

Original Article

Cite this article: Alayo I *et al.* (2025) Identifying most important predictors for suicidal thoughts and behaviours among healthcare workers active during the Spain COVID-19 pandemic: a machine-learning approach. *Epidemiology and Psychiatric Sciences* **34**, e28, 1–13. <https://doi.org/10.1017/S2045796025000198>

Received: 12 October 2024

Revised: 17 March 2025

Accepted: 23 March 2025






Keywords:

attempted suicide; interpretability; machine learning; mental health; suicidal ideation

Corresponding author: Philippe Mortier;

Email: pmortier@researchmar.net

Identifying most important predictors for suicidal thoughts and behaviours among healthcare workers active during the Spain COVID-19 pandemic: a machine-learning approach

Itxaso Alayo^{1,2,3,4} , Oriol Pujol⁵, Jordi Alonso^{1,4,6} , Montse Ferrer^{1,4,6}, Franco Amigo^{1,6}, Ana Portillo-Van Diest^{1,4,6} , Enric Aragonès^{7,8}, Andrés Aragon Peña^{9,10}, Ángel Asúnsolo Del Barco^{11,12,13}, Mireia Campos¹⁴, Meritxell Espuga¹⁵, Ana González-Pinto^{16,17}, Josep Maria Haro^{17,18}, Nieves López-Fresneña¹⁹, Alma D. Martínez de Salázar²⁰, Juan D. Molina^{17,21,22,23}, Rafael M. Ortí-Lucas²⁴, Mara Parellada^{17,19}, José Maria Pelayo-Terán^{17,25,26}, Maria João Forjaz^{3,27}, Aurora Pérez-Zapata²⁸, José Ignacio Pijoan^{3,29}, Nieves Plana^{3,30}, Elena Polentinos-Castro^{14,31,32}, Maria Teresa Puig^{33,34,35,36}, Cristina Rius^{6,37}, Ferran Sanz^{4,38,39}, Cònsol Serra^{6,40,41}, Iratxe Urreta-Barallobre^{6,42,43}, Ronny Bruffaerts⁴⁴, Eduard Vieta^{17,45} , Víctor Pérez-Solá^{17,33,46}, Philippe Mortier^{1,6} , Gemma Vilagut^{1,6} and on behalf of the *MINDCOVID* Working group

¹Hospital del Mar Research Institute, Barcelona, Spain; ²Biosistemak Institute for Health Systems Research, Bilbao, Bizkaia, Spain; ³Red de Investigación en Cronicidad, Atención Primaria y Promoción de la Salud RICAPPS-(RICORS), Instituto de Salud Carlos III (ISCIII), Madrid, Spain; ⁴Department of Medicine and Life Sciences (MELIS), Universitat Pompeu Fabra, Barcelona, Spain; ⁵Departament de Matemàtiques i Informàtica, Universitat de Barcelona, Barcelona, Spain; ⁶CIBER de Epidemiologia i Salut Pública (CIBERESP), Instituto de Salud Carlos III, Madrid, Spain; ⁷Institut d'Investigació en Atenció Primària IDIAP Jordi Gol, Barcelona, Spain; ⁸Atenció Primària Camp de Tarragona, Institut Català de la Salut, Tarragona, Spain; ⁹Epidemiology Unit, Regional Ministry of Health, Community of Madrid, Madrid, Spain; ¹⁰Fundación Investigación e Innovación Biosanitaria de AP, Comunidad de Madrid, Madrid, Spain; ¹¹Department of Surgery, Medical and Social Sciences, Faculty of Medicine and Health Sciences, University of Alcalá, Alcalá de Henares, Spain; ¹²Ramón y Cajal Institute of Sanitary Research (IRYCIS), Madrid, Spain; ¹³Department of Epidemiology and Biostatistics, Graduate School of Public Health and Health Policy, The City University of New York, New York, NY, USA; ¹⁴Service of Prevention of Labor Risks, Medical Emergencies System, Generalitat de Catalunya, Barcelona, Spain; ¹⁵Occupational Health Service, Hospital Universitari Vall d'Hebron, Barcelona, Spain; ¹⁶BIOARABA, Hospital Universitario Araba-Santiago, UPV/EHU, Vitoria-Gasteiz, Spain; ¹⁷CIBER Salud Mental (CIBERSAM), Madrid, Spain; ¹⁸Parc Sanitari Sant Joan de Déu, Institut de Recerca Sant Joan de Deu (IRSJD), Sant Boi de Llobregat, Barcelona, Spain; ¹⁹Hospital General Universitario Gregorio Marañón, Madrid, Spain; ²⁰UGC Salud Mental, Hospital Universitario Torrecárdenas, Almería, Spain; ²¹Villaverde Mental Health Center, Clinical Management Area of Psychiatry and Mental Health, Psychiatric Service, Hospital Universitario 12 de Octubre, Madrid, Spain; ²²Research Institute Hospital 12 de Octubre (i+12), Madrid, Spain; ²³Facultad de Medicina, Universidad Francisco de Vitoria, Madrid, Spain; ²⁴Servicio de Medicina Preventiva y Calidad Asistencial, Hospital Clínico Universitario de Valencia, Valencia, Spain; ²⁵Servicio de Psiquiatría y Salud Mental, Hospital el Bierzo, Gerencia de Asistencia Sanitaria del Bierzo (GASBI). Gerencia Regional de Salud de Castilla y León (SACYL), Ponferrada, León, Spain; ²⁶Area de Medicina Preventiva y Salud Pública, Departamento de Ciencias Biomédicas, Universidad de León, León, Spain; ²⁷National Center of Epidemiology, Instituto de Salud Carlos III (ISCIII), Madrid, Spain; ²⁸Hospital Universitario Príncipe de Asturias, Servicio de Prevención de Riesgos Laborales, Spain; ²⁹Clinical Epidemiology Unit-Hospital Universitario Cruces/ OSI EEC, Biobizkaia Health Research Institute, Barakaldo, Spain; ³⁰Ramón y Cajal University Hospital, IRYCIS, Department of Surgery, Medical and Social Sciences, Faculty of Medicine and Health Sciences, University of Alcalá, Alcalá de Henares, MAD, Spain; ³¹Research Unit, Primary Care Management, Madrid Health Service, Madrid, Spain; ³²Department of Medical Specialities and Public Health, King Juan Carlos University, Madrid, Spain; ³³Universitat Autònoma de Barcelona (UAB), Barcelona, Spain; ³⁴Department of Epidemiology and Public Health, Hospital de la Santa Creu i Sant Pau, Barcelona, Spain; ³⁵Biomedical Research Institute Sant Pau (IIB Sant Pau), Barcelona, Spain; ³⁶CIBER Enfermedades Cardiovasculares (CIBERCV), Madrid, Spain; ³⁷Agència de Salut Pública de Barcelona, Barcelona, Spain; ³⁸Research Programme on Biomedical Informatics (GRIB), Hospital del Mar Research Institute, Barcelona, Spain; ³⁹Instituto Nacional de Bioinformática – ELIXIR-ES, Barcelona, Spain; ⁴⁰CISAL-Centro de Investigación en Salud Laboral, Hospital del Mar Research Institute/University Pompeu Fabra, Barcelona, Spain; ⁴¹Occupational Health

© The Author(s), 2025. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

Service, Hospital del Mar, Barcelona, Spain; ⁴²Osakidetza Basque Health Service, Donostialdea Integrated Health Organisation, Donostia University Hospital, Clinical Epidemiology Unit, San Sebastián, Spain; ⁴³Biodonostia Health Research Institute, Clinical Epidemiology, San Sebastián, Spain; ⁴⁴Center for Public Health Psychiatry, Universitair Psychiatrisch Centrum, KU Leuven, Leuven, Belgium; ⁴⁵Institute of Neuroscience, Hospital Clinic, University of Barcelona, IDIBAPS, Barcelona, Spain and ⁴⁶Institute of Neuropsychiatry and Addiction (INAD), Parc de Salut Mar, Barcelona, Spain

