

Table 3
Prediction algorithm for men derived in imputed data sets.

Predictors	Coefficients	*P* value
Constant	−2.7979	< 0.0005
Age (continuous)	−0.0258	< 0.0005
Past major depressive episode	1.1226	< 0.0005
Family history of MDE	1.5775	0.003
Physician diagnosed diabetes	0.8191	0.013
Self-rated chronic stress	0.6042	< 0.0005
Childhood trauma		
Parents divorced	−0.3630	0.145
Being sent away from home	0.8897	0.006
Parents abused alcohol/drugs	0.5562	0.004
Recent life events		
Being physically attacked	0.6541	0.015
Partner had unwanted pregnancy	0.6306	0.148
Chronic stress		
Other expect too much	0.2396	0.154
Lack of money	0.3474	0.029
Feeling everything is an effort	0.7551	0.045
Took antidepressants last month	1.0073	0.001
Took sleeping pills last month	0.9553	0.003
Interaction terms		
Age x family history of MDE	−0.0236	0.084

MDE: major depressive episode.

As part of validation process, we applied the sex specific prediction algorithms in sub-populations. The results are in Table 4. The mean predicted vs. observed risk of MDE by decile groups in sub-populations are in supplementary file 1. As seen from the table and the figures, the mean predicted 4-year risk of MDE resembled the observed risk across sub-populations. The algorithms performed very well in immigrants and non-immigrants, in participants who are white, those who lived in urban areas and in individuals who lived in Ontario, Quebec and Western Canada.

We validated the sex specific algorithms in the cohort started in Cycle 6 (2004/05) to predict 4-year risk of MDE based on Cycle 7 and Cycle 8 data. Among women ($n=4418$), the mean predicted 4-year risk of MDE was 7.95%; the observed risk was 7.99%. The C statistics was 0.7400 (95% confidence interval (CI): 0.7139, 0.7662). The calibration of the algorithm in women was good (H–L chi2 = 18.11, $p=0.02$). In men ($n=3484$), the mean predicted 4-year risk of MDE was 5.16%; the observed risk was 4.31%. The C statistics was 0.7720 (95% CI: 0.7338, 0.8102). There was no evidence against that the algorithm in men had good calibration (H–L Chi2 = 5.36, $p=0.7180$). The predicted vs. observed risk by decile groups in men and in women are in Fig. 2.

4. Discussion

Using data from a large population-based longitudinal cohort, we have developed sex specific prediction algorithms for MDE. We validated the algorithms in sub-populations that are important from the practice and health policy perspectives. Moreover, we validated the algorithms in a slightly different population by predicting MDE occurrence in a different time period and the algorithms calibrated well. The sex specific prediction algorithms based on the NPHS data have acceptable discriminative power and calibration. The C statistics ranged from 0.7645 to 0.7684 for women and from 0.7920 to 0.7969 for men.

We found that age, family history of MDE and past MDE are important predictors in the algorithms for both men and women. Additionally, the sex specific prediction algorithms have few other common predictors (self-rated stress, childhood traumatic event, financial difficulties and too much being expected by others).

