BMJ Open

Development and Validation of a Nomogram for Predicting the Risk of Mental Health Problems of Factory Workers and Miners

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-057102
Article Type:	Original research
Date Submitted by the Author:	20-Sep-2021

Complete List of Authors:	Liu, Tao; Xinjiang Medical University, School of Public Health Lu, Yaoqin; Xinjiang Medical University, School of Public Health; Urumqi Center for Disease Control and Prevention Liu, Qi; Xinjiang Medical University, School of Public Health Yan, Huan; Xinjiang Medical University, Department of Nutrition and Food Hygiene; Xinjiang Autonomous Academy of Instrumental Analysis

Keywords:	MENTAL HEALTH, PREVENTIVE MEDICINE, PUBLIC HEALTH

Development and Validation of a Nomogram for Predicting the Risk of Mental Health Problems of Factory Workers and Miners

Yaoqin Lu,[1,2] Qi Liu,[1] Huan Yan[3,4], and Tao Liu[1]

[1]Department of Toxicology, School of Public Health, Xinjiang Medical University, Urumqi 830011, Xinjiang, China.
[2]Urumqi Center for Disease Control and Prevention, Urumqi 830026, Xinjiang, China.
[3]Department of Nutrition and Food Hygiene, College of Public Health, Xinjiang Medical University, Urumqi, Xinjiang, China.
[4]Xinjiang Engineering Technology Research Center for Green Processing of Nature Product Center, Xinjiang Autonomous Academy of Instrumental Analysis, Urumqi, Xinjiang, China.

Correspondence should be addressed to Tao Liu; xjmult@163.com

Abstract

Objective A nomogram for predicting the risk of mental health problems was established in a population of factory workers and miners, in order to quickly calculate the probability of a worker suffering from mental health problems.

Methods A cross-sectional survey of 7,500 factory workers and miners in Urumqi was conducted by means of an electronic questionnaire using cluster sampling method. Participants were randomly assigned to the training group (70%) and the validation group (30%). Questionnaire-based survey was conducted to collect information. A least absolute shrinkage and selection operator (LASSO) regression model was used to screen the predictors related to the risk of mental health problems of the training group. Multivariate logistic regression analysis was applied to construct the prediction model. Calibration plots and receiver operating characteristic-derived area under the curve (AUC) were used for model validation. Decision curve analysis (DCA) was applied to calculate the net benefit of the screening model.

Results A total of 7,118 participants met the inclusion criteria and the data were randomly divided into a training group (n=4,955) and a validation group (n=2,163) in a ratio of 3:1. A total of 23 characteristics were included in this study and LASSO regression selected 12 characteristics such as education, professional title, age, CMBI, ERI, asbestos dust, hypertension, diabetes, working hours per day, working years, marital status, and work schedule as predictors for the construction of the nomogram. In the validation group the Brier score was 0.176, the calibration slope was 0.970 and the calibration curve of nomogram showed a good fit. The AUC of training group and verification group were 0.785 and 0.784 respectively.

Conclusion The nomogram combining these 12 characteristics can be used to predict the risk of suffering mental health problems, providing a useful tool for quickly and accurately screening the risk of mental health problems.

Key words Mental health; Predictor; Nomogram; Risk; Factory workers and miners

Strengths and limitations of this study

1. To our knowledge, this was a new model to develop and assess the likelihood of mental health problems in a group of factory workers and miners.

2. This study provided a viable and easy-to-apply tool including factors that were closely related to factory workers and miners for identifying workers at risk of mental health problems.

3. The results of this study showed good agreement and good discrimination between predictions and observations.

4. We had considered many influential factors including demographics, job burnout, occupational stress and occupational exposure factors, but we were still not certain whether all possible influences were covered.

5. While the robustness of our nomogram was extensively validated internally in the same population, external validation was lacking for other populations in other regions and countries.

1. Introduction

The World Health Organization (WHO) defines health as a state of complete physical, mental and social well-being and not merely the absence of disease or weakness [1]. Obviously, health is an organic unity of physical and mental well-being. People with good mental health are the precondition for the normal operation of our society. However, with the acceleration of people's pace of life, people are facing an increasing risk of poor health, which has become a global public health problem [2]. Mental health problems can not only take a toll on physical health such as increasing the risk of communicable and non-communicable diseases and even causing unintentional or intentional harm to others [3], but can also have a negative impact on the economy. For example, mental health disorders represent a growing part of the global burden of disease [4], with statistics showing that nearly one billion people worldwide currently suffer from a mental disorder, and mental illness is ranked as one of the leading causes of the global burden of disease [5]. Moreover, one study has estimated that due to the impact of mental illness, the global economy loses US $1 trillion every year [6].

As researchers around the world have delved into the field of mental health, factors such as gender, income levels, environment and education have been found to be associated with people's mental health problems [7-10]. Moreover, employment is also strongly associated with quality of life, higher self-esteem and fewer psychiatric symptoms [11]. In addition, in the context of the global challenges of climate change, an increasing number of scholars have been examining the epidemiological links between mental health and environmental factors. Some studies have suggested that mental health may be influenced by ambient temperature, and an association has been found between environmental pollutants, particularly fine particulate matter, and mental health problems [12]. A relevant study shows that with short-term exposure to ambient air pollution is associated with increased emergency room visits due to depression or suicide attempts [13]. Furthermore, other factors associated with mental health include sleep, diabetes, coronary artery disease and cardiovascular disease [14-15]. It is worth noting that job burnout and occupational stress are closely linked to mental health. Job burnout is an exhaustion state of physical and psychological that

often occurs in the work environment, and has a high correlation with depression. A large study of physicians found that of the 10.3% who met criteria for a major depressive episode, 50.7% were also affected by symptoms of burnout (OR 2.99) and indicated that worsening depression leads to a higher likelihood of burnout symptoms [16]. Occupational stress refers to a work environment where non-reciprocity of effort and reward may lead to strong negative emotions and distress. Related research has shown that the combination of high effort and low reward and over-commitment increases the risk of mental health problems such as depression [17]. Apparently, it is necessary to include the CMBI and ERI in this study to predict the risk of mental health problems among factory workers and miners. In addition, the CMBI and ERI questionnaires consist of 15 and 23 items respectively, which are a smaller number of items compared to the 90 items of the Symptom Check list-90 (SCL-90) questionnaire. However, there are few studies that include these influences in a more comprehensive way in the practice of detecting mental health. Therefore, more accurate identification of mental health problems in populations requires a questionnaire that include a wider range of factors affecting factory workers and miners' mental health problems.

Factory workers and miners are a special group of workers with a relatively low overall level of education and are highly prone to suffering from mental health problems due to limited social support, excessive workload and irregular lifestyles, as well as occupational hazards such as noise and coal dust that they inevitably need to face in their working environment [18-19]. China has the world's largest group of factory workers and miners, about 6 million [20], who are regularly involved in occupational hazards. Mental health problems which need to require a long process are known to be a syndrome caused by chronic stress. Factory workers and miners, represented by those engaged in coal mining, have a mental burden rating of 8.3, one of the highest mental burdens among 150 occupations [21]. This explains the high level of mental health problems among mine workers in previous studies, making the identification and treatment of mental health problems even more important. Therefore, it is essential to provide a viable and easy-to-apply tool for identifying workers at risk of mental health problems and thus for timely interventions.

The Symptom Checklist-90 (SCL-90), which is widely used in psychiatric outpatient examinations, has a high degree of validity in evaluating various mental health surveys [22-23]. However, this questionnaire has 90 items, and in practice of our previous studies, it has been found to be complex and time-consuming to complete, requiring a high degree of patience and cooperation from the respondents. In addition, the questionnaire is slightly less targeted to the group of factory workers and miners, and lacks entries relating to the particular working environment of factory workers and miners. Nowadays, there is growing recognition that mental health plays an important role in achieving global development goals and the WHO has included mental health in the Sustainable Development Goals. However, there are currently no relevant studies that use objective indicators to form a nomogram for predicting mental health. Therefore, the aim of our study is to develop and validate an easy-to-use nomogram that combines objective information on the demographics, job burnout, occupational stress and occupational hazards to comprehensively and accurately predict the prevalence of mental health problems among factory workers and miners.

2. Materials and Methods

2.1. Participants

Participants in this cross-sectional survey were workers from factories and mining enterprises in the Urumqi region, who were recruited using a whole-group sampling method. A total of 7,500 participants in the Urumqi were surveyed from January to May 2019, covering all districts and counties in the Urumqi region, including Tianshan District, Shaibak District, Xinshi District, Shuimogou District, Toutunhe District, Dabancheng District, Middong District and Urumqi County.

The exclusion criteria were the following: (I) factory workers and miners in non-Urumqi area, (II) working history of factories and mining enterprises less than 1 year, (III) a confirmed diagnosis of a mental health problem and a history of treatment and use of psychotropic medication. Questionnaires with missing data were also excluded from the analysis based on discussion and agreement among the subject members. A total of 7,500 questionnaires were distributed and 7,315 questionnaires were returned, representing a return rate of 97.5%. After checking the validity and integrity of the questionnaires, 7,118 questionnaires were confirmed as valid, with an effective rate of 97.3%. All participants understood the purpose of the study and voluntarily participated in the study.

2.2. Research Methods

2.2.1. Assessment of mental health

The SCL-90, designed by Derogatis and his colleagues, was widely used in the mental health field [24], which contains 90 items across nine dimensions: somatization, obsessive-compulsive symptoms, interpersonal sensitivity, depression, anxiety, hostility, horror, bigotry and mental illness. The SCL-90 has been used extensively in previous studies and has relatively high reliability and validity [25]. The questionnaire uses a Likert 5-point scale, with a score of 0 point indicating none and 4 points indicating severe. A total score above 160, a score above 2 on any item, or a positive item above 43 indicates the presence of a psychological abnormality. In this survey, Cronbach α was 0.99, the half-reliability coefficient was 0.98, and the KMO was 0.994.

2.2.2. Assessment of occupational stress

This survey evaluated occupational stress in factory workers and miners through the Effort–Reward Imbalance (ERI) model developed by Siegrist [26]. The ERI scale consists of three subscales: effort (E, 6 items), reward (R, 11 items) and over commitment (6 items), for a total of 23 items. A Likert 5-level scoring method (1, "highly disagree" to 5, "highly agree") is used to grade the items in the questionnaire with the same weight for each item. The effort–return index ERI = E/R×C, where C is the adjustment coefficient, and the value is 6/11. ERI values greater than 1, equal to 1, and less than 1 correspond to high pay–low return, pay–return balance, and low pay–high return, respectively. Moreover, the higher the ERI value, the greater the occupational stress [27]. In this survey, Cronbach α was 0.94, the half-reliability coefficient was 0.93 and the KMO was 0.956.

2.2.3. Assessment of job burnout

In this survey, the Chinese Maslach Burnout Inventory (CMBI) revised by Li et al. was used to assess job burnout, which has good reliability and validity [28]. CMBI is composed of 15 items in three dimensions: emotional exhaustion (5 items), depersonalization (5 items) and reduced personal accomplishment (5 items). The score for each item ranges from 1 to 7, with 1 point indicating complete compliance and 7 points indicating complete non-compliance. According to the critical value (emotional exhaustion ≥ 25, depersonalization ≥ 11, personal achievement reduction ≥ 16), the levels of occupational burnout are divided into none (all aspects are below the critical value), mild (any one aspect is equal to or above the critical value), moderate (any two aspects are equal to or higher than the critical values), and severe (three aspects are equal to or higher than the critical values) [29]. In this survey, Cronbach α was 0.89, the half-reliability coefficient was 0.86 and the KMO was 0.919.

2.2.4. Candidate predictors

Trained investigators obtained information on demographics, job burnout, occupational stress, mental health and occupational exposure factors through on-site face-to-face collection of an electronic version of the questionnaire. Covariates included in this study: 1) demographic information: gender, ethnicity, education level, professional title, work schedule, marital status, monthly income, age, working years, labor contracts, working hours per day, and working hours per week; 2) occupational exposure factors: coal dust, silica dust, asbestos dust, benzene, lead, noise, and brucellosis; 3) questionnaires: ERI, CMBI; 4) chronic diseases: diabetes, hypertension.

Sex was defined as male or female; ethnicity was defined as Han and other; education level was defined as junior high school and below, high school, junior college or bachelor's degree or above; labor contracts was defined as signed or unsigned; professional title was defined as no, primary, middle or senior; work schedule was defined as day shift, night shift, shift or day and night shifts; marital status was defined as unmarried, married, divorced or widowed; monthly income (yuan) was defined as <3000, 3000~, 4000~, 5000~, 6000~, 7000~ or 8000~; age (years) was defined as <25, 25~, 30~, 35~, 40~ or 45~; working years was defined as ~5, 5~, 10~, 15~, 20~, 25~ or 30~; working hours per day (hours) was defined as ≤7 or > 7; working days per week (days) was defined as ≤5 or >5; exposure to coal dust, silica dust, asbestos dust, benzene, lead, noise, brucellosis were all defined as yes or no; ERI was defined as yes or no; CMBI was defined as none, mild, moderate and severe; hypertension and diabetes were both defined as yes or no.

2.3. Statistical analysis

Categorical variables were described as counts and percentages, and chi square test or Fisher exact test was used to compare categorical variables between groups. 70% of participants were randomly assigned to the training cohort and 30% to the validation cohort. Variables were screened using a least absolute shrinkage and selection operator (LASSO) regression and multivariate logistic regression models were used to estimate risk ratios and corresponding 95% confidence intervals (CI) of risk factors, from which predictive models were constructed. A nomogram for predicting was generated according to the selected characteristics. In addition, forest plot was drawn to visually depict the P-value, OR and 95% CI for the selected validations. Statistically significant predictors were applied to develop a prediction model for the risk of mental health problems among factory workers and miners by introducing all selected factors and analyzing the statistical significance levels of them. We used calibration plots and receiver operating characteristic (ROC) curves to show the calibration and discrimination of our final model. Brier scores for overall performance, calibration slopes were used to assess the predictable accuracy of the model. Decision curve analysis (DCA) was applied to calculate the net benefit of the nomogram. Statistical analysis was performed using the open-source R software Version 3.6.1 (http://www.r-project.org). The significance level (α) set at 0.05.

3. Results

3.1. Participant characteristics

A total of 7,118 participants met the inclusion criteria and the data were randomly divided into a training group (n=4,955) and a validation group (n=2,163). Over half of all participants (65.31%) were male, 57.31% of the population was over 35 years of age and 78.32% of the subjects were married, showing that factory workers and miners are generally older and most of them have spouses. The majority of them had completed high school (83.94%), while a smaller percentage had completed undergraduate education (22.98%), indicating that the group of factory workers and miners as a whole was not well educated. The total number of workers (n, %) exposed to coal dust, silica dust, asbestos dust, benzene, lead, noise and brucellosis in the factory and mining enterprises were 377 (5), 730 (10), 981 (14), 1,981 (28), 373 (5), 4,942 (69) and 121 (2) respectively, with the total number of workers exposed to noise amounting to 4,942, or 69% of the total population surveyed. The demographic, job burnout, occupational stress and occupational exposure factors for the training and validation groups are shown in Table 1. The results showed that there were no significant statistical differences between the two groups of characteristic variables, except for coal dust and CMBI, indicating that the baseline levels were largely consistent between the two groups.

Table 1 Characteristics of the study participants

Variables	Total (n = 7118)	train (n = 4955)	test (n = 2163)	*p*

Sex, n (%)				
Male	4649 (65)	3216 (65)	1433 (66)	0.284
Female	2469 (35)	1739 (35)	730 (34)	
Ethnicity, n (%)				
Han	5762 (81)	3982 (80)	1780 (82)	0.061
Other	1356 (19)	973 (20)	383 (18)	
Education level, n (%)				
Junior high school and below	1143 (16)	804 (16)	339 (16)	0.765
High school	1406 (20)	988 (20)	418 (19)	
Junior college	2933 (41)	2038 (41)	895 (41)	
Bachelor's degree or above	1636 (23)	1125 (23)	511 (24)	
Professional title, n (%)				
None	2854 (40)	1983 (40)	871 (40)	0.923
Primary	1644 (23)	1149 (23)	495 (23)	
Middle	1618 (23)	1133 (23)	485 (22)	
Senior	1002 (14)	690 (14)	312 (14)	
Work schedule, n (%)				
Day shift	3986 (56)	2801 (57)	1185 (55)	0.585
Night shift	270 (4)	187 (4)	83 (4)	
Shift	2058 (29)	1412 (28)	646 (30)	
Day and night shifts	804 (11)	555 (11)	249 (12)	
Marital status, n (%)				
Unmarried	1104 (16)	762 (15)	342 (16)	0.218
Married	5575 (78)	3906 (79)	1669 (77)	
Divorced	390 (5)	255 (5)	135 (6)	
Widowed	49 (1)	32 (1)	17 (1)	
Monthly income (yuan), n (%)				
<3000	1799 (25)	1246 (25)	553 (26)	0.966
3000~	2418 (34)	1682 (34)	736 (34)	
4000~	1600 (22)	1125 (23)	475 (22)	
5000~	752 (11)	520 (10)	232 (11)	
6000~	288 (4)	201 (4)	87 (4)	
7000~	148 (2)	106 (2)	42 (2)	
8000~	113 (2)	75 (2)	38 (2)	
Age (years), n (%)				
<25	431 (6)	297 (6)	134 (6)	0.173
25~	786 (11)	519 (10)	267 (12)	
30~	956 (13)	684 (14)	272 (13)	
35~	866 (12)	617 (12)	249 (12)	

40~	849 (12)	588 (12)	261 (12)	
45~	3230 (45)	2250 (45)	980 (45)	
Working years (years), n (%)				
<5	1170 (16)	794 (16)	376 (17)	0.248
5~	1065 (15)	736 (15)	329 (15)	
10~	997 (14)	721 (15)	276 (13)	
15~	389 (5)	273 (6)	116 (5)	
20~	763 (11)	538 (11)	225 (10)	
25~	1293 (18)	878 (18)	415 (19)	
30~	1441 (20)	1015 (20)	426 (20)	
Labor contracts, n (%)				
Signed	6641 (93)	4624 (93)	2017 (93)	0.955
Unsigned	477 (7)	331 (7)	146 (7)	
Working hours per day (hours), n (%)				
≤7	1161 (16)	814 (16)	347 (16)	0.712
>7	5957 (84)	4141 (84)	1816 (84)	
Working days per week (days), n (%)				
≤5	4442 (62)	3107 (63)	1335 (62)	0.446
>5	2676 (38)	1848 (37)	828 (38)	
Diabetes, n (%)				
Yes	429 (6)	298 (6)	131 (6)	0.988
No	6689 (94)	4657 (94)	2032 (94)	
Hypertension, n (%)				
Yes	1330 (19)	929 (19)	401 (19)	0.861
No	5788 (81)	4026 (81)	1762 (81)	
Coal dust, n (%)				
Yes	377 (5)	244 (5)	133 (6)	0.039
No	6741 (95)	4711 (95)	2030 (94)	
Silica dust, n (%)				
Yes	730 (10)	523 (11)	207 (10)	0.223
No	6388 (90)	4432 (89)	1956 (90)	
Asbestos dust, n (%)				
Yes	981 (14)	691 (14)	290 (13)	0.570
No	6137 (86)	4264 (86)	1873 (87)	
Benzene, n (%)				
Yes	1981 (28)	1360 (27)	621 (29)	0.287
No	5137 (72)	3595 (73)	1542 (71)	
Lead, n (%)				
Yes	373 (5)	246 (5)	127 (6)	0.128

No	6745 (95)	4709 (95)	2036 (94)	
Noise, n (%)				
Yes	4942 (69)	3420 (69)	1522 (70)	0.270
No	2176 (31)	1535 (31)	641 (30)	
Brucellosis, n (%)				
Yes	121 (2)	86 (2)	35 (2)	0.800
No	6997 (98)	4869 (98)	2128 (98)	
ERI, n (%)				
Yes	3147 (44)	2173 (44)	974 (45)	0.372
No	3971 (56)	2782 (56)	1189 (55)	
CMBI, n (%)				
No	959 (13)	674 (14)	285 (13)	0.033
Mild	2667 (37)	1813 (37)	854 (39)	
Moderate	2900 (41)	2031 (41)	869 (40)	
Severe	592 (8)	437 (9)	155 (7)	

3.2. Feature selection

The lambda was smallest at 0.01801 as seen from the lasso results when there were 12 characteristics, which were education, professional title, age, CMBI, ERI, asbestos dust, hypertension, diabetes, working hours per day, working years, marital status, and work schedule based on the results of the questionnaires on demographics, occupational stress, job burnout and occupational exposure factors (Figure 1).

3.3. Results of logistic regression model

The 12 features obtained from the LASSO regression were incorporated into a multivariate logistic regression model and the regression results were shown in Table 2. It was clear from the results that education, professional title, age, CMBI, ERI, asbestos dust, hypertension, diabetes, working hours per day, working years, marital status, and work schedule were independent determinants of risk for mental health problems. In addition, there was no evidence of multicollinearity between the covariates included in the model. The forest plot showed that the selected 12 features all contain items with $P < 0.05$, among which the degree of severe of CMBI (OR, 19.84; 95% CI, 13.88-28.34; $p < 0.001$) had the greatest impact on the risk of mental health problems among factory workers and miners (Figure 2).

Table 2 Predictive factors of risk for mental health problems among factory workers and miners

Variable	β	S.E.	OR(CI95%)	Wald	*P*	VIF
Intercept	-2.33	0.25	0.10(0.06,0.16)	-9.357	0	-
Education level						
Junior school and below VS High school	0.34	0.13	1.41(1.10,1.81)	2.727	0.006**	2.28

Junior school and below VS Junior college	0.44	0.11	1.56(1.24,1.95)	3.850	< 0.001***	2.79
Junior school and below VS Bachelor's degree or above	0.38	0.13	1.46(1.13,1.87)	2.953	0.003**	2.51
Professional title						
None VS Primary	0.15	0.09	1.16(0.97,1.39)	1.582	0.114	1.35
None VS Middle	0.05	0.09	1.05(0.87,1.26)	0.519	0.604	1.34
None VS Senior	0.27	0.11	1.30(1.06,1.61)	2.458	0.014*	1.32
Work schedule						
Day and night shifts VS Day shift	-0.38	0.11	0.69(0.55,0.85)	-3.364	0.001**	2.70
Day and night shifts VS Night shif	0.01	0.20	1.01(0.68,1.49)	0.044	0.965	1.30
Day and night shifts VS Shift	0.01	0.12	1.01(0.81,1.27)	0.107	0.915	2.47
Marital status						
Unmarried VS Married	0.16	0.13	1.18(0.91,1.52)	1.263	0.206	2.29
Unmarried VS Divorced	0.55	0.19	1.73(1.20,2.51)	2.918	0.004**	1.69
Unmarried VS Widowed	0.69	0.43	1.99(0.85,4.64)	1.586	0.113	1.09
Age						
~25 VS 25~	-0.02	0.20	0.98(0.66,1.47)	-0.083	0.934	3.09
~25 VS 30~	-0.02	0.22	0.98(0.64,1.50)	-0.090	0.929	4.79
~25 VS 35~	0.56	0.23	1.76(1.13,2.74)	2.503	0.012*	5.01
~25 VS 40~	0.33	0.23	1.39(0.88,2.21)	1.419	0.156	4.97
~25 VS 45~	0.23	0.22	1.26(0.81,1.95)	1.018	0.308	10.93
Working years						
~5 VS 5~	0.44	0.14	1.55(1.18,2.05)	3.114	0.002**	2.27
~5 VS 10~	0.06	0.15	1.06(0.78,1.43)	0.366	0.714	2.48
~5 VS 15~	0.06	0.20	1.06(0.72,1.56)	0.305	0.760	1.79
~5 VS 20~	0.29	0.18	1.33(0.95,1.88)	1.641	0.101	2.65
~5 VS 25~	0.48	0.17	1.61(1.15,2.25)	2.782	0.005**	3.99
~5 VS 30~	0.20	0.16	1.22(0.89,1.68)	1.239	0.216	3.90
Working hours per day						
≤7 VS >7	-0.50	0.09	0.61(0.50,0.73)	-5.363	< 0.001***	1.15
Diabetes						
No VS Yes	0.43	0.14	1.53(1.16,2.03)	2.974	0.003**	1.05
Hypertension						
No VS Yes	0.52	0.09	1.69(1.42,2.00)	5.885	< 0.001***	1.11
Asbestos dust						
No VS Yes	0.44	0.10	1.55(1.28,1.87)	4.474	< 0.001***	1.03
ERI						

	No VS Yes	0.89	0.07	2.43(2.12,2.79)	12.786	< 0.001***	1.05
CMBI							
	No VS Mild	0.26	0.12	1.30(1.03,1.64)	2.175	0.003**	2.73
	No VS Moderate	1.30	0.11	3.67(2.93,4.59)	11.361	< 0.001***	2.83
	No VS Severe	2.99	0.18	19.84(13.88,28.34)	16.41	< 0.001***	1.44

Note: β is the regression coefficient. "***" indicates P<0.001, "**" indicates P<0.01, "*" indicates P<0.05.

3.4. Development of an individualized prediction model

Based on the results of the multivariate analysis, predictors such as education, professional title, age, CMBI, ERI, asbestos dust, hypertension, diabetes, working hours per day, working years, marital status, and work schedule were included in the nomogram. A model incorporating the above independent predictors was developed and represented as a nomogram in Figure 3. Each variable in nomogram was assigned a score, and the cumulative sum of each 'point' was the 'total score'. The "total score" corresponded to the "predictable likelihood", which was the predicted probability of mental health problems among factory workers and miners as suggested by our design of the nomogram.

As an example of the use of nomogram: a randomly selected sample from the training group, one with no professional title, day shift, no diabetes or hypertension, Junior college, <5 of working years, >7 of working hours per day, married, no exposed to asbestos dust, <25 years of age, no ERI, mild of CMBI, with a calculated total score of 174 and a corresponding risk probability of 8.27% for mental health problems.

3.5 The validation of calibration

Model validation was carried out in the validation group. The prediction accuracy of the model was assessed by two aspects. (1) The Brier score for overall performance, which assessed the difference between observed and predicted values, with values closer to 0 indicating better predictive ability. (2) The calibration slope used for modal calibration, which assessed the agreement between the observed and predicted values, with values closer to 1 indicating better performance. The accuracy measurements for the bias correction were validated by the model with a Brier score of 0.176 and a calibration slope of 0.970, respectively (Figure 4). The prediction accuracy of the model was relatively high.

3.6 The validation of discrimination

ROC was plotted for the training and validation groups, and the AUC of training and the verification groups were 0.785 and 0.784, respectively (Figure 5). The AUC of training and the verification groups were both greater than 0.75, showing a good discrimination.

3.7 Decision Curve Analysis

As shown in the DCA of the risk of mental health problems nomogram in Figure 6, the model for predicting the risk of mental health problems for factory workers and miners in this study was more practically relevant if the threshold probability of patients was >10%.

4. Discussion

In this study, we developed and validated an easy-to-use nomogram model as a new method for predicting the risk of mental health problems among factory workers and miners. To the best of our knowledge, this is the first study to establish an objective indicators nomogram combination model based on mental health survey. Our study included common demographic, job burnout, occupational stress, chronic diseases and occupational exposure factors to distinguish whether the respondents suffer from mental health problems. This nomogram showed good accuracy and discrimination.

LASSO is suitable for analyzing a large number of clinical factors and avoiding over-fitting [30]. In our study, a total of 23 candidate variables were used to construct the nomogram, which were reduced to 12 potential predictor variables by using the LASSO regression method. The nomogram could be a useful tool to better identify patients with mental health problems, as it not only covered comprehensive information, including demographic information, job burnout, occupational stress, chronic diseases and occupational exposure factors closely related to factory workers and miners, but also was simple to operate and easy to use. Therefore, the possibility of early intervention for patients with high-risk mental health problems will be increased by covering multiple information and easy to use nomogram modal, especially for factory workers and miners with poor working conditions, relatively low levels of education and low patience.

Mental health problems were very common in the group of factory workers and miners, and the prevalence of mental health of them was found to be 37.08% in our study. Notably, the CMBI showed the most significant score (score = 100) and the ERI also had a high score (score = 43) in mental health problem incidence risk nomogram, which indicated that both of them were relatively important factors for mental health problems among the group of factory workers and miners. Our finding was consistent with other studies that had shown that occupational stress was a significant predictor of anxiety and was negatively associated with mental health. In addition, there is a high correlation between burnout and depression [31].

In line with previous studies, working years was also an important influential factor in this study. Related study has shown that employment could improve patients' mental health, while unemployment could lead to a deterioration in mental health [32]. In China, workers' working years is an important aspect of employment, and researchers have studied this aspect and found that precarious employment is a source of stress for individuals and predisposes them to mental health problems [33]. In addition, environmental factors were also one of the influential factors of mental health problems in our study. Relevant studies have found that exposure to air pollution is associated with increased suicide risk and depressive symptoms [34]. Hypertension and diabetes were the influential factors in this study. A study has shown that the prevalence of depression in adults with type 1 diabetes (T1D) is approximately three times higher than in the non-diabetic population [35]. Furthermore, there is a recognized association between hyperglycemia and depression, but the underlying biological mechanisms of this association are unclear [36].

