Development and validation of a prediction algorithm for the onset of common mental disorders in a working population

Ana Fernandez[1,2], Luis Salvador-Carulla[1], Isabella Choi[3], Rafael Calvo[4], Samuel B Harvey[5,6] and Nicholas Glozier[3]

Abstract

Objective: Common mental disorders are the most common reason for long-term sickness absence in most developed countries. Prediction algorithms for the onset of common mental disorders may help target indicated work-based prevention interventions. We aimed to develop and validate a risk algorithm to predict the onset of common mental disorders at 12 months in a working population.

Methods: We conducted a secondary analysis of the *Household, Income and Labour Dynamics in Australia Survey*, a longitudinal, nationally representative household panel in Australia. Data from the 6189 working participants who did not meet the criteria for a common mental disorders at baseline were non-randomly split into training and validation databases, based on state of residence. Common mental disorders were assessed with the mental component score of 36-Item Short Form Health Survey questionnaire (score ⩽45). Risk algorithms were constructed following recommendations made by the *Transparent Reporting of a multivariable prediction model for Prevention Or Diagnosis* statement.

Results: Different risk factors were identified among women and men for the final risk algorithms. In the training data, the model for women had a *C*-index of 0.73 and effect size (Hedges' *g*) of 0.91. In men, the *C*-index was 0.76 and the effect size was 1.06. In the validation data, the *C*-index was 0.66 for women and 0.73 for men, with positive predictive values of 0.28 and 0.26, respectively

Conclusion: It is possible to develop an algorithm with good discrimination for the onset identifying overall and modifiable risks of common mental disorders among working men. Such models have the potential to change the way that prevention of common mental disorders at the workplace is conducted, but different models may be required for women.

Keywords

Common mental disorders, risk algorithm, prevention

[1]Mental Health Policy Unit, Brain and Mind Centre, The University of Sydney, Camperdown, NSW, Australia
[2]Public Health Agency of Barcelona, Barcelona, Spain
[3]Brain and Mind Centre, Central Clinical School, Sydney Medical School, The University of Sydney, Camperdown, NSW, Australia
[4]School of Electrical and Information Engineering, The University of Sydney, Sydney, NSW, Australia
[5]School of Psychiatry, University of New South Wales, Sydney, NSW, Australia
[6]Black Dog Institute, Randwick, NSW, Australia

Corresponding authors:
Ana Fernandez, Mental Health Policy Unit, Brain and Mind Centre, The University of Sydney, Camperdown, NSW 2050, Australia.
Email: anafezsez@gmail.com

Nicholas Glozier, Brain and Mind Centre, Central Clinical School, Sydney Medical School, The University of Sydney, Level 5, Professor Marie Bashir Centre, Missenden Road, Camperdown, NSW 2050, Australia.
Email: nick.glozier@sydney.edu.au

Introduction

Common mental disorders (generally defined as depressive and anxiety disorders, sometimes with the addition of common forms of substance abuse) are viewed as a major and increasing public and occupational health issue as they cause high levels of distress for those who experience them and their relatives, as well as considerable financial costs to society. In almost all developed countries, common mental disorders have now taken over from musculoskeletal problems as the leading cause of incapacity benefits and sickness absence, creating a huge cost to industry, society and the individuals effected (Harvey et al., 2009).

In an effort to reduce this burden, governments have supported increases in clinical services for mental disorder. However, despite this investment, the prevalence of mental disorders has not changed (Kessler et al., 2005). Evidence suggests that even in the unlikely event of adequate treatments being available to all who need them, there will only be a 40% reduction in the disease burden of common mental disorders (Andrews et al., 2004). The prevalence of disease is a combination of incidence and duration. So, to reduce its prevalence, interventions need to focus not only on a condition's duration (i.e. through treatment) but also on reducing the number of new cases or relapses through prevention efforts (Jacka et al., 2013).

The workplace may be a good setting for prevention programmes as employers are motivated to financially support prevention activities, and it is a place where many members of society spend much of their waking hours (Mykletun and Harvey, 2012). The workplace brings together the two opposing paradigms in mental health: (a) 'toxic work-stress', and (b) 'good work is good for your health'. The former proposes that the focus of workplace mental health intervention should be on increasing the skills to manage work-related stressors or making structural and design changes to reduce the stressors, while the latter focuses on the relevance of the workplace as a source of social support and self-realization (Czabala et al., 2011; Mykletun and Harvey, 2012). Unfortunately, wide-scale implementations of mental health prevention programmes in the workplace have been impeded by the uncertainty about what programmes are most effective and how to identify those most in need of prevention efforts (Joyce et al., 2016). There is good evidence that a range of prevention interventions can decrease the incidence of depression by 25–50% (Cuijpers et al., 2012), but the vast majority of the evidence-based initiatives depend on identifying a high-risk group for selected intervention (Beekman et al., 2010). However, to date, there is no agreed way in which to quantify the overall mental health risk among working individuals.

Estimating overall risk across a range of vulnerability and protective risk factors followed by health promotion/risk profile targeting interventions or recommendations is widely accepted as the basis of much primary indicated and selected prevention in developed countries for cardiovascular and metabolic diseases (Dapp et al., 2011; Maron et al., 2008; Nagykaldi et al., 2013). Conversely, there are few risk algorithms for the onset of mental disorders (Bellon et al., 2011; King et al., 2008, 2011; Moreno-Peral et al., 2014; Wang et al., 2014), and none have been developed specifically for working people. A range of modifiable environmental and individual risk factors have been identified for poor mental health in employees, but these have almost exclusively focused on workplace factors as they have been derived from occupational cohorts (Henderson et al., 2011). Furthermore, they are subject to a number of specific selection and other biases, for example, they are often based around government employees, that limit generalisability (Glozier et al., 2010). As such, developing a potentially broadly applicable risk algorithm and profile requires data from a large population-based cohort of working people with a broad range of both work and non-work factors, adequate response rates and follow-up, and preferably participants unaware of the aim to study such risks to ameliorate response bias. The *Household, Income and Labour Dynamics in Australia Survey* (HILDA) study provides just such a resource. In this study, we aimed to develop and validate a risk prediction algorithm for the onset of common mental disorders among employed people.