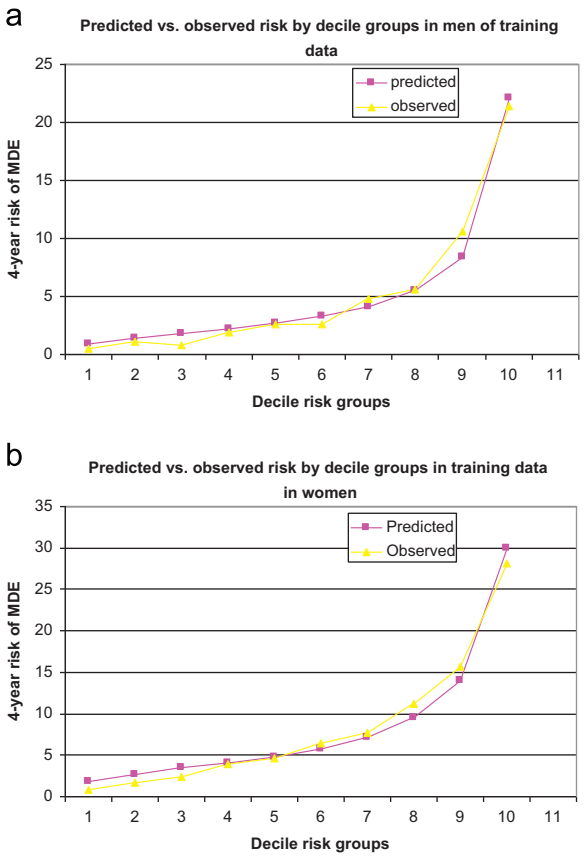


Fig. 1. Part-A: The mean predictive risk versus observed risk of MDE by decile groups in training data in men. Part-B: The mean predictive risk versus observed risk of MDE by decile groups in training data in women.

Table 4
The discrimination and collaboration characteristics of the sex specific prediction algorithms in validation data sets.

Populations	Pr.: Obs. risk %	C statistics (95% CI)	H–L Chi 2	*P* value
Men				
Immigrants	3.89: 3.85	0.7817 (0.6622, 0.9012)	11.24	0.1886
Non-immigrants	5.50: 5.28	0.8014 (0.7686, 0.8341)	9.98	0.2663
White	5.28: 4.96	0.7912 (0.7573, 0.8251)	6.44	0.5981
Non-white				
Urban areas	5.67: 5.23	0.8224 (0.7868, 0.8581)	14.44	0.0709
Rural areas	4.61: 5.00	0.7647 (0.7063, 0.8230)	6.68	0.5719
Atlantic Canada	5.07: 4.05	0.7757 (0.7059, 0.8455)	7.93	0.4400
Ontario & Quebec	5.18: 5.18	0.8013 (0.7530, 0.8497)	11.03	0.2001
Western Canada	5.63: 5.95	0.8145 (0.7641, 0.8650)	10.29	0.2453
Women				
Immigrants	6.78: 6.57	0.7760 (0.7007, 0.8513)	4.92	0.7658
Non-immigrants	8.41: 8.50	0.7647 (0.7403, 0.7891)	7.68	0.4653
White	8.16: 8.20	0.7692 (0.7454, 0.7930)	11.46	0.1768
Non-white				
Urban areas	8.39: 8.67	0.7710 (0.7428, 0.7991)	5.66	0.6856
Rural areas	7.83: 7.49	0.7562 (0.7161, 0.7964)	12.14	0.145
Atlantic Canada	7.68: 8.93	0.7715 (0.7274, 0.7857)	15.84	0.048
Ontario & Quebec	8.05: 7.82	0.7551 (0.7171, 0.7930)	5.40	0.7141
Western Canada	8.78: 8.33	0.7772 (0.7384, 0.8160)	4.80	0.7789

H–L Chi2: Hosmer–Lemeshow Chi 2 statistic.
95% CI: 95% confidence interval.
Pr.: Obs.: mean predicted risk vs. observed risk.

However, men and women appeared to differ in several factors. Different elements of ongoing stress, recent life events and childhood traumatic events seemed to matter differently for men and