Factory workers and miners were inevitably exposed to occupational hazards such as benzene and asbestos dust in their working environment. According to statistics, a total of nearly 2 million workers are exposed to various occupational hazards and over 16 million people worked in toxic and hazardous enterprises, involving more than 30 different types of operations, of which factory workers and miners is the one [37]. Similarly, the occupational hazard asbestos dust was selected as a predictor of risk for mental health problems in this study. Our study found that the work schedules of factory workers and miners were vary and the phenomenon of night shifts was very common, which inevitably affected their normal sleep. Some studies have shown that sleep problem is a risk factor for a variety of mental health and chronic diseases. Lack of sleep or poor sleep quality could lead to abnormalities in the body's self-regulatory functions and disturbances in the circadian rhythm of the biological clock, which in turn could suffer from negative emotions such as anxiety and depression [38]. Professional title and education level were also important influences on mental health issues. In the workplace, generally speaking, the higher the professional title and education level, the higher the status of the worker in the company and the greater the role played in the position. The number of studies on socio-economic status and mental health had increased in recent years. Some of these studies have shown that major depression is higher in the low socio-economic status group [39]. It has also been suggested that education itself is the best indicator of socio-economic status [40]. Marital status was one of the influential factors for mental health problems. Many studies have found an association between mental health and gender, marital status, lifestyle and working conditions, and it has been shown that poor mental health in women is associated with divorce or widowhood [41]. In this study, working more than seven hours a day was a determinant factor on mental health problems, which was consistent with other studies that had shown that long working hours could have a negative impact on employees' mental health and that excessive workloads could increase workers' fatigue, which in turn could lead to anxiety and depression [42].

In China, there are many problems in identifying people with mental health problems due to uneven and imperfect levels of medical development across regions. Some studies have shown that in mainland China, general practitioners, surgeons and primary health care workers often have little or no mental health training, which prevents them from providing basic mental health services [43]. Non-mental health professionals in general hospitals learn about mental illness on their own, rather than learning about it during their formal education[44]. Therefore, this study designed a simple and comprehensive nomogram to address the issue of timely detection and effective interventions for people with mental health problems, so that people at risk of mental health problems could easily calculate their probability of suffering from mental health problems without the help of medical staff. This study has several strengths. First, to our knowledge, this is the first model to develop and assess the likelihood of mental health problems in a group of factory workers and miners. Secondly, the nomogram in this study includes demographic information, job burnout, occupational stress, chronic illnesses, and also occupational exposure factors that are closely related to factory workers and miners, allowing for a more accurate assessment of the risk of morbidity among them, as well as providing a methodological reference for other related studies.

Patient and public involvement

Neither patients nor members of the public had any involvement in the design of this study.

Acknowledgements The authors are grateful to all participants and investigators.

Contributions Y.L., Q.L., and T.L. are responsible for conceptualization; Y.L. is responsible for methodology, software, formal analysis, resources, and visualization; Q.L. and T.L. are responsible for the original draft preparation; Q.L. and H.Y. are responsible for reviewing; Q.L. is responsible for editing; T.L. is responsible for supervision. Yaoqin Lu and Qi Liu contributed equally to this work.

Funding This study was supported by National Natural Science Foundation of China, grant number 81760581 and Public Health and Preventive Medicine, the 13th Five-Year Plan Key Subject of Xinjiang Uygur Autonomous Region.

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval The study was approved by the ethics committee of Urumqi Center for Disease Control and Prevention

Data availability statement The data used to support the findings of this study are available from the corresponding author upon request.

References

[1] WHO Terminology Information System [online glossary] http://www.who.int/health-systems-performance/docs/glossary.html.

[2] Wang Y, Liu X, Qiu J, Wang H, Liu D, Zhao Z, Song M, Song Q, Wang X, Zhou Y, Wang W. Association between ideal cardiovascular health metrics and suboptimal health status in Chinese population. *Sci Rep* 2017;7:14975.

[3] Prince M, Patel V, Saxena S, Maj M, Maselko J, Phillips MR, Rahman A. No health without mental health. *Lancet*. 2007;370:859–77.

[4] Adjaye-Gbewonyo K, Avendano M, Subramanian S.V, Kawachi I. Income inequality and depressive symptoms in South Africa: A longitudinal analysis of the National Income Dynamics Study. *Health*

Place 2016;42:37–46.

[5] Vos T, Barber R.M, Bell B, Bertozzi-Villa A, Biryukov S, Bolliger I, Charlson F, Davis A, Degenhardt L, Dicker D, *et al*. Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990–2013: A systematic analysis for the Global Burden of Disease Study 2013. *Lancet* 2015;386:743–800.

[6] Huang Z, Li T, Xu M. Are there heterogeneous impacts of national income on mental health?.*Int J Environ Res Public Health* 2020;17:7530.

[7] Asadullah M.N, Xiao S, Yeoh E. Subjective well-being in China, 2005–2010: The role of relative income, gender, and location. *China Econ Rev* 2018;48:83–101.

[8] Butterworth P, Rodgers B, Windsor T.D. Financial hardship, socio-economic position and depression: Results from the PATH Through Life Survey. *Soc Sci Med*. 2009;69:229–237.

[9] Zhang X, Zhang X, Chen X. Happiness in the Air: How Does a Dirty Sky Affect Mental Health and Subjective Well-being? *J Environ Econ Manag* 2017;85:81–94.

[10] Wahlbeck K. Public mental health: The time is ripe for translation of evidence into practice. *World Psychiatry*. 2015;14:36–42.

[11] Luciano AE, Drake RE, Bond GR, Becker DR, Carpenter-Song E, Lord S, Swarbrick P and Swanson SJ. IPS Supported employment: a review. *J Vocat Rehabil* 2014;40:1–13.

[12] Jia Z, *et al*. Exposure to ambient air particles increases the risk of mental disorder: findings from a natural experiment in Beijing. *Int J Environ Res Public Health* 2018;15:160.

[13] Szyszkowicz M, Willey J. B, Grafstein E, Rowe B. H, Colman I. Air pollution and emergency department visits for suicide attempts in vancouver, Canada. *Environ Health Insights* 2010;4:79–86.

[14] Michael J. S. International classification of sleep disorders-third edition : highlights and modifications. *Chest* 2014;146:1387–1394.

[15] AbuRuz ME, Al-Dweik G. Depressive symptoms and complications early after acute myocardial infarction: gender differences. *Open Nurs J* 2018;12:205–214.

[16] Wurm W, Vogel K, Holl A, *et al*. Depression-burnout overlap in physicians. *PLoS One* 2016;11:e0149913.

[17] Porru F, Robroek S J W, Bültmann U, *et al*. Mental health among university students: The associations of effort-reward imbalance and overcommitment with psychological distress. *J Affect Disord* 2021;282:953-961.

[18] Johnson AK, Blackstone SR, Skelly A, Simmons W. The relationship between depression, anxiety, and burnout among physician assistant students: a multi-institutional study. *Health Professions Edu* 2020;6:420–427.

[19] Hua D, Kong Y, Li W, Han Q, Zhang X, Zhu LX, Wan SW, Liu Z, Shen Q, Yang J, He HG, Zhu J. Frontline nurses' burnout, anxiety, depression, and fear statuses and their associated factors during the COVID-19 outbreak in Wuhan, China: a large-scale cross-sectional study. *EClinical Med* 2020;24:100424.

[20] Liu FD, Pan ZQ, Liu SL, Chen L, Ma JZ, Yang ML, Wang NP. The estimation of the number of underground coal miners and the annual dose to coal miners in China. *Health Phys* 2007;93:127–132.

[21] Yong X, Gao X, Zhang Z, *et al*. Associations of occupational stress with job burn-out, depression and hypertension in coal miners of xinjiang, china: A cross-sectional study. *BMJ open* 2020;10: e036087.

[22] P. Bech, J. Bille, S. B. Møller, L. C. Hellström, S. D. Østergaard. Psychometric validation of the

Hopkins Symptom Checklist (SCL-90) subscales for depression, anxiety, and interpersonal sensitivity. *J Affect Disord* 2014;160:98–103.

[23] J. Zhang, X. Zhang. Chinese college students' SCL-90 scores and their relations to the college performance. *Asian J Psychiatr* 2013;6:134–140.

[24] Derogatis L, Lipman R.S, Covi L. SCL-90: an outpatient psychiatric rating scale—preliminary report. *Psychopharmacol Bull* 1973;9:13–28.

[25] Crespo-Maraver M, Doval E, Fernã N.J, Gimã Nez-Salinas J, Prat G, Bonet P. Caregiver's health: adaption and validation in a Spanish population of the Experience of Caregiving Inventory (ECI) Gac. *Sanit* 2018;33:348–355.

[26] Y. Li, X. Sun, H. Ge, J. Liu, L. Chen. The status of occupational stress and its influence the quality of life of coppernickel miners in Xinjiang, China. *Int J Environ Res Public Health* 2019;3:353–362.

[27] Siegrist J, Wege N, Pühlhofer F, Wahrendorf M. A short generic measure of work stress in the era of globalization: Effort-reward imbalance. *Int Arch Occup Environ Health* 2009;82:1005–1013.

[28] Zhang Z, Lu Y, Yong X, *et al*. Effects of occupational radiation exposure on job stress and job burnout of medical staff in xinjiang, China: A cross-sectional study. *Med Sci Monit* 2020;26: e927848.

[29] Freudenberger HJ. Staff burnout. *J Soc Issues* 1974;30:159–165.

[30] Hepp T, Schmid M, Gefeller O, Waldmann E, Mayr A. Approaches to regularized regression - a comparison between gradient boosting and the lasso. *Methods Inf Med* 2016;55:422–430.

[31] Occupational Stress and Employees Complete Mental Health: A Cross-Cultural Empirical Study

[32] Knapp M and Wong G. Economics and mental health: the current scenario. *World Psychiatry* 2020;19:3–14.

[33] Benach, J, Vives, A, Amable, M, Vanroelen, C, Tarafa, G, & Muntaner, C. Precarious employment: understanding an emerging social determinant of health. *Annu Rev Public Health* 2014;35:229-53.

[34] Bakian A. V. *et al*. Acute air pollution exposure and risk of suicide completion. *Am J Epidemiol* 2015;181:295–303.

[35] Barnard KD, Skinner TC, Peveler R. The prevalence of co-morbid depression in adults with Type 1 diabetes: systematic literature review. *Diabet Med* 2006;23:445–448.

[36] Gilsanz P, Karter AJ, Beeri MS, Quesenberry CP, Whitmer RA. The bidirectional association between depression and severe hypoglycemic and hyperglycemic events in type 1 diabetes. *Diabetes Care* 2018;41:446–452.

[37] Lu Y, Zhang Z, Yan H, *et al*. Effects of occupational hazards on job stress and mental health of factory workers and miners: A propensity score analysis. *BioMed Res Int* 2020;2020:1754897.

[38] Shi L, Liu Y, Jiang T, *et al*. Relationship between mental health, the clock gene, and sleep quality in surgical nurses: A cross-sectional study. *BioMed Res Int* 2020;2020:4795763.

[39] Sallis JF, Saelens BE, Frank LD *et al*. Neighborhood built environment and income: examining multiple health outcomes. *Soc Sci Med* 2009;68:1285–1293.

[40] Winkleby MA, Jatulis DE, Frank E *et al*. Socioeconomic status and health: how education, income, and occupation contribute to risk-factors for cardiovascular disease. *Am J Public Health* 1992;82:816–820.

[41] Skapinakis P, Bellos S, Koupidis S, Grammatikopoulos I, Theodorakis P.N, Mavreas V. Prevalence and sociodemographic associations of common mental disorders in a nationally representative sample of the general population of Greece. *BMC Psychiatry* 2013;13:163.

[42] Virtanen M, Ferrie JE, Singh-Manoux A, Shipley MJ, Stansfeld SA, Marmot MG *et al*. Long

working hours and symptoms of anxiety and depression: a 5-year follow-up of the Whitehall II study. *Psychol Med* 2011;41:2485–2494.

[43] Phillips MR, Zhang J, Shi Q, Song Z, Ding Z, Pang S, *et al*. Prevalence, treatment, and associated disability of mental disorders in four provinces in China during 2001–05: an epidemiological survey. *Lancet* 2009;373:2041–2053.

[44] Wu Q, Luo X, Chen S, *et al*. Mental health literacy survey of non-mental health professionals in six general hospitals in hunan province of China. *PloS one* 2017;12:e0180327.

Figure legends

Fig.1. Feature selection using the LASSO binary logistic regression model. (A) Feature selection for the LASSO binary logistic regression model. The partial likelihood deviation (binomial deviation) curve was plotted against lambda by validating the optimal parameter lambda in the LASSO model. Dotted vertical lines were drawn based on 1 SE of the minimum criteria (the 1-SE criteria). (B) Feature selection was performed using the LASSO binary logistic regression model. A Coefficient profile weas plotted based on the lambda series in Figure 1(A), and 12 features with non-zero coefficients were selected by optimal lambda.

Fig.2. The forest plot of the OR of the selected feature.

Fig.3. Developed mental health problems incidence risk nomogram. The mental health problems incidence risk nomogram was developed in the array, with education, professional title, age, CMBI, ERI, asbestos dust, hypertension, diabetes, working hours per day, working years, marital status, and work schedule incorporated.

Fig.4. Calibration curves of the mental health problems incidence risk nomogram prediction in validation group. The x-axis represents the predicted risk of mental health problems. y-axis represents the actual diagnosed risk of mental health problems. The diagonal dashed line represents the perfect prediction of the ideal model. The solid lines represent the performance of the column plots, where closer to the diagonal dashed line indicates a better prediction.

Fig.5. ROC curves for training and validation groups. The y-axis represents the true positive rate of risk prediction. The x-axis represents the false positive rate of risk prediction. The ROC curves for the training and validation groups are shown in black and red.

Fig.6. Decision curve analysis for mental health problems incidence risk nomogram. The y-axis measures the net benefit. The solid red line represents nomogram of the risk of developing a mental health problem. The light blue dashed line represents the hypothesis that all participants were diagnosed with a mental health problem. The black dashed line represents the hypothesis that there is no risk of a mental health problem. The DCA showed that using this mental health problem incidence risk nomogram in the current study to predict mental health problem incidence risk increase in benefit than the intervention all patients or no intervention all patient if the threshold probability of a patients and a doctor is >10%.

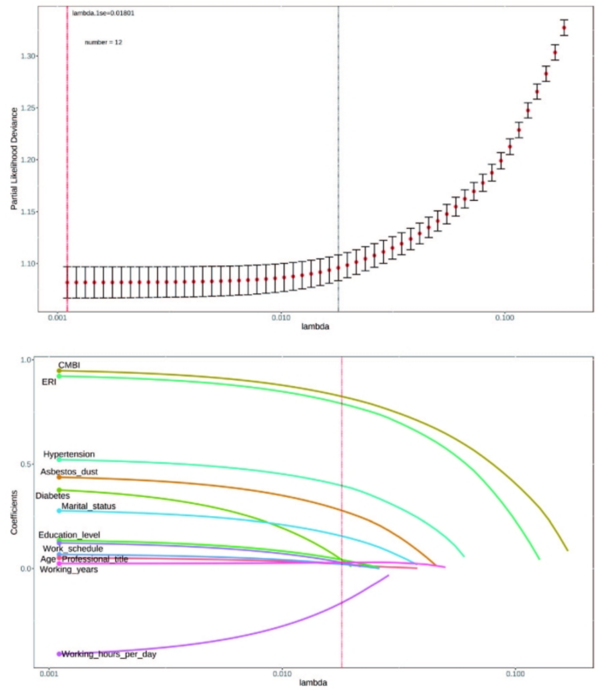
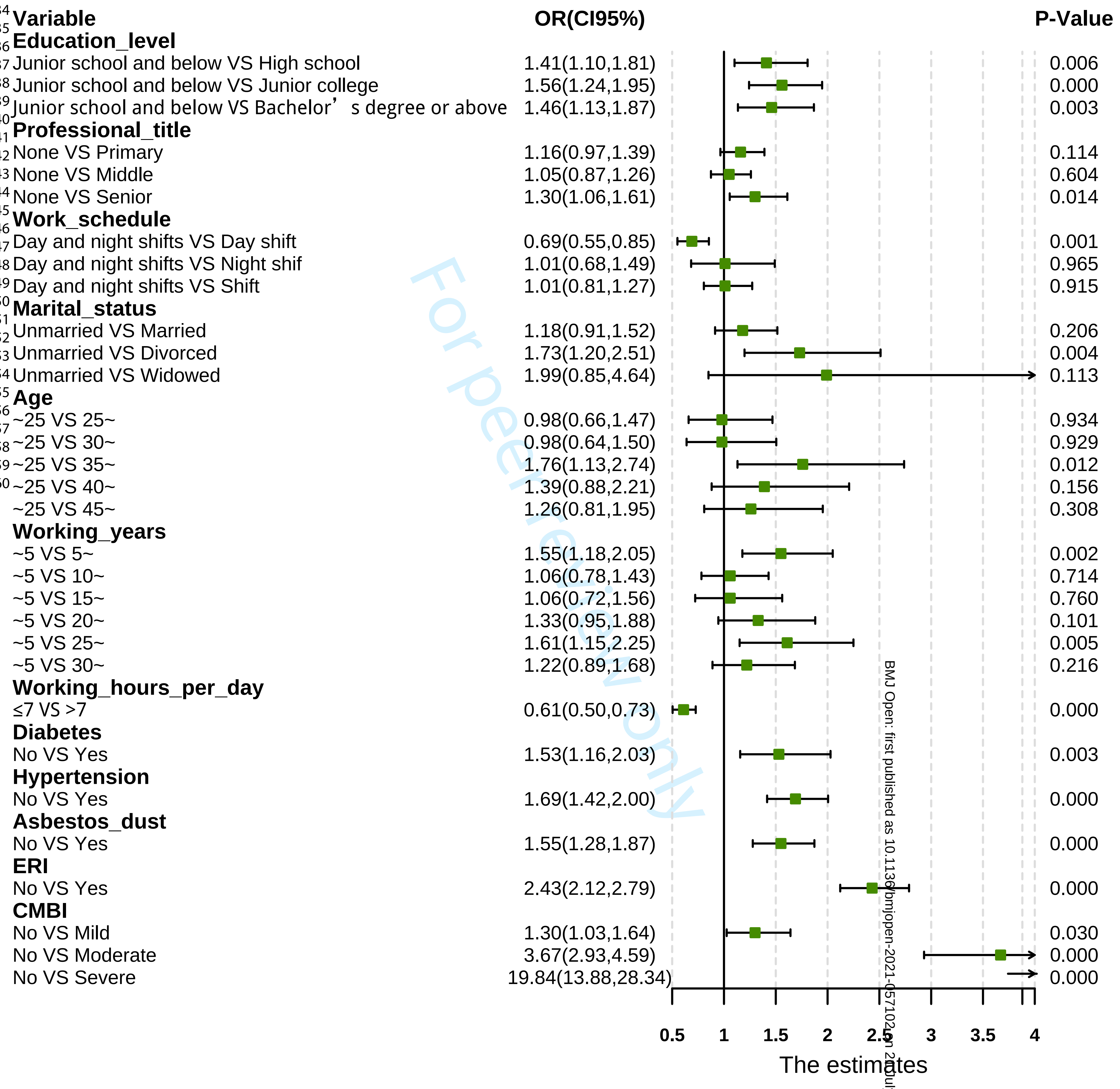


Fig.1. Feature selection using the LASSO binary logistic regression model.

210x297mm (300 x 300 DPI)


Variable
OR(CI95%)
P-Value
Education_level
Junior school and below VS High school 1.41(1.10,1.81) 0.006
Junior school and below VS Junior college 1.56(1.24,1.95) 0.000
Junior school and below VS Bachelor’s degree or above 1.46(1.13,1.87) 0.003
Professional_title
None VS Primary 1.16(0.97,1.39) 0.114
None VS Middle 1.05(0.87,1.26) 0.604
None VS Senior 1.30(1.06,1.61) 0.014
Work_schedule
Day and night shifts VS Day shift 0.69(0.55,0.85) 0.001
Day and night shifts VS Night shif 1.01(0.68,1.49) 0.965
Day and night shifts VS Shift 1.01(0.81,1.27) 0.915
Marital_status
Unmarried VS Married 1.18(0.91,1.52) 0.206
Unmarried VS Divorced 1.73(1.20,2.51) 0.004
Unmarried VS Widowed 1.99(0.85,4.64) 0.113
Age
~25 VS 25~ 0.98(0.66,1.47) 0.934
~25 VS 30~ 0.98(0.64,1.50) 0.929
~25 VS 35~ 1.76(1.13,2.74) 0.012
~25 VS 40~ 1.39(0.88,2.21) 0.156
~25 VS 45~ 1.26(0.81,1.95) 0.308
Working_years
~5 VS 5~ 1.55(1.18,2.05) 0.002
~5 VS 10~ 1.06(0.78,1.43) 0.714
~5 VS 15~ 1.06(0.72,1.56) 0.760
~5 VS 20~ 1.33(0.95,1.88) 0.101
~5 VS 25~ 1.61(1.15,2.25) 0.005
~5 VS 30~ 1.22(0.89,1.68) 0.216
Working_hours_per_day
≤7 VS >7 0.61(0.50,0.73) 0.000
Diabetes
No VS Yes 1.53(1.16,2.03) 0.003
Hypertension
No VS Yes 1.69(1.42,2.00) 0.000
Asbestos_dust
No VS Yes 1.55(1.28,1.87) 0.000
ERI
No VS Yes 2.43(2.12,2.79) 0.000
CMBI
No VS Mild 1.30(1.03,1.64) 0.030
No VS Moderate 3.67(2.93,4.59) 0.000
No VS Severe 19.84(13.88,28.34) 0.000
0.5 1 1.5 2 2.5 3 3.5 4
The estimates

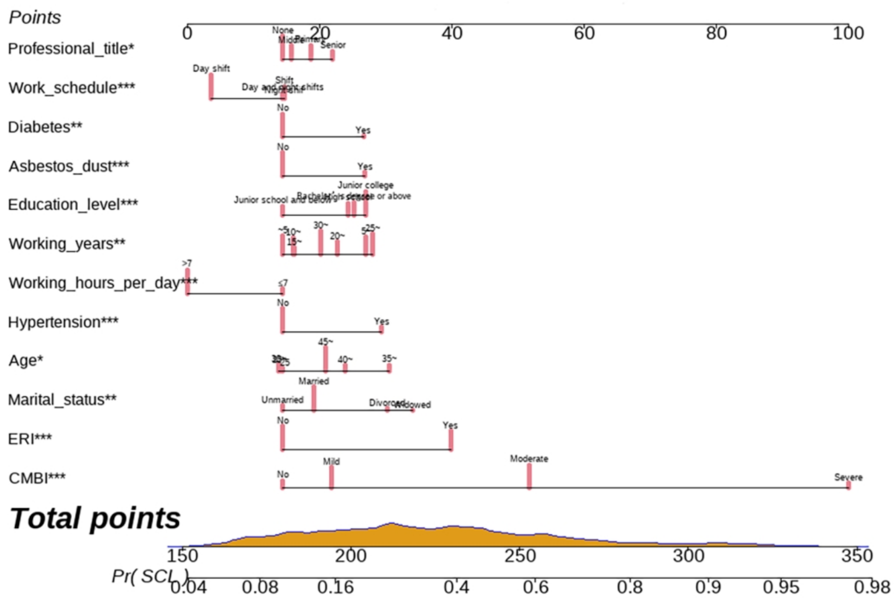
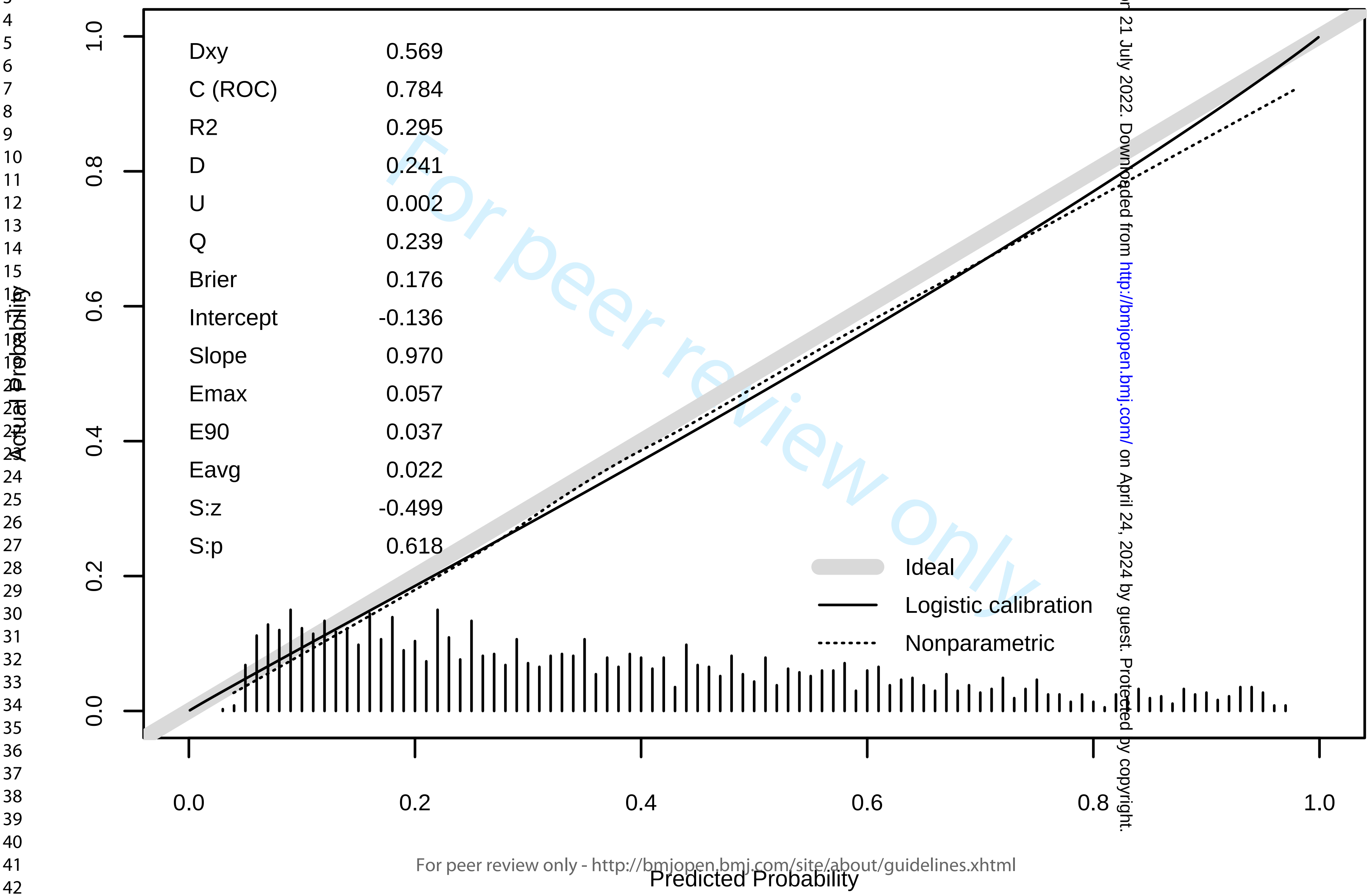
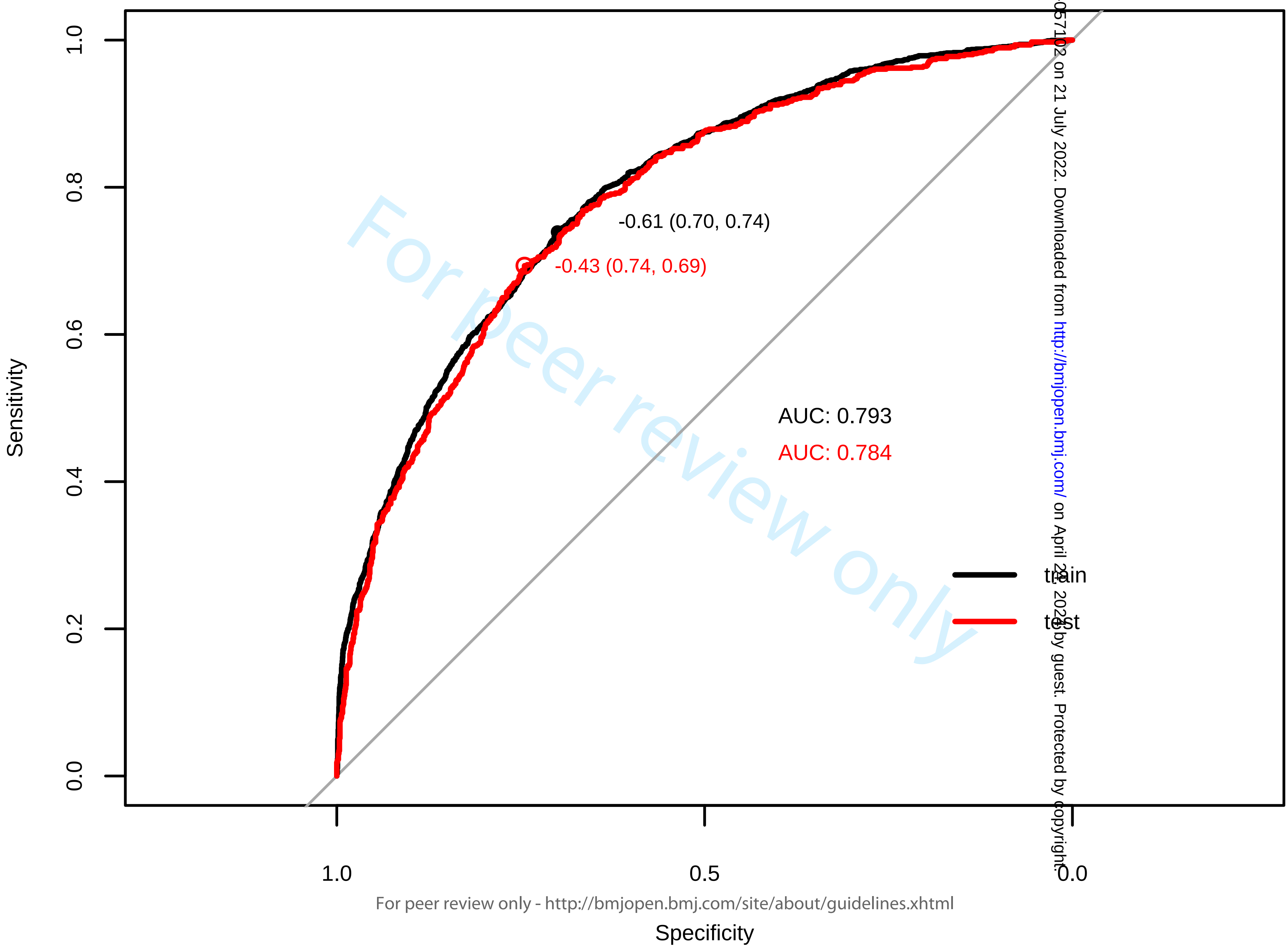
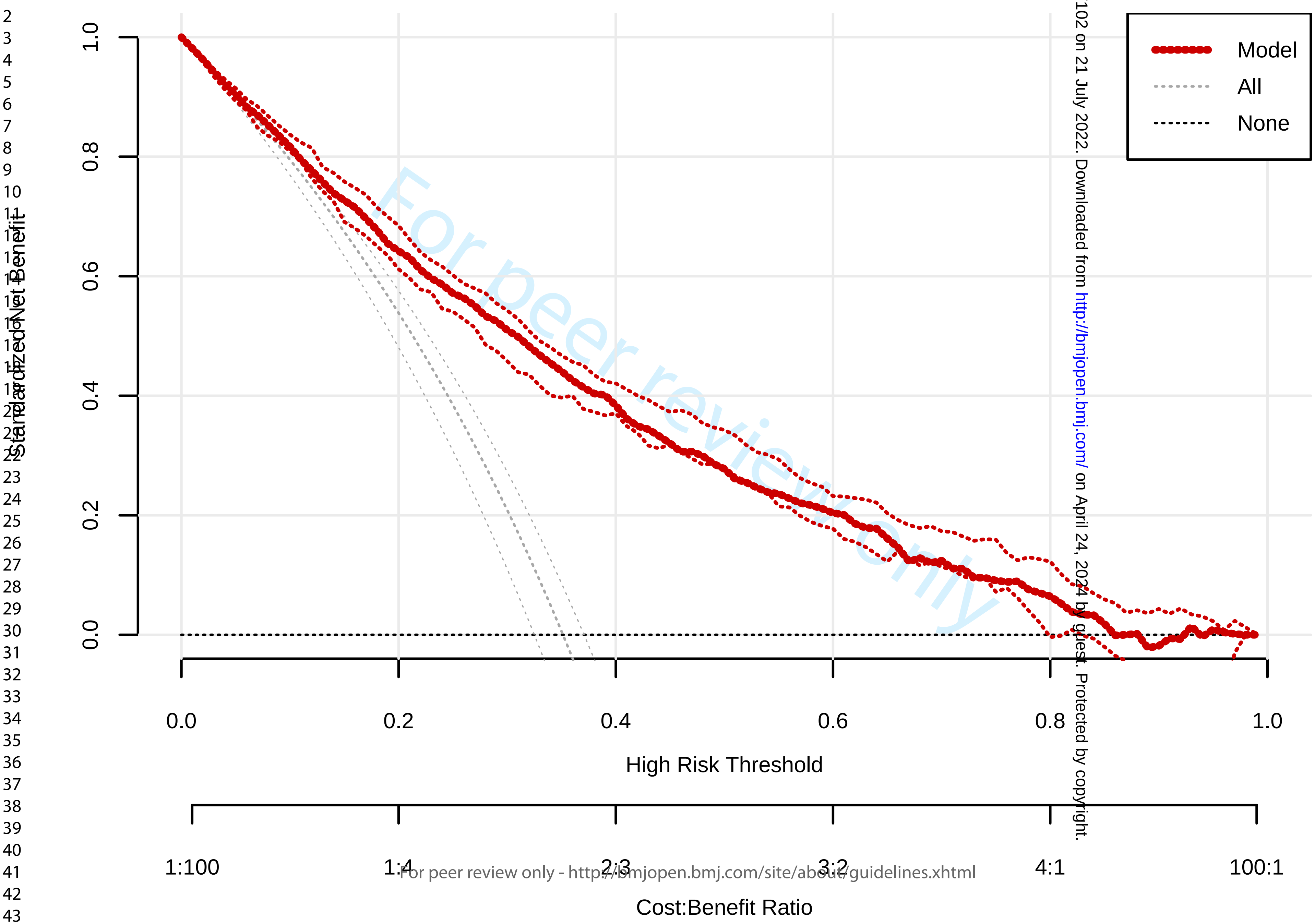