Methods

We followed the recommendations made by the TRIPOD statement (*Transparent Reporting of a multivariable prediction model for Prevention Or Diagnosis*) (Moons et al., 2015) by non-randomly splitting a sample for the development and validation of the prediction model. This process uses a frequentist approach, regression, which, although potentially falsely promoting one model of many with similar predictive properties, enables a comprehensible algorithm to be developed and then tested.

Study design and data sources

We conducted a secondary data analysis of the HILDA, a large, longitudinal, nationally representative household panel. Since 2001, it has conducted survey waves on annual basis (Watson and Wooden, 2010, 2012). Data are collected through both self-reported and a personal interview with a specific household member. It has been used primarily for socioeconomic purposes in Australian policy-making. The two most recent waves (at the time of the study, waves 11 and 12) formed the basis of this project

Participants

HILDA has two population samples: the original one that started in 2001 (wave 1) and a top-up sample that was

Figure 1. Flow diagram.

Figure 1. Flow diagram.
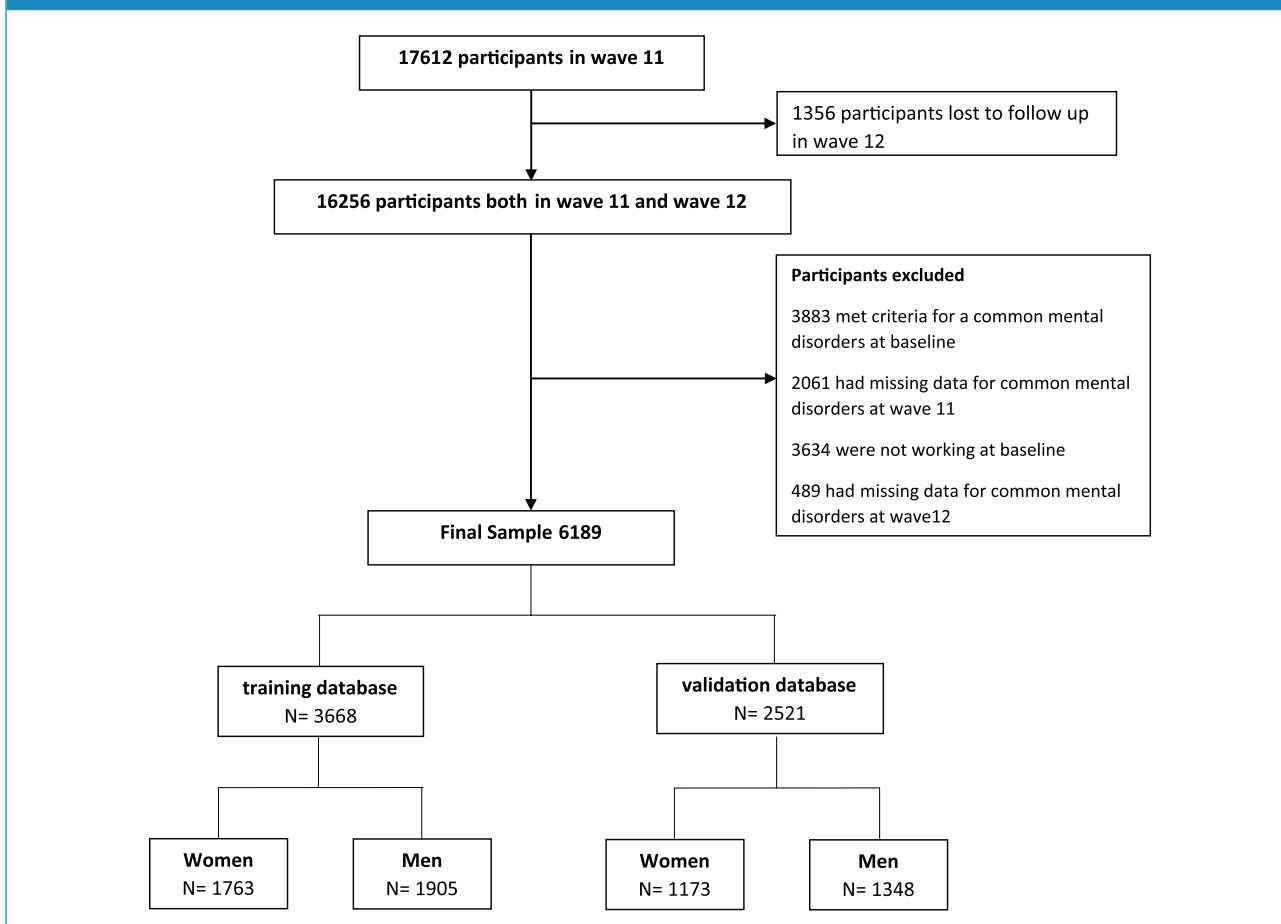
introduced in 2011 (wave 11). A total of 17,612 Australians aged 15 years and above participated in wave 11. There was a household response rate of 69% in wave 11.

Among those who answered wave 11, we selected those who (a) were followed up in wave 12, (b) did not have any common mental disorder at wave 11 as defined by the mental health score derived from the 36-Item Short Form Health Survey (SF-36) at wave 11 and (c) did not have missing data in the outcome variable (i.e. SF-36) at both baseline and follow-up. There were no gender differences at baseline between those with complete SF-36 and those with missing data, although they tend to be younger (52.27 vs 50.54 years old; $p < 0.009$): At follow-up, there were more men with missing SF-36 data (24.12% vs 21.53%; $p < 0.015$) than women, and those with missing SF-36 data were younger (52.68 vs 50.72; $p < 0.0041$). However, the differences were small and were statistically different mostly due to the large sample size.

This final sample of 6189 participants was split into two: (a) the training (development) database and (b) the validation database. Simple random selection minimises variability between the two samples, risking an overly positive

result from the validation sample (Moons et al., 2015). To avoid this, we split the sample by state/territory to maximise variability, stratified by gender. The training database included participants from New South Wales, Northern Territory, Queensland and South Australia, while people from Victoria, Western Australia, Tasmania and the Australian Capital Territory (ACT) formed the base for the validation database. As a consequence, we ended up with four subsamples (Figure 1):

1. Training database women: 1763 participants;
2. Training database men: 1905 participants;
3. Validation database women: 1173 participants;
4. Validation database men: 1348 participants.