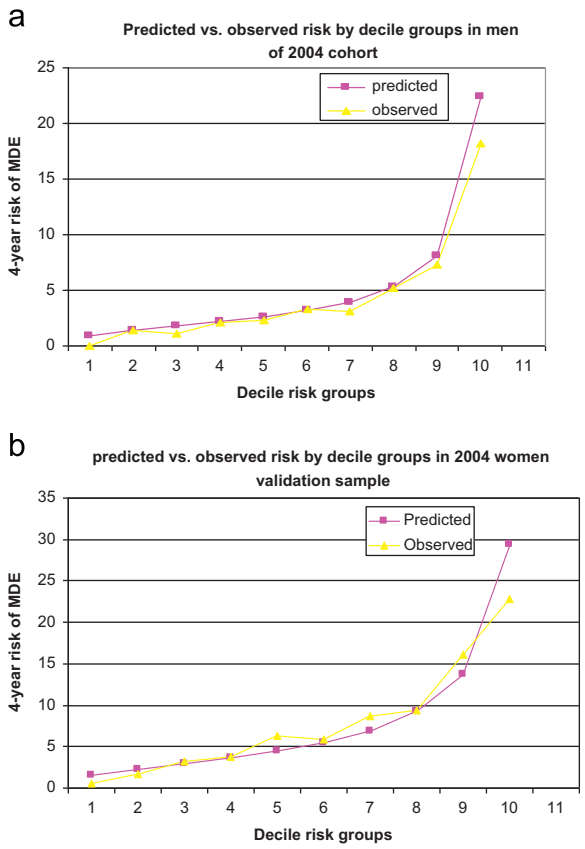


Fig. 2. Part-A: The mean predicted risk versus observed risk of MDE by decile groups in the validation data (2004 cohort) in men. Part-B: The mean predicted risk versus observed risk of MDE by decile groups in the validation data (2004 cohort) in women.

women in predicting the risk of MDE, as reflected in the algorithms. Low self-esteem, ongoing mood disturbance and general health conditions (low levels of self-rated health and activity restrictions) played an important role in the risk of MDE in women, whereas men who reported physical issues (having diabetes, fatigue and insomnia) were at higher risk of MDE. Depressed mood or lack of interest in activities and talking to health professionals about mental health issues are predictors for MDE in women; taking anti-depressants and sleeping pills were predictors for MDE in men. These may be an indication of either residual depressive symptoms from a previous episode, or ongoing depressive symptoms that may contribute to the emergence of a new episode. However, they manifested differently in men and women. Different predictors in the algorithms can be reflections of the biological and gender role differences between men and women. It is also possible that men and women share some common etiological factors for MDE, but the etiological factors for men and women are not entirely the same.

In the algorithm developed by the PredictD study (King et al., 2008), age, past depression and family history of depression were also important predictors. However, in the NPHS algorithms, education and work stress were not important predictors. Unlike our study, childhood traumatic events and ongoing life events were not retained in the final algorithms of the PredictD study. There are some notable differences between our study and the PredictD, which may contribute to the discrepancies in the results. First, we developed the algorithms by sex, whereas sex was included as a predictor in the PredictD. Second, the NPHS is a general population health study. The participants in the PredictD were primary care attendees. Third, the NPHS did not use the SF-12 and did not collect data about racial discrimination. Therefore, we could not assess whether physical and mental health summary scores and discrimination were useful predictors in the models. Fourth, the PredictD study used the full version of CID; the NPHS used the CIDI short form.

MDE has an episodic nature. One may recover from a MDE; she/he may also relapse. The developed algorithms are not specifically for predicting first onset or reoccurrence. Rather they are to be used to estimate the probability of having a MDE in the future in individuals who do not meet the criteria for MDE at the time of assessment. Individuals who have past MDE or family history of MDE alone are not necessarily in the high risk groups. We described different case scenarios and calculated the predicted risk using the algorithms (see online supplementary file 2). It shows that those who are simultaneously exposed to multiple clinical and psychosocial risk factors are at the highest risk of MDE. This indicates that population risk for MDE is concentrated in identifiable high risk groups. The implication of the results is that, with the prediction algorithms, population health planners will not only be able to monitor risk profiles, to project the number of new cases of MDE in the population, but also to implement preventive interventions in high risk individuals (Manuel et al., 2012).