Fig.3. Developed mental health problems incidence risk nomogram.

352x230mm (118 x 118 DPI)


Dxy 0.569
C (ROC) 0.784
R2 0.295
D 0.241
U 0.002
Q 0.239
Brier 0.176
Intercept -0.136
Slope 0.970
Emax 0.057
E90 0.037
Eavg 0.022
S:z -0.499
S:p 0.618
Ideal
Logistic calibration
Nonparametric
Actual Probability
0.0
0.2
0.4
0.6
0.8
1.0
Predicted Probability
0.0
0.2
0.4
0.6
0.8
1.0


1.0
0.8
0.6
0.4
0.2
0.0
Sensitivity
-0.61 (0.70, 0.74)
-0.43 (0.74, 0.69)
AUC: 0.793
AUC: 0.784
train
test
1.0
0.5
0.0
Specificity


Model
All
None
Standardized Net Benefit
0.0
0.2
0.4
0.6
0.8
1.0
High Risk Threshold
1:100
1:4
2:3
3:2
4:1
100:1
Cost:Benefit Ratio

STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No.	Recommendation	Page No.	Relevant text from manuscript
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1	Development and Validation of a Nomogram for Predicting the Risk of Mental Health Problems of Factory Workers and Miners
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	1	A total of 7,118 participants met the inclusion criteria and the data were randomly divided into a training group (n=4,955) and a validation group (n=2,163) in a ratio of 3:1. A total of 23 characteristics were included in this study and LASSO regression selected 12 characteristics such as education, professional title, age, CMBI, ERI, asbestos dust, hypertension, diabetes, working hours per day, working years, marital status, and work schedule as predictors for the construction of the nomogram. In the validation group the Brier score was 0.176, the calibration slope was 0.970 and the calibration curve of nomogram showed a

				good fit, indicating good agreement between predictions and observations. The AUC of training group and verification group were 0.785 and 0.784 respectively, which showed good discrimination. The DCA suggested that the nomogram for predicting the risk of mental health problems among factory workers and miners was more practical when the risk threshold for mental health problems was 10% for intervention.
Introduction				
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	2	Factory workers and miners are a special group of workers with a relatively low overall level of education and are highly prone to suffering from mental health problems due to limited social support, excessive workload and irregular lifestyles, as well as occupational hazards such as noise and coal dust that they inevitably need to face in their working environment[18-19]. China

				has the world's largest group of factory workers and miners, about 6 million[20], who are regularly involved in occupational hazards. Mental health problems which need to require a long process are known to be a syndrome caused by chronic stress. Factory workers and miners, represented by those engaged in coal mining, have a mental burden rating of 8.3, one of the highest mental burdens among 150 occupations[21]. This explains the high level of mental health problems among mine workers in previous studies, making the identification and treatment of mental health problems even more important. Therefore, it is essential to provide a viable and easy-to-apply tool for identifying workers at risk of mental health problems and thus for timely interventions.
Objectives	3	State specific objectives, including any prespecified hypotheses	3	Therefore, the aim of our study is to develop and validate an easy-to-use nomogram that combines objective information on the

				demographics, job burnout, occupational stress and occupational hazards to comprehensively and accurately predict the prevalence of mental health problems among factory workers and miners.
Methods				
Study design	4	Present key elements of study design early in the paper	3	The selection of participants. The quality of the questionnaires. The results of agreement and discrimination between predictions and observations in this nomogram.
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	4	Participants in this cross-sectional survey were workers from factories and mining enterprises in the Urumqi region, who were recruited using a whole-group sampling method. A total of 3,619 enterprises in the Urumqi were surveyed from January to May 2019, covering all districts and counties in the Urumqi region, including Tianshan District, Shaibak District, Xinshi District, Shuimogou District, Toutunhe District, Dabancheng District,

				Middong District and Urumqi County.
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants	4	The exclusion criteria were the following: (I) factory workers and miners in non-Urumqi area, (II) working history of factories and mining enterprises less than 1 year, (III) a confirmed diagnosis of a mental health problem and a history of treatment and use of psychotropic medication. Questionnaires with missing data were also excluded from the analysis based on discussion and agreement among the subject members.
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case		
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	4-5	2.2.1. Assessment of mental health 2.2.2. Assessment of occupational stress 2.2.3. Assessment of job burnout 2.2.4. Candidate predictors
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	6	Categorical variables were described as counts and percentages, and chi square test or Fisher exact test was used to compare categorical variables

				between groups. 70% of participants were randomly assigned to the training cohort and 30% to the validation cohort. Variables were screened using a least absolute shrinkage and selection operator (LASSO) regression and multivariate logistic regression models were used to estimate risk ratios and corresponding 95% confidence intervals (CI) of risk factors, from which predictive models were constructed. A nomogram for predicting was generated according to the selected characteristics.
Bias	9	Describe any efforts to address potential sources of bias	6	We used calibration plots and receiver operating characteristic (ROC) curves to show the calibration and discrimination of our final model. Brier scores for overall performance, calibration slopes were used to assess the predictable accuracy of the model. Decision curve analysis (DCA) was applied to calculate the net benefit of the nomogram. Statistical analysis was performed using the open-source R software Version 3.6.1 (http://www.r-

				project.org). The significance level (α) set at 0.05.
Study size	10	Explain how the study size was arrived at		

Continued on next page

Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	5	Sex was defined as male or female; ethnicity was defined as Han and other; education level was defined as junior high school and below, high school, junior college or bachelor's degree or above; labor contracts was defined as signed or unsigned; professional title was defined as no, primary, middle or senior; work schedule was defined as day shift, night shift, shift or day and night shifts; marital status was defined as unmarried, married, divorced or widowed; monthly income (yuan) was defined as <3000, 3000~, 4000~, 5000~, 6000~, 7000~ or 8000~; age (years) was defined as <25, 25~, 30~, 35~, 40~ or 45~; working years was defined as ~5, 5~, 10~, 15~, 20~, 25~ or 30~; working hours per day (hours) was defined as ≤7 or > 7; working days per week (days) was defined as ≤5 or >5; exposure to coal dust, silica dust, asbestos dust, benzene, lead, noise, brucellosis were all defined as yes or no; ERI

				was defined as yes or no; CMBI was defined as none, mild, moderate and severe; hypertension and diabetes were both defined as yes or no.
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	6	Categorical variables were described as counts and percentages, and chi square test or Fisher exact test was used to compare categorical variables between groups. 70% of participants were randomly assigned to the training cohort and 30% to the validation cohort. Variables were screened using a least absolute shrinkage and selection operator (LASSO) regression and multivariate logistic regression models were used to estimate risk ratios and corresponding 95% confidence intervals (CI) of risk factors, from which predictive models were constructed. A nomogram for predicting was generated according to the selected characteristics. In addition, forest plot was drawn to visually depict the P-value, OR and 95% CI for the selected validations.
		(*b*) Describe any methods used to examine subgroups and interactions		

		(*c*) Explain how missing data were addressed	4	A total of 7,500 questionnaires were distributed and 7,315 questionnaires were returned, representing a return rate of 97.5%. After checking the validity and integrity of the questionnaires, 7,118 questionnaires were confirmed as valid, with an effective rate of 97.3%. All participants understood the purpose of the study and voluntarily participated in the study.
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed *Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy		
		(*e*) Describe any sensitivity analyses		
Results				
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	4	7500 participants volunteered for the survey Issued a total of 7500 questionnaires Collected a total of 7315 questionnaires 7118 valid and integrated questionnaires
		(b) Give reasons for non-participation at each stage	4	
		(c) Consider use of a flow diagram	4	

Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	6	A total of 7,118 participants met the inclusion criteria and the data were randomly divided into a training group (n=4,955) and a validation group (n=2,163). Over half of all participants (65.31%) were male, 57.31% of the population was over 35 years of age and 78.32% of the subjects were married, showing that factory workers and miners are generally older and most of them have spouses. The majority of them had completed high school (83.94%), while a smaller percentage had completed undergraduate education (22.98%), indicating that the group of factory workers and miners as a whole was not well educated. The total number of workers (n, %) exposed to coal dust, silica dust, asbestos dust, benzene, lead, noise and brucellosis in the factory and mining enterprises were 377 (5), 730 (10), 981 (14), 1,981 (28), 373 (5), 4,942 (69) and 121 (2) respectively, with the total number of workers

				exposed to noise amounting to 4,942, or 69% of the total population surveyed. The demographic, job burnout, occupational stress and occupational exposure factors for the training and validation groups are shown in Table 1. The results showed that there were no significant statistical differences between the two groups of characteristic variables, except for coal dust and CMBI, indicating that the baseline levels were largely consistent between the two groups.
		(b) Indicate number of participants with missing data for each variable of interest		
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)		
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time		
		Case-control study—Report numbers in each exposure category, or summary measures of exposure		
		Cross-sectional study—Report numbers of outcome events or summary measures	6	A total of 7,118 participants met the inclusion criteria and the data were randomly divided into a training group (n=4,955) and a validation group (n=2,163). Over half of all participants (65.31%) were male, 57.31% of the population was over 35 years of age and 78.32% of the subjects were married, showing that factory workers and miners are

generally older and most of them have spouses. The majority of them had completed high school (83.94%), while a smaller percentage had completed undergraduate education (22.98%), indicating that the group of factory workers and miners as a whole was not well educated. The total number of workers (n, %) exposed to coal dust, silica dust, asbestos dust, benzene, lead, noise and brucellosis in the factory and mining enterprises were 377 (5), 730 (10), 981 (14), 1,981 (28), 373 (5), 4,942 (69) and 121 (2) respectively, with the total number of workers exposed to noise amounting to 4,942, or 69% of the total population surveyed. The demographic, job burnout, occupational stress and occupational exposure factors for the training and validation groups are shown in Table 1. The results showed that there were no significant statistical differences between the two groups of characteristic variables, except for coal dust and CMBI, indicating that the baseline levels were largely consistent between the two groups.

Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	6	Categorical variables were described as counts and percentages, and chi square test or Fisher exact test was used to compare categorical variables between groups. 70% of participants were randomly assigned to the training cohort and 30% to the validation cohort. Variables were screened using a least absolute shrinkage and selection operator (LASSO) regression and multivariate logistic regression models were used to estimate risk ratios and corresponding 95% confidence intervals (CI) of risk factors, from which predictive models were constructed.
		(*b*) Report category boundaries when continuous variables were categorized	6	
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period		

Continued on next page

Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses		
Discussion				
Key results	18	Summarise key results with reference to study objectives	13	Therefore, this study designed a simple and comprehensive nomogram to address the issue of timely detection and effective interventions for people with mental health problems, so that people at risk of mental health problems could easily calculate their probability of suffering from mental health problems without the help of medical staff.
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	13	This study also has several limitations. First, we have considered many influential factors including demographics, job burnout, occupational stress and occupational exposure factors, but we are still not certain whether all possible influences are covered. Secondly, while the robustness of our nomogram was extensively validated internally in the same population, external validation is lacking for other populations in other regions and countries. Nomogram need to be externally assessed in a wider population.

Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	13	The nomogram we proposed contains 12 characteristics related to demographics, job burnout, occupational stress and occupational hazard factors. The nomogram combining these 12 characteristics for the risk of mental health problems can be used to predict the risk of suffering mental health problems, providing a useful tool for quickly and accurately screening the risk of mental health problems among factory workers and miners.
Generalisability	21	Discuss the generalisability (external validity) of the study results		
Other information				
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	14	This work was supported by National Natural Science Foundation of China, grant number 81760581 and Public Health and Preventive Medicine, the 13th Five-Year Plan Key Subject of Xinjiang Uygur Autonomous Region.

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

Development and Validation of a Nomogram for Predicting the Risk of Mental Health Problems of Factory Workers and Miners

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-057102.R1
Article Type:	Original research
Date Submitted by the Author:	17-Mar-2022

Complete List of Authors:	Lu, Yaoqin; Xinjiang Medical University, School of Public Health; Urumqi Center for Disease Control and Prevention Liu, Qi; Xinjiang Medical University, School of Public Health Yan, Huan; Xinjiang Medical University, Department of Nutrition and Food Hygiene; Xinjiang Autonomous Academy of Instrumental Analysis Liu, Tao; Xinjiang Medical University, School of Public Health

Primary Subject Heading:	Mental health
Secondary Subject Heading:	Public health
Keywords:	MENTAL HEALTH, PREVENTIVE MEDICINE, PUBLIC HEALTH

Development and Validation of a Nomogram for Predicting the Risk of Mental Health Problems of Factory Workers and Miners

Yaoqin Lu,[1,2] Qi Liu,[1] Huan Yan[3,4], and Tao Liu[1]

[1]Department of Toxicology, School of Public Health, Xinjiang Medical University, Urumqi 830011, Xinjiang, China.
[2]Urumqi Center for Disease Control and Prevention, Urumqi 830026, Xinjiang, China.
[3]Department of Nutrition and Food Hygiene, College of Public Health, Xinjiang Medical University, Urumqi, Xinjiang, China.
[4]Xinjiang Engineering Technology Research Center for Green Processing of Nature Product Center, Xinjiang Autonomous Academy of Instrumental Analysis, Urumqi, Xinjiang, China.

Correspondence should be addressed to Tao Liu; xjmult@163.com

Yaoqin Lu and Qi Liu contributed equally to this work

Abstract

Objective A nomogram for predicting the risk of mental health problems was established in a population of factory workers and miners, in order to quickly calculate the probability of a worker suffering from mental health problems.

Methods A cross-sectional survey of 7,500 factory workers and miners in Urumqi was conducted by means of an electronic questionnaire using cluster sampling method. Participants were randomly assigned to the training group (70%) and the validation group (30%). Questionnaire-based survey was conducted to collect information. A least absolute shrinkage and selection operator (LASSO) regression model was used to screen the predictors related to the risk of mental health problems of the training group. Multivariate logistic regression analysis was applied to construct the prediction model. Calibration plots and receiver operating characteristic-derived area under the curve (AUC) were used for model validation. Decision curve analysis (DCA) was applied to calculate the net benefit of the screening model.

Results A total of 7,118 participants met the inclusion criteria and the data were randomly divided into a training group (n=4,955) and a validation group (n=2,163) in a ratio of 3:1. A total of 23 characteristics were included in this study and LASSO regression selected 12 characteristics such as education, professional title, age, CMBI, ERI, asbestos dust, hypertension, diabetes, working hours per day, working years, marital status, and work schedule as predictors for the construction of the nomogram. In the validation group the Brier score was 0.176, the calibration slope was 0.970 and the calibration curve of nomogram showed a good fit. The AUC of training group and verification group were 0.785 and 0.784 respectively.

Conclusion The nomogram combining these 12 characteristics can be used to predict the risk of suffering mental health problems, providing a useful tool for quickly and accurately screening the risk of mental health problems.

Key words Mental health; Predictor; Nomogram; Risk; Factory workers and miners

Strengths and limitations of this study

1. This is the first study to develop an easy-to-use nomogram to predict the mental health risks of factory workers and miners.
2. The AUC of training group and verification group were 0.785 and 0.784 respectively, showing moderate discriminatory and calibration power.
3. This nomogram model's variables are more comprehensive, including demographics, burnout, occupational stress and occupational hazards.
4. We had considered many influential factors, but we were still not certain whether all possible influences were covered.
5. There is a lack of external validation in other populations in other regions and countries.

1. Introduction

The World Health Organization (WHO) defines health as a state of complete physical, mental and social well-being and not merely the absence of disease or weakness [1]. Obviously, health is an organic unity of physical and mental well-being. People with good mental health are the precondition for the normal operation of our society. However, with the acceleration of people's pace of life, people are facing an increasing risk of poor health, which has become a global public health problem [2]. Mental health problems can not only take a toll on physical health such as increasing the risk of communicable and non-communicable diseases and even causing unintentional or intentional harm to others [3], but can also have a negative impact on the economy. For example, mental health disorders represent a growing part of the global burden of disease [4], with statistics showing that nearly one billion people worldwide currently suffer from a mental disorder, and mental illness is ranked as one of the leading causes of the global burden of disease [5]. Moreover, one study has estimated that due to the impact of mental illness, the global economy loses US $1 trillion every year [6].

As researchers around the world have delved into the field of mental health, factors such as gender, income levels, environment and education have been found to be associated with people's mental health problems [7-10]. Moreover, employment is also strongly associated with quality of life, higher self-esteem and fewer psychiatric symptoms [11]. In addition, in the context of the global challenges of climate change, an increasing number of scholars have been examining the epidemiological links between mental health and environmental factors. Some studies have suggested that mental health may be influenced by ambient temperature, and an association has been found between environmental pollutants, particularly fine particulate matter, and mental health problems [12]. A relevant study shows that with short-term exposure to ambient air pollution is associated with increased emergency room visits due to depression or suicide attempts [13]. Furthermore, other factors associated with mental health include sleep, diabetes, coronary artery disease and cardiovascular disease [14-15]. It is worth noting that job burnout and occupational stress are closely linked to mental health. Job burnout is an exhaustion state of physical and psychological that

often occurs in the work environment, and has a high correlation with depression. A large study of physicians found that of the 10.3% who met criteria for a major depressive episode, 50.7% were also affected by symptoms of burnout (OR 2.99) and indicated that worsening depression leads to a higher likelihood of burnout symptoms [16]. Occupational stress refers to a work environment where non-reciprocity of effort and reward may lead to strong negative emotions and distress. Related research has shown that the combination of high effort and low reward and over-commitment increases the risk of mental health problems such as depression [17]. Apparently, it is necessary to include the CMBI and ERI in this study to predict the risk of mental health problems among factory workers and miners. However, there are few studies that include these influences in a more comprehensive way in the practice of detecting mental health. Therefore, more accurate identification of mental health problems in populations requires a questionnaire that include a wider range of factors affecting factory workers and miners' mental health problems.

Factory workers and miners are a special group of workers with a relatively low overall level of education and are highly prone to suffering from mental health problems due to limited social support, excessive workload and irregular lifestyles, as well as occupational hazards such as noise and coal dust that they inevitably need to face in their working environment [18-19]. Through a review of the literature, our group found that coal dust, crystalline silica and noise pollution were common causes of health problems for workers in underground mines [20]. And, exposure to coal mine dust is a significant cause of pneumoconiosis in coal miners [21]. In addition, asbestos is one of the major occupational hazards in the daily work of workers in the construction and automotive industries [22]. China has the world's largest group of factory workers and miners, about 6 million [23], who are regularly involved in occupational hazards. Mental health problems which need to require a long process are known to be a syndrome caused by chronic stress. Factory workers and miners, represented by those engaged in coal mining, have a mental burden rating of 8.3, one of the highest mental burdens among 150 occupations [24]. This explains the high level of mental health problems among mine workers in previous studies, making the identification and treatment of mental health problems even more important. Therefore, it is essential to provide a viable and easy-to-apply tool for identifying workers at risk of mental health problems and thus for timely interventions.

There are many studies on mental health [25-26]; however, the results of previous studies lack consistency and mostly discuss factors influencing mental health, and most of them are single-center studies that focus on only certain aspects of mental health. Our study included common demographics, job burnout, occupational stress, chronic illness and occupational exposure factors to distinguish whether respondents suffered from mental health problems. In addition, there is a small body of literature that develops and validates a risk nomogram between depression and suicide to support timely intervention by clinicians. And the sample sizes of the two relevant studies were small, 474 and 273 depressed patients respectively [27-28]. Today, there is increasing recognition of the important role of mental health in achieving global development goals, and WHO has included mental health in the Sustainable Development Goals. However, there are no relevant studies that have used objective indicators for factory workers and miners to form a nomogram to predict mental health. Therefore, to bridge this gap in the literature and provide additional information for the prevention of mental health problems, we conducted a multicenter investigation to develop and validate an easy-to-use nomogram that combines objective information on demographics, job burnout, occupational stress and occupational hazards to comprehensively and accurately predict the prevalence of mental health problems among factory workers and miners.

2. Materials and Methods

2.1 Calculation of sample size

The sample size formula for the present illness rate survey, $n = \frac{z_{\alpha/2}^2 \times \mathrm{pq}}{\delta^2}$, p is the present-hazard rate, q=1-p, $\boldsymbol{\delta}$ is the tolerance error, generally taken as 0.1p, $z_{\alpha/2}$ is the significance test statistic, $z_{\alpha/2}$=1.96 for α=0.05, then the formula is calculated as, $n = 400 \times \frac{q}{p}$. A cross-sectional study in Xinjiang showed that 38.27% of factory workers and miners had mental health problems [29]。 And a study revealed that 633 out of 1675 coal miners (37.8%) suffered from mental disorders between August 2018 and June 2019[30]. In this study, we assumed a 30% prevalence of mental health problem to obtain the maximum required sample size. which would calculate a sample size of 934, taking into account non-response and a 20% loss of questionnaires, which would require approximately 1168 people.

2.2. Participants

Participants in this cross-sectional survey were factory workers and mines in the Urumqi region, and the survey covered all districts and counties in the Urumqi region to avoid selection bias as far as possible. Specifically, this survey was conducted by means of whole-group random sampling from January to May 2019, and a total of 202 enterprises were selected, including 21 in Tianshan District, 30 in Shaibak District, 24 in Xinshi District, 22 in Shuimogou District, 56 in Jingkai District, 37 in Midong District, 9 enterprises in Dabancheng District and 3 enterprises in Urumqi County.

The inclusion criteria were as follows: (1) workers working in mining enterprises or factories in Urumqi; (2) workers with a history of working for more than one year; (3) Workers with no history of mental illness and no history of taking psychotropic drugs.

The exclusion criteria were the following: (1) factory workers and miners in non-Urumqi area; (2) working history of factories and mining enterprises less than 1 year; (3) a confirmed diagnosis of a mental health problem and a history of treatment and use of psychotropic medication; (4) Questionnaires with missing data were excluded.

An online electronic questionnaire was created using the Questionnaire Star platform to collect data. The survey was conducted by trained surveyors who explained the purpose, meaning, content and requirements of the questionnaire to all participants and provided on-site instructions to ensure the return rate of the questionnaire. All participants understood the purpose of the study and were willing to participate in the study. A total of 7,500 questionnaires were distributed and 7,315 questionnaires were returned, representing a return rate of 97.5%. After checking the validity and integrity of the questionnaires, 7,118 questionnaires were confirmed as valid, with an effective rate of 97.3%. A total of 7,118 participants met the inclusion criteria and the data were randomly divided into a training group (n=4,955) and a validation group (n=2,163) (Figure 1).

2.3. Research Methods

2.3.1. Assessment of mental health

The SCL-90, designed by Derogatis and his colleagues, was widely used in the mental health field [31], which contains 90 items across nine dimensions: somatization, obsessive-compulsive symptoms, interpersonal sensitivity, depression, anxiety, hostility, horror, bigotry and mental illness. The SCL-90 has been used extensively in previous studies and has relatively high reliability and validity [32]. The questionnaire uses a Likert 5-point scale, with a score of 0 point indicating none and 4 points indicating severe. A total score above 160, a score above 2 on any item, or a positive item above 43 indicates the presence of a psychological abnormality [33]. In this survey, Cronbach α was 0.99, the half-reliability coefficient was 0.98, and the KMO was 0.994.

2.3.2. Assessment of occupational stress

This survey evaluated occupational stress in factory workers and miners through the Effort–Reward Imbalance (ERI) model developed by Siegrist [34]. The ERI scale consists of three subscales: effort (E, 6 items), reward (R, 11 items) and over commitment (6 items), for a total of 23 items. A Likert 5-level scoring method (1, "highly disagree" to 5, "highly agree") is used to grade the items in the questionnaire with the same weight for each item. The effort–return index ERI = E/R×C, where C is the adjustment coefficient, and the value is 6/11. ERI values greater than 1, equal to 1, and less than 1 correspond to high pay–low return, pay–return balance, and low pay–high return, respectively. Moreover, the higher the ERI value, the greater the occupational stress [35]. In this survey, Cronbach α was 0.94, the half-reliability coefficient was 0.93 and the KMO was 0.956.