Sample size

We did not calculate a formal sample size, as the HILDA study is an ongoing study with a fixed number of participants. However, it has been suggested that a prediction model requires at least 10 events per candidate variable and at least 100 participants with the outcome for the validation

(Moons et al., 2015). Previous analyses have estimated that incidence of a new episode of common mental disorders is between 10% and 15%. Therefore, both our sample and the number of events meet these recommendations (Moons et al., 2015).

Outcome predicted: common mental disorder

Common mental disorders have previously been shown to be ascertainable in Australia using the mental component score (MCS) of the SF-36 questionnaire (Butterworth and Crosier, 2004), which is administered each year in HILDA. To calculate the MCS, we followed standard procedures and transformed each of the eight SF-36 subscales to a 0–100 scale (McHorney et al., 1993, 1994; Ware and Sherbourne, 1992). Scoring algorithms were applied to produce the MCS, using the Australian norms (ABS, 1997). Although the time frame of SF-36 is the past 4 weeks, low scores on the mental component of SF-36 and its short form (SF-12, derived from SF-36) are highly correlated with 1-year prevalence of common mental disorders in different contexts, including Australian populations (Gill et al., 2007; Kristjansdottir et al., 2011; McCallum, 1995; Vilagut et al., 2013; Windsor et al., 2006). In a study using data from the *Australian National Survey of Mental Health and Wellbeing*, a score of 45 on the MCS was the best cut-off for depression (area under a curve [AUC]: 0.92) and a score of 50 for anxiety disorders (AUC: 0.83) (Gill et al., 2007). To avoid false-positives, and being aware that participants in HILDA tended to report slightly lower MCS scores than other Australian health surveys (Butterworth and Crosier, 2004), we followed a restrictive approach and considered that a person was experiencing a common mental disorder if his or her SF-36 MCS was ≤45.

Predictors

Most of the HILDA questions were taken or slightly modified from surveys undertaken by the Australian Bureau of Statistics, the Australian Institute of Health and Welfare or were made ad hoc in consultation with a group of experts. Based on previously published studies of mental health risk algorithms (Bellon et al., 2011; King et al., 2008, 2011; Moreno-Peral et al., 2014; Wang et al., 2014), we selected the following potential risk/protective factors.

> *Demographics and socioeconomic status*: age, education, marital status, type of employment, having English as a second language, being born outside of Australia and being from Aboriginal and Torres Strait Islander background;
> *Health*: smoking behaviour, drinking behaviour, body mass index (BMI), physical activity and self-reported chronic illness;

History of a common mental disorder in the 2 years prior to baseline defined as having met the same criteria in either of the two previous waves of HILDA (9 and 10). In the case of participants coming from the top-up sample in wave 11 (30%), this information was imputed (see missing data section).
Potential life stressors:
- *Being a carer*;
- *Stressful life events* in the past 12 months assessed with 17 questions and based on the list developed by Holmes and Rahe (Summerfield et al., 2015);
- *Financial hardship* was assessed using eight questions, assessing prosperity taking into account needs and responsibilities, and if the person had problems paying bills, rent or mortgage; or had pawned or sold something to get money, buy food or heat the home; or had to ask friends or organisations for help.
Work factors:
- *Job satisfaction* in 6 areas: pay, job security, the work itself, the numbers of hours, flexibility and total;
- *Union membership*;
- *Psychosocial work characteristics*: based on 21 questions derived from the Job Content Questionnaire (Robert and Karasek, 1979).
Satisfaction with different life domains: home, finances, safety, community, health, neighbourhood, free time, life, partner (if applicable) and the way tasks are divided at home (if applicable);
Questions related to the perception of personal control over life: assessed with the questionnaire developed by Pearlin and Schooler to try and elucidate reports of autonomy and control and/or discriminatory behaviour (Summerfield et al., 2015);
Aspects of social capital: neighbourhood trust, frequency of social contacts and being a member of a club;
Social support: assessed with 10 items, with the first 7 items coming from Henderson et al. and the last 3 items from Marshall and Barnett (Summerfield et al., 2015).

Statistical analysis

All the analysis and data imputation were performed using STATA 13SE and stratified by gender.

Missing data

As others have done when building risk algorithms (Bellon et al., 2011; King et al., 2008; Moreno-Peral et al., 2014), we excluded those participants with missing data for the outcome variable (i.e. the SF-36). Any missing value in the predictors was imputed using multiple imputations with chained equations under a missing at random (MAR) framework. We generated 30 imputed samples. Estimates for the analysis were combined using Rubin's (1987) rules.

Model building

The training databases were used to build the model and perform the internal validation. The potential variable list was too extensive to be included in the model in one step, taking into account our sample size and the guide of at least 10 events per variable. As such, for the main models, we chose baseline predictors that, in unadjusted analysis of the a priori chosen variables, were associated with the outcome with a *p* value <0.2 (Greenland, 1989). We assessed autocorrelation and, for highly correlated variables, included them in a separate, restricted, multivariate model and selected the variables that had a *p* value <0.2. This resulted in an initial list of 34 candidate variables for women and 31 for men.

The prediction algorithm was developed using a logistic regression model. We acknowledge that there are other methods for developing risk algorithms, such as classification trees, neural networks, genetic programming, random forests and vector machine learning techniques, that are receiving increasing attention. However, we decided to use regression modelling as this is the most used method (so, we will be able to compare our results with others) and because concerns have been raised related to the transparency of these methods (Moons et al., 2015). We developed the model in the imputed databases using backward procedures, removing those with the weakest association (i.e. highest *p* value) step by step to obtain a more parsimonious model, until all the remaining variables had a *p* value <0.1 (Moons et al., 2015). We then sequentially included each variable that was excluded in prior steps to test whether the final model performance could be improved. These variables were included in the final model if their *p* value was <0.1 and their presence enhanced the model, comparing AUC values (Moons et al., 2015).