It should be noted that prediction algorithms are not about identifying new risk factors. Prediction algorithms for MDE are tools that estimate the probability of having MDE in the future based on exposure to known risk factors. The algorithms are not developed to predict the first onset of MDE or "incident cases" over 4 years in the target population. The algorithms are to estimate the probability of having MDE in the future among those who do not have MDE in the past year. Some of them who use the algorithms for self-assessment and who see their family doctors may have MDE in the past. So including history of MDE in the algorithms reflects the situation in clinical settings. Although some predictors in the algorithms were not statistically significant, it should be noted that the goal of prediction algorithms is prediction rather than hypothesis testing. Therefore, it is reasonable to include non-significant predictors based on their contribution to prediction (Steyeberg, 2009).

This study has several limitations. First, the NPHS interview assesses MDE in the year before an interview, whereas the interviews are conducted 2 years apart. Thus, the proportion of MDE could have been under estimated. However, a simulation study showed that, despite the timing of the NPHS interviews, the relative risk estimates were not substantially distorted (Patten, 2012). Data with more frequent assessment of MDE would be ideal, but challenging to obtain given the required human and financial resources. Second, data about the predictors in the algorithms relied on self-report. Recall and reporting biases are possible. However, a major finding of this study is that these brief self-report items were able to generate powerful predictive algorithms. Third, although we validated the algorithms in the cohort followed from 2004 to 2008 and the participants in this validation cohort were slightly different from those in the training data, it is not an external validation. It will be ideal to validate the algorithms in populations in different regions/countries by other investigators. At present, such external validations are not feasible because population-based longitudinal studies that use similar study design and collect same information for predictors and outcome as the NPHS do not exist. Prediction algorithms are often initially developed without external validation. Once they become available, they can be truly independently validated by other investigators in different region. Fourth, like any longitudinal studies, some NPHS participants were lost to follow-up. The slightly over estimation of MDE in the highest (10th) risk group in the 2004 cohort validation (see Fig. 2) might be a reflection of the effect of attrition, rather than poor prediction of the

algorithms, because more participants who were at high risk of MDE in this risk group dropped out from the cohort. Finally, compared to the well developed algorithms in the areas of heart disease and cancer, these models are more complicated, which could be partly due to the nature and etiology of MDE. Prediction models can be presented in various formats including score charts and tables. The simplest form is the regression formula as presented in this study. Other formats can be derived from the formula, with less precision. Easy-to-understand formats such as score charts or tables are preferred by physicians. With the increasing use of computers and cell phones in offices, the regression formula can be programmed into computer and phone applications. The web version of the developed algorithms is very easy to use and can be found at: http://www.predictingdepression.com. A clinician does not have to remember the questions in order to use the model. As such, answering these questions and calculating the risk become manageable in busy clinical settings. With the computerized algorithms, self-assessment and including the predictors in general health surveys for the purpose of population health planning should not be an issue.

Despite the limitations, to our knowledge, this is the first study that developed sex specific prediction algorithms for MDE in the general population. The large cohort of participants were recruited from 10 provinces in Canada and were re-interviewed using the same diagnostic instrument every two years by Statistics Canada, which is beyond the efforts and resources of any individual research teams. The information about the predictors in the algorithms can be easily collected in general population health surveys and in primary care settings, e.g., waiting rooms or self-assessment at home before seeing a doctor, which makes the clinical and population health uses of the algorithms possible. Additional strength of this study was that sex specific prediction algorithms were developed. Sex specific prediction algorithms recognize the fact that different and common risk factors have different predictive values for MDE in men and in women. Nevertheless, we believe that more studies are needed to further develop, validate and refine the algorithms to improve their performance so that clinicians, population health planners and the general public will have better tools to guide interventions, to make decisions and to protect themselves.

Role of funding source

None.

Conflict of interest

None.

Acknowledgement

This study was funded by an operating grant from the Canadian Institutes of Health Research (grant number: MOP-114970, PI: JianLi Wang). The funder has no role in the design of this study.

Reference

American Psychiatric Association, 1994. Diagnostic and Statistical Manual of Mental Disorders. APA, Washington, D.C.