2.3.3. Assessment of job burnout

In this survey, the Chinese Maslach Burnout Inventory (CMBI) revised by Li et al. was used to assess job burnout, which has good reliability and validity [36]. CMBI is composed of 15 items in three dimensions: emotional exhaustion (5 items), depersonalization (5 items) and reduced personal accomplishment (5 items). The score for each item ranges from 1 to 7, with 1 point indicating complete compliance and 7 points indicating complete non-compliance. According to the critical value (emotional exhaustion ≥ 25, depersonalization ≥ 11, personal achievement reduction ≥ 16), the levels of occupational burnout are divided into none (all aspects are below the critical value), mild (any one aspect is equal to or above the critical value), moderate (any two aspects are equal to or higher than the critical values), and severe (three aspects are equal to or higher than the critical values) [37]. In this survey, Cronbach α was 0.89, the half-reliability coefficient was 0.86 and the KMO was 0.919.

2.3.4. Candidate predictors

Trained investigators obtained information on demographics, job burnout, occupational stress, mental health and occupational exposure factors through on-site face-to-face collection of an electronic version of the questionnaire. Covariates included in this study: 1) demographic information: gender, ethnicity, education level, professional title, work schedule, marital status, monthly income, age, working years, labor contracts, working hours per day, and working hours per week; 2) occupational exposure factors: coal dust, silica dust, asbestos dust, benzene, lead, noise, and brucellosis; 3) questionnaires: ERI, CMBI; 4) chronic diseases: diabetes, hypertension. Information on four areas, including demographic information, questionnaires, occupational hazards and chronic diseases, were filled in by participants through their own responses on the questionnaire star.

Sex was defined as male or female; ethnicity was defined as Han and other; education level was defined as junior high school and below, high school, junior college or bachelor's degree or above; labor contracts was defined as signed or unsigned; professional title was defined as no, primary, middle or senior; work schedule was defined as day shift, night shift, shift or day and night shifts; marital status was defined as unmarried, married, divorced or widowed; monthly income (yuan) was defined as <3000, 3000~, 4000~, 5000~, 6000~, 7000~ or 8000~; age (years) was defined as <25, 25~, 30~, 35~, 40~ or 45~; working years was defined as ~5, 5~, 10~, 15~, 20~, 25~ or 30~; working hours per day (hours) was defined as ≤7 or > 7; working days per week (days) was defined as ≤5 or >5; exposure to coal dust, silica dust, asbestos dust, benzene, lead, noise, brucellosis were all defined as yes or no; ERI was defined as yes or no; CMBI was defined as none, mild, moderate and severe; hypertension and diabetes were both defined as yes or no.

2.4. Statistical analysis

Categorical variables were described as counts and percentages, and chi square test or Fisher exact test was used to compare categorical variables between groups. 70% of participants were randomly assigned to the training cohort and 30% to the validation cohort. Variables were screened using a least absolute shrinkage and selection operator (LASSO) regression and multivariate logistic regression models were used to estimate risk ratios and corresponding 95% confidence intervals (CI) of risk factors, from which predictive models were constructed. A nomogram for predicting was generated according to the selected characteristics. In addition, forest plot was drawn to visually depict the P-value, OR and 95% CI for the selected validations. Statistically significant predictors were applied to develop a prediction model for the risk of mental health problems among factory workers and miners by introducing all selected factors and analyzing the statistical significance levels of them. We used calibration plots and receiver operating characteristic (ROC) curves to show the calibration and discrimination of our final model. Brier scores for overall performance, calibration slopes were used to assess the predictable accuracy of the model. Decision curve analysis (DCA) was applied to calculate the net benefit of the nomogram. Statistical analysis was performed using the open-source R software Version 3.6.1 (http://www.r-project.org). The significance level (α) set at 0.05.

3. Results

3.1. Participant characteristics

A total of 7,118 participants met the inclusion criteria and the data were randomly divided into a training group (n=4,955) and a validation group (n=2,163). Over half of all participants (65.31%) were male, 57.31% of the population was over 35 years of age and 78.32% of the subjects were married, showing that factory workers and miners are generally older and most of them have spouses. The majority of them had completed high school (83.94%), while a smaller percentage had completed undergraduate education (22.98%), indicating that the group of factory workers and miners as a whole was not well educated. The total number of workers (n, %) exposed to coal dust, silica dust, asbestos dust, benzene, lead, noise and brucellosis in the factory and mining enterprises were 377 (5.3), 730 (10.3), 981 (14), 1,981 (27.8), 373 (5.2), 4,942 (69.4) and 121 (1,7) respectively, with the total number of workers exposed to noise amounting to 4,942, or 69% of the total population surveyed. The demographic, job burnout, occupational stress and occupational exposure factors for the training and validation groups are shown in Table 1. The results showed that there were no significant statistical differences between the two groups of characteristic variables, except for coal dust and CMBI, indicating that the baseline levels were largely consistent between the two groups.

Table 1 Characteristics of the study participants

Variables	Total (n = 7118)	train (n = 4955)	test (n = 2163)	*p*
Sex, n (%)				
Male	4649 (65.3)	3216 (64.9)	1433 (66.3)	0.284
Female	2469 (34.7)	1739 (35.1)	730 (33.7)	
Ethnicity, n (%)				
Han	5762 (80.9)	3982 (80.4)	1780 (82.3)	0.061
Other	1356 (19.1)	973 (19.6)	383 (17.7)	
Education level, n (%)				
Junior high school and below	1143 (16.1)	804 (16.2)	339 (15.7)	0.765
High school	1406 (19.8)	988 (19.9)	418 (19.3)	
Junior college	2933 (41.2)	2038 (41.1)	895 (41.4)	

Bachelor's degree or above	1636 (23.0)	1125 (22.7)	511 (23.6)	
Professional title, n (%)				
None	2854 (40.1)	1983 (40.0)	871 (40.3)	0.923
Primary	1644 (23.1)	1149 (23.2)	495 (22.9)	
Middle	1618 (22.7)	1133 (22.9)	485 (22.4)	
Senior	1002 (14.1)	690 (13.9)	312 (14.4)	
Work schedule, n (%)				
Day shift	3986 (56.0)	2801 (56.5)	1185 (54.8)	0.585
Night shift	270 (3.8)	187 (3.8)	83 (3.8)	
Shift	2058 (28.9)	1412 (28.5)	646 (29.9)	
Day and night shifts	804 (11.3)	555 (11.2)	249 (11.5)	
Marital status, n (%)				
Unmarried	1104 (15.5)	762 (15.4)	342 (15.8)	0.218
Married	5575 (78.3)	3906 (78.8)	1669 (77.2)	
Divorced	390 (5.5)	255 (5.1)	135 (6.2)	
Widowed	49 (0.7)	32 (0.6)	17 (0.8)	
Monthly income (yuan), n (%)				
<3000	1799 (25.3)	1246 (25.1)	553 (25.6)	0.966
3000~	2418 (34.0)	1682 (33.9)	736 (34.0)	
4000~	1600 (22.5)	1125 (22.7)	475 (22.0)	
5000~	752 (10.6)	520 (10.5)	232 (10.7)	
6000~	288 (4.0)	201 (4.1)	87 (4.0)	
7000~	148 (2.1)	106 (2.1)	42 (1.9)	
8000~	113 (1.6)	75 (1.5)	38 (1.8)	
Age (years), n (%)				
<25	431 (6.1)	297 (6.0)	134 (6.2)	0.173
25~	786 (11.0)	519 (10.5)	267 (12.3)	
30~	956 (13.4)	684 (13.8)	272 (12.6)	
35~	866 (12.2)	617 (12.5)	249 (11.5)	
40~	849 (11.9)	588 (11.9)	261 (12.1)	
45~	3230 (45.4)	2250 (45.4)	980 (45.3)	
Working years (years), n (%)				
<5	1170 (16.4)	794 (16.0)	376 (17.4)	0.248
5~	1065 (15.0)	736 (14.9)	329 (15.2)	
10~	997 (14.0)	721 (14.6)	276 (12.8)	
15~	389 (5.5)	273 (5.5)	116 (5.4)	
20~	763 (10.7)	538 (10.9)	225 (10.4)	
25~	1293 (18.2)	878 (17.7)	415 (19.2)	
30~	1441 (20.2)	1015 (20.5)	426 (19.7)	

Labor contracts, n (%)				
Signed	6641 (93.3)	4624 (93.3)	2017 (93.3)	0.955
Unsigned	477 (6.7)	331 (6.7)	146 (6.7)	
Working hours per day (hours), n (%)				
≤7	1161 (16.3)	814 (16.4)	347 (16.0)	0.712
>7	5957 (83.7)	4141 (83.6)	1816 (84.0)	
Working days per week (days), n (%)				
≤5	4442 (62.4)	3107 (62.7)	1335 (61.7)	0.446
>5	2676 (37.6)	1848 (37.3)	828 (38.3)	
Diabetes, n (%)				
Yes	429 (6.0)	298 (6.0)	131 (6.1)	0.988
No	6689 (94.0)	4657 (94.0)	2032 (93.9)	
Hypertension, n (%)				
Yes	1330 (18.7)	929 (18.7)	401 (18.5)	0.861
No	5788 (81.3)	4026 (81.3)	1762 (81.5)	
Coal dust, n (%)				
Yes	377 (5.3)	244 (4.9)	133 (6.1)	0.039
No	6741 (94.7)	4711 (95.1)	2030 (93.9)	
Silica dust, n (%)				
Yes	730 (10.3)	523 (10.6)	207 (9.6)	0.223
No	6388 (89.7)	4432 (89.4)	1956 (90.4)	
Asbestos dust, n (%)				
Yes	981 (13.8)	691 (13.9)	290 (13.4)	0.570
No	6137 (86.2)	4264 (86.1)	1873 (86.6)	
Benzene, n (%)				
Yes	1981 (27.8)	1360 (27.4)	621 (28.7)	0.287
No	5137 (72.2)	3595 (72.6)	1542 (71.3)	
Lead, n (%)				
Yes	373 (5.2)	246 (5.0)	127 (5.9)	0.128
No	6745 (94.8)	4709 (95.0)	2036 (94.1)	
Noise, n (%)				
Yes	4942 (69.4)	3420 (69.0)	1522 (70.4)	0.270
No	2176 (30.6)	1535 (31.0)	641 (29.6)	
Brucellosis, n (%)				
Yes	121 (1.7)	86 (1.7)	35 (1.6)	0.800
No	6997 (98.3)	4869 (98.3)	2128 (98.4)	
ERI, n (%)				
Yes	3147 (44.2)	2173 (43.9)	974 (45.0)	0.372
No	3971 (55.8)	2782 (56.1)	1189 (55.0)	

CMBI, n (%)				
No	959 (13.5)	674 (13.6)	285 (13.2)	0.033
Mild	2667 (37.5)	1813 (36.6)	854 (39.5)	
Moderate	2900 (40.7)	2031 (41.0)	869 (40.2)	
Severe	592 (8.3)	437 (8.8)	155 (7.2)	

3.2. Feature selection

The lambda was smallest at 0.01801 as seen from the lasso results when there were 12 characteristics, which were education, professional title, age, CMBI, ERI, asbestos dust, hypertension, diabetes, working hours per day, working years, marital status, and work schedule based on the results of the questionnaires on demographics, occupational stress, job burnout and occupational exposure factors (Figure 2).

3.3. Results of logistic regression model

The 12 features obtained from the LASSO regression were incorporated into a multivariate logistic regression model and the regression results were shown in Table 2. It was clear from the results that education, professional title, age, CMBI, ERI, asbestos dust, hypertension, diabetes, working hours per day, working years, marital status, and work schedule were independent determinants of risk for mental health problems. In addition, there was no evidence of multicollinearity between the covariates included in the model. The forest plot showed that the selected 12 features all contain items with $P < 0.05$, among which the degree of severe of CMBI (OR, 19.84; 95% CI, 13.88-28.34; $p < 0.001$) had the greatest impact on the risk of mental health problems among factory workers and miners (Figure 3).

Table 2 Predictive factors of risk for mental health problems among factory workers and miners

Variable	β	S.E.	OR(CI95%)	Wald	*P*	VIF
Intercept	-2.33	0.25	0.10(0.06,0.16)	-9.357	0	-
Education level						
Junior school and below VS High school	0.34	0.13	1.41(1.10,1.81)	2.727	0.006**	2.28
Junior school and below VS Junior college	0.44	0.11	1.56(1.24,1.95)	3.850	< 0.001***	2.79
Junior school and below VS Bachelor's degree or above	0.38	0.13	1.46(1.13,1.87)	2.953	0.003**	2.51
Professional title						
None VS Primary	0.15	0.09	1.16(0.97,1.39)	1.582	0.114	1.35
None VS Middle	0.05	0.09	1.05(0.87,1.26)	0.519	0.604	1.34
None VS Senior	0.27	0.11	1.30(1.06,1.61)	2.458	0.014*	1.32
Work schedule						
Day and night shifts VS Day shift	-0.38	0.11	0.69(0.55,0.85)	-3.364	0.001**	2.70
Day and night shifts VS Night shif	0.01	0.20	1.01(0.68,1.49)	0.044	0.965	1.30

Day and night shifts VS Shift	0.01	0.12	1.01(0.81,1.27)	0.107	0.915	2.47
Marital status						
Unmarried VS Married	0.16	0.13	1.18(0.91,1.52)	1.263	0.206	2.29
Unmarried VS Divorced	0.55	0.19	1.73(1.20,2.51)	2.918	0.004**	1.69
Unmarried VS Widowed	0.69	0.43	1.99(0.85,4.64)	1.586	0.113	1.09
Age						
~25 VS 25~	-0.02	0.20	0.98(0.66,1.47)	-0.083	0.934	3.09
~25 VS 30~	-0.02	0.22	0.98(0.64,1.50)	-0.090	0.929	4.79
~25 VS 35~	0.56	0.23	1.76(1.13,2.74)	2.503	0.012*	5.01
~25 VS 40~	0.33	0.23	1.39(0.88,2.21)	1.419	0.156	4.97
~25 VS 45~	0.23	0.22	1.26(0.81,1.95)	1.018	0.308	10.93
Working years						
~5 VS 5~	0.44	0.14	1.55(1.18,2.05)	3.114	0.002**	2.27
~5 VS 10~	0.06	0.15	1.06(0.78,1.43)	0.366	0.714	2.48
~5 VS 15~	0.06	0.20	1.06(0.72,1.56)	0.305	0.760	1.79
~5 VS 20~	0.29	0.18	1.33(0.95,1.88)	1.641	0.101	2.65
~5 VS 25~	0.48	0.17	1.61(1.15,2.25)	2.782	0.005**	3.99
~5 VS 30~	0.20	0.16	1.22(0.89,1.68)	1.239	0.216	3.90
Working hours per day						
≤7 VS >7	-0.50	0.09	0.61(0.50,0.73)	-5.363	< 0.001***	1.15
Diabetes						
No VS Yes	0.43	0.14	1.53(1.16,2.03)	2.974	0.003**	1.05
Hypertension						
No VS Yes	0.52	0.09	1.69(1.42,2.00)	5.885	< 0.001***	1.11
Asbestos dust						
No VS Yes	0.44	0.10	1.55(1.28,1.87)	4.474	< 0.001***	1.03
ERI						
No VS Yes	0.89	0.07	2.43(2.12,2.79)	12.786	< 0.001***	1.05
CMBI						
No VS Mild	0.26	0.12	1.30(1.03,1.64)	2.175	0.003**	2.73
No VS Moderate	1.30	0.11	3.67(2.93,4.59)	11.361	< 0.001***	2.83
No VS Severe	2.99	0.18	19.84(13.88,28.34)	16.41	< 0.001***	1.44

Note: β is the regression coefficient. "***" indicates P<0.001, "**" indicates P<0.01, "*" indicates P<0.05.

3.4. Development of an individualized prediction model

Based on the results of the multivariate analysis, predictors such as education, professional title, age, CMBI, ERI, asbestos dust, hypertension, diabetes, working hours per day, working years, marital status, and work schedule were included in the nomogram. A model incorporating the above independent predictors was developed and represented as a nomogram in Figure 4. Each variable in nomogram was assigned a score, and the cumulative sum of each 'point' was the 'total score'. The "total score" corresponded to the "predictable likelihood", which was the predicted probability of mental health problems among factory workers and miners as suggested by our design of the nomogram.

As an example of the use of nomogram: a randomly selected sample from the training group, one with no professional title, day shift, no diabetes or hypertension, Junior college, <5 of working years, >7 of working hours per day, married, no exposed to asbestos dust, <25 years of age, no ERI, mild of CMBI, with a calculated total score of 174 and a corresponding risk probability of 8.27% for mental health problems.

3.5 The validation of calibration

Model validation was carried out in the validation group. The prediction accuracy of the model was assessed by two aspects. (1) The Brier score for overall performance, which assessed the difference between observed and predicted values, with values closer to 0 indicating better predictive ability. (2) The calibration slope used for modal calibration, which assessed the agreement between the observed and predicted values, with values closer to 1 indicating better performance. The accuracy measurements for the bias correction were validated by the model with a Brier score of 0.176 and a calibration slope of 0.970, respectively (Figure 5). The prediction accuracy of the model was relatively high.

3.6 The validation of discrimination

ROC was plotted for the training and validation groups, and the AUC of training and the verification groups were 0.785 and 0.784, respectively (Figure 6). The AUC of training and the verification groups were both greater than 0.75, showing a good discrimination.

3.7 Decision Curve Analysis

As shown in the DCA of the risk of mental health problems nomogram in Figure 7, the model for predicting the risk of mental health problems for factory workers and miners in this study was more practically relevant if the threshold probability of patients was >10%.

4. Discussion

To our knowledge, this is the first study to develop an easy-to-use nomogram to predict the mental health risks of factory workers and miners. The nomogram developed using the training set data contain 12 items for education, professional title, age, CMBI, ERI, asbestos dust, hypertension, diabetes, working hours per day, working years, marital status, and work schedule. In addition, validation has shown that nomogram model has good accuracy and discriminatory power. Our novel nomogram can be used in any setting to provide a rapid assessment of mental health risks and to help identify patients with mental health risks, saving time compared to previous mental health investigations and improving on the lack of entries in previous investigations related to the specific working environment of factory workers and miners. The AUC of training group and verification group were 0.785 and 0.784 respectively, showing moderate discriminatory and calibration power.

A review of the literature found that the vast majority of studies constructed nomograms to predict clinical disorders, with less literature used to predict psychological problems. In a study to predict the correlates of suicide attempts in a Chinese population with major depressive disorder, the C-index was 0.715 and the C-index in the internal validation set was 0.703, and the calibration curve of the column line plot also showed good agreement between the predicted and observed risk of suicide attempts. The variables in the nomogram included socio-demographic information and clinical variables including age, duration, number of episodes, age at onset, number of hospitalizations, characteristics of anxiety and psychiatric symptoms, marital status, income, education level and employment status [27]. In another study that created a nomogram to predict the risk of psychosocial and behavioral problems in children and adolescents during the COVID-19 pandemic, the C index exceeded 0.800 and the calibration curve also showed good predictive accuracy. The variables covered three subject areas, namely demographic information, the psychosocial impact of the epidemic such as homework time and sedentary time, and the Child Behaviour Checklist score (CBCL) for the evaluation of psychological problems [38]. In this study, 7,118 participants were randomly divided into a training group (n=4,955) and a validation group (n=2,163) in a ratio of 3:1, involving a total of 23 features, and 12 features were selected by LASSO regression. The nomogram could be a useful tool to better identify patients with mental health problems, as it not only covered comprehensive information, including demographic information, job burnout, occupational stress, chronic diseases and occupational exposure factors closely related to factory workers and miners, but also was simple to operate and easy to use. In the validation group the Brier score was 0.176, the calibration slope was 0.970 and the calibration curve of nomogram showed a good fit. The AUC of training group and verification group were 0.785 and 0.784 respectively. Compared to the two studies above, our nomogram showed good accuracy and discrimination, and more comprehensive coverage in this nomogram model. Therefore, the possibility of early intervention for patients with high-risk mental health problems will be increased by covering multiple information and easy to use nomogram modal, especially for factory workers and miners with poor working conditions, relatively low levels of education and low patience.

Mental health problems were very common in the group of factory workers and miners, and the prevalence of mental health of them was found to be 37.08% in our study. Notably, the CMBI showed the most significant score (score = 100) and the ERI also had a high score (score = 43) in mental health problem incidence risk nomogram, which indicated that both of them were relatively important factors for mental health problems among the group of factory workers and miners. Our finding was consistent with other studies that had shown that occupational stress was a significant predictor of anxiety and was negatively associated with mental health. In addition, there is a high correlation between burnout and depression [39].

In line with previous studies, working years was also an important influential factor in this study. Related study has shown that employment could improve patients' mental health, while unemployment could lead to a deterioration in mental health [40]. In China, workers' working years is an important aspect of employment, and researchers have studied this aspect and found that precarious employment is a source of stress for individuals and predisposes them to mental health problems [41]. In addition, environmental factors were also one of the influential factors of mental health problems in our study. Relevant studies have found that exposure to air pollution is associated with increased suicide risk and depressive symptoms [42]. Hypertension and diabetes were the influential factors in this study. A study has shown that the prevalence of depression in adults with type 1 diabetes (T1D) is approximately three times higher than in the non-diabetic population [43]. Furthermore, there is a recognized association between hyperglycemia and depression, but the underlying biological mechanisms of this association are unclear [44].

Factory workers and miners were inevitably exposed to occupational hazards such as benzene and asbestos dust in their working environment. According to statistics, a total of nearly 2 million workers are exposed to various occupational hazards and over 16 million people worked in toxic and hazardous enterprises, involving more than 30 different types of operations, of which factory workers and miners is the one [45]. Similarly, the occupational hazard asbestos dust was selected as a predictor of risk for mental health problems in this study. Our study found that the work schedules of factory workers and miners were vary and the phenomenon of night shifts was very common, which inevitably affected their normal sleep. Some studies have shown that sleep problem is a risk factor for a variety of mental health and chronic diseases. Lack of sleep or poor sleep quality could lead to abnormalities in the body's self-regulatory functions and disturbances in the circadian rhythm of the biological clock, which in turn could suffer from negative emotions such as anxiety and depression [46]. Professional title and education level were also important influences on mental health issues. In the workplace, generally speaking, the higher the professional title and education level, the higher the status of the worker in the company and the greater the role played in the position. The number of studies on socio-economic status and mental health had increased in recent years. Some of these studies have shown that major depression is higher in the low socio-economic status group [47]. It has also been suggested that education itself is the best indicator of socio-economic status [48]. Marital status was one of the influential factors for mental health problems. Many studies have found an association between mental health and gender, marital status, lifestyle and working conditions, and it has been shown that poor mental health in women is associated with divorce or widowhood [49]. In this study, working more than seven hours a day was a determinant factor on mental health problems, which was consistent with other studies that had shown that long working hours could have a negative impact on employees' mental health and that excessive workloads could increase workers' fatigue, which in turn could lead to anxiety and depression [50].

In China, there are many problems in identifying people with mental health problems due to uneven and imperfect levels of medical development across regions. Some studies have shown that in mainland China, general practitioners, surgeons and primary health care workers often have little or no mental health training, which prevents them from providing basic mental health services [51]. Non-mental health professionals in general hospitals learn about mental illness on their own, rather than learning about it during their formal education [52]. Therefore, this study designed a simple and comprehensive nomogram to address the issue of timely detection and effective interventions for people with mental health problems, so that people at risk of mental health problems could easily calculate their probability of suffering from mental health problems without the help of medical staff. This study has several strengths. First, to our knowledge, this is the first model to develop and assess the likelihood of mental health problems in a group of factory workers and miners. Secondly, the nomogram in this study includes demographic information, job burnout, occupational stress, chronic illnesses, and also occupational exposure factors that are closely related to factory workers and miners, allowing for a more accurate assessment of the risk of morbidity among them, as well as providing a methodological reference for other related studies.

5. Limitations

This study also has several limitations. Firstly, we have considered many influential factors including demographics, job burnout, occupational stress and occupational exposure factors, but we are still not certain whether all possible influences are covered. Secondly, while the robustness of our nomogram was extensively validated internally in the same population, external validation is lacking for other populations in other regions and countries. Nomogram need to be externally assessed in a wider population.

Patient and public involvement

Neither patients nor members of the public had any involvement in the design of this study.

Acknowledgements The authors are grateful to all participants and investigators.

Contributions Y.L., Q.L., and T.L. are responsible for conceptualization; Y.L. is responsible for methodology, software, formal analysis, resources, and visualization; Q.L. and T.L. are responsible for the original draft preparation; Q.L. and H.Y. are responsible for reviewing; Q.L. is responsible for editing; T.L. is responsible for supervision. Yaoqin Lu and Qi Liu contributed equally to this work.

Funding This work was supported by the Natural Science Foundation of Xinjiang Uygur Autonomous Region (grant number 2020D01A27), the Postgraduate Innovation Project of Xinjiang Uyghur Autonomous Region (grant number XJ2021G215), the Outstanding Young Scientist Training Program of Urumqi Science and Technology Talent Project (grant number N/A), the Public Health and Preventive Medicine, the 13th Five-Year Plan Key Subject of Xinjiang Uygur Autonomous Region (grant number

N/A), The funders were not involved in the conception, design, analysis or interpretation of this study.

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval The study was approved by the ethics committee of Urumqi Center for Disease Control and Prevention (20181123)

Data availability statement Data are available on reasonable request. The data used in this study are available from the corresponding authors on reasonable request.

References

[1] WHO Terminology Information System [online glossary] http://www.who.int/health-systems-performance/docs/glossary.html.

[2] Wang Y, Liu X, Qiu J, Wang H, Liu D, Zhao Z, Song M, Song Q, Wang X, Zhou Y, Wang W. Association between ideal cardiovascular health metrics and suboptimal health status in Chinese population. *Sci Rep* 2017;7:14975.

[3] Prince M, Patel V, Saxena S, Maj M, Maselko J, Phillips MR, Rahman A. No health without mental health. *Lancet*. 2007;370:859–77.

[4] Adjaye-Gbewonyo K, Avendano M, Subramanian S.V, Kawachi I. Income inequality and depressive symptoms in South Africa: A longitudinal analysis of the National Income Dynamics Study. *Health Place* 2016;42:37–46.

[5] Vos T, Barber R.M, Bell B, Bertozzi-Villa A, Biryukov S, Bolliger I, Charlson F, Davis A, Degenhardt L, Dicker D, *et al*. Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990–2013: A systematic analysis for the Global Burden of Disease Study 2013. *Lancet* 2015;386:743–800.

[6] Huang Z, Li T, Xu M. Are there heterogeneous impacts of national income on mental health?.*Int J Environ Res Public Health* 2020;17:7530.

[7] Asadullah M.N, Xiao S, Yeoh E. Subjective well-being in China, 2005–2010: The role of relative income, gender, and location. *China Econ Rev* 2018;48:83–101.

[8] Butterworth P, Rodgers B, Windsor T.D. Financial hardship, socio-economic position and depression: Results from the PATH Through Life Survey. *Soc Sci Med*. 2009;69:229–237.

[9] Zhang X, Zhang X, Chen X. Happiness in the Air: How Does a Dirty Sky Affect Mental Health and Subjective Well-being? *J Environ Econ Manag* 2017;85:81–94.

[10] Wahlbeck K. Public mental health: The time is ripe for translation of evidence into practice. *World Psychiatry*. 2015;14:36–42.

[11] Luciano AE, Drake RE, Bond GR, Becker DR, Carpenter-Song E, Lord S, Swarbrick P and Swanson SJ. IPS Supported employment: a review. *J Vocat Rehabil* 2014;40:1–13.

[12] Jia Z, *et al*. Exposure to ambient air particles increases the risk of mental disorder: findings from a natural experiment in Beijing. *Int J Environ Res Public Health* 2018;15:160.

[13] Szyszkowicz M, Willey J. B, Grafstein E, Rowe B. H, Colman I. Air pollution and emergency department visits for suicide attempts in vancouver, Canada. *Environ Health Insights* 2010;4:79–86.

[14] Michael J. S. International classification of sleep disorders-third edition : highlights and modifications. *Chest* 2014;146:1387–1394.

[15] AbuRuz ME, Al-Dweik G. Depressive symptoms and complications early after acute myocardial infarction: gender differences. *Open Nurs J* 2018;12:205–214.

[16] Wurm W, Vogel K, Holl A, *et al.* Depression-burnout overlap in physicians. *PLoS One* 2016;11:e0149913.

[17] Porru F, Robroek S J W, Bültmann U, *et al*. Mental health among university students: The associations of effort-reward imbalance and overcommitment with psychological distress. *J Affect Disord* 2021;282:953-961.

[18] Johnson AK, Blackstone SR, Skelly A, Simmons W. The relationship between depression, anxiety, and burnout among physician assistant students: a multi-institutional study. *Health Professions Edu* 2020;6:420–427.

[19] Hua D, Kong Y, Li W, Han Q, Zhang X, Zhu LX, Wan SW, Liu Z, Shen Q, Yang J, He HG, Zhu J. Frontline nurses' burnout, anxiety, depression, and fear statuses and their associated factors during the COVID-19 outbreak in Wuhan, China: a large-scale cross-sectional study. *EClinical Med* 2020;24:100424.

[20] Armah E K, Adedeji J A, Boafo B B, *et al*. Underground gold miner exposure to noise, diesel particulate matter and crystalline silica dust. *J Health Pollut* 2021; 11(29):210301.

[21] Hall N B, Blackley D J, Halldin C N, *et al*. Current review of pneumoconiosis among US coal miners. *Curr Environ Health Rep*, 2019; 6(3):137-147.

[22] Wickramatillake B A, Fernando M A, Frank A L. Prevalence of asbestos-related disease among workers in sri lanka. *Ann Glob Health*, 2019, 85(1):108.

[23] Liu FD, Pan ZQ, Liu SL, Chen L, Ma JZ, Yang ML, Wang NP. The estimation of the number of underground coal miners and the annual dose to coal miners in China. *Health Phys* 2007;93:127–132.

[24] Yong X, Gao X, Zhang Z, *et al*. Associations of occupational stress with job burn-out, depression and hypertension in coal miners of Xinjiang, China: A cross-sectional study. *BMJ open* 2020;10: e036087.

[25] P. Bech, J. Bille, S. B. Møller, L. C. Hellström, S. D. Østergaard. Psychometric validation of the Hopkins Symptom Checklist (SCL-90) subscales for depression, anxiety, and interpersonal sensitivity. *J Affect Disord* 2014;160:98–103.

[26] J. Zhang, X. Zhang. Chinese college students' SCL-90 scores and their relations to the college performance. *Asian J Psychiatr* 2013;6:134–140.

[27] Liang S, Zhang J, Zhao Q, *et al*. Incidence trends and risk prediction nomogram for suicidal attempts in patients with major depressive disorder. *Front Psychiatry*, 2021, 12: 644038.

[28] Kan S K, Chen N N, Zhang Y L. Predicting the risk of suicide attempt in a depressed population: Development and assessment of an efficient predictive nomogram. *Psychiatry Res*, 2022, 310: 114436.

[29] Lu Y, Zhang Z, Yan H, *et al*. Effects of occupational hazards on job stress and mental health of factory workers and miners: A propensity score analysis. *BioMed Res Int*, 2020, 2020: 1754897.

[30] Li X, Jiang T, Sun X, *et al*. The relationship between occupational stress, musculoskeletal disorders and the mental health of coal miners: The interaction between bdnf gene, tph2 gene polymorphism and the environment. *J Psychiatr Res*, 2021, 135: 76-85.

[31] Derogatis L, Lipman R.S, Covi L. SCL-90: an outpatient psychiatric rating scale—preliminary

report. *Psychopharmacol Bull* 1973;9:13–28.

[32] Crespo-Maraver M, Doval E, Fernã N.J, Gimã Nez-Salinas J, Prat G, Bonet P. Caregiver's health: adaption and validation in a Spanish population of the Experience of Caregiving Inventory (ECI) Gac. *Sanit* 2018;33:348–355.

[33] Dang W, Xu Y, Ji J, *et al*. Study of the scl-90 scale and changes in the chinese norms. *Front Psychiatry*, 2020, 11: 524395.

[34] Siegrist J. Adverse health effects of high-effort/low-reward conditions. *J Occup. Health Psychol*. 1996;1:27–41.

[35] Siegrist J, Wege N, Pühlhofer F, Wahrendorf M. A short generic measure of work stress in the era of globalization: Effort-reward imbalance. *Int Arch Occup Environ Health* 2009;82:1005–1013.

[36] Zhang Z, Lu Y, Yong X, *et al*. Effects of occupational radiation exposure on job stress and job burnout of medical staff in xinjiang, China: A cross-sectional study. *Med Sci Monit* 2020;26: e927848.

[37] Freudenberger HJ. Staff burnout. *J Soc Issues* 1974;30:159–165.

[38] Wang L, Chen L, Jia F, *et al*. Risk factors and prediction nomogram model for psychosocial and behavioural problems among children and adolescents during the covid-19 pandemic: A national multicentre study: Risk factors of childhood psychosocial problems. *J Affect Disord*, 2021, 294: 128-136.

[39] Occupational Stress and Employees Complete Mental Health: A Cross-Cultural Empirical Study

[40] Knapp M and Wong G. Economics and mental health: the current scenario. *World Psychiatry* 2020;19:3–14.

[41] Benach, J, Vives, A, Amable, M, Vanroelen, C, Tarafa, G, & Muntaner, C. Precarious employment: understanding an emerging social determinant of health. *Annu Rev Public Health* 2014;35:229-53.

[42] Bakian A. V. *et al*. Acute air pollution exposure and risk of suicide completion. *Am J Epidemiol* 2015;181:295–303.

[43] Barnard KD, Skinner TC, Peveler R. The prevalence of co-morbid depression in adults with Type 1 diabetes: systematic literature review. *Diabet Med* 2006;23:445–448.

[44] Gilsanz P, Karter AJ, Beeri MS, Quesenberry CP, Whitmer RA. The bidirectional association between depression and severe hypoglycemic and hyperglycemic events in type 1 diabetes. *Diabetes Care* 2018;41:446–452.

[45] Lu Y, Zhang Z, Yan H, *et al*. Effects of occupational hazards on job stress and mental health of factory workers and miners: A propensity score analysis. *BioMed Res Int* 2020;2020:1754897.

[46] Shi L, Liu Y, Jiang T, *et al*. Relationship between mental health, the clock gene, and sleep quality in surgical nurses: A cross-sectional study. *BioMed Res Int* 2020;2020:4795763.

[47] Sallis JF, Saelens BE, Frank LD *et al*. Neighborhood built environment and income: examining multiple health outcomes. *Soc Sci Med* 2009;68:1285–1293.

[48] Winkleby MA, Jatulis DE, Frank E *et al*. Socioeconomic status and health: how education, income, and occupation contribute to risk-factors for cardiovascular disease. *Am J Public Health* 1992;82:816–820.

[49] Skapinakis P, Bellos S, Koupidis S, Grammatikopoulos I, Theodorakis P.N, Mavreas V. Prevalence and sociodemographic associations of common mental disorders in a nationally representative sample of the general population of Greece. *BMC Psychiatry* 2013;13:163.

[50] Virtanen M, Ferrie JE, Singh-Manoux A, Shipley MJ, Stansfeld SA, Marmot MG *et al*. Long working hours and symptoms of anxiety and depression: a 5-year follow-up of the Whitehall II

study. *Psychol Med* 2011;41:2485–2494.

[51] Phillips MR, Zhang J, Shi Q, Song Z, Ding Z, Pang S, *et al*. Prevalence, treatment, and associated disability of mental disorders in four provinces in China during 2001–05: an epidemiological survey. *Lancet* 2009;373:2041–2053.

[52] Wu Q, Luo X, Chen S, *et al*. Mental health literacy survey of non-mental health professionals in six general hospitals in hunan province of China. *PloS one* 2017;12:e0180327.

Figure legends

Fig.1. Flow diagram of the participants involved in this study

Fig.2. Feature selection using the LASSO binary logistic regression model. (A) Feature selection for the LASSO binary logistic regression model. The partial likelihood deviation (binomial deviation) curve was plotted against lambda by validating the optimal parameter lambda in the LASSO model. Dotted vertical lines were drawn based on 1 SE of the minimum criteria (the 1-SE criteria). (B) Feature selection was performed using the LASSO binary logistic regression model. A Coefficient profile weas plotted based on the lambda series in Figure 1(A), and 12 features with non-zero coefficients were selected by optimal lambda.

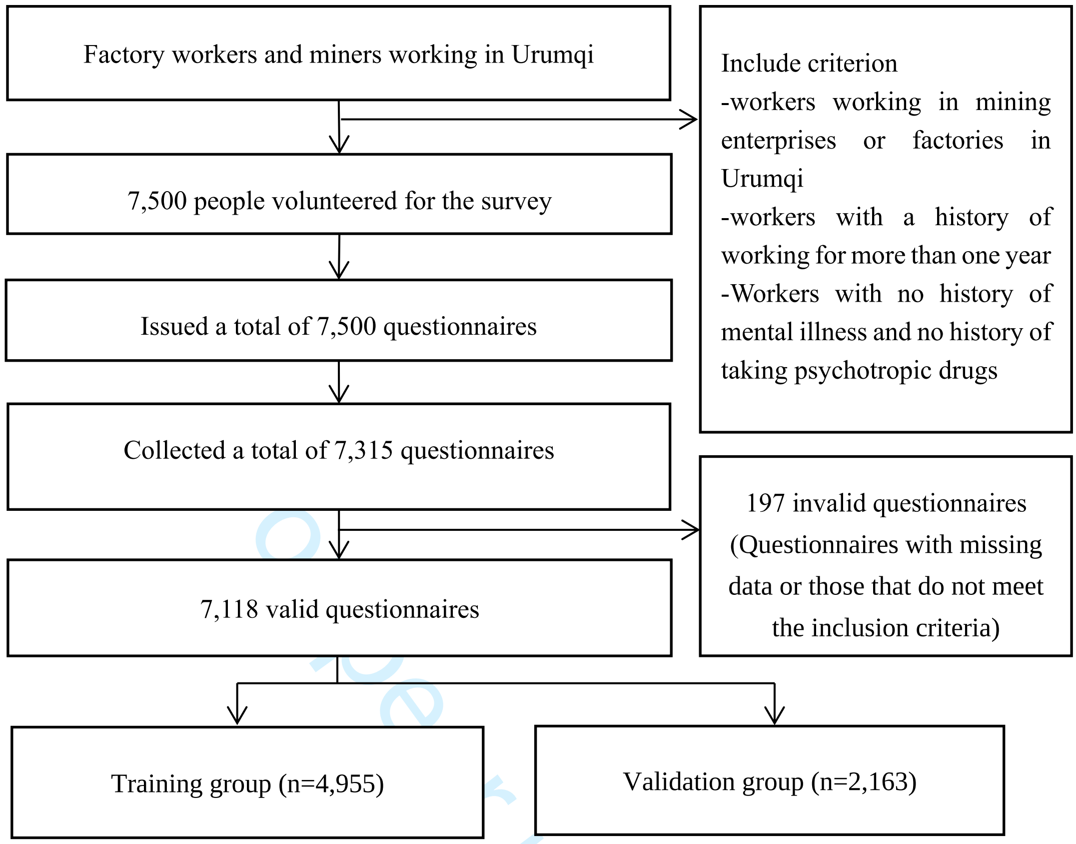
Fig.3. The forest plot of the OR of the selected feature.

Fig.4. Developed mental health problems incidence risk nomogram. The mental health problems incidence risk nomogram was developed in the array, with education, professional title, age, CMBI, ERI, asbestos dust, hypertension, diabetes, working hours per day, working years, marital status, and work schedule incorporated.

Fig.5. Calibration curves of the mental health problems incidence risk nomogram prediction in validation group. The x-axis represents the predicted risk of mental health problems. y-axis represents the actual diagnosed risk of mental health problems. The diagonal dashed line represents the perfect prediction of the ideal model. The solid lines represent the performance of the column plots, where closer to the diagonal dashed line indicates a better prediction.

Fig.6. ROC curves for training and validation groups. The y-axis represents the true positive rate of risk prediction. The x-axis represents the false positive rate of risk prediction. The ROC curves for the training and validation groups are shown in black and red.

Fig.7. Decision curve analysis for mental health problems incidence risk nomogram. The y-axis measures the net benefit. The solid red line represents nomogram of the risk of developing a mental health problem. The light blue dashed line represents the hypothesis that all participants were diagnosed with a mental health problem. The black dashed line represents the hypothesis that there is no risk of a mental health problem. The DCA showed that using this mental health problem incidence risk nomogram in the current study to predict mental health problem incidence risk increase in benefit than the intervention all patients or no intervention all patient if the threshold probability of a patients and a doctor is >10%.


Factory workers and miners working in Urumqi
7,500 people volunteered for the survey
Issued a total of 7,500 questionnaires
Collected a total of 7,315 questionnaires
7,118 valid questionnaires
Include criterion
-workers working in mining enterprises or factories in Urumqi
-workers with a history of working for more than one year
-Workers with no history of mental illness and no history of taking psychotropic drugs
197 invalid questionnaires (Questionnaires with missing data or those that do not meet the inclusion criteria)
Training group (n=4,955)
Validation group (n=2,163)

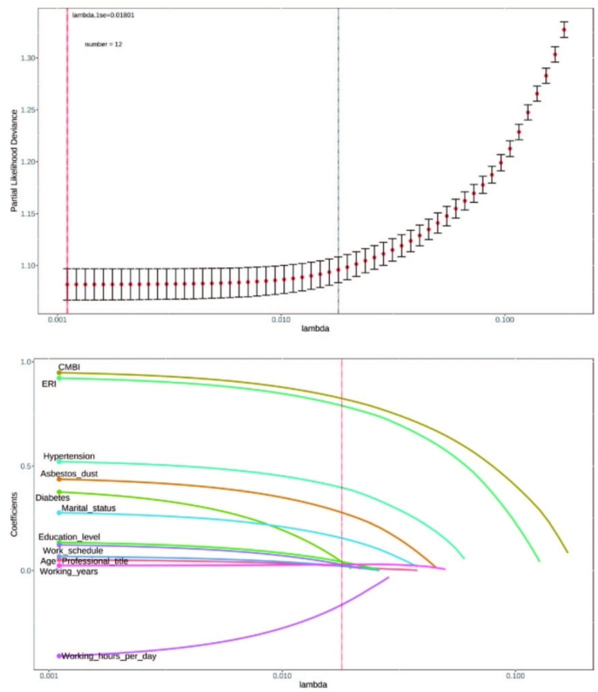
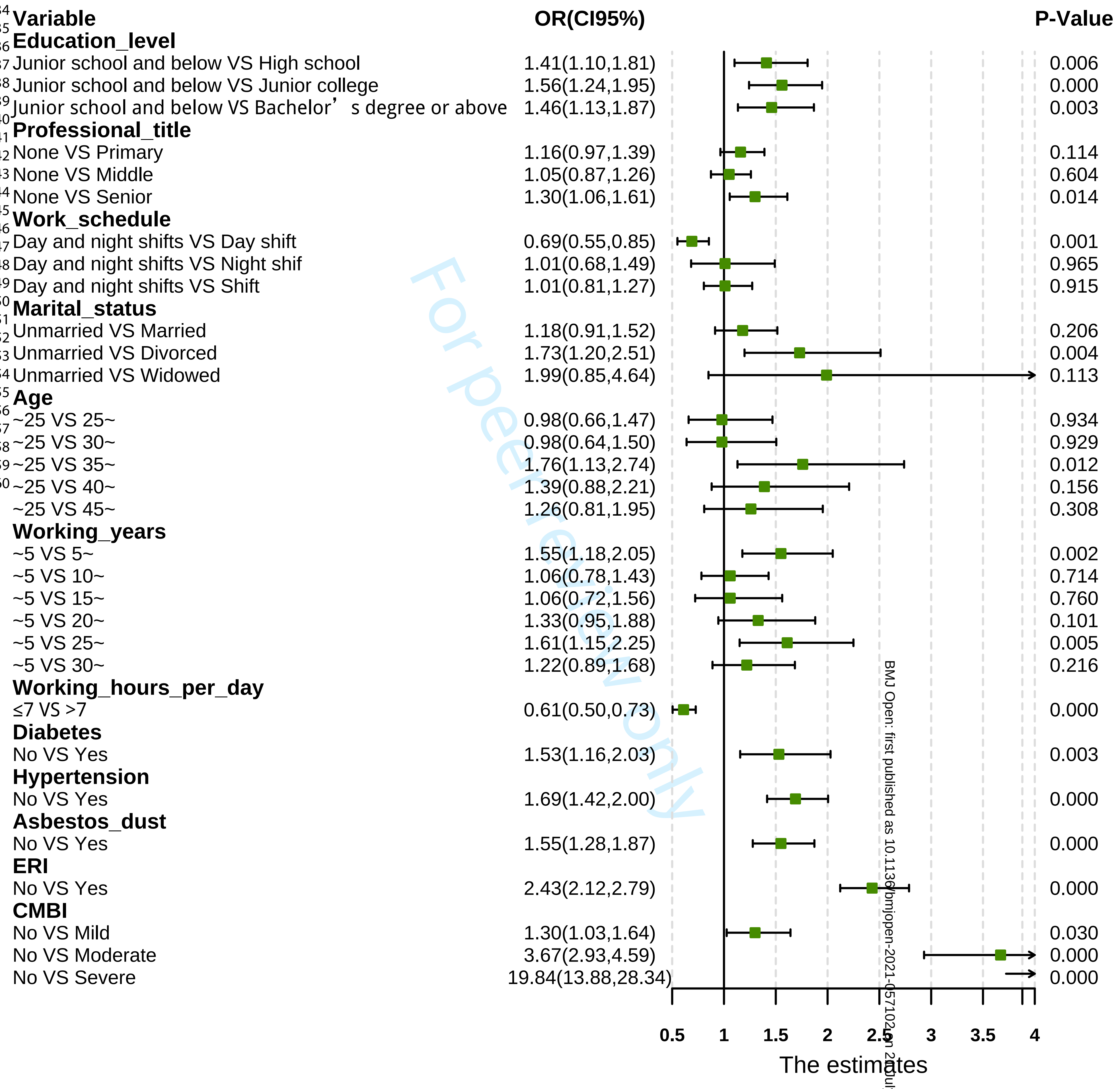
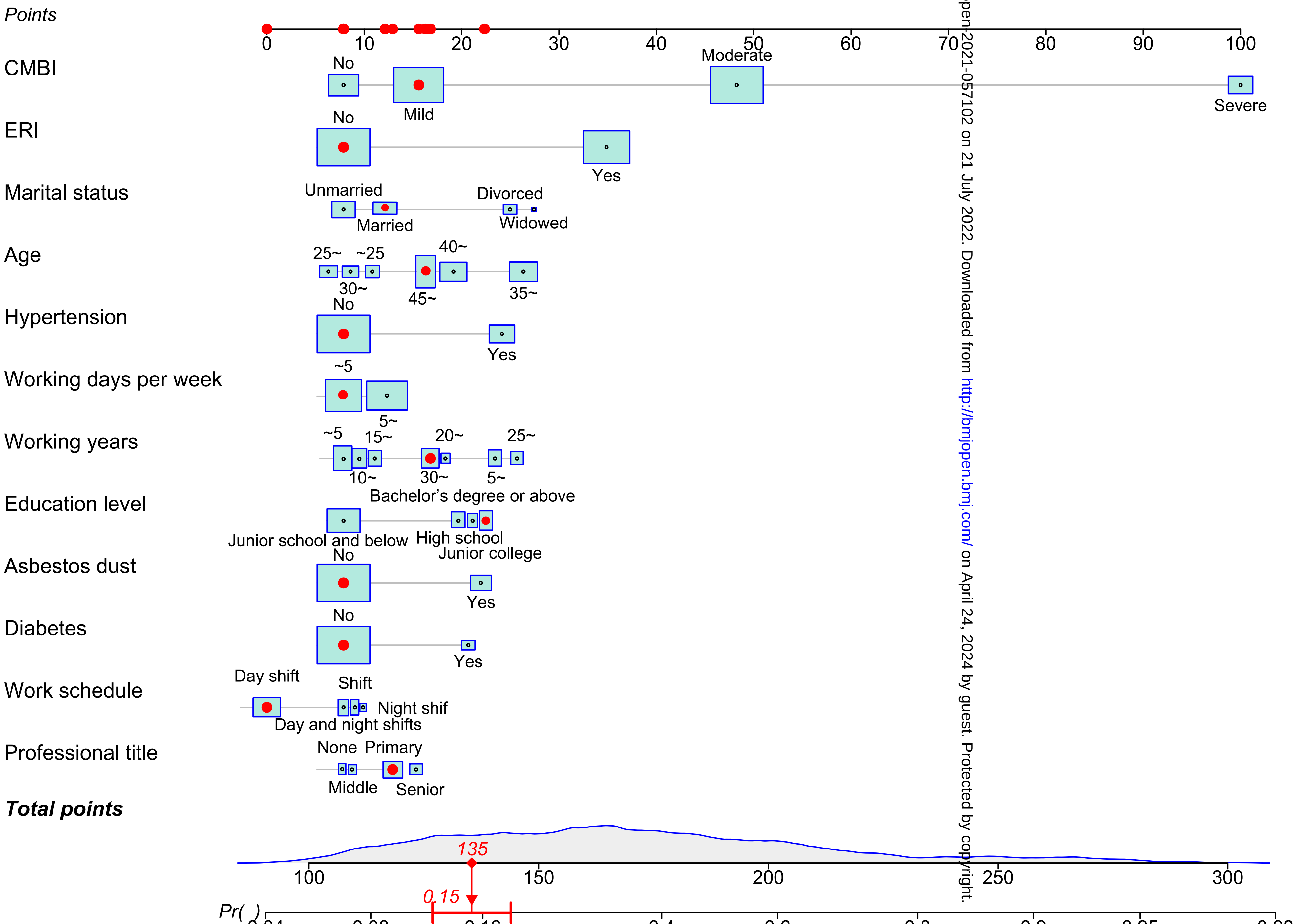
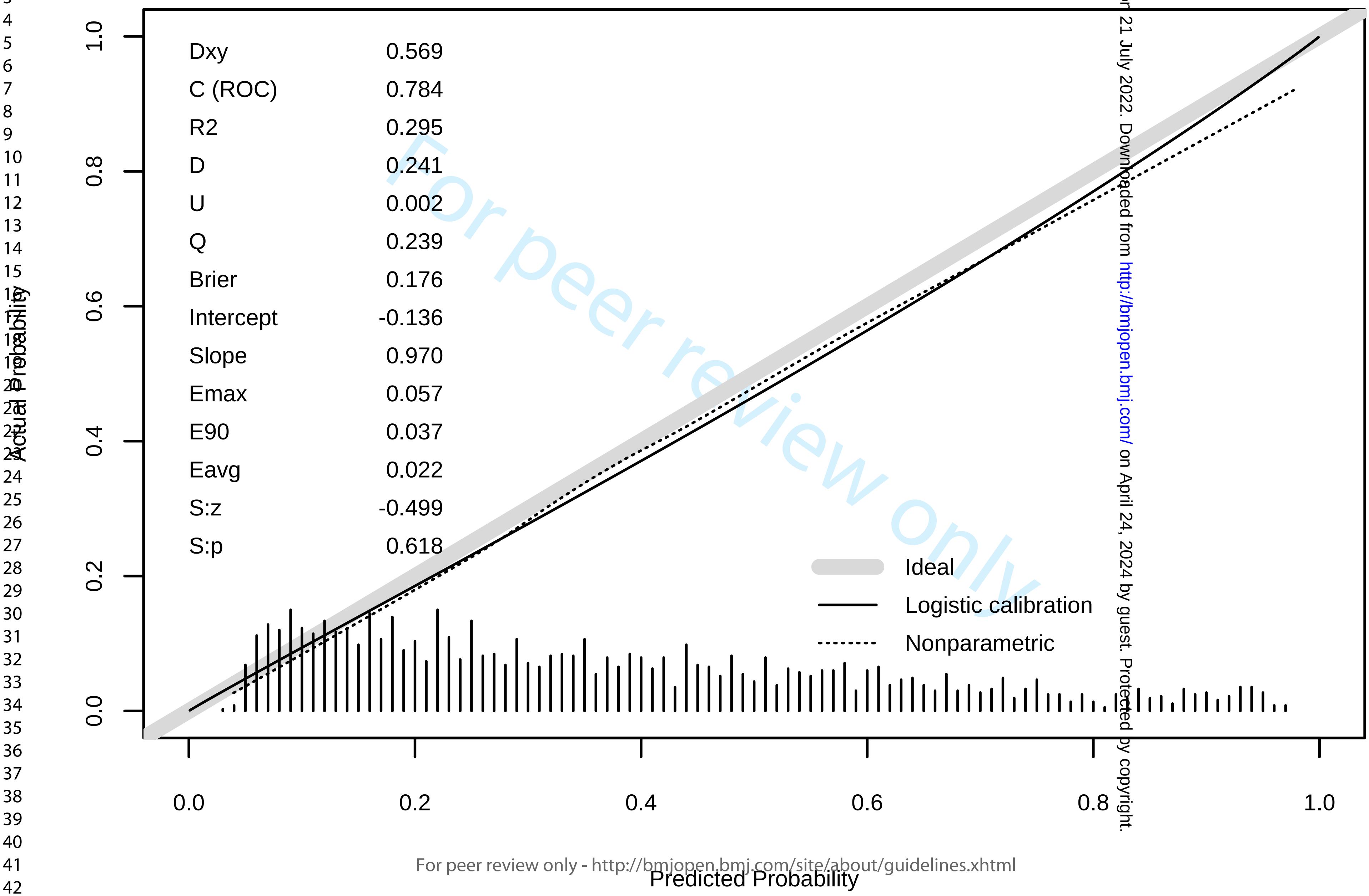


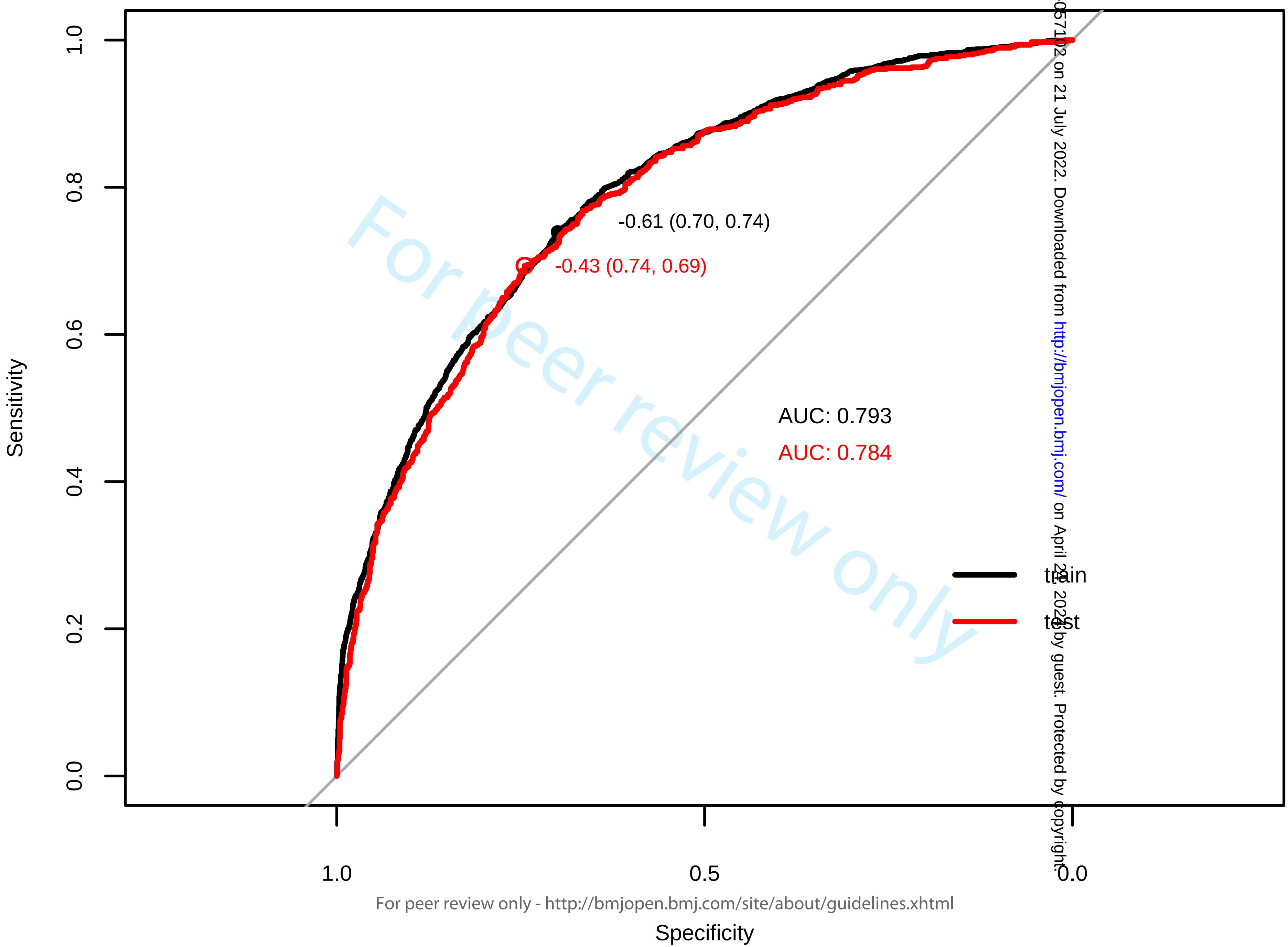
Fig.2. Feature selection using the LASSO binary logistic regression model. (A) Feature selection for the LASSO binary logistic regression model. The partial likelihood deviation (binomial deviation) curve was plotted against lambda by validating the optimal parameter lambda in the LASSO model. Dotted vertical lines were drawn based on 1 SE of the minimum criteria (the 1-SE criteria). (B) Feature selection was performed using the LASSO binary logistic regression model. A Coefficient profile weas plotted based on the lambda series in Figure 1(A), and 12 features with non-zero coefficients were selected by optimal lambda.