The two core psychometrics when developing a risk algorithm are (a) discrimination and (b) calibration. *Discrimination*, which measures how well an algorithm distinguishes (discriminates) individuals with and without the outcome of interest, is commonly assessed by *C*-index (Tripepi et al., 2010a), which, for a dichotomous outcome, is the same as the receiver operating characteristic (ROC). We calculated the *C*-index as proposed by Copas (1983) to adjust for over-fitting of our prediction model. *Calibration* is the ability of a risk model to correctly estimate the probability of a given event across the whole range of prognostic estimates (Tripepi et al., 2010b). We assessed the goodness of fit of the final risk model by grouping individuals into deciles of risk and comparing the observed probability of developing common mental disorders within these groups with the average risk. We calculated the effect sizes using Hedges' *g* index (Cooper et al., 1994) for the difference in log odds of predicted probability between participants in each

decile who were later observed to develop a common mental disorder and those who were not. We report the threshold values of risk score and the associated sensitivity and specificity values. We also calculated Youden's index (J = sensitivity + specificity − 1). We used the *C*-index, Hedges' *g* and a comparison of predicted vs observed probability of common mental disorders to evaluate the performance of the trained model in the validation databases.

Results

The participant flow and inclusion are shown in Figure 1. Wave 11 had 17,612 participants. A total of 1356 (8%) persons with information in wave 11 were lost to follow-up in wave 12. Those lost to follow-up were more likely to be younger, male, divorced or widowed; have lower education; not working; and from an Aboriginal and/or Torres Strait Islander (ATSI) background. However, those lost to follow-up in wave 12 had a lower probability of meeting the criteria for a common mental disorder in wave 11 (i.e. they were healthier). Indeed, although not statistically significant, those with a chronic health condition had a lower probability of be missed at follow-up.

The removal of those lost to follow-up resulted in a sample of 16,256 participants that we were able to link from wave 11 to wave 12. In total, 10,067 did not meet inclusion criteria of working, had no common mental disorder at wave 1, and had common mental disorder data at wave 2 (Figure 1). This resulted in a final sample of 6189 participants, which was split into (a) the training database (development) and (2) the validation database and stratified by gender.

Participants

The characteristics of the participants in the training and validation datasets, stratified by gender, are shown in Table 1. There were fewer married women than men and a higher proportion of women with a high educational degree. More men were working in primary industries (e.g. farming, mining) or construction, while more women were working in the health care, education or professional sector. Men and women were similar in age, ATSI background, being born overseas or having English as a second language. For both men and women, there were no systematic differences between participants in the development and validation datasets. In the women's development dataset, 242 (13.7%) developed a common mental disorder. In the validation data, 182 (15.0%) developed a common mental disorder over the 1-year follow-up period. A total of 203 (10.7%) and 140 (10.4%) developed a common mental disorder in the men's training and validation database, respectively.

Table 1. Demographic characteristics of the participants.

	Women		Men	
	Training database (N = 1763)	Validation database (N = 1173)	Training database (N = 1905)	Validation database (N = 1348)
Mean age (years)	41.62 (40.97−42.26)	40.63 (39.84−41.40)	41.38 (40.75−42.00)	41.83 (41.10−42.57)
Marital status (%)				
Married or living with someone	62.63 (60.37−64.89)	65.55 (62.83−68.28)	70.88 (68.84−72.93)	70.15 (67.71−72.60)
Divorced or separated	14.48 (12.84−16.13)	13.04 (11.11−14.97)	9.41 (8.09−10.71)	9.42 (7.86−10.99)
Never married and living alone	20.50 (18.61−22.38)	20.71 (18.39−23.04)	19.39 (17.61−21.16)	19.60 (17.48−21.72)
Widowed	2.38 (1.67−3.09)	0.68 (0.21−1.15)	0.31 (0.63−0.56)	0.81 (0.34−1.30)
Education (%)				
Degree	32.96 (30.76−35.15)	36.83 (34.06−39.59)	25.58 (23.61−27.54)	29.52 (27.09−31.96)
Certificate or diploma	28.76 (26.64−30.87)	24.89 (22.41−27.37)	40.80 (38.60−43.02)	38.28 (35.68−40.88)
Year 12	17.07 (15.31−18.83)	18.92 (16.68−21.17)	14.60 (13.01−16.19)	15.65 (13.71−17.59)
Year 11 or below	21.21 (19.30−23.12)	19.35 (17.08−21.62)	19.01 (17.25−20.78)	16.54 (14.56−18.53)
Occupation (%)				
Manufacturing, construction, primary	8.04 (6.77−9.32)	9.68 (7.98−11.38)	33.84 (31.71−35.98)	33.48 (30.95−36.02)
Service industries/retail	35.94 (33.69−38.18)	32.39 (29.70−35.08)	35.53 (33.37−37.70)	33.56 (31.02−36.09)
Health care, professional, education	56.02 (53.69−58.34)	57.93 (55.09−60.76)	30.62 (28.54−32.70)	32.96 (30.43−35.48)
English as a second language (%)	9.98 (8.58−11.38)	9.71 (8.02−11.42)	9.71 (8.38−11.04)	8.68 (7.17−10.18)
Born overseas (%)	20.36 (18.48−22.24)	19.18 (16.93−21.43)	21.15 (19.32−22.99)	22.10 (19.88−24.33)
Aboriginal and/or Torres Strait Islander Background (%)	1.53 (0.95−2.11)	1.19 (0.57−1.82)	1.15 (0.67−1.64)	1.78 (1.07−2.48)

Prediction models

Women. The final prediction model for women included eight variables (Table 2). The *C*-index (discrimination) and effect size (Hedges' *g*, calibration) were 0.73 (95% confidence interval [CI] = [0.72, 0.74] and 0.91 (95% CI = [0.86, 0.96]), respectively. The calibration showed an accurate goodness of fit, except in groups 7 and 8 (Figure 1 in Supplemental File 1). The predicted probability cut-point of 14.26% was associated with the greatest Youden's *J* statistic which had limited sensitivity (62.7%) but greater specificity (74.8%) and a positive predictive value (PPV) of 0.28 (see Table 3).

We evaluated the prediction algorithm developed for women in the validation database using the shrunk regression coefficients. The shrinkage factor was 0.95 (see Table 2). The *C*-index and effect size (Hedges' *g*) were 0.66 (95% CI = [0.64, 0.68]) and 0.64 (95% CI = [0.57, 0.70]), respectively. This suggests that the algorithm did not work as well in the validation database. We visually compared the predicted versus the observed risk of common mental disorders by decile risk groups (Figure 2 in Supplemental File

1). In general, the predicted risk did not agree with the observed risk, except in the highest decile groups 8−10, where accuracy improved.

Men. The final model for men included 13 predictors (Table 2). The *C*-index and effect size (Hedges' *g*) were 0.76 (95% CI = [0.74, 0.77]) and 1.06 (95% CI = [1.01, 1.12]), respectively. In this case, the calibration showed a good fit, with almost perfect agreement in all groups except decile 9 (Figure 3 in Supplemental File 1). The predicted probability with the greatest Youden's *J* statistic was 11.99%, which again had limited sensitivity (63.87%), good specificity (78.32%), and a PPV of 0.26 (see Table 3).

The shrinkage factor in the men's model was 0.89 (see Table 2). We applied the shrunken coefficients in the validation database: the *C*-index was 0.73 (95% CI = [0.72, 0.75]) and Hedges' *g* was 1.04 (95% CI = [0.97, 1.11]). The calibration plot showed limited agreement between the observed and the predicted value in the lowest decile groups, but good agreement in the highest decile groups (Figure 4 in Supplemental File 1).

Table 2. Prediction model.

	B coefficient	p > t	Shrunk B coefficient
Women			
Age	−0.0223	0	−0.0212
Being from an Aboriginal and/or Torres Strait Islander Background	0.8463	0.071	0.8040
Satisfaction with the hours worked	−0.0829	0.024	−0.0788
History of common mental disorders	1.2344	<0.0001	1.1727
Satisfaction with employment opportunities	−0.1118	0.005	−0.1062
Satisfaction with health	−0.1487	0.003	−0.1413
Agree with the statement 'I am pushed around'	0.3463	0.133	0.3290
Agree with the statement 'I don't have anyone that I can confide in'	0.8236	<0.0001	0.7824
Constant	1.1809	0.028	1.1197
Men			
Age	−0.0217	0.002	−0.0194
Being a carer	1.0361	0.023	0.9252
Agree with the statement 'I have lot of freedom to decide at work'	−0.3254	0.107	−0.2906
Doing physical activity 3 or more times per week	−0.3904	0.018	−0.3486
Drinking between 1 and 4 drinks per week (ref. none)	0.2535	0.207	0.2264
Drinking more than 4 drinks per week (ref. none)	0.5658	0.026	0.5053
History of common mental disorders	1.2098	<0.0001	1.0804
Satisfaction with health	−0.1516	0.011	−0.1354
Satisfaction with the neighbourhood	−0.0920	0.083	−0.0821
Neutral satisfaction with partner (ref. satisfied with partner)	0.8426	0.004	0.7525
Dissatisfied with partner (ref. satisfied with partner)	0.8717	0.098	0.7785
Not having a partner (ref. satisfied with partner)	0.6574	0.087	0.5871
Neutral satisfaction with the way tasks are divided at home (ref. satisfied)	0.1261	0.633	0.1126
Dissatisfied with the way tasks are divided at home (ref. satisfied)	0.5083	0.294	0.4539
Not having to divide tasks at home (ref. satisfied)	−0.9950	0.006	−0.8885
Agree with the statement 'I don't have anyone that I can confide in'	0.7672	<0.0001	0.6851
Agree with the statement 'I am pushed around'	0.6999	0.004	0.6250
English as a second language	0.5192	0.037	0.4636
Constant	0.3167	0.643	0.2828

Scenarios. Table 4 shows some examples of the varieties of participants scoring at increasing levels of predicted probability of common mental disorders on the score algorithm and the impact on predicted risk by changing mutable factors.

Discussion

To the best of our knowledge this is the first prediction algorithm for the onset of common mental disorders developed specifically for a working population. The prediction algorithm developed for men had an adequate performance in both the training and the validation samples. In contrast, the model developed in women, although showing relatively fair indices in the training database, did not work well in the validation database. Despite this, the *C*-statistic is similar to well-known, widely used and, in some cases, health service mandated cardiovascular algorithms such as Framingham and SCORE, which have reported *C*-statistics from validation databases ranging from 0.57 to 0.91 (Damen et al., 2016). The PPVs at risks with the best Youden's indices (those set at risk values of 0.14 for women and 0.11 for men) were also equal to the PPVs derived for the cardiovascular risks of 20% or more using the

Table 3. Specificity and sensitivity values of different cut-offs.

Predicted probability	Specificity	Sensitivity	Youden's index
Women			
≥0.069052	0.292196	0.900215	0.192411
≥0.142609	**0.748321**	**0.627121**	**0.375442**
≥0.180336	0.819621	0.500673	0.320294
≥0.21112	0.860159	0.41826	0.278419
≥0.30731	0.938647	0.236332	0.174979
Men			
≥0.035166	0.131949	0.962736	0.094685
≥0.119759	**0.783218**	**0.638651**	**0.421869**
≥0.15029	0.852032	0.514922	0.366954
≥0.20205	0.915685	0.391031	0.306716
≥0.300572	0.96472	0.236651	0.201371

Bold indicates the value with the best Youden's index.

Framingham Index for cardiovascular disease (Artigao-Rodenas et al., 2013).