Anderson, K.M., Wilson, P.W., Odell, P.M., Kannel, W.B., 1991. An updated coronary risk profile. A statement for health professionals. Circulation 83, 356–362.

Bellon, J.A., de Dios, L.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, M.M.-G., 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. Psychological Medicine 41, 2075–2088.

Blazer, D.G., Hybels, C.F., 2005. Origins of depression in later life. Psychological Medicine 35, 1241–1252.

Catlin, G., Will, P., 1992. The National Population Health Survey: highlights of initial developments. Health Reports. 4, 313–319.

Chen, J., Pee, D., Ayyagari, R., Graubard, B., Schairer, C., Byrne, C., Benichou, J., Gail, M.H., 2006. Projecting absolute invasive breast cancer risk in white women with a model that includes mammographic density. Journal of the National Cancer Institute 98, 1215–1226.

D'Agostino R.B., Griffith J., Schmidt C., Terrin N. Measures for evaluating model performance. In: Annual Meeting of the American Statistical Association. Alexandria, Va: American Statistical Association, 1997. pp. 253–258.

D'Agostino Sr., R.B., Grundy, S., Sullivan, L.M., Wilson, P., 2001. Validation of the Framingham coronary heart disease prediction scores: results of a multiple ethnic groups investigation. Journal of the American Medical Association 286, 180–187.

Decarli, A., Calza, S., Masala, G., Specchia, C., Palli, D., Gail, M.H., 2006. Gail model for prediction of absolute risk of invasive breast cancer: independent evaluation in the Florence-European Prospective Investigation Into Cancer and Nutrition cohort. Journal of the National Cancer Institute 98, 1686–1693.

Goldberg, D., 2006. The aetiology of depression. Psychological Medicine 36, 1341–1347.

Hope, S., Power, C., Rodgers, B., 1992. Does financial hardship account for elevated psychological distress in lone mothers? Social Science & Medicine 49, 1637–1649.

Janzing, J.G., de, G.R., ten, H.M., Vollebergh, W.A., Verhagen, M., Buitelaar, J.K., 2009. Familiality of depression in the community; associations with gender and phenotype of major depressive disorder. Social Psychiatry and Psychiatric Epidemiology 44, 1067–1074.

Karasek, R., Brisson, C., Kawakami, N., Houtman, I., Bongers, P., Amick, B., 1998. The Job Content Questionnaire (JCQ): an instrument for internationally comparative assessments of psychosocial job characteristics. Journal of Occupational Health Psychology 3, 322–355.

Kendler, K.S., Hettema, J.M., Butera, F., Gardner, C.O., Prescott, C.A., 2003. Life event dimensions of loss, humiliation, entrapment, and danger in the prediction of onsets of major depression and generalized anxiety. Archives of General Psychiatry 60, 789–796.

Kessing, L.V., Agerbo, E., Mortensen, P.B., 2003. Does the impact of major stressful life events on the risk of developing depression change throughout life? Psychological Medicine 33, 1177–1184.

Kessler, R., Andrews, G., Mroczek, D., Ustun, T.B., Wittchen, H.U., 1998. The world health organization composite international diagnostic interview short-form (CIDI-SF). International Journal of Methods in Psychiatric Research 7, 171–185.

Kessler, R.C., Andrews, G., Colpe, L.J., Hiripi, E., Mroczek, D.K., Normand, S.L., Walters, E.E., Zaslavsky, A.M., 2002. Short screening scales to monitor population prevalences and trends in non-specific psychological distress. Psychological Medicine 32, 959–976.

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. Archives of General Psychiatry 65, 1368–1376.

Libby, A.M., Orton, H.D., Novins, D.K., Beals, J., Manson, S.M., 2005. Childhood physical and sexual abuse and subsequent depressive and anxiety disorders for two American Indian tribes. Psychological Medicine 35, 329–340.

Lloyd-Jones, D.M., 2010. Cardiovascular risk prediction: basic concepts, current status, and future directions. Circulation 121, 1768–1777.