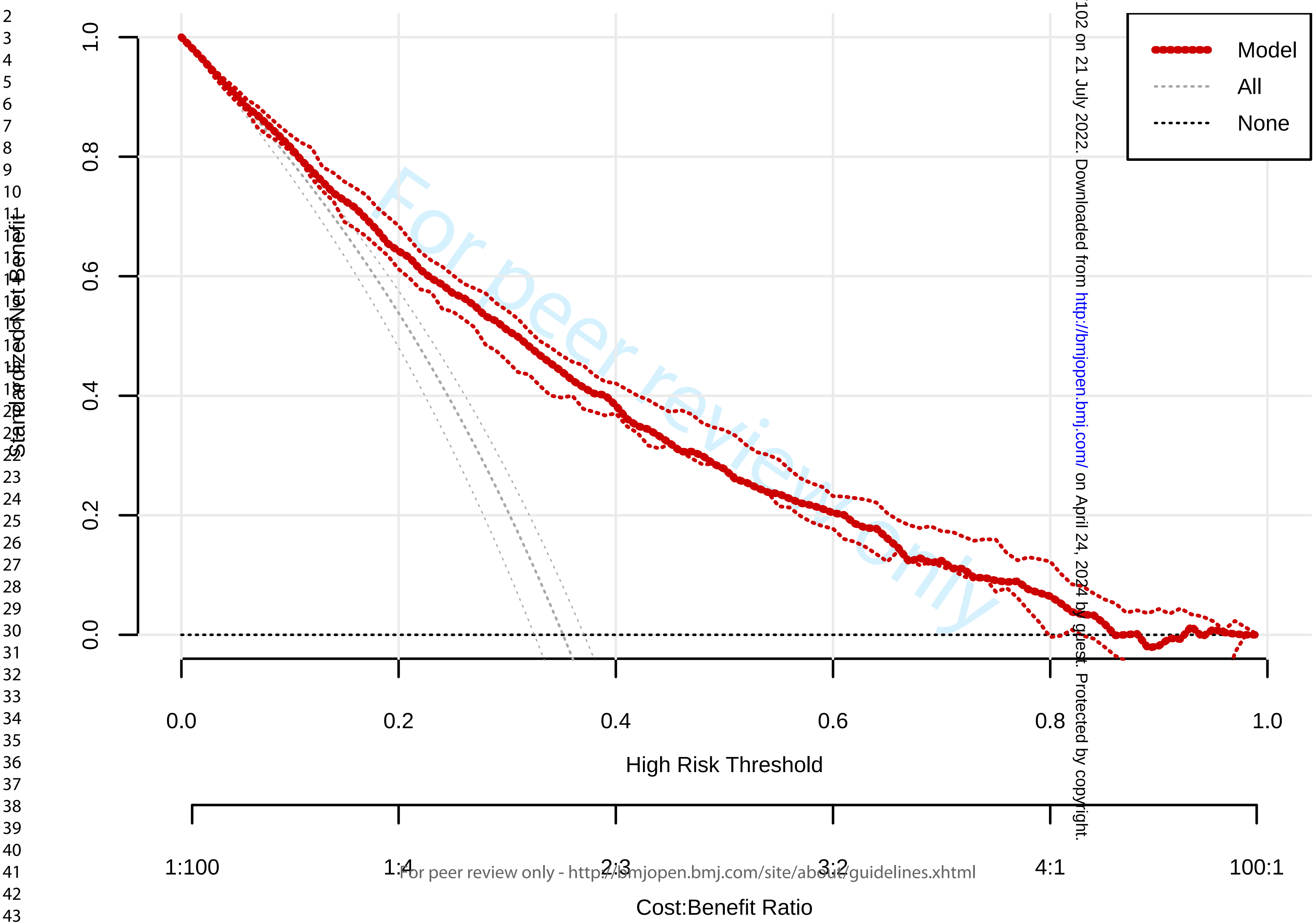
210x297mm (300 x 300 DPI)


Variable
OR(CI95%)
P-Value
Education_level
Junior school and below VS High school 1.41(1.10,1.81) 0.006
Junior school and below VS Junior college 1.56(1.24,1.95) 0.000
Junior school and below VS Bachelor’s degree or above 1.46(1.13,1.87) 0.003
Professional_title
None VS Primary 1.16(0.97,1.39) 0.114
None VS Middle 1.05(0.87,1.26) 0.604
None VS Senior 1.30(1.06,1.61) 0.014
Work_schedule
Day and night shifts VS Day shift 0.69(0.55,0.85) 0.001
Day and night shifts VS Night shif 1.01(0.68,1.49) 0.965
Day and night shifts VS Shift 1.01(0.81,1.27) 0.915
Marital_status
Unmarried VS Married 1.18(0.91,1.52) 0.206
Unmarried VS Divorced 1.73(1.20,2.51) 0.004
Unmarried VS Widowed 1.99(0.85,4.64) 0.113
Age
~25 VS 25~ 0.98(0.66,1.47) 0.934
~25 VS 30~ 0.98(0.64,1.50) 0.929
~25 VS 35~ 1.76(1.13,2.74) 0.012
~25 VS 40~ 1.39(0.88,2.21) 0.156
~25 VS 45~ 1.26(0.81,1.95) 0.308
Working_years
~5 VS 5~ 1.55(1.18,2.05) 0.002
~5 VS 10~ 1.06(0.78,1.43) 0.714
~5 VS 15~ 1.06(0.72,1.56) 0.760
~5 VS 20~ 1.33(0.95,1.88) 0.101
~5 VS 25~ 1.61(1.15,2.25) 0.005
~5 VS 30~ 1.22(0.89,1.68) 0.216
Working_hours_per_day
≤7 VS >7 0.61(0.50,0.73) 0.000
Diabetes
No VS Yes 1.53(1.16,2.03) 0.003
Hypertension
No VS Yes 1.69(1.42,2.00) 0.000
Asbestos_dust
No VS Yes 1.55(1.28,1.87) 0.000
ERI
No VS Yes 2.43(2.12,2.79) 0.000
CMBI
No VS Mild 1.30(1.03,1.64) 0.030
No VS Moderate 3.67(2.93,4.59) 0.000
No VS Severe 19.84(13.88,28.34) 0.000
0.5 1 1.5 2 2.5 3 3.5 4
The estimates


Points
0
10
20
30
40
50
60
70
80
90
100
CMBI
No
Mild
Moderate
Severe
ERI
No
Yes
Marital status
Unmarried
Married
Divorced
Widowed
Age
25~
~25
30~
40~
45~
35~
Hypertension
No
Yes
Working days per week
~5
5~
Working years
~5
15~
20~
25~
10~
30~
5~
Education level
Bachelor's degree or above
Junior school and below
High school
Junior college
Asbestos dust
No
Yes
Diabetes
No
Yes
Work schedule
Day shift
Shift
Night shif
Day and night shifts
Professional title
None
Primary
Middle
Senior
Total points
135
100
150
200
250
300
0.15
Pr( )
0.04
0.08
0.16
0.4
0.6
0.8
0.9
0.95
0.98


Dxy 0.569
C (ROC) 0.784
R2 0.295
D 0.241
U 0.002
Q 0.239
Brier 0.176
Intercept -0.136
Slope 0.970
Emax 0.057
E90 0.037
Eavg 0.022
S:z -0.499
S:p 0.618
Ideal
Logistic calibration
Nonparametric
Actual Probability
Predicted Probability
0.0
0.2
0.4
0.6
0.8
1.0


1.0
0.8
0.6
0.4
0.2
0.0
Sensitivity
-0.61 (0.70, 0.74)
-0.43 (0.74, 0.69)
AUC: 0.793
AUC: 0.784
train
test
1.0
0.5
0.0
Specificity


Standardized Net Benefit
1.0
0.8
0.6
0.4
0.2
0.0
Model
All
None
0.0
0.2
0.4
0.6
0.8
1.0
High Risk Threshold
1:100
1:4
2:3
3:2
4:1
100:1
Cost:Benefit Ratio

STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No.	Recommendation	Page No.	Relevant text from manuscript
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1	
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	1	
Introduction				
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	2	
Objectives	3	State specific objectives, including any prespecified hypotheses	3	
Methods				
Study design	4	Present key elements of study design early in the paper	4	
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	4	
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants	4	
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case		
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	4	
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	4	
Bias	9	Describe any efforts to address potential sources of bias	4	
Study size	10	Explain how the study size was arrived at	4	

Continued on next page

Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	6
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	6
		(*b*) Describe any methods used to examine subgroups and interactions	
		(*c*) Explain how missing data were addressed	4
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed *Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy	4
		(*e*) Describe any sensitivity analyses	
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	7
		(b) Give reasons for non-participation at each stage	7
		(c) Consider use of a flow diagram	7
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	7
		(b) Indicate number of participants with missing data for each variable of interest	4
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time	
		Case-control study—Report numbers in each exposure category, or summary measures of exposure	
		Cross-sectional study—Report numbers of outcome events or summary measures	10
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	
		(*b*) Report category boundaries when continuous variables were categorized	
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	

Continued on next page

Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	
Discussion			
Key results	18	Summarise key results with reference to study objectives	12
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	15
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	13
Generalisability	21	Discuss the generalisability (external validity) of the study results	13
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	15

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

Development and Validation of a Nomogram for Predicting the Risk of Mental Health Problems of Factory Workers and Miners

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-057102.R2
Article Type:	Original research
Date Submitted by the Author:	24-Jun-2022
Complete List of Authors:	Lu, Yaoqin; Xinjiang Medical University, School of Public Health; Urumqi Center for Disease Control and Prevention Liu, Qi; Xinjiang Medical University Affiliated First Hospital, Postgraduate Education Management Section,; Xinjiang Medical University, School of Public Health Yan, Huan; Xinjiang Medical University, Department of Nutrition and Food Hygiene; Xinjiang Autonomous Academy of Instrumental Analysis Liu, Tao; Xinjiang Medical University, School of Public Health
Primary Subject Heading:	Mental health
Secondary Subject Heading:	Public health
Keywords:	MENTAL HEALTH, PREVENTIVE MEDICINE, PUBLIC HEALTH

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Development and Validation of a Nomogram for Predicting the Risk of Mental Health Problems of Factory Workers and Miners

Yaoqin Lu,[1,2] Qi Liu,[1,5] Huan Yan[3,4], and Tao Liu[1]

[1]Department of Toxicology, School of Public Health, Xinjiang Medical University, Urumqi 830011, Xinjiang, China.
[2]Urumqi Center for Disease Control and Prevention, Urumqi 830026, Xinjiang, China.
[3]Department of Nutrition and Food Hygiene, College of Public Health, Xinjiang Medical University, Urumqi, Xinjiang, China.
[4]Xinjiang Engineering Technology Research Center for Green Processing of Nature Product Center, Xinjiang Autonomous Academy of Instrumental Analysis, Urumqi, Xinjiang, China.
[5]Graduate Education Management Section of Teaching and Scientific Research Department of the First Affiliated Hospital of Xinjiang Medical University, Urumqi 830017, Xinjiang, China.

Correspondence should be addressed to Tao Liu; xjmult@163.com

Yaoqin Lu and Qi Liu contributed equally to this work

Abstract

Objective A nomogram for predicting the risk of mental health problems was established in a population of factory workers and miners, in order to quickly calculate the probability of a worker suffering from mental health problems.

Methods A cross-sectional survey of 7,500 factory workers and miners in Urumqi was conducted by means of an electronic questionnaire using cluster sampling method. Participants were randomly assigned to the training group (70%) and the validation group (30%). Questionnaire-based survey was conducted to collect information. A least absolute shrinkage and selection operator (LASSO) regression model was used to screen the predictors related to the risk of mental health problems of the training group. Multivariate logistic regression analysis was applied to construct the prediction model. Calibration plots and receiver operating characteristic-derived area under the curve (AUC) were used for model validation. Decision curve analysis (DCA) was applied to calculate the net benefit of the screening model.

Results A total of 7,118 participants met the inclusion criteria and the data were randomly divided into a training group (n=4,955) and a validation group (n=2,163) in a ratio of 3:1. A total of 23 characteristics were included in this study and LASSO regression selected 12 characteristics such as education, professional title, age, CMBI, ERI, asbestos dust, hypertension, diabetes, working hours per day, working years, marital status, and work schedule as predictors for the construction of the nomogram. In the validation group the Brier score was 0.176, the calibration slope was 0.970 and the calibration curve of nomogram showed a good fit. The AUC of training group and verification group were 0.785 and 0.784 respectively.

Conclusion The nomogram combining these 12 characteristics can be used to predict the risk of suffering mental health problems, providing a useful tool for quickly and accurately screening the risk of mental health problems.

Key words Mental health; Predictor; Nomogram; Risk; Factory workers and miners

Strengths and limitations of this study

1. This is the first study to develop an easy-to-use nomogram to predict the mental health risks of factory workers and miners.
2. The AUC of training group and verification group were 0.785 and 0.784 respectively, showing moderate discriminatory and calibration power.
3. This nomogram model's variables are more comprehensive, including demographics, burnout, occupational stress and occupational hazards.
4. We had considered many influential factors, but we were still not certain whether all possible influences were covered.
5. There is a lack of external validation in other populations in other regions and countries.

1. Introduction

The World Health Organization (WHO) defines health as a state of complete physical, mental and social well-being and not merely the absence of disease or weakness [1]. Obviously, health is an organic unity of physical and mental well-being. People with good mental health are the precondition for the normal operation of our society. However, with the acceleration of people's pace of life, people are facing an increasing risk of poor health, which has become a global public health problem [2]. Mental health problems can not only take a toll on physical health such as increasing the risk of communicable and non-communicable diseases and even causing unintentional or intentional harm to others [3], but can also have a negative impact on the economy. For example, mental health disorders represent a growing part of the global burden of disease [4], with statistics showing that nearly one billion people worldwide currently suffer from a mental disorder, and mental illness is ranked as one of the leading causes of the global burden of disease [5]. Moreover, one study has estimated that due to the impact of mental illness, the global economy loses US $1 trillion every year [6].

As researchers around the world have delved into the field of mental health, factors such as gender, income levels, environment and education have been found to be associated with people's mental health problems [7-10]. Moreover, employment is also strongly associated with quality of life, higher self-esteem and fewer psychiatric symptoms [11]. In addition, in the context of the global challenges of climate change, an increasing number of scholars have been examining the epidemiological links between mental health and environmental factors. Some studies have suggested that mental health may be influenced by ambient temperature, and an association has been found between environmental pollutants, particularly fine particulate matter, and mental health problems [12]. A relevant study shows that with short-term exposure to ambient air pollution is associated with increased emergency room visits due to depression or suicide attempts [13]. Furthermore, other factors associated with mental health include sleep, diabetes, coronary artery disease and cardiovascular disease [14-15]. It is worth noting that job burnout and occupational stress are closely linked to mental health. Job burnout is an exhaustion state of physical and psychological that

often occurs in the work environment, and has a high correlation with depression. A large study of physicians found that of the 10.3% who met criteria for a major depressive episode, 50.7% were also affected by symptoms of burnout (OR 2.99) and indicated that worsening depression leads to a higher likelihood of burnout symptoms [16]. Occupational stress refers to a work environment where non-reciprocity of effort and reward may lead to strong negative emotions and distress. Related research has shown that the combination of high effort and low reward and over-commitment increases the risk of mental health problems such as depression [17]. Apparently, it is necessary to include the CMBI and ERI in this study to predict the risk of mental health problems among factory workers and miners. However, there are few studies that include these influences in a more comprehensive way in the practice of detecting mental health. Therefore, more accurate identification of mental health problems in populations requires a questionnaire that include a wider range of factors affecting factory workers and miners' mental health problems.

Factory workers and miners are a special group of workers with a relatively low overall level of education and are highly prone to suffering from mental health problems due to limited social support, excessive workload and irregular lifestyles, as well as occupational hazards such as noise and coal dust that they inevitably need to face in their working environment [18-19]. Through a review of the literature, our group found that coal dust, crystalline silica and noise pollution were common causes of health problems for workers in underground mines [20]. And, exposure to coal mine dust is a significant cause of pneumoconiosis in coal miners [21]. In addition, asbestos is one of the major occupational hazards in the daily work of workers in the construction and automotive industries [22]. China has the world's largest group of factory workers and miners, about 6 million [23], who are regularly involved in occupational hazards. Mental health problems which need to require a long process are known to be a syndrome caused by chronic stress. Factory workers and miners, represented by those engaged in coal mining, have a mental burden rating of 8.3, one of the highest mental burdens among 150 occupations [24]. This explains the high level of mental health problems among mine workers in previous studies, making the identification and treatment of mental health problems even more important. Therefore, it is essential to provide a viable and easy-to-apply tool for identifying workers at risk of mental health problems and thus for timely interventions.

There are many studies on mental health [25-26]; however, the results of previous studies lack consistency and mostly discuss factors influencing mental health, and most of them are single-center studies that focus on only certain aspects of mental health. Our study included common demographics, job burnout, occupational stress, chronic illness and occupational exposure factors to distinguish whether respondents suffered from mental health problems. In addition, there is a small body of literature that develops and validates a risk nomogram between depression and suicide to support timely intervention by clinicians. And the sample sizes of the two relevant studies were small, 474 and 273 depressed patients respectively [27-28]. Today, there is increasing recognition of the important role of mental health in achieving global development goals, and WHO has included mental health in the Sustainable Development Goals. However, there are no relevant studies that have used objective indicators for factory workers and miners to form a nomogram to predict mental health. Therefore, to bridge this gap in the literature and provide additional information for the prevention of mental health problems, we conducted a multicenter investigation to develop and validate an easy-to-use nomogram that combines objective information on demographics, job burnout, occupational stress and occupational hazards to comprehensively and accurately predict the prevalence of mental health problems among factory workers and miners.

2. Materials and Methods

2.1 Calculation of sample size

The sample size formula for the present illness rate survey, $n = \frac{z_{\alpha/2}^2 \times \mathrm{pq}}{\delta^2}$, p is the present-hazard rate, q=1-p, **δ** is the tolerance error, generally taken as 0.1p, $z_{\alpha/2}$ is the significance test statistic, $z_{\alpha/2}$=1.96 for α=0.05, then the formula is calculated as,$n = 400 \times \frac{q}{p}$. A cross-sectional study in Xinjiang showed that 38.27% of factory workers and miners had mental health problems [29]。 And a study revealed that 633 out of 1675 coal miners (37.8%) suffered from mental disorders between August 2018 and June 2019[30]. In this study, we assumed a 30% prevalence of mental health problem to obtain the maximum required sample size. which would calculate a sample size of 934, taking into account non-response and a 20% loss of questionnaires, which would require approximately 1168 people.

2.2. Participants

Participants in this cross-sectional survey were factory workers and mines in the Urumqi region, and the survey covered all districts and counties in the Urumqi region to avoid selection bias as far as possible. Specifically, this survey was conducted by means of whole-group random sampling from January to May 2019, and a total of 202 enterprises were selected, including 21 in Tianshan District, 30 in Shaibak District, 24 in Xinshi District, 22 in Shuimogou District, 56 in Jingkai District, 37 in Midong District, 9 enterprises in Dabancheng District and 3 enterprises in Urumqi County.

The inclusion criteria were as follows: (1) workers working in mining enterprises or factories in Urumqi; (2) workers with a history of working for more than one year; (3) Workers with no history of mental illness and no history of taking psychotropic drugs.

The exclusion criteria were the following: (1) factory workers and miners in non-Urumqi area; (2) working history of factories and mining enterprises less than 1 year; (3) a confirmed diagnosis of a mental health problem and a history of treatment and use of psychotropic medication; (4) Questionnaires with missing data were excluded.

An online electronic questionnaire was created using the Questionnaire Star platform to collect data. In the introductory section of the electronic questionnaire, we provide a paragraph stating that volunteers can choose to continue answering the survey if they wish to participate and the relevant data will be used for scientific research, or refuse to answer if they do not wish to participate in the survey. In addition, this survey was conducted by trained surveyors who explained the purpose, meaning, content and requirements of the questionnaire to all participants and provided on-site instructions to ensure the return rate of the questionnaire. All participants understood the purpose of the study and were willing to participate in the study. A total of 7,500 questionnaires were distributed and 7,315 questionnaires were returned, representing a return rate of 97.5%. After checking the validity and integrity of the questionnaires, 7,118 questionnaires were confirmed as valid, with an effective rate of 97.3%. A total of 7,118 participants met the inclusion criteria and the data were randomly divided into a training group (n=4,955) and a validation group (n=2,163) (Figure 1).

2.3. Research Methods

2.3.1. Assessment of mental health

The SCL-90, designed by Derogatis and his colleagues, was widely used in the mental health field [31], which contains 90 items across nine dimensions: somatization, obsessive-compulsive symptoms, interpersonal sensitivity, depression, anxiety, hostility, horror, bigotry and mental illness. The SCL-90 has been used extensively in previous studies and has relatively high reliability and validity [32]. The questionnaire uses a Likert 5-point scale, with a score of 0 point indicating none and 4 points indicating severe. A total score above 160, a score above 2 on any item, or a positive item above 43 indicates the presence of a psychological abnormality [33]. In this survey, Cronbach α was 0.99, the half-reliability coefficient was 0.98, and the KMO was 0.994.

2.3.2. Assessment of occupational stress

This survey evaluated occupational stress in factory workers and miners through the Effort–Reward Imbalance (ERI) model developed by Siegrist [34]. The ERI scale consists of three subscales: effort (E, 6 items), reward (R, 11 items) and over commitment (6 items), for a total of 23 items. A Likert 5-level scoring method (1, "highly disagree" to 5, "highly agree") is used to grade the items in the questionnaire with the same weight for each item. The effort–return index $ERI = E/R \times C$, where C is the adjustment coefficient, and the value is 6/11. ERI values greater than 1, equal to 1, and less than 1 correspond to high pay–low return, pay–return balance, and low pay–high return, respectively. Moreover, the higher the ERI value, the greater the occupational stress [35]. In this survey, Cronbach α was 0.94, the half-reliability coefficient was 0.93 and the KMO was 0.956.

2.3.3. Assessment of job burnout

In this survey, the Chinese Maslach Burnout Inventory (CMBI) revised by Li et al. was used to assess job burnout, which has good reliability and validity [36]. CMBI is composed of 15 items in three dimensions: emotional exhaustion (5 items), depersonalization (5 items) and reduced personal accomplishment (5 items). The score for each item ranges from 1 to 7, with 1 point indicating complete compliance and 7 points indicating complete non-compliance. According to the critical value (emotional exhaustion ≥ 25, depersonalization ≥ 11, personal achievement reduction ≥ 16), the levels of occupational burnout are divided into none (all aspects are below the critical value), mild (any one aspect is equal to or above the critical value), moderate (any two aspects are equal to or higher than the critical values), and severe (three aspects are equal to or higher than the critical values) [37]. In this survey, Cronbach α was 0.89, the half-reliability coefficient was 0.86 and the KMO was 0.919.

2.3.4. Candidate predictors

Trained investigators obtained information on demographics, job burnout, occupational stress, mental health and occupational exposure factors through on-site face-to-face collection of an electronic version of the questionnaire. Covariates included in this study: 1) demographic information: gender, ethnicity, education level, professional title, work schedule, marital status, monthly income, age, working years, labor contracts, working hours per day, and working hours per week; 2) occupational exposure factors: coal dust, silica dust, asbestos dust, benzene, lead, noise, and brucellosis; 3) questionnaires: ERI, CMBI; 4) chronic diseases: diabetes, hypertension. Information on four areas, including demographic information, questionnaires, occupational hazards and chronic diseases, were filled in by participants through their own responses on the questionnaire star.

Sex was defined as male or female; ethnicity was defined as Han and other; education level was defined as junior high school and below, high school, junior college or bachelor's degree or above; labor contracts was defined as signed or unsigned; professional title was defined as no, primary, middle or senior; work schedule was defined as day shift, night shift, shift or day and night shifts; marital status was defined as unmarried, married, divorced or widowed; monthly income (yuan) was defined as <3000, 3000~, 4000~, 5000~, 6000~, 7000~ or 8000~; age (years) was defined as <25, 25~, 30~, 35~, 40~ or 45~; working years was defined as ~5, 5~, 10~, 15~, 20~, 25~ or 30~; working hours per day (hours) was defined as ≤ 7 or > 7; working days per week (days) was defined as ≤ 5 or >5; exposure to coal dust, silica dust, asbestos dust, benzene, lead, noise, brucellosis were all defined as yes or no; ERI was defined as yes or no; CMBI was defined as none, mild, moderate and severe; hypertension and diabetes were both defined as yes or no.

2.4. Statistical analysis

Categorical variables were described as counts and percentages, and chi square test or Fisher exact test was used to compare categorical variables between groups. 70% of participants were randomly assigned to the training cohort and 30% to the validation cohort. Variables were screened using a least absolute shrinkage and selection operator (LASSO) regression and multivariate logistic regression models were used to estimate risk ratios and corresponding 95% confidence intervals (CI) of risk factors, from which predictive models were constructed. A nomogram for predicting was generated according to the selected characteristics. In addition, forest plot was drawn to visually depict the P-value, OR and 95% CI for the selected validations. Statistically significant predictors were applied to develop a prediction model for the risk of mental health problems among factory workers and miners by introducing all selected factors and analyzing the statistical significance levels of them. We used calibration plots and receiver operating characteristic (ROC) curves to show the calibration and discrimination of our final model. Brier scores for overall performance, calibration slopes were used to assess the predictable accuracy of the model. Decision curve analysis (DCA) was applied to calculate the net benefit of the nomogram. Statistical analysis was performed using the open-source R software Version 3.6.1 (http://www.r-project.org). The significance level (α) set at 0.05.

2.5. Patient and public involvement

Neither patients nor members of the public had any involvement in the design of this study.

3. Results

3.1. Participant characteristics

A total of 7,118 participants met the inclusion criteria and the data were randomly divided into a training group (n=4,955) and a validation group (n=2,163). Over half of all participants (65.31%) were male, 57.31% of the population was over 35 years of age and 78.32% of the subjects were married, showing that factory workers and miners are generally older and most of them have spouses. The majority of them had completed high school (83.94%), while a smaller percentage had completed undergraduate education (22.98%), indicating that the group of factory workers and miners as a whole was not well educated. The total number of workers (n, %) exposed to coal dust, silica dust, asbestos dust, benzene, lead, noise and brucellosis in the factory and mining enterprises were 377 (5.3), 730 (10.3), 981 (14), 1,981 (27.8), 373 (5.2), 4,942 (69.4) and 121 (1,7) respectively, with the total number of workers exposed to noise amounting to 4,942, or 69% of the total population surveyed. The demographic, job burnout, occupational stress and occupational exposure factors for the training and validation groups are shown in Table 1. The results showed that there were no significant statistical differences between the two groups of characteristic variables, except for coal dust and CMBI, indicating that the baseline levels were largely consistent between the two groups.