While the algorithms classified those with higher risk well, which is the main aim of such models, they had some problems adequately classifying those with lower risk. In other words, the algorithms had better specificity than sensitivity. This may be useful for ruling in, as a positive result when the specificity is high signifies a high probability of developing a common mental disorder (for instance, scores >0.20, where specificity is more than 0.85 in women and 0.90 in men). However, following this same cut-off, a score <0.20 would not rule out the possibility that the person would develop a common mental disorder. In any case, as these algorithms are intended to be used in non-clinical settings, such as the workplace, we have preferred to minimise the false-positives. In practical terms, however, the use of such algorithms allows three approaches. By providing an individual with their modifiable risks, this may empower them to promote behavioural change. If the overall risk level is high, this person may be more motivated if the risk is communicated effectively as the high specificity of our algorithm suggests it is highly probable that the person would develop a common mental disorder. Identification of higher risk individuals could enable targeting of individual/group secondary prevention interventions, for which a range of interventions, such as cognitive behavioural therapy–based stress management, appear to be effective (Joyce et al., 2016). Given there will false-negatives and false-positives, particularly in females, organisations can use the estimations of the proportion of their workforce with a risk profile in aggregated ways to determine how to allocate resources to more universal or targeted approaches.

We found that age, the presence of a common mental disorder in the past 2 years, satisfaction with health, not having someone to confide in and feeling 'pushed around' in life were common predictors of the onset common mental disorder among both women and men. Health and social support have been consistently linked with better mental health in numerous studies. The strong predictor of being 'pushed around' would seem to be a good proxy for either experiences of discrimination (a strong risk factor in Predict (King et al., 2008)) or for the degree of autonomy one has in life generally, the latter linked to aspects often termed 'resilience' at work (Weinstein and Ryan, 2011). However, men and women differed in several factors: health-related habits such as higher alcohol intake (risk) and greater frequency of physical activity (protective) were significant for men but not for women. Similarly, and somewhat surprisingly, being a carer of someone, satisfaction with one's partner and the way that tasks are divided at home were factors associated with the onset of common mental disorders in men, but not in women. These factors seem to be contrary to expectations but may relate to changes in traditional/stereotypical roles. Alternatively, it may be that such situations are still expected for women, but not for men, making their presence more of a risk among men as they impose a role beyond the archetypal norms. Different psychosocial work risks also affected men and women: While for men having freedom to decide what to do at work, but not, for example, hours worked, appeared an important factor, women dissatisfied with the number of hours worked and the employment opportunities were at greater risk. Finally, English as a second language was a risk factor in men. Although being from an ATSI background was only

Table 4. Examples.

	Women	Men
Low risk	55 years old Not from an ATSI background Satisfaction with the hours worked (0–10): 8 Without past history of common mental disorders Satisfaction with employment opportunities (0–10): 8 Satisfaction with health (0–10): 8 Does not feel she is pushed around Has someone that she can confide in *Risk: 0.06 (6%)*	40 years old Does not actively care for someone Has freedom to decide at job Does 3 or more times physical activity per week Drinks 2 times per week Without past history of common mental disorders Satisfaction with health (0–10): 8 Satisfaction with neighbourhood (0–10): 8 Satisfied with his partner Satisfied with the way tasks are divided at home Has someone to confide on Does not feel pushed around English as a first language *Risk: 0.049 (4.9%)*
Moderate risk	40 years old Not from an ATSI background Satisfaction with the hours worked (0–10): 5 Without past history of common mental disorders Satisfaction with employment opportunities (0–10): 8 Satisfaction with health (0–10): 5 Does not feel she is pushed around Has someone that she can confide in *Risk: 0.157 (15.7%)* *She starts to take care of her health and increases her satisfaction to health to 8. In addition, she talks with her boss about her situation at job and increases her satisfaction to 7* *New risk: 0.094 (9.4%)*	30 years old Does not actively care for someone Does not have freedom to decide at job Does not do 3 or more times physical activity per week Does not drink Without past history of common mental disorders Satisfaction with health (0–10): 7 Satisfaction with neighbourhood (0–10): 8 Does not have a partner Lives alone (does not have to divide tasks at home) Has someone to confide on Does not feel pushed around English as a second language *Risk: 0.125 (12.5%)* *He increases physical activity, feels better and also increases his satisfaction with health (8)* *New risk: 0.076 (7.6%)*
High risk	25 years old From an ATSI background Satisfaction with the hours worked (0–10): 5 Without a past history of common mental disorders Satisfaction with employment opportunities (0–10): 5 Satisfaction with health (0–10): 5 Does not feel she is pushed around Has anyone that she can confide in *Risk: 0.441 (44.1%)* *After a problem-solving process, she decides to change her job, which she does not like. She finds another one where she feels better (satisfaction with hours worked changes to 7 and opportunities to 8). In addition, she starts to do more physical activity, so her satisfaction with health improves (to 7)* *New risk: 0.27 (27%)* *Unfortunately, being from an ATSI background dramatically increases the risk*	55 years old Does actively cares someone Does have freedom to decide at job Does not do 3 or more times physical activity per week Drink >4 times per week With a past history of common mental disorders Satisfaction with health (0–10): 5 Satisfaction with neighbourhood (0–10): 8 Satisfied with his partner Dissatisfied with the way tasks are divided at home Has someone to confide on Does not feel pushed around English as a first language *Risk: 0.63 (63%)* *He decreases the number of drinks per week to 2 and starts to do physical activity. That makes him feel better and also increases his satisfaction with health (to 6). In addition, he negotiates new ways to divide the tasks, so he feels neutral about that.* *New risk: 0.331 (33.1%)* *As he has a past history of common mental disorders, his basal risk is quite high*

ATSI: Aboriginal and/or Torres Strait Islander.

found to be a risk factor for women, we did not have enough variability in the sample to introduce this variable in the equation, so we are not sure about the overall impact. People from an ATSI background report a higher prevalence of common mental disorder than other Australians, which can be a reflection of the cultural disruption experimented by Aboriginal societies (Parker, 2010), so this factor needs to be addressed in future projects.

The risk factors identified in this study of working individuals are somewhat different from those previously identified in studies of other adult populations. In the algorithms developed from the Predict study in European primary care (Bellon et al., 2011; King et al., 2008; Moreno-Peral et al., 2014) and from the Canadian National Population Health Service (NPHS) (Wang et al., 2014), age, past history of common mental disorders and satisfaction with health were also important. In the Predict study, education was an important predictor for both the onset of depression and anxiety, while in our study it was not. Wang et al., however, did not include education in their model. Surprisingly, neither our study nor the Predict study found life events to be an important enough predictor of common mental disorders to be included in the final algorithms. However, in the NPHS, a major financial crisis and changing a job for a worse one were important predictors for women, while being physically attacked and a partner having an unwanted pregnancy were for men. The major differences between this study and previous risk algorithms developed were our ability to assess the impact of a wide range of job-related predictors, which have been consistently identified as significant risks for common mental disorder (Harvey et al., 2017; Stansfeld and Candy, 2006) and that our sample were all employed. This may account for some differences, as low education and negative life events are more common in those not working. Having different risk factors appears not to be the exception, but the rule, when developing prediction models even in other areas such as cardiovascular health and diabetes (Damen et al., 2016).

Limitations

This study has several limitations. First, we performed a secondary analysis of a study that did not aim to develop a prediction algorithm for the onset of common mental disorders, so not all the predictors suggested by the literature were available. For example, we did not have good measures of discrimination or personality factors, although some measures, as identified above, might be seen as proxies. In addition, our dependent measure, common mental disorder, was derived from the SF-36 MCS, which is not a disease-specific assessment or a screener specifically developed to screen for mental disorders. However, it has shown good psychometric properties in ascertaining common mental disorders in the Australian population (Gill et al., 2007) and may be more applicable to population health than clinically

derived measures. Conversely, secondary analysis of available databases is an efficient way of generating new evidence and alleviates the potential response bias inherent in specific studies. Second, data about the potential predictors were collected by means of self-report and therefore are highly subjective. Nevertheless, how people feel and perceive reality is core to the development of common mental disorders, such as anxiety and depression, and has, in the case of work-related factors, been shown to be more relevant than objective measures (Glozier et al., 2010). Third, there were a number of participants lost to follow-up. Although we have imputed all the missing values for any predictor, we elected not to impute the dependent variable. The relative poor performance of the algorithm in the lowest and medium risk groups in the external validation may be related to the fact that in our study, those lost to follow-up were, indeed, those who were healthier at baseline.

Implications

The predicted probability of one person developing common mental disorders can be calculated from the equations in Table 2, using the constant and the shrunk coefficients. However, this is complicated to do unless automated. One option for implementation would be to embed the risk algorithm for the onset of common mental disorders in employed people, together with recommendations according to the different level and profile of risk, into a personal eHealth or decision support tool platforms, much as with clinical desktop tools for cardiovascular disease risk. This could empower people to implement behavioural change to reduce their risk, by addressing modifiable factors, without fear of being stigmatised or need for help seeking, similar to some on-line self-help strategies to prevent depression (Buntrock et al., 2016; Sander et al., 2016). A decision support tool could be used by occupational health or organisational professionals in order to know more about the level and profile of risk for common mental disorders of their clients, giving them the opportunity to target selective preventive strategies. Indeed, a prevention strategy based on a similar risk algorithm has recently shown to be effective to reduce the incidence of major depression episodes in primary care attendees (Bellon et al., 2016). It is hoped that the risk algorithm presented in this paper could be used as the basis for similar interventions in the workplace. This approach based on the individual risk, however, needs to be combined with universal measures that target the social determinants of mental health, such as workplace culture. De-identified aggregate risk data may also be useful to employers in considering whether and how they address the challenges of enabling a mentally healthy workplace.

These are the first prediction algorithms for the onset of common mental disorders in Australia and the first, worldwide, specific for workers. Further studies are needed to refine these algorithms (especially for women) and to

evaluate their usefulness for workplace-based prevention initiatives. These potential different uses, by workers seeking to take more control over their risk, occupational health professionals, or organisations, raise ethical and privacy challenges that we are addressing in current implementation studies.

References

ABS (1997) *4399.0 – National Health Survey: SF-36 Population Norms, Australia, 1995*. Canberra, ACT, Australia: Australian Bureau of Statistics.

Andrews G, Issakidis C, Sanderson K, et al. (2004) Utilising survey data to inform public policy: Comparison of the cost-effectiveness of treatment of ten mental disorders. *British Journal of Psychiatry* 184: 526–533.

Artigao-Rodenas LM, Carbayo-Herencia JA, Divison-Garrote JA, et al. (2013) Framingham risk score for prediction of cardiovascular diseases: A population-based study from southern Europe. *PLoS ONE* 8: e73529.

Beekman AT, Smit F, Stek ML, et al. (2010) Preventing depression in high-risk groups. *Current Opinion in Psychiatry* 23: 8–11.

Bellon JA, Conejo-Ceron S, Moreno-Peral P, et al. (2016) Intervention to prevent major depression in primary care: A cluster randomized trial. *Annals of Internal Medicine* 164: 656–665.

Bellon JA, De Dios Luna J, King M, et al. (2011) Predicting the onset of major depression in primary care: International validation of a risk prediction algorithm from Spain. *Psychological Medicine* 41: 2075–2088.

Buntrock C, Ebert DD, Lehr D, et al. (2016) Effect of a web-based guided self-help intervention for prevention of major depression in adults with subthreshold depression: A randomized clinical trial. *Journal of the American Medical Association* 315: 1854–1863.

Butterworth P and Crosier T (2004) The validity of the SF-36 in an Australian National Household Survey: Demonstrating the applicability of the Household Income and Labour Dynamics in Australia (HILDA) Survey to examination of health inequalities. *BMC Public Health* 4: 44.

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

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

Cuijpers P, Beekman AT and Reynolds CF (2012) Preventing depression: A global priority. *Journal of the American Medical Association* 307: 1033–1034.

Czabala C, Charzynska K and Mroziak B (2011) Psychosocial interventions in workplace mental health promotion: An overview. *Health Promotion International* 26(Suppl. 1): i70–84.

Damen JA, Hooft L, Schuit E, et al. (2016) Prediction models for cardiovascular disease risk in the general population: Systematic review. *British Medical Journal* 353: i2416.

Dapp U, Anders JAM, Von Renteln-Kruse W, et al. (2011) A randomized trial of effects of health risk appraisal combined with group sessions or home visits on preventive behaviors in older adults. *Journals of Gerontology Series A: Biological Sciences and Medical Sciences* 66: 591–598.

Gill SC, Butterworth P, Rodgers B, et al. (2007) Validity of the mental health component scale of the 12-item Short-Form Health Survey (MCS-12) as measure of common mental disorders in the general population. *Psychiatry Research* 152: 63–71.

Glozier N, Overland S, Greenberg N, et al. (2010) Work and psychiatric disorder: An evidence based approach. In: Baxter PJ, Aw T-C, Cockcroft A, et al. (ed.) *Hunters Diseases of Occupations*. London: Edward Arnold Publishers, pp. 833–857.

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

Harvey SB, Henderson M, Lelliott P, et al. (2009) Mental health and employment: Much work still to be done. *British Journal of Psychiatry* 194: 201–203.

Harvey SB, Modini M, Joyce S, et al. (2017) Can work make you mentally ill? A systematic meta-review of work-related risk factors for common mental health problems. *Occupational and Environmental Medicine* 74: 301–310.

Henderson M, Harvey SB, Overland S, et al. (2011) Work and common psychiatric disorders. *Journal of the Royal Society of Medicine* 104: 198–207.

Jacka FN, Reavley NJ, Jorm AF, et al. (2013) Prevention of common mental disorders: What can we learn from those who have gone before and where do we go next? *Australian and New Zealand Journal of Psychiatry* 47: 920–929.

Joyce S, Modini M, Christensen H, et al. (2016) Workplace interventions for common mental disorders: A systematic meta-review. *Psychological Medicine* 46: 683–697.

Kessler RC, Demler O, Frank RG, et al. (2005) Prevalence and treatment of mental disorders, 1990 to 2003. *New England Journal of Medicine* 352: 2515–2523.

King M, Bottomley C, Bellon-Saameno JA, et al. (2011) An international risk prediction algorithm for the onset of generalized anxiety and panic syndromes in general practice attendees: PredictA. *Psychological Medicine* 41: 1625–1639.

King M, Walker C, Levy G, et al. (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.

Kristjansdottir J, Olsson GI, Sundelin C, et al. (2011) Could SF-36 be used as a screening instrument for depression in a Swedish youth population? *Scandinavian Journal of Caring Sciences* 25: 262–268.

McCallum J (1995) The SF-36 in an Australian sample: Validating a new, generic health status measure. *Australian Journal of Public Health* 19: 160–166.

McHorney CA, Ware JE and Raczek AE (1993) The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of

validity in measuring physical and mental health constructs. *Medical Care* 31: 247–263.

McHorney CA, Ware JE, Lu JF, et al. (1994) The MOS 36-item Short-Form Health Survey (SF-36): III. Tests of data quality, scaling assumptions, and reliability across diverse patient groups. *Medical Care* 32: 40–66.

Maron DJ, Forbes BL, Groves JR, et al. (2008) Health-risk appraisal with or without disease management for worksite cardiovascular risk reduction. *Journal of Cardiovascular Nursing* 23: 513–518.

Moons KG, Altman DG, Reitsma JB, et al. (2015) Transparent reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): Explanation and elaboration. *Annals of Internal Medicine* 162: W1–73.

Moreno-Peral P, Luna JDE, Marston L, et al. (2014) Predicting the onset of anxiety syndromes at 12 months in primary care attendees. The PredictA-Spain study. *PLoS ONE* 9: e106370.

Mykletun A and Harvey SB (2012) Prevention of mental disorders: A new era for workplace mental health. *Occupational and Environmental Medicine* 69: 868–869.

Nagykaldi ZJ, Voncken-Brewster V, Aspy CB, et al. (2013) Novel computerized health risk appraisal may improve longitudinal health and wellness in primary care. *Applied Clinical Informatics* 4: 75–87.

Parker R (2010) Australia's aboriginal population and mental health. *Journal of Nervous and Mental Disease* 198: 3–7.

Robert A and Karasek J (1979) Job demands, job decision latitude, and mental strain: Implications for job redesign. *Administrative Science Quarterly* 24: 285–308.

Rubin D (1987) *Multiple Imputations for Nonresponse in Surveys*. New York: John Wiley & Sons.

Sander L, Rausch L and Baumeister H (2016) Effectiveness of internet-based interventions for the prevention of mental disorders: A systematic review and meta-analysis. *JMIR Mental Health* 3: e38.

Stansfeld S and Candy B (2006) Psychosocial work environment and mental health – A meta-analytic review. *Scandinavian Journal of Work Environment & Health* 32: 443–462.

Summerfield M, Freidin S, Hahn M, et al. (2015) *HILDA User Manual – Release 14*. Melbourne, VIC, Australia: Melbourne Institute of Applied Economic and Social Research, University of Melbourne.

Tripepi G, Jager KJ, Dekker FW, et al. (2010a) Statistical methods for the assessment of prognostic biomarkers (Part I): Discrimination. *Nephrology Dialysis Transplantation* 25: 1399–1401.

Tripepi G, Jager KJ, Dekker FW, et al. (2010b) Statistical methods for the assessment of prognostic biomarkers (Part II): Calibration and re-classification. *Nephrology Dialysis Transplantation* 25: 1402–1405.

Vilagut G, Forero CG, Pinto-Meza A, et al. (2013) The mental component of the short-form 12 health survey (SF-12) as a measure of depressive disorders in the general population: Results with three alternative scoring methods. *Value in Health* 16: 564–573.

Wang J, Sareen J, Patten S, et al. (2014) A prediction algorithm for first onset of major depression in the general population: Development and validation. *Journal of Epidemiology and Community Health* 68: 418–424.

Ware JE and Sherbourne CD (1992) The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Medical Care* 30: 473–483.

Watson N and Wooden M (2010) The HILDA survey: Progress and future developments. *Australian Economic Review* 43: 326–336.

Watson N and Wooden M (2012) The HILDA survey: A case study in the design and development of a successful household panel study. *Longitudinal and Life Course Studies* 3: 369–381.

Weinstein N and Ryan RM (2011) A self-determination theory approach to understanding stress incursion and responses. *Stress and Health* 27: 4–17.

Windsor TD, Rodgers B, Butterworth P, et al. (2006) Measuring physical and mental health using the SF-12: Implications for community surveys of mental health. *Australian and New Zealand Journal of Psychiatry* 40: 797–803.