Manuel, D.G., Rosella, L.C., 2010. Commentary: assessing population (baseline) risk is a cornerstone of population health planning–looking forward to address new challenges. International Journal of Epidemiology 39, 380–382.

Manuel, D.G., Rosella, L.C., Hennessy, D., Sanmartin, C., Wilson, K., 2012. Predictive risk algorithms in a population setting: an overview. Journal of Epidemiology & Community Health.

Murray, C.J., Lopez, A.D., 1996. The Global Burden of Disease: A Comprehensive Assessment of Mortality and Disability from Disease, Injuries, and Risk Factors in 1990 and Projected to 2020. Harvard University Press.

Park, Y., Freedman, A.N., Gail, M.H., Pee, D., Hollenbeck, A., Schatzkin, A., Pfeiffer, R. M., 2009. Validation of a colorectal cancer risk prediction model among white patients age 50 years and older. Journal of Clinical Oncology 27, 694–698.

Patel, V., Araya, R., de, L.M., Ludermir, A., Todd, C., 1999. Women, poverty and common mental disorders in four restructuring societies. Social Science & Medicine 49, 1461–1471.

Patten, S.B., 2012. The National Population Health Survey's assessment of depression risk factor associations: a simulation study assessing vulnerability to bias. Chronic Diseases and Injuries in Canada 32, 70–75.

Patten, S.B., Beck, C.A., Kassam, A., Williams, J.V., Barbui, C., Metz, L.M., 2005. Long-term medical conditions and major depression: strength of association for specific conditions in the general population. Canadian Journal of Psychiatry 50, 195–202.

Patton, G.C., Coffey, C., Posterino, M., Carlin, J.B., Bowes, G., 2003. Life events and early onset depression: cause or consequence? Psychological Medicine 33, 1203–1210.

Pearlin, L., Schooler, C., 1981. The Structure of Coping. Journal of Health and Social Behaviour.

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

Rubin, D., 1987. Multiple Imputation for Non-Response in Surveys. John Wiley and Sons, New York.
Skapinakis, P., Weich, S., Lewis, G., Singleton, N., Araya, R., 2006. Socio-economic position and common mental disorders Longitudinal study in the general population in the UK. British Journal of Psychiatry 189, 109–117.
StataCorp. Stata Statistical Software: Release 11.0. In: College Station, TX: Stata Corporation, 2010.
Steyeberg, E., 2009. Clinical Prediction Models. A Practical Approach to Development, Validation, and Updating. Springer.
Steyerberg, E.W., Vickers, A.J., Cook, N.R., Gerds, T., Gonen, M., Obuchowski, N., Pencina, M.J., Kattan, M.W., 2010. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology 21, 128–138.
Tambay, J.L., Catlin, G., 1995. Sample design of the national population health survey. Health Reports. 7, 29–42.
Wang, J., 2005. Work stress as a risk factor for major depressive episode(s). Psychological Medicine 35, 865–871.
Wang, J., Patten, S.B., Currie, S., Sareen, J., Schmitz, N.A., 2012. Population-based Longitudinal Study on Work Environmental Factors and the risk of major depressive disorder. American Journal of Epidemiology 176, 52–59.
Wang, J.L., Schmitz, N., Dewa, C.S., 2010. Socioeconomic status and the risk of major depression: the Canadian National Population Health Survey. Journal of Epidemiology & Community Health 64, 447–452.
Weich, S., Lewis, G., 1998a. Poverty, unemployment, and common mental disorders: population based cohort study. British Medical Journal 317, 115–119.
Weich, S., Lewis, G., 1998b. Material standard of living, social class, and the prevalence of the common mental disorders in Great Britain. Journal of Epidemiology & Community Health 52, 8–14.
Weich, S., Sloggett, A., Lewis, G., 1998. Social roles and gender difference in the prevalence of common mental disorders. British Journal of Psychiatry 173, 489–493.