Table 1 Characteristics of the study participants

Variables		Total (n = 7118)	train (n = 4955)	test (n = 2163)	*p*
Sex, n (%)	Male	4649 (65.3)	3216 (64.9)	1433 (66.3)	0.284
	Female	2469 (34.7)	1739 (35.1)	730 (33.7)	
Ethnicity, n (%)	Han	5762 (80.9)	3982 (80.4)	1780 (82.3)	0.061
	Other	1356 (19.1)	973 (19.6)	383 (17.7)	
Education level, n (%)	Junior high school and below	1143 (16.1)	804 (16.2)	339 (15.7)	0.765
	High school	1406 (19.8)	988 (19.9)	418 (19.3)	
	Junior college	2933 (41.2)	2038 (41.1)	895 (41.4)	
	Bachelor's degree or above	1636 (23.0)	1125 (22.7)	511 (23.6)	
Professional title, n (%)	None	2854 (40.1)	1983 (40.0)	871 (40.3)	0.923
	Primary	1644 (23.1)	1149 (23.2)	495 (22.9)	
	Middle	1618 (22.7)	1133 (22.9)	485 (22.4)	
	Senior	1002 (14.1)	690 (13.9)	312 (14.4)	
Work schedule, n (%)	Day shift	3986 (56.0)	2801 (56.5)	1185 (54.8)	0.585
	Night shift	270 (3.8)	187 (3.8)	83 (3.8)	
	Shift	2058 (28.9)	1412 (28.5)	646 (29.9)	
	Day and night shifts	804 (11.3)	555 (11.2)	249 (11.5)	
Marital status, n (%)	Unmarried	1104 (15.5)	762 (15.4)	342 (15.8)	0.218
	Married	5575 (78.3)	3906 (78.8)	1669 (77.2)	
	Divorced	390 (5.5)	255 (5.1)	135 (6.2)	
	Widowed	49 (0.7)	32 (0.6)	17 (0.8)	
Monthly income (yuan), n (%)	<3000	1799 (25.3)	1246 (25.1)	553 (25.6)	0.966
	3000~	2418 (34.0)	1682 (33.9)	736 (34.0)	
	4000~	1600 (22.5)	1125 (22.7)	475 (22.0)	
	5000~	752 (10.6)	520 (10.5)	232 (10.7)	
	6000~	288 (4.0)	201 (4.1)	87 (4.0)	
	7000~	148 (2.1)	106 (2.1)	42 (1.9)	
	8000~	113 (1.6)	75 (1.5)	38 (1.8)	
Age (years), n (%)	<25	431 (6.1)	297 (6.0)	134 (6.2)	0.173
	25~	786 (11.0)	519 (10.5)	267 (12.3)	
	30~	956 (13.4)	684 (13.8)	272 (12.6)	
	35~	866 (12.2)	617 (12.5)	249 (11.5)	
	40~	849 (11.9)	588 (11.9)	261 (12.1)	
	45~	3230 (45.4)	2250 (45.4)	980 (45.3)	
Working years (years), n (%)	<5	1170 (16.4)	794 (16.0)	376 (17.4)	0.248
	5~	1065 (15.0)	736 (14.9)	329 (15.2)	

	10~	997 (14.0)	721 (14.6)	276 (12.8)	
	15~	389 (5.5)	273 (5.5)	116 (5.4)	
	20~	763 (10.7)	538 (10.9)	225 (10.4)	
	25~	1293 (18.2)	878 (17.7)	415 (19.2)	
	30~	1441 (20.2)	1015 (20.5)	426 (19.7)	
Labor contracts, n (%)	Signed	6641 (93.3)	4624 (93.3)	2017 (93.3)	0.955
	Unsigned	477 (6.7)	331 (6.7)	146 (6.7)	
Working hours per day (hours), n (%)	≤7	1161 (16.3)	814 (16.4)	347 (16.0)	0.712
	>7	5957 (83.7)	4141 (83.6)	1816 (84.0)	
Working days per week (days), n (%)	≤5	4442 (62.4)	3107 (62.7)	1335 (61.7)	0.446
	>5	2676 (37.6)	1848 (37.3)	828 (38.3)	
Diabetes, n (%)	Yes	429 (6.0)	298 (6.0)	131 (6.1)	0.988
	No	6689 (94.0)	4657 (94.0)	2032 (93.9)	
Hypertension, n (%)	Yes	1330 (18.7)	929 (18.7)	401 (18.5)	0.861
	No	5788 (81.3)	4026 (81.3)	1762 (81.5)	
Coal dust, n (%)	Yes	377 (5.3)	244 (4.9)	133 (6.1)	0.039
	No	6741 (94.7)	4711 (95.1)	2030 (93.9)	
Silica dust, n (%)	Yes	730 (10.3)	523 (10.6)	207 (9.6)	0.223
	No	6388 (89.7)	4432 (89.4)	1956 (90.4)	
Asbestos dust, n (%)	Yes	981 (13.8)	691 (13.9)	290 (13.4)	0.570
	No	6137 (86.2)	4264 (86.1)	1873 (86.6)	
Benzene, n (%)	Yes	1981 (27.8)	1360 (27.4)	621 (28.7)	0.287
	No	5137 (72.2)	3595 (72.6)	1542 (71.3)	
Lead, n (%)	Yes	373 (5.2)	246 (5.0)	127 (5.9)	0.128
	No	6745 (94.8)	4709 (95.0)	2036 (94.1)	
Noise, n (%)	Yes	4942 (69.4)	3420 (69.0)	1522 (70.4)	0.270
	No	2176 (30.6)	1535 (31.0)	641 (29.6)	
Brucellosis, n (%)	Yes	121 (1.7)	86 (1.7)	35 (1.6)	0.800
	No	6997 (98.3)	4869 (98.3)	2128 (98.4)	
ERI, n (%)	Yes	3147 (44.2)	2173 (43.9)	974 (45.0)	0.372
	No	3971 (55.8)	2782 (56.1)	1189 (55.0)	
CMBI, n (%)	No	959 (13.5)	674 (13.6)	285 (13.2)	0.033
	Mild	2667 (37.5)	1813 (36.6)	854 (39.5)	
	Moderate	2900 (40.7)	2031 (41.0)	869 (40.2)	
	Severe	592 (8.3)	437 (8.8)	155 (7.2)	

3.2. Feature selection

The lambda was smallest at 0.01801 as seen from the lasso results when there were 12 characteristics, which were education, professional title, age, CMBI, ERI, asbestos dust, hypertension, diabetes, working hours per day, working years, marital status, and work schedule based on the results of the questionnaires on demographics, occupational stress, job burnout and occupational exposure factors (Figure 2).

3.3. Results of logistic regression model

The 12 features obtained from the LASSO regression were incorporated into a multivariate logistic regression model and the regression results were shown in Table 2. It was clear from the results that education, professional title, age, CMBI, ERI, asbestos dust, hypertension, diabetes, working hours per day, working years, marital status, and work schedule were independent determinants of risk for mental health problems. In addition, there was no evidence of multicollinearity between the covariates included in the model. The forest plot showed that the selected 12 features all contain items with $P < 0.05$, among which the degree of severe of CMBI (OR, 19.84; 95% CI, 13.88-28.34; $P < 0.001$) had the greatest impact on the risk of mental health problems among factory workers and miners (Figure 3).

Table 2 Predictive factors of risk for mental health problems among factory workers and miners

Variable	β	S.E.	OR(CI95%)	Wald	*P*	VIF
Intercept	-2.33	0.25	0.10(0.06,0.16)	-9.357	0	-
Education level						
Junior school and below VS High school	0.34	0.13	1.41(1.10,1.81)	2.727	0.006**	2.28
Junior school and below VS Junior college	0.44	0.11	1.56(1.24,1.95)	3.850	< 0.001***	2.79
Junior school and below VS Bachelor's degree or above	0.38	0.13	1.46(1.13,1.87)	2.953	0.003**	2.51
Professional title						
None VS Primary	0.15	0.09	1.16(0.97,1.39)	1.582	0.114	1.35
None VS Middle	0.05	0.09	1.05(0.87,1.26)	0.519	0.604	1.34
None VS Senior	0.27	0.11	1.30(1.06,1.61)	2.458	0.014*	1.32
Work schedule						
Day and night shifts VS Day shift	-0.38	0.11	0.69(0.55,0.85)	-3.364	0.001**	2.70
Day and night shifts VS Night shif	0.01	0.20	1.01(0.68,1.49)	0.044	0.965	1.30
Day and night shifts VS Shift	0.01	0.12	1.01(0.81,1.27)	0.107	0.915	2.47
Marital status						
Unmarried VS Married	0.16	0.13	1.18(0.91,1.52)	1.263	0.206	2.29
Unmarried VS Divorced	0.55	0.19	1.73(1.20,2.51)	2.918	0.004**	1.69
Unmarried VS Widowed	0.69	0.43	1.99(0.85,4.64)	1.586	0.113	1.09
Age						
~25 VS 25~	-0.02	0.20	0.98(0.66,1.47)	-0.083	0.934	3.09
~25 VS 30~	-0.02	0.22	0.98(0.64,1.50)	-0.090	0.929	4.79
~25 VS 35~	0.56	0.23	1.76(1.13,2.74)	2.503	0.012*	5.01

~25 VS 40~	0.33	0.23	1.39(0.88,2.21)	1.419	0.156	4.97
~25 VS 45~	0.23	0.22	1.26(0.81,1.95)	1.018	0.308	10.93
Working years						
~5 VS 5~	0.44	0.14	1.55(1.18,2.05)	3.114	0.002**	2.27
~5 VS 10~	0.06	0.15	1.06(0.78,1.43)	0.366	0.714	2.48
~5 VS 15~	0.06	0.20	1.06(0.72,1.56)	0.305	0.760	1.79
~5 VS 20~	0.29	0.18	1.33(0.95,1.88)	1.641	0.101	2.65
~5 VS 25~	0.48	0.17	1.61(1.15,2.25)	2.782	0.005**	3.99
~5 VS 30~	0.20	0.16	1.22(0.89,1.68)	1.239	0.216	3.90
Working hours per day						
≤7 VS >7	-0.50	0.09	0.61(0.50,0.73)	-5.363	< 0.001***	1.15
Diabetes						
No VS Yes	0.43	0.14	1.53(1.16,2.03)	2.974	0.003**	1.05
Hypertension						
No VS Yes	0.52	0.09	1.69(1.42,2.00)	5.885	< 0.001***	1.11
Asbestos dust						
No VS Yes	0.44	0.10	1.55(1.28,1.87)	4.474	< 0.001***	1.03
ERI						
No VS Yes	0.89	0.07	2.43(2.12,2.79)	12.786	< 0.001***	1.05
CMBI						
No VS Mild	0.26	0.12	1.30(1.03,1.64)	2.175	0.003**	2.73
No VS Moderate	1.30	0.11	3.67(2.93,4.59)	11.361	< 0.001***	2.83
No VS Severe	2.99	0.18	19.84(13.88,28.34)	16.41	< 0.001***	1.44

Note: β is the regression coefficient. "***" indicates P<0.001, "**" indicates P<0.01, "*" indicates P<0.05.

3.4. Development of an individualized prediction model

Based on the results of the multivariate analysis, predictors such as education, professional title, age, CMBI, ERI, asbestos dust, hypertension, diabetes, working hours per day, working years, marital status, and work schedule were included in the nomogram. A model incorporating the above independent predictors was developed and represented as a nomogram in Figure 4. Each variable in nomogram was assigned a score, and the cumulative sum of each 'point' was the 'total score'. The "total score" corresponded to the "predictable likelihood", which was the predicted probability of mental health problems among factory workers and miners as suggested by our design of the nomogram.

As an example of the use of nomogram: a randomly selected sample from the training group, one with no professional title, day shift, no diabetes or hypertension, Junior college, <5 of working years, >7 of working hours per day, married, no exposed to asbestos dust, <25 years of age, no ERI, mild of CMBI, with a calculated total score of 174 and a corresponding risk probability of 8.27% for mental health problems.

3.5 The validation of calibration

Model validation was carried out in the validation group. The prediction accuracy of the model was assessed by two aspects. (1) The Brier score for overall performance, which assessed the difference between observed and predicted values, with values closer to 0 indicating better predictive ability. (2) The calibration slope used for modal calibration, which assessed the agreement between the observed and predicted values, with values closer to 1 indicating better performance. The accuracy measurements for the bias correction were validated by the model with a Brier score of 0.176 and a calibration slope of 0.970, respectively (Figure 5). The prediction accuracy of the model was relatively high.

3.6 The validation of discrimination

ROC was plotted for the training and validation groups, and the AUC of training and the verification groups were 0.785 and 0.784, respectively (Figure 6). The AUC of training and the verification groups were both greater than 0.750, showing a good discrimination.

3.7 Decision Curve Analysis

As shown in the DCA of the risk of mental health problems nomogram in Figure 7, the model for predicting the risk of mental health problems for factory workers and miners in this study was more practically relevant if the threshold probability of patients was >10%.

4. Discussion

To our knowledge, this is the first study to develop an easy-to-use nomogram to predict the mental health risks of factory workers and miners. The nomogram developed using the training set data contain 12 items for education, professional title, age, CMBI, ERI, asbestos dust, hypertension, diabetes, working hours per day, working years, marital status, and work schedule**.** In addition, validation has shown that nomogram model has good accuracy and discriminatory power. Our novel nomogram can be used in any setting to provide a rapid assessment of mental health risks and to help identify patients with mental health risks, saving time compared to previous mental health investigations and improving on the lack of entries in previous investigations related to the specific working environment of factory workers and miners. The AUC of training group and verification group were 0.785 and 0.784 respectively, showing moderate discriminatory and calibration power.

A review of the literature found that the vast majority of studies constructed nomograms to predict clinical disorders, with less literature used to predict psychological problems. In a study to predict the correlates of suicide attempts in a Chinese population with major depressive disorder, the C-index was 0.715 and the C-index in the internal validation set was 0.703, and the calibration curve of the column line plot also showed good agreement between the predicted and observed risk of suicide attempts. The variables in the nomogram included socio-demographic information and clinical variables including age, duration, number of episodes, age at onset, number of hospitalizations, characteristics of anxiety and psychiatric symptoms, marital status, income, education level and employment status [27]. In another study that created a nomogram to predict the risk of psychosocial and behavioral problems in children and adolescents during the COVID-19 pandemic, the C index exceeded 0.800 and the calibration curve also showed good predictive accuracy. The variables covered three subject areas, namely demographic information, the psychosocial impact of the epidemic such as homework time and sedentary time, and the Child Behaviour Checklist score (CBCL) for the evaluation of psychological problems [38]. In this study, 7,118 participants were randomly divided into a training group (n=4,955) and a validation group (n=2,163) in a ratio of 3:1, involving a total of 23 features, and 12 features were selected by LASSO regression. The nomogram could be a useful tool to better identify patients with mental health problems, as it not only covered comprehensive information, including demographic information, job burnout, occupational stress, chronic diseases and occupational exposure factors closely related to factory workers and miners, but also was simple to operate and easy to use. In the validation group the Brier score was 0.176, the calibration slope was 0.970 and the calibration curve of nomogram showed a good fit. The AUC of training group and verification group were 0.785 and 0.784 respectively. Compared to the two studies above, our nomogram showed good accuracy and discrimination, and more comprehensive coverage in this nomogram model. Therefore, the possibility of early intervention for patients with high-risk mental health problems will be increased by covering multiple information and easy to use nomogram modal, especially for factory workers and miners with poor working conditions, relatively low levels of education and low patience.

Mental health problems were very common in the group of factory workers and miners, and the prevalence of mental health of them was found to be 37.08% in our study. Notably, the CMBI showed the most significant score (score = 100) and the ERI also had a high score (score = 43) in mental health problem incidence risk nomogram, which indicated that both of them were relatively important factors for mental health problems among the group of factory workers and miners. Our finding was consistent with other studies that had shown that occupational stress was a significant predictor of anxiety and was negatively associated with mental health. In addition, there is a high correlation between burnout and depression [39].

In line with previous studies, working years was also an important influential factor in this study. Related study has shown that employment could improve patients' mental health, while unemployment could lead to a deterioration in mental health [40]. In China, workers' working years is an important aspect of employment, and researchers have studied this aspect and found that precarious employment is a source of stress for individuals and predisposes them to mental health problems [41]. In addition, environmental factors were also one of the influential factors of mental health problems in our study. Relevant studies have found that exposure to air pollution is associated with increased suicide risk and depressive symptoms [42]. Hypertension and diabetes were the influential factors in this study. A study has shown that the prevalence of depression in adults with type 1 diabetes (T1D) is approximately three times higher than in the non-diabetic population [43]. Furthermore, there is a recognized association between hyperglycemia and depression, but the underlying biological mechanisms of this association are unclear [44].

Factory workers and miners were inevitably exposed to occupational hazards such as benzene and asbestos dust in their working environment. According to statistics, a total of nearly 2 million workers are exposed to various occupational hazards and over 16 million people worked in toxic and hazardous enterprises, involving more than 30 different types of operations, of which factory workers and miners is the one [45]. Similarly, the occupational hazard asbestos dust was selected as a predictor of risk for mental health problems in this study. Our study found that the work schedules of factory workers and miners were vary and the phenomenon of night shifts was very common, which inevitably affected their normal sleep. Some studies have shown that sleep problem is a risk factor for a variety of mental health and chronic diseases. Lack of sleep or poor sleep quality could lead to abnormalities in the body's self-regulatory functions and disturbances in the circadian rhythm of the biological clock, which in turn could suffer from negative emotions such as anxiety and depression [46]. Professional title and education level were also important influences on mental health issues. In the workplace, generally speaking, the higher the professional title and education level, the higher the status of the worker in the company and the greater the role played in the position. The number of studies on socio-economic status and mental health had increased in recent years. Some of these studies have shown that major depression is higher in the low socio-economic status group [47]. It has also been suggested that education itself is the best indicator of socio-economic status [48]. Marital status was one of the influential factors for mental health problems. Many studies have found an association between mental health and gender, marital status, lifestyle and working conditions, and it has been shown that poor mental health in women is associated with divorce or widowhood [49]. In this study, working more than seven hours a day was a determinant factor on mental health problems, which was consistent with other studies that had shown that long working hours could have a negative impact on employees' mental health and that excessive workloads could increase workers' fatigue, which in turn could lead to anxiety and depression [50].

In China, there are many problems in identifying people with mental health problems due to uneven and imperfect levels of medical development across regions. Some studies have shown that in mainland China, general practitioners, surgeons and primary health care workers often have little or no mental health training, which prevents them from providing basic mental health services [51]. Non-mental health professionals in general hospitals learn about mental illness on their own, rather than learning about it during their formal education [52]. Therefore, this study designed a simple and comprehensive nomogram to address the issue of timely detection and effective interventions for people with mental health problems, so that people at risk of mental health problems could easily calculate their probability of suffering from mental health problems without the help of medical staff. This study has several strengths. First, to our knowledge, this is the first model to develop and assess the likelihood of mental health problems in a group of factory workers and miners. Secondly, the nomogram in this study includes demographic information, job burnout, occupational stress, chronic illnesses, and also occupational exposure factors that are closely related to factory workers and miners, allowing for a more accurate assessment of the risk of morbidity among them, as well as providing a methodological reference for other related studies.

5. Limitations

This study also has several limitations. Firstly, we have considered many influential factors including demographics, job burnout, occupational stress and occupational exposure factors, but we are still not certain whether all possible influences are covered. Secondly, while the robustness of our nomogram was extensively validated internally in the same population, external validation is lacking for other populations in other regions and countries. Nomogram need to be externally assessed in a wider population.

Acknowledgements The authors are grateful to all participants and investigators.

Contributions Y.L., Q.L., and T.L. are responsible for conceptualization; Y.L. is responsible for methodology, software, formal analysis, resources, and visualization; Q.L. and T.L. are responsible for the original draft preparation; Q.L. and H.Y. are responsible for reviewing; Q.L. is responsible for editing; T.L. is responsible for supervision. Yaoqin Lu and Qi Liu contributed equally to this work.

Funding This work was supported by the Natural Science Foundation of Xinjiang Uygur Autonomous Region (grant number 2020D01A27), the Postgraduate Innovation Project of Xinjiang Uyghur Autonomous Region (grant number XJ2021G215), the Outstanding Young Scientist Training Program of Urumqi Science and Technology Talent Project (grant number N/A), the Public Health and Preventive Medicine, the 13th Five-Year Plan Key Subject of Xinjiang Uygur Autonomous Region (grant number N/A), The funders were not involved in the conception, design, analysis or interpretation of this study.

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval The study was approved by the ethics committee of Urumqi Center for Disease Control and Prevention (20181123)

Data availability statement Data are available on reasonable request. The data used in this study are available from the corresponding authors on reasonable request.

References

[1] WHO Terminology Information System [online glossary] http://www.who.int/health-systems-

performance/docs/glossary.html.

[2] Wang Y, Liu X, Qiu J, Wang H, Liu D, Zhao Z, Song M, Song Q, Wang X, Zhou Y, Wang W. Association between ideal cardiovascular health metrics and suboptimal health status in Chinese population. *Sci Rep* 2017;7:14975.

[3] Prince M, Patel V, Saxena S, Maj M, Maselko J, Phillips MR, Rahman A. No health without mental health. *Lancet*. 2007;370:859–77.

[4] Adjaye-Gbewonyo K, Avendano M, Subramanian S.V, Kawachi I. Income inequality and depressive symptoms in South Africa: A longitudinal analysis of the National Income Dynamics Study. *Health Place* 2016;42:37–46.

[5] Vos T, Barber R.M, Bell B, Bertozzi-Villa A, Biryukov S, Bolliger I, Charlson F, Davis A, Degenhardt L, Dicker D, *et al*. Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990–2013: A systematic analysis for the Global Burden of Disease Study 2013. *Lancet* 2015;386:743–800.

[6] Huang Z, Li T, Xu M. Are there heterogeneous impacts of national income on mental health?.*Int J Environ Res Public Health* 2020;17:7530.

[7] Asadullah M.N, Xiao S, Yeoh E. Subjective well-being in China, 2005–2010: The role of relative income, gender, and location. *China Econ Rev* 2018;48:83–101.

[8] Butterworth P, Rodgers B, Windsor T.D. Financial hardship, socio-economic position and depression: Results from the PATH Through Life Survey. *Soc Sci Med*. 2009;69:229–237.

[9] Zhang X, Zhang X, Chen X. Happiness in the Air: How Does a Dirty Sky Affect Mental Health and Subjective Well-being? *J Environ Econ Manag* 2017;85:81–94.

[10] Wahlbeck K. Public mental health: The time is ripe for translation of evidence into practice. *World Psychiatry*. 2015;14:36–42.

[11] Luciano AE, Drake RE, Bond GR, Becker DR, Carpenter-Song E, Lord S, Swarbrick P and Swanson SJ. IPS Supported employment: a review. *J Vocat Rehabil* 2014;40:1–13.

[12] Jia Z, *et al*. Exposure to ambient air particles increases the risk of mental disorder: findings from a natural experiment in Beijing. *Int J Environ Res Public Health* 2018;15:160.

[13] Szyszkowicz M, Willey J. B, Grafstein E, Rowe B. H, Colman I. Air pollution and emergency department visits for suicide attempts in vancouver, Canada. *Environ Health Insights* 2010;4:79–86.

[14] Michael J. S. International classification of sleep disorders-third edition : highlights and modifications. *Chest* 2014;146:1387–1394.

[15] AbuRuz ME, Al-Dweik G. Depressive symptoms and complications early after acute myocardial infarction: gender differences. *Open Nurs J* 2018;12:205–214.

[16] Wurm W, Vogel K, Holl A, *et al*. Depression-burnout overlap in physicians. *PLoS One* 2016;11:e0149913.

[17] Porru F, Robroek S J W, Bültmann U, *et al*. Mental health among university students: The associations of effort-reward imbalance and overcommitment with psychological distress. *J Affect Disord* 2021;282:953-961.

[18] Johnson AK, Blackstone SR, Skelly A, Simmons W. The relationship between depression, anxiety, and burnout among physician assistant students: a multi-institutional study. *Health Professions Edu* 2020;6:420–427.

[19] Hua D, Kong Y, Li W, Han Q, Zhang X, Zhu LX, Wan SW, Liu Z, Shen Q, Yang J, He HG, Zhu J. Frontline nurses' burnout, anxiety, depression, and fear statuses and their associated factors during the COVID-19 outbreak in Wuhan, China: a large-scale cross-sectional study. *EClinical Med*

2020;24:100424.
[20] Armah E K, Adedeji J A, Boafo B B, *et al*. Underground gold miner exposure to noise, diesel particulate matter and crystalline silica dust. *J Health Pollut* 2021; 11(29):210301.
[21] Hall N B, Blackley D J, Halldin C N, *et al*. Current review of pneumoconiosis among US coal miners. *Curr Environ Health Rep*, 2019; 6(3):137-147.
[22] Wickramatillake B A, Fernando M A, Frank A L. Prevalence of asbestos-related disease among workers in sri lanka. *Ann Glob Health*, 2019, 85(1):108.
[23] Liu FD, Pan ZQ, Liu SL, Chen L, Ma JZ, Yang ML, Wang NP. The estimation of the number of underground coal miners and the annual dose to coal miners in China. *Health Phys* 2007;93:127–132.
[24] Yong X, Gao X, Zhang Z, *et al*. Associations of occupational stress with job burn-out, depression and hypertension in coal miners of Xinjiang, China: A cross-sectional study. *BMJ open* 2020;10: e036087.
[25] P. Bech, J. Bille, S. B. Møller, L. C. Hellström, S. D. Østergaard. Psychometric validation of the Hopkins Symptom Checklist (SCL-90) subscales for depression, anxiety, and interpersonal sensitivity. *J Affect Disord* 2014;160:98–103.
[26] J. Zhang, X. Zhang. Chinese college students' SCL-90 scores and their relations to the college performance. *Asian J Psychiatr* 2013;6:134–140.
[27] Liang S, Zhang J, Zhao Q, *et al*. Incidence trends and risk prediction nomogram for suicidal attempts in patients with major depressive disorder. *Front Psychiatry*, 2021, 12: 644038.
[28] Kan S K, Chen N N, Zhang Y L. Predicting the risk of suicide attempt in a depressed population: Development and assessment of an efficient predictive nomogram. *Psychiatry Res*, 2022, 310: 114436.
[29] Lu Y, Zhang Z, Yan H, *et al*. Effects of occupational hazards on job stress and mental health of factory workers and miners: A propensity score analysis. *BioMed Res Int*, 2020, 2020: 1754897.
[30] Li X, Jiang T, Sun X, *et al*. The relationship between occupational stress, musculoskeletal disorders and the mental health of coal miners: The interaction between bdnf gene, tph2 gene polymorphism and the environment. *J Psychiatr Res*, 2021, 135: 76-85.
[31] Derogatis L, Lipman R.S, Covi L. SCL-90: an outpatient psychiatric rating scale—preliminary report. *Psychopharmacol Bull* 1973;9:13–28.
[32] Crespo-Maraver M, Doval E, Fernã N.J, Gimã Nez-Salinas J, Prat G, Bonet P. Caregiver's health: adaption and validation in a Spanish population of the Experience of Caregiving Inventory (ECI) Gac. *Sanit* 2018;33:348–355.
[33] Dang W, Xu Y, Ji J, *et al*. Study of the scl-90 scale and changes in the chinese norms. *Front Psychiatry*, 2020, 11: 524395.
[34] Siegrist J. Adverse health effects of high-effort/low-reward conditions. *J Occup. Health Psychol*. 1996;1:27–41.
[35] Siegrist J, Wege N, Pühlhofer F, Wahrendorf M. A short generic measure of work stress in the era of globalization: Effort-reward imbalance. *Int Arch Occup Environ Health* 2009;82:1005–1013.
[36] Zhang Z, Lu Y, Yong X, *et al*. Effects of occupational radiation exposure on job stress and job burnout of medical staff in xinjiang, China: A cross-sectional study. *Med Sci Monit* 2020;26: e927848.
[37] Freudenberger HJ. Staff burnout. *J Soc Issues* 1974;30:159–165.
[38] Wang L, Chen L, Jia F, *et al*. Risk factors and prediction nomogram model for psychosocial and

behavioural problems among children and adolescents during the covid-19 pandemic: A national multicentre study: Risk factors of childhood psychosocial problems. *J Affect Disord*, 2021, 294: 128-136.

[39] Occupational Stress and Employees Complete Mental Health: A Cross-Cultural Empirical Study

[40] Knapp M and Wong G. Economics and mental health: the current scenario. *World Psychiatry* 2020;19:3–14.

[41] Benach, J, Vives, A, Amable, M, Vanroelen, C, Tarafa, G, & Muntaner, C. Precarious employment: understanding an emerging social determinant of health. *Annu Rev Public Health* 2014;35:229-53.

[42] Bakian A. V. *et al*. Acute air pollution exposure and risk of suicide completion. *Am J Epidemiol* 2015;181:295–303.

[43] Barnard KD, Skinner TC, Peveler R. The prevalence of co-morbid depression in adults with Type 1 diabetes: systematic literature review. *Diabet Med* 2006;23:445–448.

[44] Gilsanz P, Karter AJ, Beeri MS, Quesenberry CP, Whitmer RA. The bidirectional association between depression and severe hypoglycemic and hyperglycemic events in type 1 diabetes. *Diabetes Care* 2018;41:446–452.

[45] Lu Y, Zhang Z, Yan H, *et al*. Effects of occupational hazards on job stress and mental health of factory workers and miners: A propensity score analysis. *BioMed Res Int* 2020;2020:1754897.

[46] Shi L, Liu Y, Jiang T, *et al*. Relationship between mental health, the clock gene, and sleep quality in surgical nurses: A cross-sectional study. *BioMed Res Int* 2020;2020:4795763.

[47] Sallis JF, Saelens BE, Frank LD *et al*. Neighborhood built environment and income: examining multiple health outcomes. *Soc Sci Med* 2009;68:1285–1293.

[48] Winkleby MA, Jatulis DE, Frank E *et al*. Socioeconomic status and health: how education, income, and occupation contribute to risk-factors for cardiovascular disease. *Am J Public Health* 1992;82:816–820.

[49] Skapinakis P, Bellos S, Koupidis S, Grammatikopoulos I, Theodorakis P.N, Mavreas V. Prevalence and sociodemographic associations of common mental disorders in a nationally representative sample of the general population of Greece. *BMC Psychiatry* 2013;13:163.

[50] Virtanen M, Ferrie JE, Singh-Manoux A, Shipley MJ, Stansfeld SA, Marmot MG *et al*. Long working hours and symptoms of anxiety and depression: a 5-year follow-up of the Whitehall II study. *Psychol Med* 2011;41:2485–2494.

[51] Phillips MR, Zhang J, Shi Q, Song Z, Ding Z, Pang S, *et al*. Prevalence, treatment, and associated disability of mental disorders in four provinces in China during 2001–05: an epidemiological survey. *Lancet* 2009;373:2041–2053.

[52] Wu Q, Luo X, Chen S, *et al*. Mental health literacy survey of non-mental health professionals in six general hospitals in hunan province of China. *PloS one* 2017;12:e0180327.

Figure legends

Fig.1. Flow diagram of the participants involved in this study

Fig.2. Feature selection using the LASSO binary logistic regression model. (A) Feature selection for the LASSO binary logistic regression model. The partial likelihood deviation (binomial deviation) curve was plotted against lambda by validating the optimal parameter lambda in the LASSO model. Dotted vertical lines were drawn based on 1 SE of the minimum criteria (the 1-SE criteria). (B) Feature selection was performed using the LASSO binary logistic regression model. A Coefficient profile weas plotted based on the lambda series in Figure 1(A), and 12

features with non-zero coefficients were selected by optimal lambda.

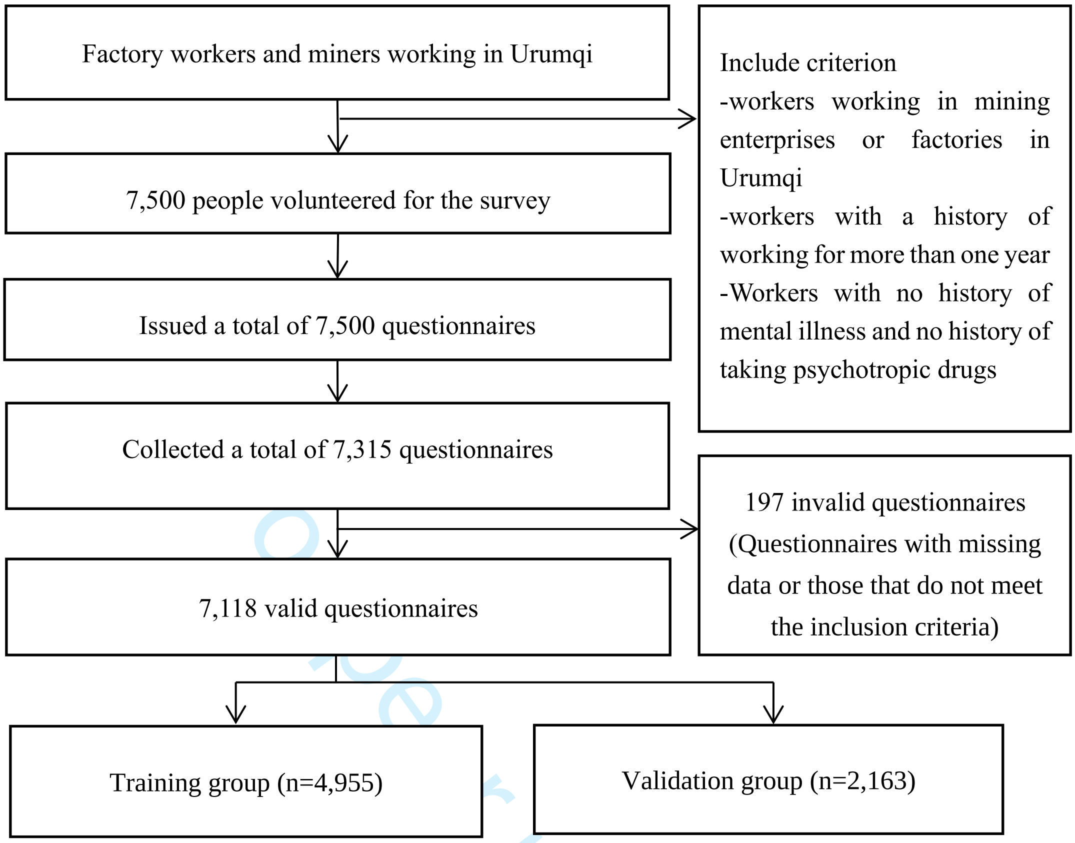
Fig.3. The forest plot of the OR of the selected feature.

Fig.4. Developed mental health problems incidence risk nomogram. The mental health problems incidence risk nomogram was developed in the array, with education, professional title, age, CMBI, ERI, asbestos dust, hypertension, diabetes, working hours per day, working years, marital status, and work schedule incorporated.

Fig.5. Calibration curves of the mental health problems incidence risk nomogram prediction in validation group. The x-axis represents the predicted risk of mental health problems. y-axis represents the actual diagnosed risk of mental health problems. The diagonal dashed line represents the perfect prediction of the ideal model. The solid lines represent the performance of the column plots, where closer to the diagonal dashed line indicates a better prediction.

Fig.6. ROC curves for training and validation groups. The y-axis represents the true positive rate of risk prediction. The x-axis represents the false positive rate of risk prediction. The ROC curves for the training and validation groups are shown in black and red.

Fig.7. Decision curve analysis for mental health problems incidence risk nomogram. The y-axis measures the net benefit. The solid red line represents nomogram of the risk of developing a mental health problem. The light blue dashed line represents the hypothesis that all participants were diagnosed with a mental health problem. The black dashed line represents the hypothesis that there is no risk of a mental health problem. The DCA showed that using this mental health problem incidence risk nomogram in the current study to predict mental health problem incidence risk increase in benefit than the intervention all patients or no intervention all patient if the threshold probability of a patients and a doctor is >10%.


Factory workers and miners working in Urumqi
7,500 people volunteered for the survey
Issued a total of 7,500 questionnaires
Collected a total of 7,315 questionnaires
7,118 valid questionnaires
Include criterion
-workers working in mining enterprises or factories in Urumqi
-workers with a history of working for more than one year
-Workers with no history of mental illness and no history of taking psychotropic drugs
197 invalid questionnaires
(Questionnaires with missing data or those that do not meet the inclusion criteria)
Training group (n=4,955)
Validation group (n=2,163)

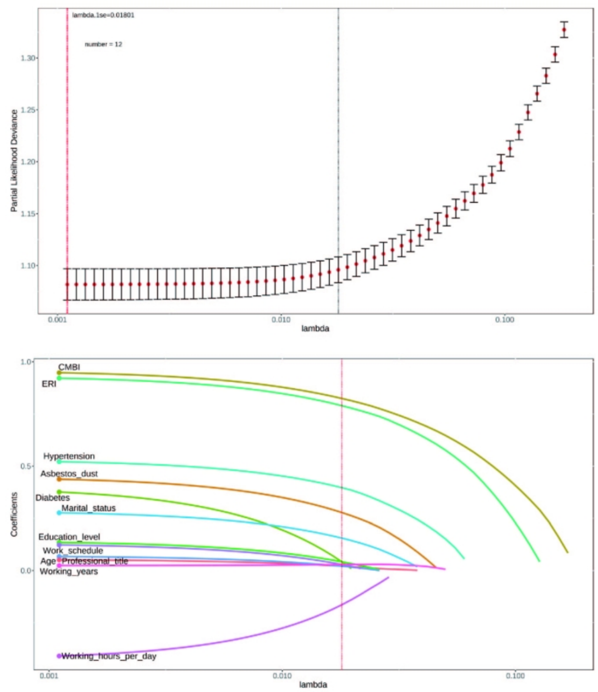
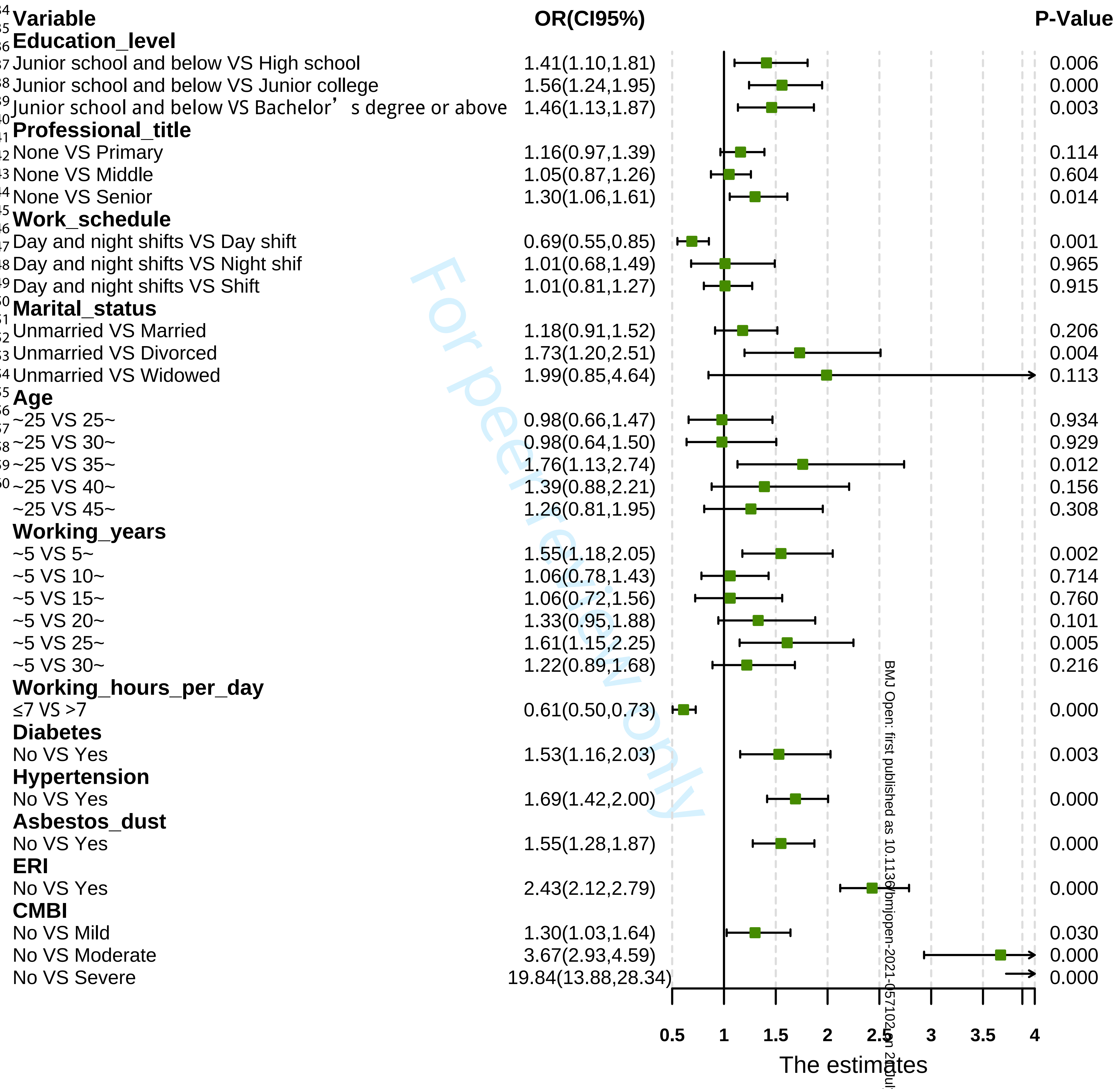
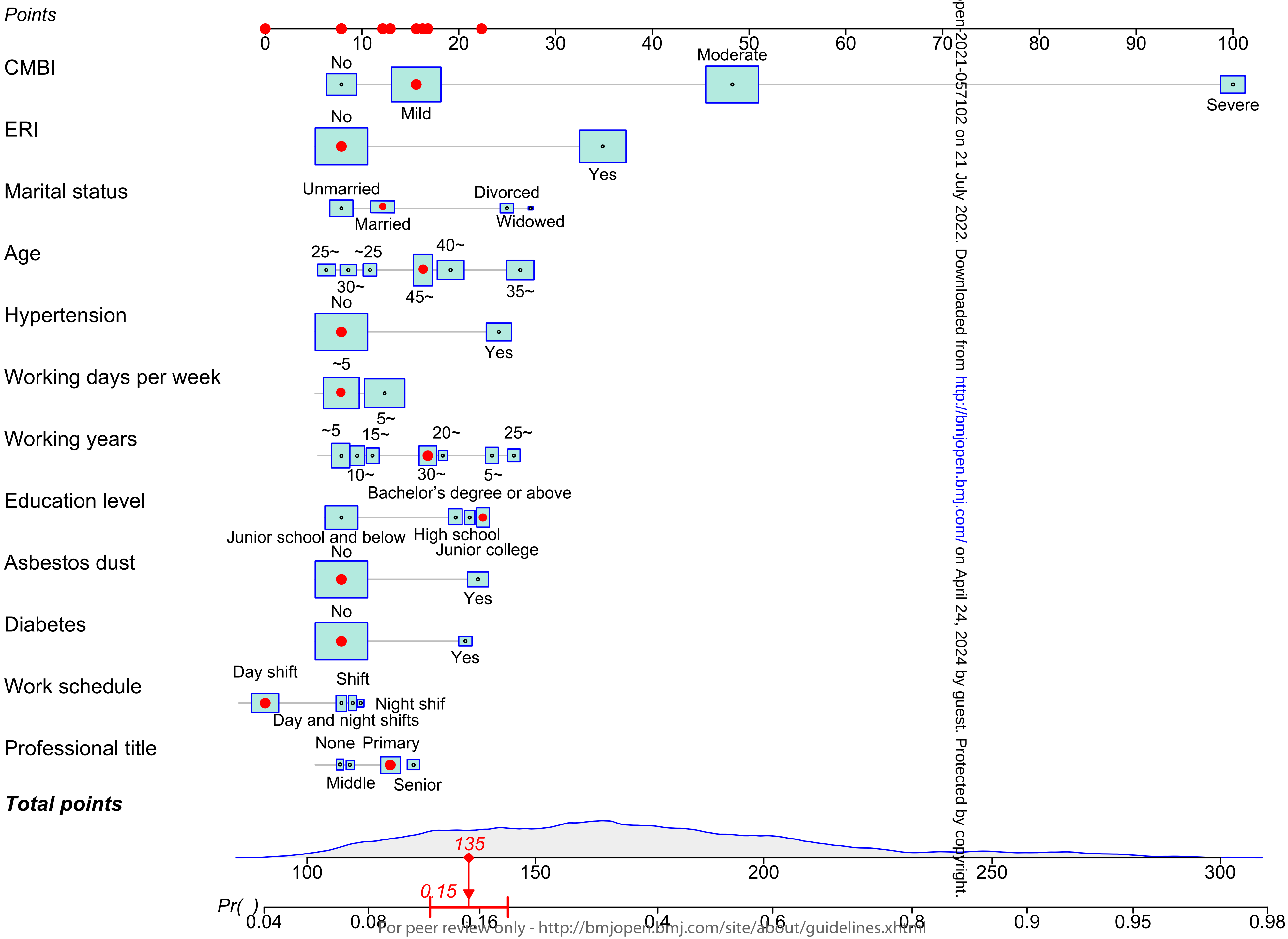
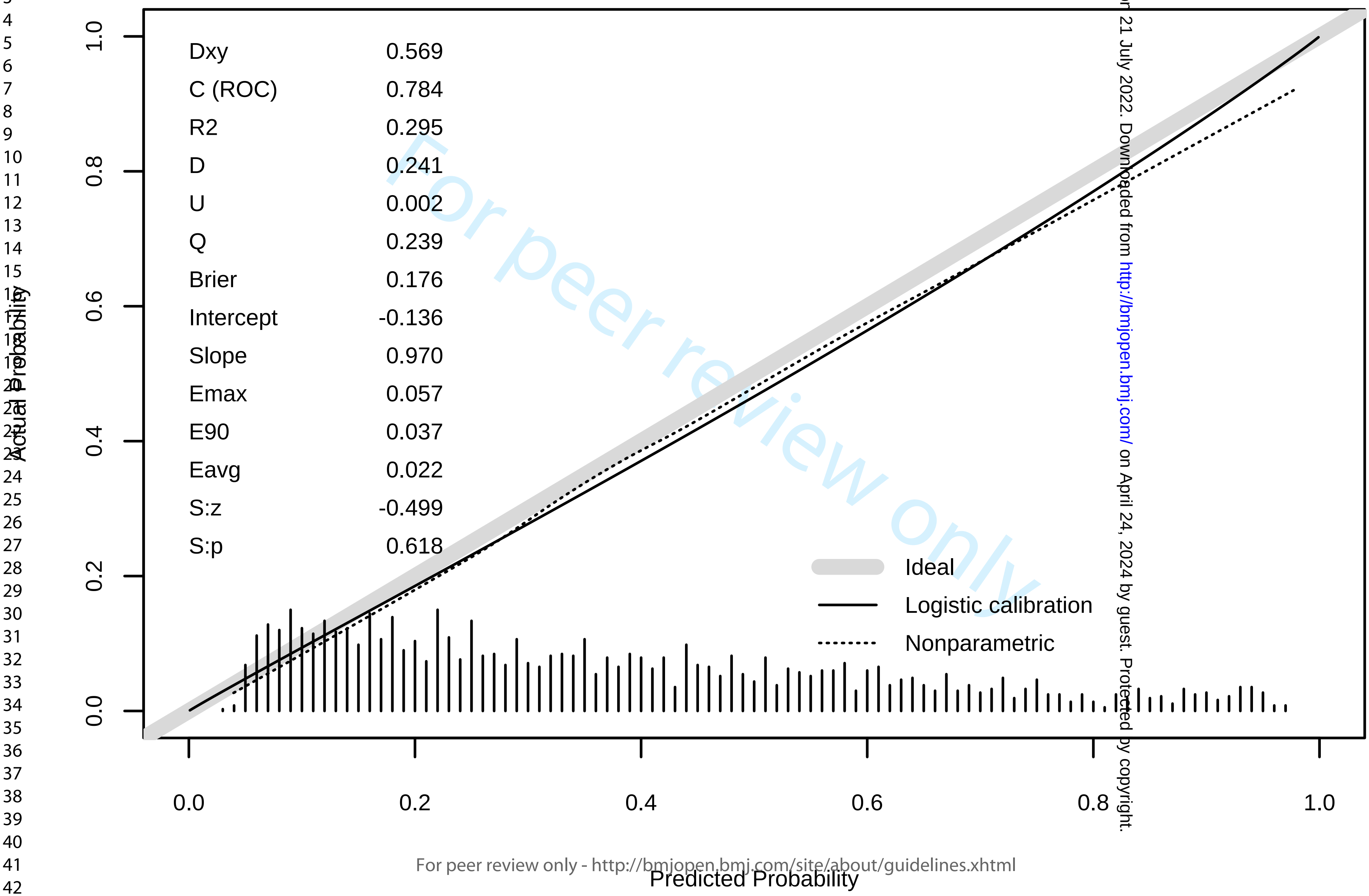


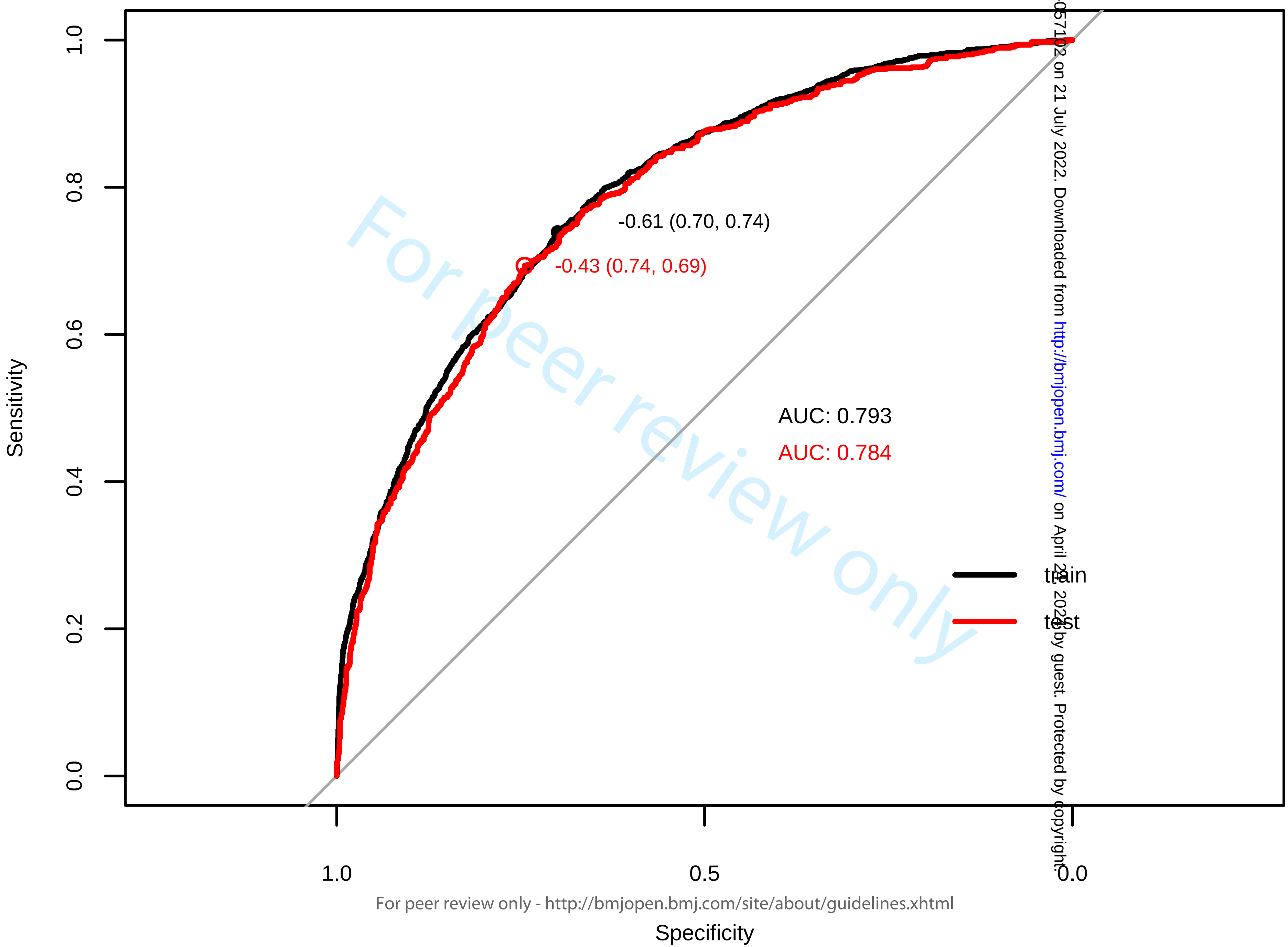
Fig.2. Feature selection using the LASSO binary logistic regression model. (A) Feature selection for the LASSO binary logistic regression model. The partial likelihood deviation (binomial deviation) curve was plotted against lambda by validating the optimal parameter lambda in the LASSO model. Dotted vertical lines were drawn based on 1 SE of the minimum criteria (the 1-SE criteria). (B) Feature selection was performed using the LASSO binary logistic regression model. A Coefficient profile weas plotted based on the lambda series in Figure 1(A), and 12 features with non-zero coefficients were selected by optimal lambda.

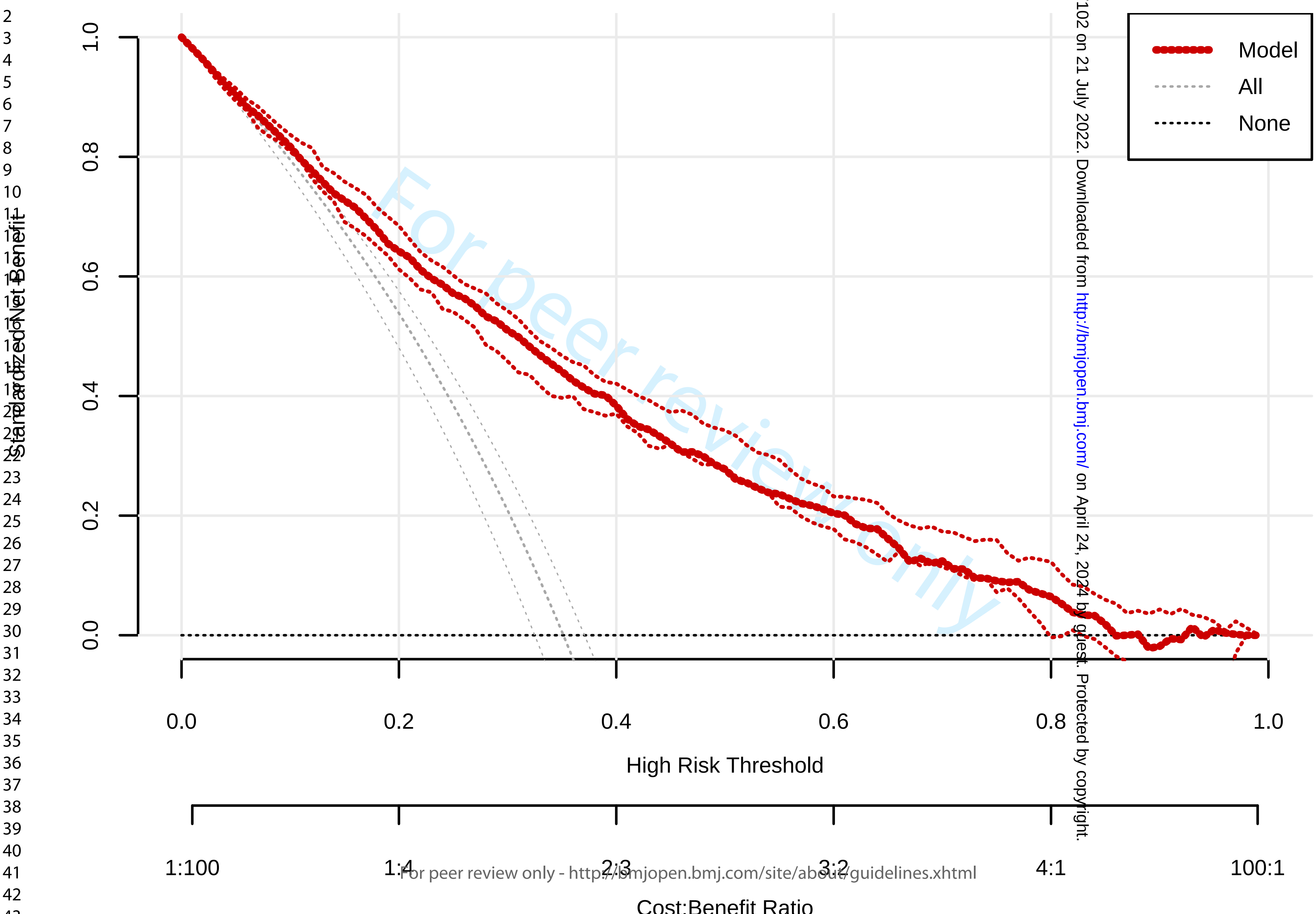
210x297mm (300 x 300 DPI)


Variable
OR(CI95%)
P-Value
Education_level
Junior school and below VS High school 1.41(1.10,1.81) 0.006
Junior school and below VS Junior college 1.56(1.24,1.95) 0.000
Junior school and below VS Bachelor's degree or above 1.46(1.13,1.87) 0.003
Professional_title
None VS Primary 1.16(0.97,1.39) 0.114
None VS Middle 1.05(0.87,1.26) 0.604
None VS Senior 1.30(1.06,1.61) 0.014
Work_schedule
Day and night shifts VS Day shift 0.69(0.55,0.85) 0.001
Day and night shifts VS Night shif 1.01(0.68,1.49) 0.965
Day and night shifts VS Shift 1.01(0.81,1.27) 0.915
Marital_status
Unmarried VS Married 1.18(0.91,1.52) 0.206
Unmarried VS Divorced 1.73(1.20,2.51) 0.004
Unmarried VS Widowed 1.99(0.85,4.64) 0.113
Age
~25 VS 25~ 0.98(0.66,1.47) 0.934
~25 VS 30~ 0.98(0.64,1.50) 0.929
~25 VS 35~ 1.76(1.13,2.74) 0.012
~25 VS 40~ 1.39(0.88,2.21) 0.156
~25 VS 45~ 1.26(0.81,1.95) 0.308
Working_years
~5 VS 5~ 1.55(1.18,2.05) 0.002
~5 VS 10~ 1.06(0.78,1.43) 0.714
~5 VS 15~ 1.06(0.72,1.56) 0.760
~5 VS 20~ 1.33(0.95,1.88) 0.101
~5 VS 25~ 1.61(1.15,2.25) 0.005
~5 VS 30~ 1.22(0.89,1.68) 0.216
Working_hours_per_day
≤7 VS >7 0.61(0.50,0.73) 0.000
Diabetes
No VS Yes 1.53(1.16,2.03) 0.003
Hypertension
No VS Yes 1.69(1.42,2.00) 0.000
Asbestos_dust
No VS Yes 1.55(1.28,1.87) 0.000
ERI
No VS Yes 2.43(2.12,2.79) 0.000
CMBI
No VS Mild 1.30(1.03,1.64) 0.030
No VS Moderate 3.67(2.93,4.59) 0.000
No VS Severe 19.84(13.88,28.34) 0.000
0.5
1
1.5
2
2.5
3
3.5
4
The estimates


Points
0 10 20 30 40 50 60 70 80 90 100
CMBI
No
Mild
Moderate
Severe
ERI
No
Yes
Marital status
Unmarried
Married
Divorced
Widowed
Age
25~
~25
30~
40~
45~
35~
Hypertension
No
Yes
Working days per week
~5
5~
Working years
~5
15~
20~
25~
10~
30~
5~
Education level
Bachelor's degree or above
Junior school and below
High school
Junior college
Asbestos dust
No
Yes
Diabetes
No
Yes
Work schedule
Day shift
Shift
Night shif
Day and night shifts
Professional title
None
Primary
Middle
Senior
Total points
135
100 150 200 250 300
Pr( )
0.15
0.04 0.08 0.16 0.4 0.6 0.8 0.9 0.95 0.98


Dxy 0.569
C (ROC) 0.784
R2 0.295
D 0.241
U 0.002
Q 0.239
Brier 0.176
Intercept -0.136
Slope 0.970
Emax 0.057
E90 0.037
Eavg 0.022
S:z -0.499
S:p 0.618
Ideal
Logistic calibration
Nonparametric
Actual Probability
Predicted Probability
0.0
0.2
0.4
0.6
0.8
1.0


1.0
0.8
0.6
0.4
0.2
0.0
Sensitivity
-0.61 (0.70, 0.74)
-0.43 (0.74, 0.69)
AUC: 0.793
AUC: 0.784
train
test
1.0
0.5
0.0
Specificity


Standardized Net Benefit
1.0
0.8
0.6
0.4
0.2
0.0
Model
All
None
0.0
0.2
0.4
0.6
0.8
1.0
High Risk Threshold
1:100
1:4
2:3
3:2
4:1
100:1
Cost:Benefit Ratio