The MINDCOVID Working Group is formed by Jordi Alonso, Itxaso Alayo, Manuel Alonso, Mar Álvarez, Benedikt Amann, Franco F. Amigo, Gerard Anmella, Andres Aragón, Nuria Aragonés, Enric Aragonès, Ana Isabel Arizón, Angel Asunsolo, Alfons Ayora, Laura Ballester, Puri Barbas, Josep Basora, R. Bausà, Elena Bereciartua, Inés Bravo, Alberto Cotillas, Andres Cuartero, Concha de Paz, Isabel del Cura, Maria Jesus del Yerro, Domingo Diaz, Jose Luis Domingo, Jose I. Emparanza, Mireia Espallargues, Meritxell Espuga, Patricia Estevan, M. Isabel Fernandez, Tania Fernandez, Montse Ferrer, Yolanda Ferreres, Giovanna Fico, M. Joao Forjaz, Rosa Garcia Barranco, J. Manuel Garcia Torrecillas, C. Garcia-Ribera, Araceli Garrido, Elisa Gil, Marta Gomez, Javier Gomez, Ana Gonzalez Pinto, Josep Maria Haro, Margarita Hernando, Maria Giola Insigna, Milagros Iriberrri, Nuria Jimenez, Xavi Jimenez, Amparo Larrauri, Fernando Leon, Nieves Lopez-Fresneña, Carmen Lopez, Mayte Lopez-Atanes, Juan Antonio Lopez-Rodriguez, German Lopez-Cortacans, Alba Marcos, Jesus Martin, Vicente Martin, Mercedes Martinez-Cortés, Raquel Martinez-Martinez, Alma D. Martinez de Salazar, Isabel Martinez, Marco Marzola, Nelva Mata, Josep Maria Molina, Juan de Dios Molina, Emilia Molinero, Philippe Mortier, Carmen Muñoz, Andrea Murru, L. Navarro, Jorge Olmedo, Rafael M. Ortí, Rafael Padrós, Meritxell Pallejà, Raul Parra, Julio Pascual, Jose Maria Pelayo, Rosa Pla, Nieves Plana, Coro Perez Aznar, Beatriz Perez Gomez, Aurora Perez Zapata, Jose Ignacio Pijoan, Elena Polentinos, Beatriz Puertolas, Maria Teresa Puig, Alex Quílez, M. Jesus Quintana, Antonio Quiroga, David Rentero, Cristina Rey, Cristina Rius, Carmen Rodriguez-Blazquez, M. Jose Rojas, Yamina Romero, Gabriel Rubio, Mercedes Rumayor, Pedro Ruiz, Margarita Saenz, Jesus Sanchez, Ignacio Sanchez-Arcilla, Ferran Sanz, Consol Serra, Victoria Serra-Sutton, Manuela Serrano, Silvia Sola, Sara Solera, Miguel Soto, Alejandra Tarrago, Natividad Tolosa, Mireia Vazquez, Margarita Viciola, Eduard Vieta, Gemma Vilagut, Sara Yago, Jesus Yañez, Yolanda Zapico, Luis Maria Zorita, Iñaki Zorrilla, Saioa L. Zurbano and Víctor Perez-Solá.

Abstract

Aims. Studies conducted during the COVID-19 pandemic found high occurrence of suicidal thoughts and behaviours (STBs) among healthcare workers (HCWs). The current study aimed to (1) develop a machine learning-based prediction model for future STBs using data from a large prospective cohort of Spanish HCWs and (2) identify the most important variables in terms of contribution to the model's predictive accuracy.

Methods. This is a prospective, multicentre cohort study of Spanish HCWs active during the COVID-19 pandemic. A total of 8,996 HCWs participated in the web-based baseline survey (May–July 2020) and 4,809 in the 4-month follow-up survey. A total of 219 predictor variables were derived from the baseline survey. The outcome variable was any STB at the 4-month follow-up. Variable selection was done using an L1 regularized linear Support Vector Classifier (SVC). A random forest model with 5-fold cross-validation was developed, in which the Synthetic Minority Oversampling Technique (SMOTE) and undersampling of the majority class balancing techniques were tested. The model was evaluated by the area under the Receiver Operating Characteristic (AUROC) curve and the area under the precision–recall curve. Shapley's additive explanatory values (SHAP values) were used to evaluate the overall contribution of each variable to the prediction of future STBs. Results were obtained separately by gender.

Results. The prevalence of STBs in HCWs at the 4-month follow-up was 7.9% (women = 7.8%, men = 8.2%). Thirty-four variables were selected by the L1 regularized linear SVC. The best results were obtained without data balancing techniques: AUROC = 0.87 (0.86 for women and 0.87 for men) and area under the precision–recall curve = 0.50 (0.55 for women and 0.45 for men). Based on SHAP values, the most important baseline predictors for any STB at the 4-month follow-up were the presence of passive suicidal ideation, the number of days in the past 30 days with passive or active suicidal ideation, the number of days in the past 30 days with binge eating episodes, the number of panic attacks (women only) and the frequency of intrusive thoughts (men only).

Conclusions. Machine learning-based prediction models for STBs in HCWs during the COVID-19 pandemic trained on web-based survey data present high discrimination and classification capacity. Future clinical implementations of this model could enable the early detection of HCWs at the highest risk for developing adverse mental health outcomes.

Study registration. NCT04556565

Introduction

Suicide is a major public health issue and one of the leading causes of preventable death worldwide (World Health Organization, 2024). Pre-pandemic studies showed consistently that both

physicians (Dutheil *et al.*, 2019) and nurses (Davis *et al.*, 2021) are at high risk for suicide compared to other employed people (Milner *et al.*, 2013), in part related to high access to lethal means and low willingness to seek help (Harvey *et al.*, 2021). An important risk factor for suicide is suicidal thought and behaviour (STB) (Ribeiro *et al.*, 2016). Studies carried out during the pandemic found high levels of STBs among healthcare workers (HCWs) compared to the pre-pandemic period (Greenberg *et al.*, 2020; Mediavilla *et al.*, 2023; Mortier *et al.*, 2021b; Murata *et al.*, 2021; Sahimi *et al.*, 2021; Xiaoming *et al.*, 2020; Xu *et al.*, 2021; Zhou *et al.*, 2020).

Risk factors identified in these studies span various risk domains and include pre-pandemic lifetime factors, current mental disorders and emotional problems (e.g., burn-out, traumatic stress, anxiety and depression), loneliness and social isolation, financial stress, and pandemic-specific factors, such as having been in quarantine, moral injury, interpersonal and health-related stress (Eyles *et al.*, 2021; García-Iglesias *et al.*, 2022; Mortier *et al.*, 2022, 2021a).

Identifying individuals at the highest risk for future STBs is a significant challenge in the field of mental health research, especially given the relatively low occurrence of STBs. Over the past five decades, traditional statistical approaches have been predominant in predicting STB risk (Nordin *et al.*, 2023), which often, due to their limited capacity as to including a wide range of predictor variables, require the researcher to define a priori a limited set of predictors to be included in the models. This approach has been criticized because variable selection is often based on predefined theoretical frameworks that only consider some of the potentially relevant predictors for STBs (Franklin *et al.*, 2017), resulting in relatively simple models with limited predictive accuracy (Boudreaux *et al.*, 2021).

Advanced analytical approaches, such as machine learning (ML) models, have demonstrated higher predictive accuracy of STBs than traditional statistical approaches (e.g. linear regression, generalized linear models or analysis of variance) including a limited number of variables based on predefined theoretical frameworks (Schafer *et al.*, 2021). ML models handle complex interactions and high-dimensional data effectively by capturing non-linear relationships and efficiently processing and analysing large volumes of data with multiple variables, overcoming the limitations of traditional approaches (Bennett *et al.*, 2022). They also allow a better understanding of the complex patterns and non-evident relationships among a very large set of STB-related variables including not only commonly considered factors such as mental health and family history, but also contextual aspects such as lifestyles, access to healthcare, adverse childhood experiences and social and economic environments, among others (Favril *et al.*, 2022). Thus, these advanced approaches can provide a more comprehensive and accurate understanding of the factors contributing to the risk of STBs. The key is to integrate traditional approaches with the empirical power of data-driven techniques (Schafer *et al.*, 2021). An increased focus on the prediction of adverse mental health, including identification of predictors that mostly contribute to increased prediction accuracy, may lead to new hypotheses about causal associations, and ultimately, a better understanding and prevention of these outcomes, including STBs (Yarkoni and Westfall, 2017).

Despite ML being increasingly used for STB risk prediction, to the best of our knowledge, there is no previous study using ML to develop a prediction model for STBs among HCWs. Such models could help with early identification and intervention for at-risk HCWs and also provide valuable insights into the complex

interplay of factors contributing to suicidal ideation in this population. In addition, although there are clear gender differences in both absolute STB risk and the distribution of risk factors (Gradus *et al.*, 2021; Jiang *et al.*, 2021; Miranda-Mendizabal *et al.*, 2019a, 2019b), ML-based studies often do not take these differences into consideration, leading to a lack of gender-specific STB prediction models.

The aim of the current study is to develop an ML-based prediction model for future STBs using data from the MINDCOVID project, a large prospective cohort study of Spanish HCWs (Alonso *et al.*, 2021; MINDCOVID, 2020). The HCW cohort was recruited just after the height of the first wave of the Spanish COVID-19 pandemic and was followed up 4 months later, including a reassessment of STBs. Importantly, predictor variables to develop the prediction model were created using all information included in the baseline survey. Although the information collected in the baseline survey was not exhaustive with regard to including all factors potentially related to HCWs' STB in the literature, it spanned various relevant risk factor domains for adverse mental health and STB, including depression, anxiety and post-traumatic stress disorder (PTSD). Variable selection techniques were employed to avoid manual selection of predictors. In addition, we aim to identify predictors that are the most important contributors to the model's predictive accuracy, separately for men and women.

Methods

Recruitment

Data for this study come from the MINDCOVID project (Alonso *et al.*, 2021; MINDCOVID, 2020), a multicentre, prospective, observational cohort study of Spanish HCWs, representing a convenience sample of 18 healthcare institutions (hospitals, primary care and public healthcare centres) from six Autonomous Communities in Spain and included all types of HCWs (medical doctors, nurses, auxiliary nurses, other professions involved in patient care and professions not directly involved in patient care). The cohort was assessed at two time points using web-based self-report surveys. The first assessment (T1) was conducted from 5 May through 7 September 2020, i.e., just after the height of the first wave of the Spain COVID-19 pandemic. The follow-up assessment (T2) was conducted 4 months (mean = 120.1 days [SD = 22.2]) after the T1 assessment.

Recruitment for the T1 survey was done by healthcare representatives who contacted all employed HCWs in each participating healthcare centre using administrative email distribution lists (i.e., census sampling). A total of 8,996 HCWs participated at T1, representing a mean weighted response rate across healthcare centres (weighted by achieved sample size) of 11.7% (unweighted mean response rate of 12.8%). A total of 4,809 T1 participants also participated at T2 (53.5%). For both surveys, two reminder emails were sent within 2–4 weeks after the initial invitation. For the current study, we included data from the 4,809 HCWs described previously (Alonso *et al.*, 2022; Mortier *et al.*, 2022) that participated in both T1 and T2 assessments.

Informed consent was obtained from all participants. The study complies with the principles established by national and international regulations, including the Declaration of Helsinki and the Code of Ethics. The study was approved by the Research Integrity and Good Scientific Practices Committee of IMIM-Parc de Salut Mar, Barcelona, Spain (2020/9203/I), and by all participating centres' institutional review boards.

Table 1. Sociodemographic and work characteristics of Spanish healthcare workers during the COVID-19 pandemic assessed at T1 ($N = 4,809$)

	Total ($N = 4,809$)		WOMEN ($n = 3,988$)		MEN ($n = 910$)	
	<i>n</i>	% (SE) or mean (SD)	<i>n</i>	% (SE) or Mean (SD)	<i>n</i>	% (SE) or Mean (SD)
Women	3,899	81.1 (0.8)	–	–	–	–
Age		45.8 (11.0)		45.5 (11.0)		47.1 (11.25)
Country of birth						
Spain	4,582	95.3 (0.4)	3,717	95.3 (0.5)	864	95.0 (0.6)
Other	227	4.7 (0.4)	182	4.7 (0.5)	46	5.0 (0.6)
Living with partner	3,465	72.1 (1.2)	2,764	70.9 (1.3)	701	77.0 (1.2)
Marital status						
Single	1,756	36.5 (3.1)	1,452	37.2 (3.3)	304	33.4 (2.9)
Married	2,537	52.7 (2.6)	2,007	51.5 (2.9)	530	58.2 (1.7)
Divorced or legally separated	459	9.6 (0.5)	387	9.9 (0.4)	72	7.9 (1.2)
Widowed	57	1.2 (0.2)	53	1.4 (0.2)	4	0.4 (0.3)
Children in care	1,995	41.5 (0.9)	1,631	41.8 (0.9)	364	40.0 (1.7)
Healthcare profession						
Physician	1,650	34.3 (4.3)	1,217	31.2 (4.7)	433	47.6 (3.4)
Nurse	1,406	29.2 (1.8)	1,261	32.3 (2.2)	145	15.9 (1.0)
Auxiliary nurse	387	8.1 (1.9)	341	8.7 (2.1)	46	5.1 (2.0)
Other profession involved in patient care	555	11.5 (0.9)	441	11.3 (0.9)	114	12.5 (1.4)
Other profession not involved in patient care	812	16.9 (1.8)	640	16.4 (1.7)	172	18.9 (2.7)
Type of workplace						
Hospital	2,818	58.6 (13.2)	2,273	58.3 (13.4)	545	59.9 (12.4)
Primary care	1,581	32.9 (14.3)	1,313	33.7 (14.5)	268	29.5 (13.6)
Others	410	8.5 (2.6)	313	8.0 (2.4)	97	10.7 (3.3)
Had close contact with someone who has died from COVID-19 because of work	2,190	45.5 (3.3)	1,781	45.7 (3.3)	409	44.9 (4.1)
Positive test or medical diagnosis of COVID-19	936	19.5 (2.2)	752	19.3 (2.1)	184	20.2 (2.9)

SE: standard error; SD: standard deviation.

Measures

Primary outcome

The study's primary outcome was any 30-day STBs at the 4-month follow-up (T2), assessed using a modified version of four selected items from the Columbia Suicide Severity Rating Scale (Posner *et al.*, 2011), each with dichotomous response options (yes or no). The items assess passive suicidal ideation (SI) ('wish you were dead or would go to sleep and never wake up'), active SI ('have thoughts of killing yourself'), suicide plans ('think about how you might kill yourself [e.g. taking pills, shooting yourself] or work out a plan of how to kill yourself') and suicide attempts in the past 30 days ('make a suicide attempt [i.e. purposefully hurt yourself with at

least some intent to die]'). Following previous studies, the primary outcome labelled as 'any STB' was created as a dichotomous variable indicating the presence of any of the four STB outcomes (Mortier *et al.*, 2021a; Nock *et al.*, 2014).

Baseline predictor variables

The baseline survey (T1) contains 207 items that were used to create the 219 predictor variables for STBs (items with non-ordinal categorical answers were converted into as many dummy variables as the number of categories minus one). The items were organized into eight different sections based on their contents (see Supplementary Table 1 for the list of items). Due to

space constraints, we provide here a short description of predictor variables and corresponding T1 survey sections: (1) eight sociodemographic variables (e.g., age, gender and marital status); (2) 14 variables related to COVID-19 exposure, infection status and perceived risk for COVID-19 infection (e.g., having received a positive COVID-19 test and having been hospitalized for COVID-19); (3) 55 items related to mental disorders, including a checklist for pre-pandemic lifetime mental disorders and screening scale items spanning five common current mental disorders, i.e., Major Depressive Disorder (PHQ-8; Kroenke *et al.*, 2009), Generalized Anxiety Disorder 7-item (Spitzer *et al.*, 2006), 30-day panic attacks (item adapted from the Composite International Diagnostic Interview (CIDI) screening scale; Kessler *et al.*, 2013), 30-day traumatic stress symptoms (four-item abbreviated form of the PTSD Checklist, PCL-5; Zuromski *et al.*, 2019) and substance use disorder (four-item version of the CAGE Adapted to Include Drugs (CAGE-AID); Hinkin *et al.*, 2001). In addition, any 30-day STB was assessed (Posner *et al.*, 2011), as well as burnout (six-item personal burnout subscale of the Copenhagen Burnout Inventory; Kristensen *et al.*, 2005), 30-day psychotic symptoms (items taken from the prodromal questionnaire; Loewy *et al.*, 2011) and 30-day obsessive compulsive disorder symptoms (the three-item obsessing subscale of the obsessive compulsive inventory revised; Foa *et al.*, 2002); (4) 30 items assessing treatment use, including healthcare service and psychotropic medication use for emotional or substance use problems, as well as barriers for treatment use; (5) 27 items assessing relevant work-related variables (e.g., type of HCWs, type of workplace, income, perceived risk for COVID-19 at work, perceived lack of healthcare centre preparedness and moral injury); (6) three items about isolation, quarantine and confinement due to COVID-19; (7) 35 items assessing 12-month serious stressful events, perceived stress (adapted peri life events scale; Dohrenwend *et al.*, 1978), resilience (Connor–Davidson resilience scale; Connor and Davidson, 2003) and healthy habits and (8) 35 items assessing social support (Oslo social support scale; Husain *et al.*, 2016), loneliness (UCLA three-item loneliness scale; Hughes *et al.*, 2004), use of social media, family functioning (Brief Assessment of Family Functioning Scale; Mansfield *et al.*, 2019), parental stress (items taken from the Parental Stress Scale; Berry and Jones, 1995), quality of life (five-level version of EQ-5D; Herdman *et al.*, 2011), somatic comorbidity (self-administered comorbidity questionnaire (Sangha *et al.*, 2003) and role impairment (Sheehan disability scales; Sheehan *et al.*, 1996).

Statistical analysis

The percentage of missing values across all variables analysed was moderate, with a mean missing rate of 6.5% and a median value of less than 1% (see Supplementary Table 2 for the percentage of missing values for each variable). Multiple imputation by chained equations with 10 iterations per imputation and 12 imputed datasets was used to impute missing item-level data (Van Buuren, 2018) using R's mice package (Buuren and Groothuis-Oudshoorn, 2011). The choice of 12 imputations provided a reasonable trade-off between statistical accuracy and computational efficiency, following recommendations that 5–20 imputations are generally sufficient under moderate missingness (Van Buuren, 2018; White *et al.*, 2011) and that the number of imputations should be at least equal to 100 times the fraction of missing information, which in our study was below 0.1 (White *et al.*, 2011) for key performance measures such as AUC.

The regularization path of linear Support Vector Classifier (SVC) with L1 penalty (Dai and Zhao, 2020) was implemented to

select the most critical predictor variables out of the 219 candidates by forcing some coefficients to be exactly zero, aiming to improve the accuracy and efficiency of predictive models (Montesinos López & Crossa, 2022). SVC for variable selection was applied to the 12 imputed datasets. Variables that were selected in at least 7 of the 12 imputed datasets were included in the final prediction model. This decision is justified by the work of Zhao and Long (Zhao and Long, 2017) who propose to perform variable selection separately in each imputed dataset and then include variables that are selected with a frequency above a defined threshold. In this study, the threshold chosen was 7 out of 12 imputations to ensure consistency across more than 50% of the imputations. According to Wood *et al.* (2008), this strategy improves the robustness of variable selection and coefficient estimation in regression models.

To address imbalanced data (i.e., 7.9% prevalence of STBs at the 4-month follow-up), which can lead to poor minority class classification (Rezvani and Wang, 2023), two different techniques were compared: (1) Synthetic Minority Oversampling Technique (SMOTE) (Chawla *et al.*, 2002) and (2) majority class undersampling (Devi *et al.*, 2021).

A random forest (RF) classifier was used to develop prediction models for STBs at the 4-month follow-up (Hammelrath *et al.*, 2023; Navarro *et al.*, 2021). Different values of the predefined hyperparameters specifying the number of decision trees to be included in the RF (*n_estimators*: [20, 25, 50, 75, 100]) and the maximum depth allowed for each decision tree (*max_depth*: [7, 8, 9, 10, 11]) were tested using a grid combination and 5-fold cross-validation. The grid combination was tested on a single stacked dataset of all 12 imputed datasets for both balancing methods (Seki *et al.*, 2021). The model with the selected hyperparameters was then independently trained and tested for each of the 12 imputed datasets. To decrease the risk of overfitting, 5-fold cross-validation was used in each imputed dataset. We aggregated the results of the predicted probabilities of each of the imputations into a single dataset to obtain performance metrics. The *RandomForestClassifier* function of the *sklearn* library in Python version 3.8 was used (Pedregosa *et al.*, 2012).

Model characteristics were assessed on the test dataset through the area under the Receiver Operating Characteristic (AUROC) curve and the area under the precision (positive predictive value [PPV])–recall (sensitivity) curve. The precision–recall curve is particularly useful for imbalanced datasets (Saito and Rehmsmeier, 2015). These curves allowed us to evaluate recall, specificity and precision for different cut-off points. The model was applied to each of the imputed datasets and predictions from each dataset were aggregated to obtain the overall metrics values (Seki *et al.*, 2021). All metrics were also obtained separately for men and women.

To quantify variable importance, the Shapley Additive Explanations (SHAP) (Lundberg and Lee, 2017) method was used. SHAP values represent the contribution that each variable had in the final model prediction. As SHAP values can display variability across imputations, they were obtained separately for each of the 12 imputed datasets and a combined representation of the contributions of each variable was obtained as the mean of these values (Seki *et al.*, 2021). Although the model is the same for both genders, the SHAP values have been obtained separately for the subsamples of men and women. SHAP summary plots are provided. In this plot, variables are ordered according to their influence on the predictions of the model. Each dot represents an individual's SHAP value, plotted along the horizontal axis. The dots are collared based on the variable's value, ranging from low (blue) to high (pink). If pink dots appear on the right side and blue

dots on the left, it indicates that the risk increases as the value of the variable rises. Analyses were conducted with the SHAP library in Python version 3.8.

Results

Sample characteristics

Table 1 shows the descriptive characteristics of the study sample by gender at T1. Ages ranged from 18 to 71 years, with a mean age of 45.8 (SD = 11.0). The country of birth was Spain for 95.3% of the sample. Healthcare professionals were mostly women (81.1%). About one-third (34.3%) were physicians, 29.2% nurses, 8.1% auxiliary nurses, 11.5% other profession involved in patient care and 16.9% other profession not involved in patient care. A total of 19.5% had a positive test or medical diagnosis of COVID-19 (women = 19.3%, men = 20.2%). A total of 381 subjects (7.9%; women = 7.8%, men = 8.2%) reported having had 30-day STBs at T2.

Of the 7.9% of subjects with any STB at the 12-month follow-up, 0.9% reported a suicide attempt, 26.3% a suicide plan without attempt, 10.7% active ideation without plan or attempt and 62.1% passive ideation without active ideation, plan or attempt.

Variables selection

Out of the initial 219 candidate predictor variables, 34 variables were selected by the linear SVM (Supplementary Figure 1). Although the gender variable was not selected, it was included because gender is a key variable due to significant differences in STB risk factors between men and women (Miranda-Mendizabal et al., 2019; Schrijvers et al., 2012), and to be able to assess relevant risk factors within each gender. This leads to a total of 35 variables being included in the model. The inclusion in the RF model is also justified by the ability of the RF model to account for complex interactions between variables (Auret and Aldrich, 2012).

Random forest

The selected hyperparameters specified 50 decision trees to be included in the RF ($n_{\text{estimators}} = 50$) and a maximum depth of 9 allowed for each decision tree ($\text{max_depth} = 9$).

Figure 1 presents AUROC and precision–recall curves resulting from the RF fitted on the test sample using the 35 baseline selected variables. The AUROC with and without balancing techniques is higher than 0.80. In the total sample, the best result was obtained with the model without data balancing (AUROC = 0.87).

Regarding the area under the precision–recall curve, large differences are observed between balancing methods, being the model without data balancing the one with the best result (area under the precision–recall curve = 0.52).

Figure 2 shows that when the goodness of fit of the model without data balancing is assessed separately by gender, the good metric properties are maintained. The AUROC curve is 0.84 and 0.86 for men and women, respectively. In the case of the area under the precision–recall curve, the values obtained are 0.45 for men and 0.54 for women.

SHAP values

In the summary plot of the SHAP values (Figures 3 and 4), the variables are sorted according to their importance in the prediction model. The colour of each point on the graph represents the value

of the corresponding variable: pink indicates high values and blue indicates low values. The horizontal axis (x -axis) represents the SHAP value: having values above 0 indicates that these experiences are potentially important predictor variables in predicting future suicidal ideation. Figure 3 shows the most important variables in the prediction of STBs. The ranking is headed by the number of days in the past 30 days with suicidal ideation (passive or active) followed by passive suicidal ideation, the number of days in the past 30 days with binge eating episodes and having intrusive thoughts (i.e., nasty thoughts and having difficulty in getting rid of them). Another relevant factor identified is concentration problems. Some of the factors associated with COVID-19 infection that have been identified include: having been in isolation or quarantine, fear of personal or loved ones' infection, work-related factors and experiences during the initial pandemic outbreak, such as perceived lack of supervision at work, not getting along with co-workers, stress related to having to prioritize care among patients and work-related role impairment. Financial stress also appears as a risk factor for STBs.

Among men (Figure 4a) the most important variables for predicting STBs at T2 included the number of days in the past 30 days with suicidal ideation (passive or active), passive suicidal ideation and frequency of intrusive thoughts at the T1 assessment.

Among women (Figure 4b), the most important baseline variables for predicting STBs (at T2) included the number of days in the past 30 days with suicidal ideation (passive or active), passive suicidal ideation and number of panic attacks.

Discussion

In this study, we developed and validated a predictive model for STBs within a 4-month period using survey data collected from HCWs during the COVID-19 pandemic in Spain. The ML-based model showed robust predictive performance for STBs and identified, out of a total of 219 variables, the 35 key predictive variables associated with STBs. Our model showed very good metric characteristics with an AUROC of 0.86 (0.86 in women and 0.84 in men) and an area under the precision–recall curve of 0.52 (0.54 and 0.45 in women and men, respectively). The results align with previous studies, as shown in the systematic review by Somé et al. (2024), which found a mean AUROC of 0.81 in 84 studies, a mean recall of 0.68 in 64 studies and a mean precision of 0.41 in 46 studies. Our results improve precision metric (Nock et al., 2022), which is challenging due to the low prevalence of STBs. ML models developed to predict STBs in previous studies have been criticized for having low precision (often below 1%) and thus producing too many false positives (Nock et al., 2022). Our model achieved a precision of 50% with a recall of 60%. With recalls of 80%, the precision is greater than 20%. The fact that the model's cut-off point is not predetermined allows for their selection based on the required recall and precision, depending on the objective or application of the predictive models. As the data were unbalanced (92.1% of the subjects in one category), balancing techniques were tested, but these techniques did not improve the results of the models.

STB at T1 was identified as a strong predictor at T2, consistent with previous literature (Ribeiro et al., 2016). While the association between mental disorders and suicidal ideation is well established (Franklin et al., 2017), a key contribution of our study is the identification of specific mental disorder symptoms as independent predictors of STBs among HCWs active during the COVID-19 pandemic, including binge eating, panic attacks, intrusive thoughts and concentration problems. These results align with studies in

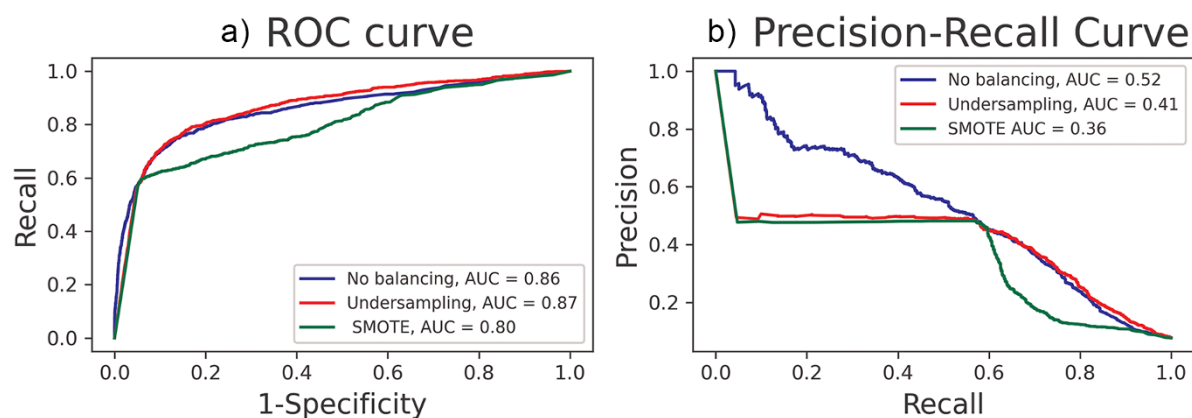


Figure 1. (a) The receiver operating characteristics curve and the area under the receiver operating characteristic curve (AUROC) for suicidal thoughts and behavior prediction. The results of the prediction using different balancing test are shown (left). (b) The precision-recall curve and the area under the precision-recall curve of the models. The results of the prediction using different balancing test are shown (right). Abbreviations: ROC: receiver operating characteristics; AUC: area under the curve; No balancing: no balancing technique was used; Undersampling: undersampling of the majority class technique was used; SMOTE: Synthetic Minority Oversampling Technique was used.

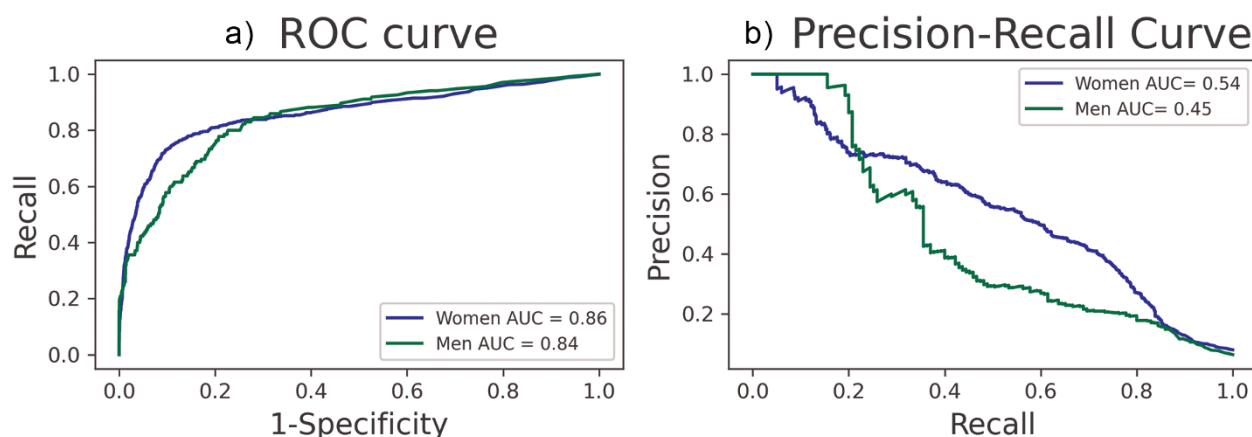


Figure 2. (a) The receiver operating characteristics curve and the area under the receiver operating characteristic (AUROC) curve of the models for men and women. (b) The precision-recall curve and the area under the precision-recall curve of the models for men and women. Abbreviations: ROC: receiver operating characteristics; AUC: area under the curve.

non-HCW populations linking STBs to eating disorders (Brown *et al.*, 2018; Sohn *et al.*, 2023), panic disorder (Zhang *et al.*, 2022), obsessive-compulsive disorder (Pellegrini *et al.*, 2021) and concentration difficulties (Lo *et al.*, 2023). These findings highlight the critical need for early identification and screening of mental health symptoms in HCWs, as well as the challenge of ensuring access to timely, evidence-based mental health care (Jain *et al.*, 2024). Notably, our study also found that practical barriers to seeking treatment and disruptions in psychiatric or psychological care due to the COVID-19 pandemic were significant predictors of future STBs, a particularly concerning issue given the low treatment utilization among HCWs (Braquehais *et al.*, 2022; Dellazizzo *et al.*, 2021; Mortier *et al.*, 2024; Rogoža *et al.*, 2021).

Consistent with prior research (Du *et al.*, 2023; Kavukcu and Akdeniz, 2021), our study found that COVID-19-related experiences, including isolation or quarantine and fear of personal or familial infection, predicted STBs at follow-up, likely due to their traumatic and stressful nature (Portillo-Van Diest *et al.*, 2023). Additionally, work-related disruptions during the initial pandemic outbreak, such as perceived lack of supervision, interpersonal conflicts with co-workers, stress from prioritizing patient care and role impairment, emerged as significant predictors of later suicidal

ideation. These findings underscore the need for systemic workplace reforms, including improved healthcare centre preparedness for viral outbreaks through enhanced equipment, staffing, training and protocols. Moreover, fostering supportive work environments, encouraging the reporting of interpersonal conflicts (Alshammari and Dayrit, 2017) and implementing effective communication and conflict resolution strategies (Jerng *et al.*, 2017) are essential. Future research should focus on delineating causal pathways underlying STBs among HCWs to inform targeted prevention efforts, addressing the critical gap in evidence-based interventions for mental health issues in this population at both individual and organizational levels (Petrie *et al.*, 2019).

Gender was not selected as a relevant predictor when the Support Vector Machine (SVM) model was applied. This result was unexpected as gender has been shown in the mental health and STB literature to be a key variable in identifying significant risk factors. The ability of the RF model to capture complex interactions, identify non-linear dependencies and consider multivariate relationships between variables (Auret and Aldrich, 2012), together with the recognized clinical and theoretical relevance of gender in mental disorders and STB – given that risk factors differ significantly between men and women (Miranda-Mendizabal *et al.*,



Figure 3. Shapley additive explanation (SHAP) summary graph. Each point on the graph is a SHAP value for one variable. The color represents the value of the variable from low (blue) to high (pink).

Notes: The color of each point on the graph represents the value of the corresponding variable: pink indicates high values and blue indicates low values. The horizontal axis (x-axis) represents the SHAP value.

Abbreviations: C-SSRS: Suicidal thoughts and behaviors screen; PHQ-9: Patient Health Questionnaire 9 item; CIDI: Composite International Diagnostic Interview; Prime-MD: Primary Care Evaluation of Mental Disorders; BAFFS: Brief Assessment of Family Functioning Scale; CD-RISC: Connor-Davidson Resilience Scale; OSS3: Oslo Social Support Scale; OCI-R: Obsessive Compulsive Inventory revised.

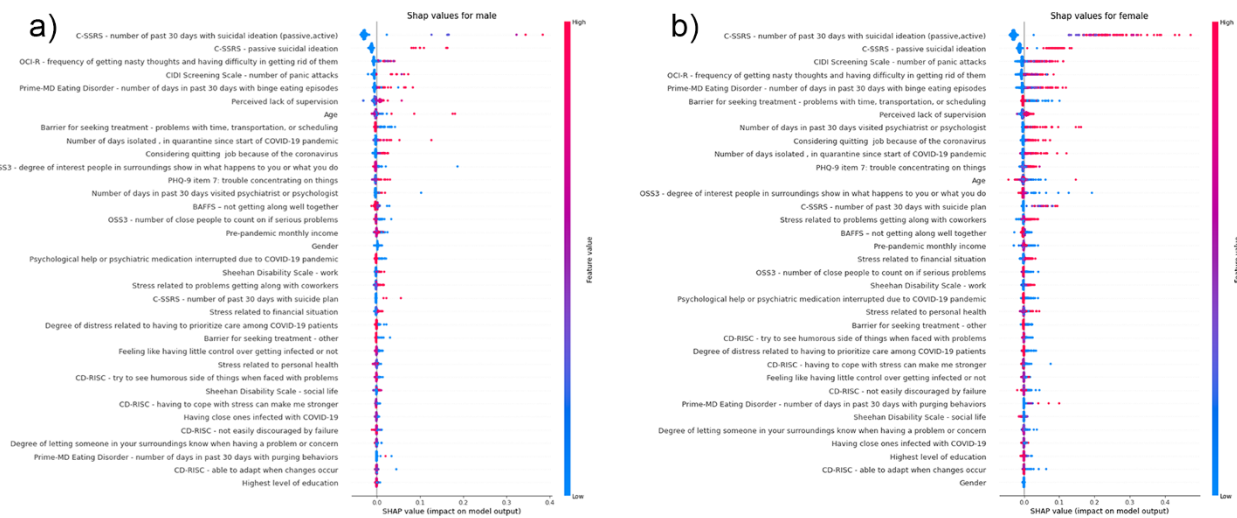


Figure 4. Shapley additive explanation (SHAP) summary graph. for men (a) and women (b).

Notes: The color of each point on the graph represents the value of the corresponding variable: pink indicates high values and blue indicates low values. The horizontal axis (x-axis) represents the SHAP value.

Abbreviations: C-SSRS: Suicidal thoughts and behaviors screen; PHQ-9: Patient Health Questionnaire 9 item; CIDI: Composite International Diagnostic Interview; Prime-MD: Primary Care Evaluation of Mental Disorders; BAFFS: Brief Assessment of Family Functioning Scale; CD-RISC: Connor-Davidson Resilience Scale; OSS3: Oslo Social Support Scale; OCI-R: Obsessive Compulsive Inventory revised.

2019; Schrijvers *et al.*, 2012) – justifies the inclusion of the gender variable in the model.

Considering the evident gender differences in both the absolute risk of STBs and the prevalence of associated risk factors (Gradus

et al., 2021; Jiang *et al.*, 2021; Miranda-Mendizabal *et al.*, 2019a, 2019b), the accuracy of the model was assessed separately for men and women, and the model proved to be a good fit for both genders. For both genders, the most important factors were the number of

suicidal thoughts (passive or active) and passive suicidal thoughts in the last 30 days. For men, the second most important factor was the frequency of unpleasant thoughts and the difficulty in getting rid of them, while for women it was the number of panic attacks. This is an important finding suggesting that there are gender differences in the relative importance of risk factors for STBs. Previous research has shown significant interactions between gender and certain risk factors for STBs (Miranda-Mendizabal *et al.*, 2019a). These gender differences in risk factors have been linked to variations in the prevalence of internalizing and externalizing disorders between genders, as well as differences in coping strategies, including the frequency of help-seeking. These differences may be attributed to gender socialization.

Strengths and limitations

This study has some limitations. First, due to low numbers of suicide plans and suicide attempts at the 4-month follow-up, we operationalized the study outcome as any STB (i.e., having passive or active suicidal ideation with or without plan or attempt), as all four separate outcomes indicate the presence of at least passive suicidal ideation. This is in line with previous work by our group (Mortier *et al.*, 2021a) and others (Benjet *et al.*, 2022; Nock *et al.*, 2022). Second, a convenience sample was used and results need to be validated using external samples. This limitation is partially addressed by obtaining a random and heterogeneous census sample of HCWs from 18 healthcare institutions in six Autonomous Communities. Third, STB is complex and difficult to predict; therefore, psychosocial and environmental factors cannot be easily excluded (Ati *et al.*, 2021). However, the number of variables used to predict was very large, all 219 variables collected in the survey were used. Fourth, although an SVM with an L1 regularization model was used for a large and objective selection of variables, the analysis is based on a predefined survey with a specific number of items. This implies that, although the model allows for a greater inclusion of predictor variables, there is still the limitation of not including all possible variables relevant in the context of STBs. Fifth, we have exclusively used the RF algorithm to predict STBs. Although the predictive capacity of this algorithm was effective in our study, other models may also provide meaningful and complementary results that could improve the accuracy of our results analysis.

Conclusion

In this study of Spanish HCWs during the COVID-19 pandemic, we have developed a predictive model of the risk of STBs. Our results show that RF ML algorithm has a high prediction performance for STBs (AUROC = 0.86; 0.86 in women and 0.84 in men). Importantly, our study improves the precision compared to previous research. The results generated by the proposed model help to identify and explain risk factors for STBs and contribute to the development of a first comprehensive conceptual framework for understanding STB occurrence in major epidemics and other disasters with high impact on essential workers. The most important predictors contributing to the prediction of suicide ideation in healthcare professionals were ideation frequency in the last 30 days, passive suicidal ideation and the number of days with binge eating episodes in the last 30 days.

Supplementary material. The supplementary material for this article can be found at <https://doi.org/10.1017/S2045796025000198>.

Availability of data and materials. The de-identified participant data as well as the study protocol, statistical analysis plan and data dictionaries used for this study are available as from publication and upon reasonable request from the corresponding author (P.M.; pmortier@researchmar.net) as long as the main objective of the data-sharing request is replicating the analysis and findings as reported in this paper.

Acknowledgements. The authors would like to sincerely thank all HCWs participating in the study. They also thank Puri Barbas for the management of the project.

Author Contributions. I.A., G.V., P.M. and O.P. reviewed the literature. I.A., J.A., G.V., P.M., O.P., M.F., E.A., V.P.S., J.M.H. and R.B. conceived and designed the study. E.A., J.D.M., N.L.F., T.P., J.M.P.-T., J.I.P., M.E., M.N.P., A.G.-P., C.R., E.A., N.N.A., M.C., A.P.-Z., E.V., C.S. and V.P.-S. acquired the data. G.V., I.A., F.A. and P.M. cleaned and analysed the data. I.A., G.V., O.P. and P.M. drafted the initial version of the manuscript. All authors reviewed the initial draft, made a critical contribution to the interpretation of the data and approved the manuscript. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Financial support. This work was supported by grants from the Instituto de Salud Carlos III (ISCIII)/Ministerio de Ciencia e Innovación/FEDER COV20/00711 (J.A.); ISCIII-FEDER (J.A., grant number PI17/00521); Miguel Servet grant (P.M., CP21/00078) co-financed by the ISCIII and co-funded by the European Union; ISCIII Sara Borrell CD18/00049 (P.M.); ISCIII co-financed by the European Union through the European Social Fund Plus PFIS grant FI23/00004 (A.P.-V.D.); the Secretaria d'Universitats i Recerca del Departament d'Economia i Coneixement of the Generalitat de Catalunya AGAUR 2021 SGR 00624 (J.A.); PERIS, Departament de Salut SLT017/20/000009 (I.A.); and CIBER of Epidemiology & Public Health, ISCIII CB06/02/0046. Additional partial funding was received from the Gerencia Regional de Salud de Castilla y León (SACYL) GRS COVID 32/A/20 (J.M.P.-T.).

Competing interests. E.A. reports personal fees from Lundbeck, Esteve and Boehringer-Ingelheim, outside the submitted work. E.V. has received grants and served as a consultant, advisor or Continuing Medical Education (CME) speaker for the following entities: AB-Biotics, AbbVie, Adamed, Alcediag, Angelini, Biogen, Beckley-Psytech, Biohaven, Boehringer-Ingelheim, Celon Pharma, Compass, Dainippon Sumitomo Pharma, Ethypharm, Ferrer, Gedeon Richter, GH Research, Glaxo-Smith Kline, HMNC, Idorsia, Johnson & Johnson, Lundbeck, Luye Pharma, Medincell, Merck, Newron, Novartis, Orion Corporation, Organon, Otsuka, Roche, Rovi, Sage, Sanofi-Aventis, Sunovion, Takeda, Teva and Viatris, outside the submitted work. A.G.-P. has received grants and served as a consultant, advisor or CME speaker for the following entities: Janssen-Cilag, Lundbeck, Otsuka, Alter, Angelini, Novartis, Rovi, Takeda, the Spanish Ministry of Science and Innovation (CIBERSAM), the Ministry of Science (Carlos III Institute), the Basque Government and the European Framework Program of Research. J.M.H. has served as a consultant, advisor or CME speaker for the following entities: Boehringer-Ingelheim, Eli Lilly and Co. and Lundbeck, outside the submitted work. J.M.P.-T. has served as a consultant, advisor or CME speaker for the following entities: Angelini, Boehringer-Ingelheim, Eli Lilly and Co, Johnson & Johnson Lundbeck, Otsuka and Rovi, outside the submitted work. All other authors reported no conflict of interest. R.B. reports fees from Janssen-Cilag, outside the submitted work.

Ethical standards. The authors assert that all procedures contributing to this study comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2000.

References

Alonso J, Vilagut G, Alayo I, Ferrer M, Amigo F, Aragón-Peña A, Aragonès E, Campos M, Del Cura-gonzález I, Urreta I, Espuga M, González Pinto A, Haro JM, López Fresneña N, Martínez de Salazar A, Molina JD, Ortí Lucas RM, Parellada M, Pelayo-Terán JM, Pérez Zapata A, Pijoan JI, Plana N, Puig MT, Rius C, Rodríguez-Blázquez C,

- Sanz F, Serra C, Kessler RC, Bruffaerts R, Vieta E, Pérez-Solà V and Mortier P (2022) Mental impact of Covid-19 among Spanish healthcare workers. A large longitudinal survey. *Epidemiology and Psychiatric Sciences* 31, e28. doi:10.1017/S2045796022000130
- Alonso J, Vilagut G, Mortier P, Ferrer M, Alayo I, Aragón-Peña A, Aragonès E, Campos M, Cura-González ID, Emparanza JI, Espuga M, Forjaz MJ, González-Pinto A, Haro JM, López-Fresneña N, Salazar ADMD, Molina JD, Ortí-Lucas RM, Parellada M, Pelayo-Terán JM, Pérez-Zapata A, Pijoan JI, Plana N, Puig MT, Rius C, Rodríguez-Blázquez C, Sanz F, Serra C, Kessler RC, Bruffaerts R, Vieta E and Pérez-Solà V (2021) Mental health impact of the first wave of COVID-19 pandemic on Spanish healthcare workers: a large cross-sectional survey. *Revista de Psiquiatria y Salud Mental*. 14, 90–105. doi:10.1016/j.rpsm.2020.12.001
- Alshammari HF and Dayrit RDJ (2017) Conflict and conflict resolution among the medical and nursing personnel of selected hospitals in Hail City. *IOSR Journal of Nursing and Health Science* 06, 45–60. doi:10.9790/1959-0603014560
- Ati NAL, Paraswati MD and Windarwati HD (2021) What are the risk factors and protective factors of suicidal behavior in adolescents? A systematic review. *Journal of Child and Adolescent Psychiatric Nursing* 34, 7–18. doi:10.1111/jcap.12295
- Auret L and Aldrich C (2012) Interpretation of nonlinear relationships between process variables by use of random forests. *Minerals Engineering* 35, 27–42. doi:10.1016/j.mineng.2012.05.008
- Benjet C, Borges G, Miah S, Albor Y, Gutiérrez-García RA, Zavala Berben A, Guzmán R, Vargas-Contreras E, Hermsillo de la Torre AE, Hernández Uribe PC, Quevedo G, Covarrubias Díaz A, Martínez Ruiz S, Valdés-García KP, Martínez Jerez AM and Mortier P (2022) One-year incidence, predictors, and accuracy of prediction of suicidal thoughts and behaviors from the first to second year of university. *Depression and Anxiety* 39, 727–740. doi:10.1002/da.23278
- Bennett M, Kleczyk EJ, Hayes K and Mehta R (2022) Evaluating similarities and differences between machine learning and traditional statistical modeling in healthcare analytics. In Aceves-Fernández MA and Travieso-Gonzalez CM (eds), *Artificial Intelligence Annual Volume 2022*. Rijeka: IntechOpen. doi:10.5772/intechopen.105116
- Berry JO and Jones WH (1995) The parental stress scale: initial psychometric evidence. *Journal of Social and Personal Relationships* 12, 463–472. doi:10.1177/0265407595123009
- Boudreaux ED, Rundensteiner E, Liu F, Wang B, Larkin C, Agu E, Ghosh S, Semeter J, Simon G and Davis-Martin RE (2021) Applying machine learning approaches to suicide prediction using healthcare data: overview and future directions. *Frontiers in Psychiatry*, 12. doi:10.3389/fpsy.2021.707916
- Braquehais MD, Gómez-Duran EL, Nieva G, Valero S, Ramos-Quiroga JA and Bruguera E (2022) Help seeking of highly specialized mental health treatment before and during the COVID-19 pandemic among health professionals. *International Journal of Environmental Research and Public Health* 19, 3665. doi:10.3390/ijerph19063665
- Brown KL, LaRose JG and Mezuk B (2018) The relationship between body mass index, binge eating disorder and suicidality. *BMC Psychiatry* 18, 196. doi:10.1186/s12888-018-1766-z
- Buuren SV and Groothuis-Oudshoorn K (2011) Mice: multivariate imputation by chained equations in R. *Journal of Statistical Software* 45. doi:10.18637/jss.v045.i03
- Chawla NV, Bowyer KW, Hall LO and Kegelmeyer WP (2002) SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research* 16, 321–357. doi:10.1613/jair.953
- Connor KM and Davidson JRT (2003) Development of a new resilience scale: the Connor-Davidson Resilience Scale (CD-RISC). *Depression and Anxiety* 18, 76–82. doi:10.1002/da.10113
- Dai Y and Zhao P (2020) A hybrid load forecasting model based on support vector machine with intelligent methods for feature selection and parameter optimization. *Applied Energy* 279, 115332. doi:10.1016/j.apenergy.2020.115332
- Davis MA, Cher BAY, Friese CR, Bynum JPW (2021) Association of US nurse and physician occupation with risk of suicide. *JAMA Psychiatry* 78, 651. doi:10.1001/jamapsychiatry.2021.0154
- Dellazzio L, Léveillé N, Landry C and Dumais A (2021) Systematic review on the mental health and treatment impacts of COVID-19 on neurocognitive disorders. *Journal of Personalized Medicine*. 11, 746. doi:10.3390/jpm11080746
- Devi D, Biswas S and Purkayastha B (2021) A Review on Solution to Class Imbalance Problem: Undersampling Approaches. In Paul S and Verma JK (eds), *Proceedings of the 2021 International Conference on Computational Performance Evaluation (ComPE 2021)*. Piscataway, NJ: IEEE, pp. 626–631. doi:10.1109/ComPE52596.2021.00022
- Dohrenwend BS, Askenasy AR, Krasnoff L and Dohrenwend BP (1978) Exemplification of a method for scaling life events: the PERI life events scale. *Journal of Health and Social Behavior* 19, 205. doi:10.2307/2136536
- Du W, Jia YJ, Hu FH, Ge MW, Cheng YJ, Qu X and Chen HL (2023) Prevalence of suicidal ideation and correlated risk factors during the COVID-19 pandemic: a meta-analysis of 113 studies from 31 countries. *Journal of Psychiatric Research* 166, 147–168. doi:10.1016/j.jpsychires.2023.07.040
- Dutheil F, Aubert C, Pereira B, Dambrun M, Moustafa F, Mermillod M, Baker JS, Trousselard M, Lesage F-X and Navel V (2019) Suicide among physicians and health-care workers: a systematic review and meta-analysis. *PLoS One* 14, e0226361. doi:10.1371/journal.pone.0226361
- Eyles E, Moran P, Okolie C, Dekel D, Macleod-Hall C, Webb RT, Schmidt L, Knipe D, Sinyor M, McGuinness LA, Arensman E, Hawton K, O'Connor RC, Kapur N, O'Neill S, Olorisade B, Cheng HY, Higgins JPT, John A and Gunnell D (2021) Systematic review of the impact of the COVID-19 pandemic on suicidal behaviour amongst health and social care workers across the world. *Journal of Affective Disorders Reports* 6, 100271. doi:10.1016/j.jadr.2021.100271.
- Favril L, Yu R, Uyar A, Sharpe M and Fazel S (2022) Risk factors for suicide in adults: systematic review and meta-analysis of psychological autopsy studies. *Evidence Based Mental Health* 25, 148–155. doi:10.1136/ebmental-2022-300549
- Foa EB, Huppert JD, Leiberg S, Langner R, Kichic R, Hajcak G and Salkovskis PM (2002) The Obsessive-Compulsive Inventory: development and validation of a short version. *Psychol Assessment* 14, 485–496. doi:10.1037/1040-3590.14.4.485
- Franklin JC, Ribeiro JD, Fox KR, Bentley KH, Kleiman EM, Huang X, Musacchio KM, Jaroszewski AC, Chang BP and Nock MK (2017) Risk factors for suicidal thoughts and behaviors: a meta-analysis of 50 years of research. *Psychological Bulletin* 143, 187–232. doi:10.1037/bul0000084
- García-Iglesias JJ, Gómez-Salgado J, Fernández-Carrasco FJ, Rodríguez-Díaz L, Vázquez-Lara JM, Prieto-Callejero B and Allande-Cussó R (2022). Suicidal ideation and suicide attempts in healthcare professionals during the COVID-19 pandemic: A systematic review. *Frontiers in Public Health* 10, 1043216. doi:10.3389/fpubh.2022.952250
- Gradus JL, Rosellini AJ, Horváth-Puhó E, Jiang T, Street AE, Galatzer-Levy I, Lash TL and Sørensen HT (2021) Predicting sex-specific nonfatal suicide attempt risk using machine learning and data from Danish national registries. *American Journal of Epidemiology* 190, 2517–2527. doi:10.1093/aje/kwab112
- Greenberg N, Docherty M, Gnanapragasam S and Wessely S (2020) Managing mental health challenges faced by healthcare workers during covid-19 pandemic. *BMJ* 368, m1211. doi:10.1136/bmj.m1211
- Hammelrath L, Hilbert K, Heinrich M, Zagorscak P and Knaevelsrud C (2023) Select or adjust? How information from early treatment stages boosts the prediction of non-response in internet-based depression treatment. *Psychological Medicine* 54, 1641–1650. doi:10.1017/S0033291723003537
- Harvey SB, Epstein RM, Glozier N, Petrie K, Strudwick J, Gayed A, Dean K and Henderson M (2021) Mental illness and suicide among physicians. *The Lancet* 398, 920–930. doi:10.1016/S0140-6736(21)01596-8
- Herdman M, Gudex C, Lloyd A, Janssen MF, Kind P, Parkin D, Bonsel G and Badia X (2011) Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Quality of Life Research* 20, 1727–1736. doi:10.1007/s11136-011-9903-x
- Hinkin CH, Castellon SA, Dickson-Fuhrman E, Daum G, Jaffe J and Jarvik L (2001) Screening for drug and alcohol abuse among older adults using a modified version of the CAGE. *The American Journal on*

- Addictions/American Academy of Psychiatrists in Alcoholism and Addictions* 10, 319–326. doi:10.1111/j.1521-0391.2001.tb00521.x
- Hughes ME, Waite LJ, Hawkey LC and Cacioppo JT (2004) A short scale for measuring loneliness in large surveys. *Psychological Medicine* 26, 655–672. doi:10.1177/0164027504268574
- Husain N, Mukherjee I, Notiar A, Alavi Z, Tomenson B, Hawa F, Malik A, Ahmed A and Chaudhry N (2016) Prevalence of common mental disorders and its association with life events and social support in mothers attending a well-child clinic. *Sage Open* 6, 215824401667732. doi:10.1177/2158244016677324
- Jain L, Sarfraz Z, Karlapati S, Kazmi S, Nasir MJ, Atiq N, Ansari D, Shah D, Aamir U, Zaidi K, Shakil Zubair A and Jyotsana P (2024) Suicide in healthcare workers: an umbrella review of prevalence, causes, and preventive strategies. *Journal of Primary Care & Community Health* 15. doi:10.1177/21501319241273242
- Jerng J-S, Huang S-F, Liang H-W, Chen L-C, Lin C-K, Huang H-F, Hsieh M-Y and Sun J-S (2017) Workplace interpersonal conflicts among the healthcare workers: retrospective exploration from the institutional incident reporting system of a university-affiliated medical center. *PLoS One* 12, e0171696. doi:10.1371/journal.pone.0171696
- Jiang T, Rosellini AJ, Horváth-Puhó E, Shiner B, Street AE, Lash TL, Sørensen HT and Gradus JL (2021) Using machine learning to predict suicide in the 30 days after discharge from psychiatric hospital in Denmark. *The British Journal of Psychiatry* 219, 440–447. doi:10.1192/bjp.2021.19
- Kavukcu E and Akdeniz M (2021) Tsunami after the novel coronavirus (COVID-19) pandemic: a global wave of suicide? *International Journal of Social Psychiatry* 67, 197–199. doi:10.1177/0020764020946348
- Kessler RC, Santiago PN, Colpe LJ, Dempsey CL, First MB, Heeringa SG, Stein MB, Fullerton CS, Gruber MJ, Naifeh JA, Nock MK, Sampson NA, Schoenbaum M, Zaslavsky AM and Ursano RJ (2013) Clinical reappraisal of the Composite International Diagnostic Interview Screening Scales (CIDI-SC) in the Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS). *International Journal of Methods in Psychiatric Research* 22, 303–321. doi:10.1002/mpr.1398
- Kristensen TS, Borritz M, Villadsen E and Christensen KB (2005) The Copenhagen Burnout Inventory: a new tool for the assessment of burnout. *Work & Stress* 19, 192–207. doi:10.1080/02678370500297720
- Kroenke K, Strine TW, Spitzer RL, Williams JBW, Berry JT and Mokdad AH (2009) The PHQ-8 as a measure of current depression in the general population. *Journal of Affective Disorders* 114, 163–173. doi:10.1016/j.jad.2008.06.026
- Lo HKY, Chan JKN, Yip EWC, Chui EMC, Fung VSC, Wong CSM, Chu RST, So YK, Chan JMT, Chung AKK, