Journal of Affective Disorders 355 (2024) 363–370

Contents lists available at ScienceDirect

Journal of Affective Disorders

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

Research paper

Development and validation of a prediction score to assess the risk of depression in primary care

Francesco Lapi [a,*,1], Giovanni Castellini [b,1], Valdo Ricca [b], Iacopo Cricelli [c], Ettore Marconi [a], Claudio Cricelli [d]

[a] Health Search, Italian College of General Practitioners and Primary Care, Florence, Italy
[b] Psychiatric Unit, Department of Health Sciences, University of Florence, Italy
[c] Genomedics SRL, Florence, Italy
[d] Italian College of General Practitioners and Primary Care, Florence, Italy

ARTICLE INFO

Keywords:
Depression
Score
Prediction model
Primary healthcare
General practitioner

ABSTRACT

Background: Major depression is the most frequent psychiatric disorder and primary care is a crucial setting for its early recognition. This study aimed to develop and validate the DEP-HScore as a tool to predict depression risk in primary care and increase awareness and investigation of this condition among General Practitioners (GPs).
Methods: The DEP-HScore was developed using data from the Italian Health Search Database (HSD). A cohort of 903,748 patients aged 18 years or older was selected and followed until the occurrence of depression, death or end of data availability (December 2019). Demographics, somatic signs/symptoms and psychiatric/medical comorbidities were entered in a multivariate Cox regression to predict the occurrence of depression. The coefficients formed the DEP-HScore for individual patients. Explained variance (pseudo-R^2), discrimination (AUC) and calibration (slope estimating predicted-observed risk relationship) assessed the prediction accuracy.
Results: The DEP-HScore explained 18.1 % of the variation in occurrence of depression and the discrimination value was equal to 67 %. With an event horizon of three months, the slope and intercept were not significantly different from the ideal calibration.
Limitations: The DEP-HScore has not been tested in other settings. Furthermore, the model was characterized by limited calibration performance when the risk of depression was estimated at the 1-year follow-up.
Conclusions: The DEP-HScore is reliable tool that could be implemented in primary care settings to evaluate the risk of depression, thus enabling prompt and suitable investigations to verify the presence of this condition.

1. Introduction

Major depression is the most frequent mental disorder in Western countries (Alonso et al., 2004; Aragonès et al., 2004; Bellón et al., 2013; King et al., 2013), with an annual prevalence of 4 % (Alonso et al., 2004). Mood disorders are associated with the highest economic burden among mental disorders (King et al., 2008; Trautmann et al., 2016), with € 118 billion annual cost for depression estimated in European countries in 2004 (Sobocki et al., 2006). Primary care is an important setting for early recognition of depression; indeed, there is a central role of general practitioners (GPs) in systematic and opportunistic screening of depressive disorder (Barry et al., 2023; Medina et al., 2020). As per Lech and coworkers, the cooperation of GPs with mental health specialists should be fostered in identification and treatment of depression according to clinical guidelines. Among GPs, perceived usefulness of the clinical guidelines was positively associated (4.7-fold higher) with the usage of the guidelines themselves (Lech et al., 2022).

Most people seeking help for depressive symptoms are treated in primary care (Australian Bureau of Statistics 4329.0, 2011) and almost 25 % of primary care attendees report current depressive symptoms. Indeed, it has been demonstrated in the primary care setting that the first presentation of a mood disorder is often preceded or accompanied by somatic signs/symptoms, such as gastrointestinal or urological functional symptoms and/or headache (Castellini et al., 2016a).

The need for fine-tuned indicators of signs and/or symptoms suggestive for depression in primary care is demonstrated because no

* Corresponding author at: Health Search, Italian College of General Practitioners and Primary Care, Via del Sansovino 179, 50142 Florence, Italy.
E-mail address: lapi.francesco@simg.it (F. Lapi).
[1] Contributed equally.

https://doi.org/10.1016/j.jad.2024.03.160
Received 7 April 2023; Received in revised form 25 March 2024; Accepted 26 March 2024
Available online 27 March 2024

relevant reduction of prevalence rate has been observed for this condition, despite the availability of treatment of proven efficacy and the improvement in screening and detection of depressive symptomatology (Ludman et al., 2000). Although early onset and treatment of depression does not seem necessarily associated with an improvement of disease outcome (Herzog et al., 2021), the use of prediction score supporting the case finding is considered helpful to initiate an individualized treatment approach (Kraus et al., 2020).

Several studies demonstrated the high burden of depression, in terms of impact on general functioning, mortality rate and comorbid disorders (Alonso et al., 2004; Aragonès et al., 2004; Castellini et al., 2016a; King et al., 2013). Indeed, poor physical health is considered an important risk factor for depression in later life (Ormel et al., 1994). On the other hand, major depressive disorder is associated with several chronic medical conditions (Moussavi et al., 2007), including cardiovascular, metabolic (Ormel et al., 1994) and urological (Castellini et al., 2016b) diseases. Despite improvements in treatment interventions and the whole healthcare process for depression (Bellón et al., 2013), primary care needs prevention strategies to reduce the burden of this condition. Indeed, there are several issues related to the recognition of depressive disorder in the primary care setting. In particular, the lack of routine assessment as well as the incomplete evaluation of risk factors might lead to a delay in syndrome recognition (Ludman et al., 2000). Bellón et al. developed an algorithm for personalized prevention of major depression in primary care attendees (Bellón et al., 2011). Namely, the predicD algorithm was built according to several variables, including gender, age, education, childhood physical abuse, lifetime depression and other demographic variables (Bellón et al., 2011; King et al., 2013). Compared to other clinical specialized settings, a lower range of specific clinical variables is available in primary care. Nevertheless, primary care data have the advantage of including large representative and heterogenous populations with several years of follow-up, with lifestyle and clinical variables that are longitudinally updated over time. For this reason, a reliable and easy-to-use algorithm to detect depression in primary care might allow for the disorder to be suspected or diagnosed in the early phases, staging of the risk, verification of the diagnosis and related severity and the prevention of relapses. Indeed, it has been reported that between 10 % and 21 % of GPs attendees with subthreshold symptoms and no history of depression develop major depression over 6 months (Davidson et al., 2015) and 2 years (Karsten et al., 2011), respectively. We therefore hypothesized that a clinical decision support system (CDSS) including this type of algorithm should improve GPs' capacity to recognize cases of depression. By using this tool, GPs can direct their focus towards particular subgroups of patients who are at a higher risk of developing depression, while also managing their workload effectively. After identifying these high-risk patients, further assessments can be conducted using specialized tools, such as the PHQ-2/9,(Levis et al., 2020) to confirm the diagnosis, evaluate its severity, and determine whether a specialist referral is necessary.

We therefore aimed to develop and validate the DEPression Health Search score (DEP-HScore) for predicting the risk of depression in primary care.

2. Methods

2.1. Data source

Data for the development of the DEP-HScore were collected from the Health Search Database (HSD), an Italian general practice database that includes patients' records for a group of over 1000 GPs homogenously distributed across Italy. The details of the HSD structure are extensively reported elsewhere (Castellini et al., 2016a; Guglielmi et al., 2017; Lapi et al., 2012). In brief, GPs voluntarily agreed to collect patients' information after attending training courses for data entry. In order to include GPs in the present study, they should meet standard quality criteria pertaining to the levels of coding, prevalence of well-known diseases, mortality rates and years of recording (Cricelli et al., 2003). The present study included 800 GPs homogenously distributed across all areas of Italy, covering a patient population of 1,163,855 individuals.

Patients' demographic details are linked with an encrypted code to clinical records (diagnoses, referrals, tests prescriptions, and results), drug prescriptions (drug name, date of filled prescription and number of days' supply), lifestyle-related records (i.e., body mass index [BMI], smoking and alcohol consumption), hospital admissions and date of death. Diagnoses and medications are coded using internationally recognized codes, such as the 9th Revision of the International Classification of Diseases, Clinical Modification version (ICD-9-CM), and the ATC classification, respectively.

The study protocol was approved by the Scientific Committee of the Italian College of General Practitioners and Primary Care.

2.2. Study population

A cohort of patients aged 18 years or older during the period between 1 January 2002 and 31 December 2018 was selected. The date of the first GP's visit within the eligibility period was the study *index date*. Patients were followed until the occurrence of one of these events, whichever came first: depression (i.e., *event date*), death from any cause, end of data registration with GP or end of study period (31 December 2019). Patients diagnosed with depression (i.e., as defined below) in the overall period preceding or on the index date were excluded. Then, the cohort was randomly divided into two sub-cohorts containing approximately two-thirds and one-third of patients, respectively; these were referred to as the development and (internal) validation cohorts.

We adopted the same selection criteria to form three temporal (as proxy for external validity) validation cohorts, which were operationally defined within the HSD in 2020, 2021 and 2022. These three time-windows differ from those adopted to develop and internally validate the DEP-HScore. Indeed, when this these algorithms are developed through a representative data source with the aim to apply this score in the same setting, the “internal” validity is sufficient (Ramspek et al., 2021). Nevertheless, the changes over time of patients' characteristics sustain temporal validation as advisable for any score intended to be implemented in clinical practice, especially when a longitudinal data collection is considered (Steyerberg et al., 2010b).

2.3. Event definition

We captured the first-ever episode of depression coded via ICD-9-CM codes (i.e., 311, 296.2, 296.3) registered by GPs during the study follow-up. We also identified those cases with a specialist referral (i.e., psychiatrist, neurologist) within 1 month before or after the event date. When the date of the specialist referral preceded the date of the ICD-9-CM code record, it was adopted as the event date. It is expected that Italian GPs did not code a depression diagnosis without a specialist's confirmation. Nevertheless, the HSD could miss some information on the prescription of the specialist's referral. The aforementioned procedures therefore allowed us to identify those cases that were more reliable and specific in terms of event definition and these cases were then used to perform sensitivity analyses.

2.4. Determinants for depression onset

We considered all determinants, along with age (centered on the cohort's mean age) and gender, known to affect the onset of depression. They were identified according to the medical literature (Barkow et al., 2003; Bellón et al., 2011; King et al., 2011), including our prior work (Castellini et al., 2016a) and clinical rationale. Namely, we selected those diagnoses coded via the ICD-9-CM (and free-text wherever necessary) or other measures of interest that were registered in the overall period preceding the index date, inclusive. We identified lifestyle, variables coding for somatic signs/symptoms and psychiatric/

medical comorbidities. Specifically, the lifestyle variables comprised smoking (i.e., current, former, non-smoker), alcohol abuse and alcohol-related diseases, and also obesity (BMI $\geq$ 30 kg/m^2). Variables coding for somatic signs/symptoms included the presence of insomnia, hypersomnia, weight loss, joint pain, dizziness, headache or migraine, back pain, neck pain, amnesic syndrome, fatigue, psychalgia, irritable bowel syndrome, constipation, abdominal pain and pelvic inflammatory disease.

Psychiatric comorbidities were as follows: neurotic disorders, stress-related disorders, substance abuse, anorexia and bulimia nervosa, and also eating disorders not otherwise specified.

Furthermore, medical comorbidities included congestive heart failure, ischemic heart disease, peripheral vascular disease, chronic respiratory disease, rheumatic diseases, fibromyalgia, peptic ulcer disease, mild liver disease, severe liver disease, diabetes with and without chronic complications, hemiplegia or paraplegia, kidney disease (also referred to any value of GFR $<$ 60 mL/min/1.73 m^2; last measurement prior or on the index date), dialysis and/or kidney transplantation or free text "dialysis" or "kidney transplant"), any form of cancer, metastatic cancer, ischemic stroke, transient ischemic attack, celiac disease and senile dementia. Finally, we included the number of contacts with GPs in the years before the index date (1–5 contacts or $\geq$ 6 contacts; i.e., nominally below or above the mean values).

2.5. Data analyses

To ensure the appropriate conduct and reporting of prediction/diagnostic studies (Bossuyt et al., 2003), development and validation of the DEP-HScore was conducted according to the TRIPOD statements (Collins et al., 2015) and the PROBAST tool (Wolff et al., 2017).

We reported descriptive statistics for continuous (mean and standard deviation [SD]) and categorical values (% and related 95 % confidence interval [CI]). We estimated the incidence rate of depression by dividing the number of events by the person-years cumulated during follow-up.

The association between demographics, somatic signs/symptoms, comorbidities and the onset of depression was quantified in the development cohort. Namely, along with age (centered on the mean value) and gender, all the aforementioned candidate determinants were entered in a multivariate Cox regression model according to a backward stepwise approach ($p < 0.10$ for entering and < 0.15 for exiting variables). By doing so, we entered the covariates in the model according to clinical and statistical bases (Wang et al., 2008). We therefore obtained a regression beta coefficient for each determinant. Thereafter, a patient-specific score, named the DEP-HScore, was computed through the linear combination of individual coefficients, excluding those estimated for missing categories. The DEP-HScore was therefore categorized in deciles. Using the internal validation cohort, we evaluated the accuracy of the DEP-HScore by calculating the explained variance (pseudo-R^2) as a performance measure, the area under the curve (AUC) as a discrimination measure and the predicted/observed ratio as a calibration measure (where a ratio of 1 indicates perfect calibration). Pseudo-R^2 indicates how the selected determinants are able to explain the variation of depression occurrence; the AUC provides an indication of the score capacity to distinguish cases versus non-cases of depression; and the score calibration allowed quantification of the distance between predicted and observed risk against perfect (i.e., ideal bisector) calibration (Moons et al., 2015; Steyerberg et al., 2010b).

We calculated the predicted risk of incident depression at 3, 6 and 12 months by combining the regression coefficients of the DEP-HScore with the baseline hazard function for the diagnosis of depression. By doing so, we were able to evaluate the different distribution of incident cases of depression during the first year of follow-up.

Furthermore, we evaluated the presence of "moderate" calibration using a flexible calibration curve. In this respect, we used a spline-based tool with confidence intervals to plot the predicted versus observed risk of depression (Van Calster et al., 2016; Van Hoorde et al., 2015). The calibration was formally tested at 3, 6 and 12 months of follow-up by calculating the calibration slope and related intercept for all these intervals.

In the temporal "external" cohorts, defined in 2020, 2021 and 2022, we calculated pseudo-R^2 as a performance measure and AUC as a discrimination measure. For the same years, calibration slope and calibration-in-the-large (CITL) (Steyerberg et al., 2010b) were obtained and formally tested by contrasting expected and observed risk deciles of the DEP-HScore. For these analyses, every determinant of incident depression was operationally defined on 1 January of each year.

We conducted two sensitivity analyses. First, we tested the burden of prevalent cases of depression on the score performance, given that depression is frequently defined as a "period" prevalent disease (Lim et al., 2018). Namely, we re-ran the primary models, limiting the exclusion of patients previously diagnosed with depression to those presenting the event in the 2 years preceding or on the index date. By doing so, we accounted for the burden of depression relapses in the primary care setting. Second, to account for event misclassification, we repeated the primary analysis by limiting cases of depression to those coupled with psychiatric/neurological referrals. All these analyses were conducted in the internal validation cohort by calculating pseudo-R^2, the AUC, and the predicted/observed ratio and. All analyses were conducted using Stata Version 14.0. To report the effect size of the individual determinants forming the DEP-HScore, a significance level was set to $\alpha < 0.05$.

3. Results

A total of 1,355,623 patients met the study inclusion criteria. The mean age was 47.61 years old and there were more females (around 53 %) than males forming the development (n = 903,748) and internal validation (n = 451,875) cohorts. During follow-up, we identified 42,006 and 21,031 patients with depression, yielding overall incidence rates of 50.74 (95 % CI = 50.26–51.23) and 50.81 (95 % CI = 50.13–51.5) per 10,000 person-years in the development and internal validation cohort, respectively. The rate of depression was twofold in females compared to males. In addition, the rate of depression increased with age, achieving the greatest rate in patients aged 75–84 years (Table 1).

Table 2 depicts the degree of association with onset of depression for each determinant across the observation period for the development cohort. According to the backward stepwise approach, 30 out of 42 were kept in the final models (beta coefficients are depicted in Supplementary Table S1). As far as demographic and lifestyle characteristics are concerned, female gender, obesity and current smokers showed a positive association with a higher incidence rate of depression, increasing the risk of depression by 1.72-, 1.11- and 1.35-fold, respectively. Among the somatic signs/symptoms included in the final model, the strongest associations with the occurrence of depression were found for fatigue (hazard ratio [HR] = 1.32; 95 % CI = 1.25–1.4) and insomnia (HR = 1.24; 95 % CI = 1.01–1.52). The other determinants showed HRs ranging from 1.21 to 1.17. Only psychalgia had a non-significant association with the occurrence of depression. Regarding "medical comorbidities", the strongest associations were found for celiac disease and hemiplegia/paraplegia, with 1.45- and 1.42-fold increased risk of depression, respectively. The other risk factors showed an increase in the risk of occurrence of depression ranging from 1.07 to 1.29. In contrast, the presence of kidney disease was associated with a reduction in the risk of depression (HR = 0.92; 95 % CI = 0.89–0.95). Among the "psychiatric comorbidities", eating disorders showed a positive association with the occurrence of depression, increasing the disease risk by 1.69-fold. In contrast, neurotic disorders and stress-related disorders seemed to be associated with a reduced risk of depression: 37 % and 27 %, respectively.

The individual DEP-HScore was determined by combining the 30 beta-coefficients (corresponding to the HRs; see Supplementary Table 1

Table 1
Characteristics of patients in the development and in the internal validation cohort.

	Development Cohort (n = 903,748)	Validation Cohort (n = 451,875)
	n (%)	n (%)
Sex (Male)		
Female	479,471 (53.1)	240,249 (53.2)
Age categories *(18–34)*		
35–44	183,999 (20.4)	92,205 (20.4)
45–54	147,608 (16.3)	73,609 (16.3)
55–64	124,197 (13.7)	62,157 (13.8)
65–74	101,999 (11.3)	51,373 (11.4)
75–84	66,518 (7.4)	32,973 (7.3)
≥85	21,229 (2.4)	10,551 (2.3)
Age (years, mean [±SD])	47.61 ± 18.27	47.61 ± 18.25
Smoking		
Smokers	48,041 (5.3)	24,044 (5.3)
Ex-smokers	27,269 (3.0)	13,433 (3.0)
Missing	750,788 (83.1)	375,251 (83.0)
Obesity		
Obesity	25,116 (2.8)	12,714 (2.8)
Missing	789,443 (87.4)	394,689 (87.3)
Alcohol consumption	3821 (0.4)	1816 (0.4)
Somatic signs/symptoms		
Weight loss	1186 (0.1)	550 (0.1)
Joint pain	11,428 (1.3)	5733 (1.3)
Insomnia	3117 (0.3)	1524 (0.3)
Hypersomnia	4941 (0.6)	2477 (0.6)
Dizziness	19,877 (2.2)	9816 (2.2)
Migraine/headache	26,246 (2.9)	13,273 (2.9)
Back pain	47,660 (5.3)	23,831 (5.3)
Neck pain	15,285 (1.7)	7501 (1.7)
Amnesic syndrome	15 (0.0)	–
Fatigue	19,848 (2.2)	9966 (2.2)
Psychalgia	2396 (0.3)	1204 (0.3)
Irritable bowel syndrome	10,715 (1.2)	5392 (1.2)
Constipation	7690 (0.9)	3765 (0.8)
Abdominal pain	17,604 (2.0)	8965 (2.0)
Pelvic inflammation disease	210 (0.0)	93 (0.0)
Medical comorbidities		
Congestive heart failure	4487 (0.5)	2255 (0.5)
Ischemic heart disease	7507 (0.8)	3629 (0.8)
Peripheral vascular disease	18,964 (2.1)	9348 (2.1)
Chronic respiratory disease	50,337 (5.6)	25,046 (5.5)
Rheumatic disease	4274 (0.5)	2154 (0.5)
Peptic ulcer disease	17,719 (2)	9070 (2.0)
Mild liver disease	6185 (0.7)	3061 (0.7)
Severe liver disease	303 (0.0)	147 (0.0)
Diabetes without complications	38,630 (4.3)	19,342 (4.3)
Diabetes with complications	215 (0.0)	124 (0.0)
Hemiplegia or paraplegia	718 (0.1)	343 (0.1)
Kidney disease	63,121 (7.0)	31,483 (7.0)
Cancer	27,225 (3.0)	13,577 (3.0)
Metastatic cancer	332 (0.0)	178 (0.0)
Celiac disease	856 (0.1)	456 (0.1)
Ischemic Stroke	7062 (0.8)	3534 (0.8)
Transient ischemic attack	3683 (0.4)	1863 (0.4)
Fibromyalgia	2006 (0.2)	968 (0.2)
Psychiatric comorbidities		
Neurotic disorders	42,883 (4.8)	21,492 (4.8)
Stress-related disorders	2545 (0.3)	1190 (0.3)
Eating disorder[a]	699 (0.1)	360 (0.1)
Substance abuse	6806 (0.8)	3448 (0.8)
Contacts		
≥6	335,197 (37.1)	167,653 (37.1)
Mean [±SD]	6.26 ± 7.53	6.27 ± 7.51

SD = standard deviation.
[a] Anorexia and bulimia.

Table 2
Determinants of depression occurrence in the development cohort.

	HR (95 % CI)[b]	p[c]
Sex (Male)		
Female	1.72 (1.68–1.75)	< 0.001***
Age (years)	1.02 (1.02–1.02)	< 0.001***
Smoking (Non-smokers)		
Smokers	1.35 (1.28–1.43)	< 0.001***
Ex-smokers	1.09 (1.02–1.16)	0.007**
Missing	1.2 (1.15–1.25)	< 0.001***
Obesity *(No)*		
Obesity	1.11 (1.05–1.18)	0.001**
Missing	1.04 (1–1.08)	0.053
Alcohol consumption	1.22 (1.06–1.4)	0.006**
Somatic signs/symptoms		
Insomnia	1.24 (1.01–1.52)	0.038*
Hypersomnia	1.21 (1.03–1.42)	0.022*
Migraine/headache	1.26 (1.2–1.33)	< 0.001***
Fatigue	1.32 (1.25–1.4)	< 0.001***
Psychalgia	1.14 (0.96–1.34)	0.132
Irritable bowel syndrome	1.17 (1.09–1.26)	< 0.001***
Abdominal pain	1.21 (1.14–1.29)	< 0.001***
Medical comorbidities		
Congestive heart failure	1.15 (1.03–1.28)	0.015*
Ischemic heart disease	1.07 (0.98–1.17)	0.146
Chronic respiratory disease	1.14 (1.1–1.18)	< 0.001***
Rheumatic disease	1.23 (1.11–1.36)	< 0.001***
Peptic ulcer disease	1.18 (1.11–1.25)	< 0.001***
Mild liver disease	1.3 (1.18–1.42)	< 0.001***
Severe liver disease	1.38 (0.9–2.1)	0.135
Diabetes without complications	1.12 (1.07–1.16)	< 0.001***
Hemiplegia or paraplegia	1.42 (1.07–1.88)	0.014*
Kidney disease	0.92 (0.89–0.95)	< 0.001***
Cancer	1.17 (1.12–1.23)	< 0.001***
Celiac disease	1.45 (1.09–1.93)	0.010*
Ischemic Stroke	1.21 (1.11–1.32)	< 0.001***
Transient ischemic attack	1.14 (1.01–1.28)	0.031*
Fibromyalgia	1.29 (1.1–1.5)	0.001**
Psychiatric comorbidities		
Neurotic disorders	0.73 (0.7–0.77)	< 0.001***
Stress-related disorders	0.63 (0.51–0.76)	< 0.001***
Eating disorders[a]	1.69 (1.29–2.22)	< 0.001***
Contacts *(1–5)*		
≥6	1.36 (1.33–1.38)	< 0.001***

HR = Hazard Ratio; CI = Confidence interval.
[a] Anorexia and bulimia.
[b] Multivariate Cox regression: covariates are included in the model according to stepwise backward (p values <0.10 for entering and < 0.15 for exiting variables).
[c] Significance: p-values refer to the Wald test whether the beta (β) coefficient of a given variable is statistically significantly different from 0. The related HRs are the exponentiated coefficients.
* $p < 0.05$.
** $p < 0.01$.
*** $p < 0.001$.

for specific values) that were associated with the aforementioned determinants. Each subject was assessed for their predicted risk of depression at 3 (6 and 12) months of follow-up using a Cox regression that included a baseline hazard function (intercept) and the relevant beta-coefficients based on the presence or absence of risk factors, with 1 (i.e. the product with beta coefficient was therefore equal to the beta values) indicating the presence of covariates and 0 (i.e. the product with beta coefficient was therefore null) indicating their absence.

When the DEP-HScore was categorized in deciles and applied to the internal validation cohort, it was able to explain 18.1 % (95 % CI = 17.7–18.49) of the variation in depression onset. In terms of discrimination, the AUC was 0.67 (95 % CI = 0.66–0.68). In order to evaluate the calibration performance, spline regression fitting for predicted versus observed risk of overall depression episodes showed some over- and underestimation. As a whole, the greater proportion of patients was associated with a predicted risk of approximately 0–0.02 (Fig. 1). Calibration slope and related intercept were formally tested at 3, 6 and 12 months of follow-up. As shown in Table 3, the equivalence hypothesis was not rejected with 3 months of follow-up.

With regard to temporal (as proxy of external) validity, pseudo-R^2 was 25 %, 24 %, and 28 % in 2020 and 2021, and 2022, respectively; AUC was 0.68 (95 % CI = 0.66–0.69) and 0.71 (95 % CI = 0.70–0.74) for 2020/2021 and 2022, respectively. In the 2020, 2021 and 2022 cohorts,

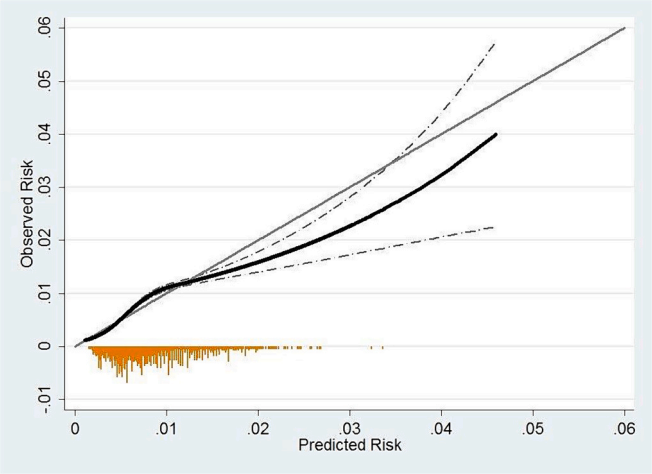

Fig. 1. Calibration curve showing 1-year predicted vs. observed risk of depression according to deciles of DEP-HScore: internal validation sample.

Table 3
Calibration slope and intercept between observed and predicted risks of depression, at 3, 6 and 12 months of follow-up: internal validation cohort.

t = 3 months	
Slope (p value)	0.92 (0.126)
Intercept (p value)	0.005 (0.878)
t = 6 months	
Slope (p value)	0.91 (0.034)
Intercept (p value)	−0.05 (0.05)
t = 12 months	
Slope (p value)	0.91 (0.007)
Intercept (p value)	−0.07 (< 0.001)

t = time of follow-up.
Significance: test of equivalence for calibration slope = 1 and intercept = 0.

the calibration slope did not reject the equivalence hypothesis, while tests on calibration intercepts were statistically significant (Table 4).

Concerning the sensitivity analysis, when we limited the exclusion of those patients diagnosed with depression in the 2 years preceding the index date, the score was able to explain 17.3 % (95 % CI = 17.0–18.0) of the variation for depression, while in terms of discrimination the AUC was equal to 0.67 (95 % CI = 0.67–0.69). When cases of depression were limited to those coupled with psychiatric referral, the score explained 19 % (95 % CI = 18.3–20) of the variation and AUC was equal to 0.68 (95 % CI = 0.66–0.69); for all the sensitivity analyses, calibration values were consistent with those obtained for the primary analysis.

Table 4
Calibration slope and Calibration-In-The-Large (CITL) between observed and predicted risks of depression, at 3 months of follow-up: temporal "external" validation cohorts.

2020	
Slope (p value)	0.98 (0.904)
CITL (p value)	−0.83 (< 0.001)
2021	
Slope (p value)	1.09 (0.460)
CITL (p value)	−0.24 (0.006)
2022	
Slope (p value)	0.85 (0.151)
CITL (p value)	−0.60 (< 0.001)

CITL = Calibration-In-The-Large.
Significance: test of equivalence for calibration slope = 1 and CITL = 0.

4. Discussion

The results of the present study demonstrated that the DEP-HScore has good accuracy as a prediction tool to identify depression onset in primary care. The design of the study to develop the model was able to identify predictors that were well defined and reproducible so increasing the generalizability of the model, as confirmed by the score calibrations calculated in 2020, 2021 and 2022. Unlike many prognostic models developed in hospital-based settings and then applied to primary care (Wynants et al., 2020), our model was specifically developed for use in general practice using data collected from general practice attendees.

Overall, the incidence rate of depression in the present population was found to be similar to those detected in other studies with a similar design (Vilagut et al., 2013; Wittchen et al., 2003). Considering that a large proportion of patients with psychiatric symptoms receive a first diagnosis of mood disorder by GPs (Castellini et al., 2016a), we attempted to overcome the aforementioned shortcomings using primary care data. The novelty of the present approach was based on several lifestyle and clinical variables dynamically updated over time and not limited to specialists or hospitalization records.

In accordance with previous observations, we confirmed the importance of somatic signs/symptoms as predictors of depression onset in the mid-term (Castellini et al., 2016a). Indeed, a significant association between medical comorbidities and onset of depression was also observed (McIntyre et al., 2007), confirming that somatic signs/symptoms can be considered as the first signals of a latent mood disorder (Castellini et al., 2016a). On the other hand, the data of the present study add further evidence that mood disorders possibly share common biological underpinnings with several medical conditions (Pariante, 2021), such as cardiovascular diseases (Joynt et al., 2003), metabolic syndromes (McIntyre et al., 2007) and urinary problems (Castellini et al., 2016b). The relationship between somatic signs/symptoms, organic comorbidity and psychopathology is of particular interest in primary care (Hüsing et al., 2018). Indeed, in clinical practice, GPs often deal with medically unexplained symptoms and somatoform disorders (Hiller and Fichter, 2004; Smits et al., 2009). Furthermore, it has been reported that, in primary care, 40 % of patients presenting with at least one somatic symptom not directly associated with a specific medical comorbidity suffer from somatoform disorder (Haller et al., 2015; Hiller and Fichter, 2004) and these patients generally show clinically relevant depression levels in over 50 % (Löwe et al., 2008). Considering that depressive combined with somatic signs/symptoms are often associated with active suicide ideations and a relevant disability (Wiborg et al., 2013), proper identification of this particular condition is a relevant issue for GPs' activity. A valid instrument for the identification of psychopathology hidden by somatic signs/symptoms is a matter of concern in primary care, considering the consequences of a lack of recognition of this relationship. Indeed, as a result of the overuse of somatic and underuse of mental healthcare, impairment remains high while treatment duration and costs continue to rise (Hüsing et al., 2018).

Overall, the DEP-HScore has an acceptable accuracy and its good properties have been confirmed by the explained variance, discrimination and calibration measures obtained in the validation datasets. It is of note that higher predictive capacity was observed for females and older patients. Compared to previous investigations (Dowrick et al., 2011; Rubenstein et al., 2007) that developed prognostic models such as the THREAD study to predict depression, the present model was based on a

larger population and a standardized assessment. The above-mentioned studies were insufficiently robust to use in the clinical prediction tool because of low prognostic accuracy (Chondros et al., 2018) and the development sample only included participants with mild to moderate depression and thus could not be generalized to new primary care patients who present with potential severe depression. The development of the Diagnostic Prognostic Index (Rubenstein et al., 2007) included a larger dataset (including 1471 primary care attendees) but was also unsuitable because the development sample excluded patients with subthreshold depression. Given that primary care subthreshold depression makes up the largest group of patients presenting with depressive symptoms, the prognostic model would not be generalizable to this population (Chondros et al., 2018). Furthermore, the length of the Diagnostic Prognostic Index (60 items) would limit its usability and usefulness in routine clinical practice (Toll et al., 2008). Finally, recent papers have reported the development and validation of prediction algorithms obtained using machine learning and electroencephalogram (EEG) data (Jan et al., 2022; Liu et al., 2022; Shahabi et al., 2023) Although these tools showed good performances, the use of an EEG is not easily applicable to primary care records, thus limiting its use to specialist centers.

The optimal early treatment strategy for patients with depression remains a topic of ongoing debate (Herzog et al., 2021; Rost et al., 2023). Nonetheless, GPs play a crucial role in depression screening, case finding, and early recognition which are relevant aspects to better decide on individualized approach for patient's treatment. By using a CDSS based on DEP-HScore, GPs can direct their focus towards particular subgroups of patients who are at a higher risk of developing depression, while also managing their workload effectively. After identifying these high-risk patients, further assessments can be conducted using specialized tools, such as the PHQ-2/9,(Levis et al., 2020) to confirm the diagnosis, evaluate its severity, and determine whether a specialist referral is necessary. Thus, the essential collaboration between GPs and mental health specialists should be facilitated.

This study has several strengths. First, the dimension of the HSD allowed us to develop and validate a score in a population whose features can be generalized to other Italian regular citizens. Second, most of the determinants of depression onset are represented by chronic diseases. Only in a primary care setting with a 20-year or longer look-back period, we were able to sensibly capture most of these conditions. Third, compared to prior studies, the DEP-HScore model can be used to predict future onset of depression, stratifying patients into different risk categories. This systematic approach would protect GPs from the issue of over-treating patients with subthreshold depression (Davidson et al., 2015) or the lack of efficacious treatment for major depression (Wang et al., 2007). Fourth, the length of the follow-up is longer compared to previous observations (e.g., 18 months for the predictD study; Fernández et al., 2018). Finally, our model should be easily implemented as CDSS in primary care, using electronic health records, which are mandatory for Italian GPs to provide patient's care (D.M. 4 March 2009 [G.U. n. 146, 26 June 2009]; D.P.C.M. 26 March 2008 [G.U. n. 124, 28 May 2008]). As stated above, such a tool should optimize the GPs' workload. According to the European Medical Device Regulation (MDR; EU 2017/745), these findings would be part of Clinical Evidence Report (CER) as part of any conformity assessment.

There are also limitations that should be mentioned. First, the results obtained might be applicable to primary care attendees but cannot necessarily be generalized to other settings. A tool like this could potentially be implemented in other primary care settings internationally, but it would require recalibration to consider the unique demographics of the patient population overseen by GPs. Indeed, primary care is an ideal setting for prevention, with more frequent attendees generally (because of health and social issues) at greater risk of depression compared with the general population. Second, given that some information on psychiatric referrals might be missing, the operational definition of depression might be prone to some misclassification. Reassuringly, when we rerun the analyses by limiting cases to those combined with psychiatric/neurological referrals, we found results that were consistent with those found in the primary analyses. Third, a pseudo-R^2 value of 18 % might suggest that some determinants are still missing. Indeed, variables concerning socioeconomic status and emotional and traumatic experiences are not generally available in clinical data sources. Nevertheless, the fact that we adopted a number of determinants (n = 30 out of 42), along with the presence of good discrimination and calibration,(de Hond et al., 2022) meant that GPs' opportunity to identify the stable and/or transitory psychological impairments of their patients in the family context seems reassuring for the application of the DEP-HScore in clinical practice. Fourth, the predicted risk showed an underestimation when plotted against observed risk. Nevertheless, the confidence intervals for spline-based regression mostly overlapped the line of perfect calibration, and the greatest proportion of patients spread up to a predicted risk of 0.02, where the spline curve was comparable with the line of perfect calibration. This better accuracy for patients belonging to the lower risk categories is due to the fact that they represent most of the patients being cared for by GPs. The primary care setting indeed has the advantage of identifying most of the patients at low/moderate risk of depression. This is the largest category for which intervention strategies might have the most relevant implications in terms of public health. The greater risk categories of depression showed more unstable estimates (as shown in Fig. 1), probably related to the reduced sample size. However, from a clinical perspective, it is hard to imagine that very-high-risk patients were frequently encountered by GPs and, whenever suspected, they would be referred to psychiatrists. Fifth, the dilution of incidence rate during the first year of follow-up led to rejection of the equivalence hypothesis when the risk of depression was estimated at 1-year follow-up. Nevertheless, for both calibration slope and intercept, the equivalence hypothesis was not rejected with a 3-month event horizon. The fact that CITL was tested as statistically significant in the external (i.e., temporal) cohorts was likely due to reduced power. Reassuringly, when the calibration slopes were formally tested, they did not reject the equivalence hypothesis in 2020, 2021 and 2022. Finally, the score was not formally tested in other data sources and/or settings. Nonetheless, when these algorithms are developed through a representative data source with the aim to apply this score in the same setting, the “internal” validity is sufficient (Ramspek et al., 2021). Additionally, we established consistent findings through temporal validation, which could be functioned as a proxy for external validation (Steyerberg et al., 2010a).

5. Conclusions

These findings indicate that the DEP-HScore can be implemented in primary care for risk prediction of depression. GPs could therefore benefit from a decision tool that automatically notifies the risk of depression, along with visualizing this prediction in a dedicated dashboard. Thus, the DEP-HScore could be implemented in a busy GP practice with the goal of increasing the suspicion of depression and facilitating timely and appropriate investigations to ascertain the presence of this condition.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jad.2024.03.160.

Ethical considerations

The study protocol was approved by the Scientific Committee of the Italian College of General Practitioners and Primary Care. This study followed the principles of the Declaration of Helsinki and was compliant with the TRIPOD statements (Collins et al., 2015).

Funding

This research was supported by the Italian College of General

Practitioners and Primary Care.

CRediT authorship contribution statement

Francesco Lapi: Writing – original draft, Validation, Supervision, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Giovanni Castellini:** Writing – original draft, Supervision, Methodology, Investigation, Conceptualization. **Valdo Ricca:** Writing – review & editing, Methodology, Investigation, Conceptualization. **Iacopo Cricelli:** Writing – review & editing, Software, Data curation, Conceptualization. **Ettore Marconi:** Writing – review & editing, Validation, Software, Data curation. **Claudio Cricelli:** Writing – review & editing, Supervision, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

F.L. and E.M. provided consultancies in protocol preparation for epidemiological studies and data analyses for Angelini, GSK, Lundbeck and Takeda; C.C. provided clinical consultancies for Angelini, GSK, Lundbeck and Takeda; I.C. is an employee of Genomedics Srl; and G.C. and V.R. have no conflict of interest to disclose.

Acknowledgments

We would like to thank all the GPs that formed the HSN. We also thank Dr. Serena Pecchioli MSc and Dr. Elisa Bianchini MSc, PhD, who revised the statistical analysis plan and supervised its execution.

References

Alonso, J., Angermeyer, M.C., Bernert, S., Bruffaerts, R., Brugha, T.S., Bryson, H., de Girolamo, G., Graaf, R., Demyttenaere, K., Gasquet, I., Haro, J.M., Katz, S.J., Kessler, R.C., Kovess, V., Lépine, J.P., Ormel, J., Polidori, G., Russo, L.J., Vilagut, G., Almansa, J., Arbabzadeh-Bouchez, S., Autonell, J., Bernal, M., Buist-Bouwman, M. A., Codony, M., Domingo-Salvany, A., Ferrer, M., Joo, S.S., Martínez-Alonso, M., Matschinger, H., Mazzi, F., Morgan, Z., Morosini, P., Palacín, C., Romera, B., Taub, N., Vollebergh, W.A.M., 2004. Prevalence of mental disorders in Europe: results … [Acta Psychiatr Scand Suppl. 2004] - PubMed result. Acta Psychiatr. Scand. 21–27. https://doi.org/10.1111/j.1600-0047.2004.00327.x. Suppl.

Aragonès, E., Piñol, J.L., Labad, A., Masdéu, R.M., Pino, M., Cervera, J., 2004. Prevalence and determinants of depressive disorders in primary care practice in Spain. Int. J. Psychiatry Med. 34, 21–35. https://doi.org/10.2190/C25N-W4NY-BN8W-TXN2.

Australian Bureau of Statistics 4329.0, 2011. Characteristics of people using mental health services and prescription medication [WWW Document]. Aust. Inst. Heal. Welfare 2015. Aust. Inst. Heal. Welf. Ment. Heal. Serv.

Barkow, K., Maier, W., Ustün, T.B., Gänsicke, M., Wittchen, H.-U., Heun, R., 2003. Risk factors for depression at 12-month follow-up in adult primary health care patients with major depression: an international prospective study. J. Affect. Disord. 76, 157–169.

Barry, M.J., Nicholson, W.K., Silverstein, M., Chelmow, D., Coker, T.R., Davidson, K.W., Davis, E.M., Donahue, K.E., Jaén, C.R., Li, L., Ogedegbe, G., Pbert, L., Rao, G., Ruiz, J.M., Stevermer, J.J., Tsevat, J., Underwood, S.M., Wong, J.B., 2023. Screening for depression and suicide risk in adults: US preventive services task force recommendation statement. JAMA 329, 2057–2067. https://doi.org/10.1001/jama.2023.9297.

Bellón, J.Á., de Dios Luna, J., King, M., Moreno-Küstner, B., Nazareth, I., Montón-Franco, C., GildeGómez-Barragán, M.J., Sánchez-Celaya, M., Díaz-Barreiros, M.Á., Vicens, C., Cervilla, J.A., Svab, I., Maaroos, H.-I., Xavier, M., Geerlings, M.I., Saldivia, S., Gutiérrez, B., Motrico, E., Martínez-Cañavate, M.T., Oliván-Blázquez, B., Sánchez-Artiaga, M.S., March, S., del Mar Muñoz-García, M., Vázquez-Medrano, A., Moreno-Peral, P., Torres-González, F., 2011. Predicting the onset of major depression in primary care: international validation of a risk prediction algorithm from Spain. Psychol. Med. 41, 2075–2088. https://doi.org/10.1017/S0033291711000468.

Bellón, J.Á., Conejo-Cerón, S., Moreno-Peral, P., King, M., Nazareth, I., Martín-Pérez, C., Fernández-Alonso, C., Ballesta-Rodríguez, M.I., Fernández, A., Aiarzaguena, J.M., Montón-Franco, C., Ibanez-Casas, I., Rodríguez-Sánchez, E., Rodríguez-Bayón, A., Serrano-Blanco, A., Gómez, M.C., LaFuente, P., Del Mar Muñoz-García, M., Mínguez-Gonzalo, P., Araujo, L., Palao, D., Espinosa-Cifuentes, M., Zubiaga, F., Navas-Campaña, D., Mendive, J., Aranda-Regules, J.M., Rodriguez-Morejón, A., Salvador-Carulla, L., de Dios Luna, J., 2013. Preventing the onset of major depression based on the level and profile of risk of primary care attendees: protocol of a cluster randomised trial (the predictD-CCRT study). BMC Psychiatry 13, 171. https://doi.org/10.1186/1471-244X-13-171.

Bossuyt, P.M., Reitsma, J.B., Bruns, D.E., Gatsonis, C.A., Glasziou, P.P., Irwig, L.M., Lijmer, J.G., Moher, D., Rennie, D., De Vet, H.C.W., 2003. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. BMJ 326, 41–44. https://doi.org/10.1136/BMJ.326.7379.41.

Castellini, G., Pecchioli, S., Cricelli, I., Mazzoleni, F., Cricelli, C., Ricca, V., Hudziak, J.J., Brignoli, O., Lapi, F., 2016a. How to early recognize mood disorders in primary care: a nationwide, population-based, cohort study. Eur. Psychiatry 37. https://doi.org/10.1016/j.eurpsy.2016.04.003.

Castellini, Giovanni, Wu, F.C.W., Finn, J.D., O'Neill, T.W., Lean, M.E.J., Pendleton, N., Rastrelli, G., Rutter, M.K., Gacci, M., Ricca, V., Maggi, M., EMAS Study Group, 2016b. Interactions between depression and lower urinary tract symptoms: the role of adverse life events and inflammatory mechanisms. Results from the European male ageing study. Psychosom. Med. 78, 758–769. https://doi.org/10.1097/PSY.0000000000000328.

Chondros, P., Davidson, S., Wolfe, R., Gilchrist, G., Dowrick, C., Griffiths, F., Hegarty, K., Herrman, H., Gunn, J., 2018. Development of a prognostic model for predicting depression severity in adult primary patients with depressive symptoms using the diamond longitudinal study. J. Affect. Disord. 227, 854–860. https://doi.org/10.1016/j.jad.2017.11.042.

Collins, G.S., Reitsma, J.B., Altman, D.G., Moons, K.G.M., 2015. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. J. Clin. Epidemiol. 68, 134–143. https://doi.org/10.1016/j.jclinepi.2014.11.010.

Cricelli, C., Mazzaglia, G., Samani, F., Marchi, M., Sabatini, A., Nardi, R., Ventriglia, G., Caputi, A.P., 2003. Prevalence estimates for chronic diseases in Italy: exploring the differences between self-report and primary care databases. J. Public Health Med. 25, 254–257.

Davidson, S.K., Harris, M.G., Dowrick, C.F., Wachtler, C.A., Pirkis, J., Gunn, J.M., 2015. Mental health interventions and future major depression among primary care patients with subthreshold depression. J. Affect. Disord. 177, 65–73. https://doi.org/10.1016/j.jad.2015.02.014.

Dowrick, C., Flach, C., Leese, M., Chatwin, J., Morriss, R., Peveler, R., Gabbay, M., Byng, R., Moore, M., Tylee, A., Kendrick, T., THREAD Study Group, 2011. Estimating probability of sustained recovery from mild to moderate depression in primary care: evidence from the THREAD study. Psychol. Med. 41, 141–150. https://doi.org/10.1017/S0033291710000437.

Fernández, A., Mendive, J.M., Conejo-Cerón, S., Moreno-Peral, P., King, M., Nazareth, I., Martín-Pérez, C., Fernández-Alonso, C., Rodríguez-Bayón, A., Aiarzaguena, J.M., Montón-Franco, C., Serrano-Blanco, A., Ibañez-Casas, I., Rodríguez-Sánchez, E., Salvador-Carulla, L., Garay, P.B., Ballesta-Rodríguez, M.I., LaFuente, P., Del Mar Muñoz-García, M., Mínguez-Gonzalo, P., Araujo, L., Palao, D., Gómez, M.C., Zubiaga, F., Navas-Campaña, D., Aranda-Regules, J.M., Rodriguez-Morejón, A., de Dios Luna, J., Bellón, J.Á., 2018. A personalized intervention to prevent depression in primary care: cost-effectiveness study nested into a clustered randomized trial. BMC Med. 16, 28. https://doi.org/10.1186/s12916-018-1005-y.

Guglielmi, V., Bellia, A., Bianchini, E., Medea, G., Cricelli, I., Sbraccia, P., Lauro, D., Cricelli, C., Lapi, F., 2017. Drug interactions in users of tablet vs. oral liquid levothyroxine formulations: a real-world evidence study in primary care. Endocrine. https://doi.org/10.1007/s12020-017-1412-4.

Haller, H., Cramer, H., Lauche, R., Dobos, G., 2015. Somatoform Disorders and Medically Unexplained Symptoms in Primary Care. Dtsch. Aerzteblatt Online. https://doi.org/10.3238/arztebl.2015.0279.

Herzog, D.P., Wagner, S., Engelmann, J., Treccani, G., Dreimüller, N., Müller, M.B., Tadic, A., Murck, H., Lieb, K., 2021. Early onset of depression and treatment outcome in patients with major depressive disorder. J. Psychiatr. Res. 139, 150–158. https://doi.org/10.1016/j.jpsychires.2021.05.048.

Hiller, W., Fichter, M.M., 2004. High utilizers of medical care: a crucial subgroup among somatizing patients. J. Psychosom. Res. 56, 437–443. https://doi.org/10.1016/S0022-3999(03)00628-7.

de Hond, A.A.H., Steyerberg, E.W., van Calster, B., 2022. Interpreting area under the receiver operating characteristic curve. Lancet Digit. Heal. 4, e853–e855. https://doi.org/10.1016/s2589-7500(22)00188-1.

Hüsing, P., Löwe, B., Piontek, K., Shedden-Mora, M., 2018. Somatoform disorder in primary care: the influence of co-morbidity with anxiety and depression on health care utilization. J. Eval. Clin. Pract. 24, 892–900. https://doi.org/10.1111/jep.12898.

Jan, D., de Vega, M., López-Pigüi, J., Padrón, I., 2022. Applying deep learning on a few EEG electrodes during resting state reveals depressive states. A data driven study. Brain Sci. 12 https://doi.org/10.3390/brainsci12111506.

Joynt, K.E., Whellan, D.J., O'Connor, C.M., 2003. Depression and cardiovascular disease: mechanisms of interaction. Biol. Psychiatry. https://doi.org/10.1016/S0006-3223(03)00568-7.

Karsten, J., Hartman, C.A., Smit, J.H., Zitman, F.G., Beekman, A.T.F., Cuijpers, P., Van Der Does, A.J.W., Ormel, J., Nolen, W.A., Penninx, B.W.J.H., 2011. Psychiatric history and subthreshold symptoms as predictors of the occurrence of depressive or anxiety disorder within 2 years. Br. J. Psychiatry 198, 206–212. https://doi.org/10.1192/bjp.bp.110.080572.

King, M., Nazareth, I., Levy, G., Walker, C., Morris, R., Weich, S., Bellón-Saameño, J.Á., Moreno, B., Švab, I., Rotar, D., Rifel, J., Maaroos, H.I., Aluoja, A., Kalda, R., Neeleman, J., Geerlings, M.I., Xavier, M., De Almeida, M.C., Correa, B., Torres-Gonzalez, F., 2008. Prevalence of common mental disorders in general practice attendees across Europe. Br. J. Psychiatry 192, 362–367. https://doi.org/10.1192/bjp.bp.107.039966.

King, M., Bottomley, C., Bellón-Saameño, J.A., Torres-Gonzalez, F., Švab, I., Rifel, J., Maaroos, H.-I., Aluoja, A., Geerlings, M.I., Xavier, M., Carraça, I., Vicente, B., Saldivia, S., Nazareth, I., 2011. An international risk prediction algorithm for the onset of generalized anxiety and panic syndromes in general practice attendees:

predictA. Psychol. Med. 41, 1625–1639. https://doi.org/10.1017/S0033291710002400.
King, M., Bottomley, C., Bellón-Saameño, J., Torres-Gonzalez, F., Svab, I., Rotar, D., Xavier, M., Nazareth, I., 2013. Predicting onset of major depression in general practice attendees in Europe: extending the application of the predictD risk algorithm from 12 to 24 months. Psychol. Med. 43, 1929–1939. https://doi.org/10.1017/S0033291712002693.
Kraus, C., Kadriu, B., Lanzenberger, R., Zarate, C.A., Kasper, S., 2020. Prognosis and improved outcomes in major depression: a review. Focus (Am. Psychiatr. Publ). 18, 220–235. https://doi.org/10.1176/appi.focus.18205.
Lapi, F., Simonetti, M., Michieli, R., Pasqua, A., Brandi, M.L., Frediani, B., Cricelli, C., Mazzaglia, G., 2012. Assessing 5-year incidence rates and determinants of osteoporotic fractures in primary care. Bone 50. https://doi.org/10.1016/j.bone.2011.09.048.
Lech, S., Herrmann, W., Trautmann, S., Schwantes, U., Gellert, P., Behr, J., Buspavanich, P., 2022. Depression in primary care and the role of evidence-based guidelines: cross-sectional data from primary care physicians in Germany. BMC Health Serv. Res. 22, 1–10. https://doi.org/10.1186/S12913-022-08631-W/TABLES/4.
Levis, B., Sun, Y., He, C., Wu, Y., Krishnan, A., Bhandari, P.M., Neupane, D., Imran, M., Brehaut, E., Negeri, Z., Fischer, F.H., Benedetti, A., Thombs, B.D., 2020. Accuracy of the PHQ-2 alone and in combination with the PHQ-9 for screening to detect major depression: systematic review and meta-analysis. JAMA - J. Am. Med. Assoc. 323, 2290–2300. https://doi.org/10.1001/jama.2020.6504.
Lim, G.Y., Tam, W.W., Lu, Y., Ho, C.S., Zhang, M.W., Ho, R.C., 2018. Prevalence of depression in the community from 30 countries between 1994 and 2014. Sci. Rep. 8, 2861. https://doi.org/10.1038/s41598-018-21243-x.
Liu, W., Jia, K., Wang, Z., Ma, Z., 2022. A depression prediction algorithm based on spatiotemporal feature of EEG signal. Brain Sci. 12 https://doi.org/10.3390/brainsci12050630.
Löwe, B., Spitzer, R.L., Williams, J.B.W., Mussell, M., Schellberg, D., Kroenke, K., 2008. Depression, anxiety and somatization in primary care: syndrome overlap and functional impairment. Gen. Hosp. Psychiatry 30, 191–199. https://doi.org/10.1016/j.genhosppsych.2008.01.001.
Ludman, E., Von Korff, M., Katon, W., Lin, E., Simon, G., Walker, E., Unützer, J., Bush, T., Wahab, S., 2000. The design, implementation, and acceptance of a primary care-based intervention to prevent depression relapse. Int. J. Psychiatry Med. 30, 229–245. https://doi.org/10.2190/44LK-28E9-RRJ5-KQVW.
McIntyre, R.S., Soczynska, J.K., Konarski, J.Z., Woldeyohannes, H.O., Law, C.W.Y., Miranda, A., Fulgosi, D., Kennedy, S.H., 2007. Should depressive syndromes be reclassified as "metabolic syndrome type II"? Ann. Clin. Psychiatry 19, 257–264. https://doi.org/10.1080/10401230701653377.
Medina, J.C., Schmelefske, E., Hébert, C., Drapeau, M., 2020. European clinical practice guidelines for depression in adults: are they good enough? J. Affect. Disord. 263, 382–385. https://doi.org/10.1016/j.jad.2019.12.005.
Moons, K.G.M., Altman, D.G., Reitsma, J.B., Ioannidis, J.P.A., Macaskill, P., Steyenberg, E.W., Vickers, A.J., Ransohoff, D., Collins, G.S., 2015. Transparent reporting of a multivariable prediction model for individual prognosis or disagnosis (TRIPOD): explanantion and elaboration. Ann. Intern. Med. 162, W1–73. https://doi.org/10.7326/M14-0698.
Moussavi, S., Chatterji, S., Verdes, E., Tandon, A., Patel, V., Ustun, B., 2007. Depression, chronic diseases, and decrements in health. Lancet 370, 851–858. https://doi.org/10.1016/S0140-6736(07)61415-9.
Ormel, J., VonKorff, M., Ustun, T.B., Pini, S., Korten, A., Oldehinkel, T., 1994. Common mental disorders and disability across cultures. Results from the WHO collaborative study on psychological problems in general health care. JAMA 272, 1741–1748. https://doi.org/10.1001/jama.272.22.1741.
Pariante, C.M., 2021. Increased inflammation in depression: a little in all, or a lot in a few? Am. J. Psychiatry 178, 1077–1079. https://doi.org/10.1176/APPI.AJP.2021.21101043.
Ramspek, C.L., Jager, K.J., Dekker, F.W., Zoccali, C., van Diepen, M., 2021. External validation of prognostic models: what, why, how, when and where? Clin. Kidney J. 14, 49–58. https://doi.org/10.1093/ckj/sfaa188.
Rost, N., Binder, E.B., Brückl, T.M., 2023. Predicting treatment outcome in depression: an introduction into current concepts and challenges. Eur. Arch. Psychiatry Clin. Neurosci. 273, 113–127. https://doi.org/10.1007/s00406-022-01418-4.
Rubenstein, L.V., Rayburn, N.R., Keeler, E.B., Ford, D.E., Rost, K.M., Sherbourne, C.D., 2007. Predicting outcomes of primary care patients with major depression: development of a depression prognosis index. Psychiatr. Serv. 58, 1049–1056. https://doi.org/10.1176/appi.ps.58.8.1049.
Shahabi, M.S., Shalbaf, A., Rostami, R., Kazemi, R., 2023. A convolutional recurrent neural network with attention for response prediction to repetitive transcranial magnetic stimulation in major depressive disorder. Sci. Rep. 13 https://doi.org/10.1038/s41598-023-35545-2.
Smits, F.T.M., Brouwer, H.J., ter Riet, G., van Weert, H.C.P., 2009. Epidemiology of frequent attenders: a 3-year historic cohort study comparing attendance, morbidity and prescriptions of one-year and persistent frequent attenders. BMC Public Health 9, 36. https://doi.org/10.1186/1471-2458-9-36.
Sobocki, P., Jönsson, B., Angst, J., Rehnberg, C., 2006. Cost of depression in Europe. J. Ment. Health Policy Econ. 9, 87–98.
Steyerberg, E.W., Vickers, A.J., Cook, N.R., Gerds, T., Gonen, M., Obuchowski, N., Pencina, M.J., Kattan, M.W., 2010a. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology 21, 128–138. https://doi.org/10.1097/EDE.0b013e3181c30fb2.
Steyerberg, E.W., Vickers, A.J., Cook, N.R., Gerds, T., Obuchowski, N., Pencina, M.J., Kattan, M.W., 2010b. Assessing the performance of prediction models: a framework for some traditional and novel measures. Epidemiology 21, 128–138. https://doi.org/10.1097/EDE.0b013e3181c30fb2.Assessing.
Toll, D.B., Janssen, K.J.M., Vergouwe, Y., Moons, K.G.M., 2008. Validation, updating and impact of clinical prediction rules: a review. J. Clin. Epidemiol. https://doi.org/10.1016/j.jclinepi.2008.04.008.
Trautmann, S., Rehm, J., Wittchen, H., 2016. The economic costs of mental disorders: do our societies react appropriately to the burden of mental disorders? EMBO Rep. 17, 1245–1249. https://doi.org/10.15252/EMBR.201642951.
Van Calster, B., Nieboer, D., Vergouwe, Y., De Cock, B., Pencina, M.J., Steyerberg, E.W., 2016. A calibration hierarchy for risk models was defined: from utopia to empirical data. J. Clin. Epidemiol. 74, 167–176. https://doi.org/10.1016/j.jclinepi.2015.12.005.
Van Hoorde, K., Van Huffel, S., Timmerman, D., Bourne, T., Van Calster, B., 2015. A spline-based tool to assess and visualize the calibration of multiclass risk predictions. J. Biomed. Inform. 54, 283–293. https://doi.org/10.1016/j.jbi.2014.12.016.
Vilagut, G., Forero, C.G., Pinto-Meza, A., Haro, J.M., De Graaf, R., Bruffaerts, R., Kovess, V., De Girolamo, G., Matschinger, H., Ferrer, M., Alonso, J., 2013. The mental component of the short-form 12 health survey (SF-12) as a measure of depressive disorders in the general population: results with three alternative scoring methods. Value Health 16, 564–573. https://doi.org/10.1016/j.jval.2013.01.006.
Wang, P.S., Aguilar-Gaxiola, S., Alonso, J., Angermeyer, M.C., Borges, G., Bromet, E.J., Bruffaerts, R., de Girolamo, G., de Graaf, R., Gureje, O., Haro, J.M., Karam, E.G., Kessler, R.C., Kovess, V., Lane, M.C., Lee, S., Levinson, D., Ono, Y., Petukhova, M., Posada-Villa, J., Seedat, S., Wells, J.E., 2007. Use of mental health services for anxiety, mood, and substance disorders in 17 countries in the WHO world mental health surveys. Lancet 370, 841–850. https://doi.org/10.1016/S0140-6736(07)61414-7.
Wang, Q., Koval, J.J., Mills, C.A., Lee, K.I.D., 2008. Determination of the Selection Statistics and Best Significance Level in Backward Stepwise Logistic Regression, 37, pp. 62–72. https://doi.org/10.1080/03610910701723625.
Wiborg, J.F., Gieseler, D., Fabisch, A.B., Voigt, K., Lautenbach, A., Löwe, B., 2013. Suicidality in primary care patients with somatoform disorders. Psychosom. Med. 75, 800–806. https://doi.org/10.1097/PSY.0000000000000013.
Wittchen, H.U., Mühlig, S., Pezawas, L., 2003. Natural course and burden of bipolar disorders. Int. J. Neuropsychopharmacol. https://doi.org/10.1017/S146114570300333X.
Wolff, R., Moons, K., Riley, R., Whiting, P., Westwood, M., Collins, G., Reitsma, J., Kleijnen, J., Mallett, S., 2017. PROBAST – A Risk-of-bias Tool for Prediction-Modelling Studies. Glob. Evid, Summit.
Wynants, L., Van Calster, B., Collins, G.S., Riley, R.D., Heinze, G., Schuit, E., Bonten, M.M.J., Damen, J.A.A., Debray, T.P.A., De Vos, M., Dhiman, P., Haller, M.C., Harhay, M.O., Henckaerts, L., Kreuzberger, N., Lohmann, A., Luijken, K., Ma, J., Andaur Navarro, C.L., Reitsma, J.B., Sergeant, J.C., Shi, C., Skoetz, N., Smits, L.J.M., Snell, K.I.E., Sperrin, M., Spijker, R., Steyerberg, E.W., Takada, T., Van Kuijk, S.M.J., Van Royen, F.S., Wallisch, C., Hooft, L., Moons, K.G.M., Van Smeden, M., 2020. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. BMJ 369. https://doi.org/10.1136/bmj.m1328.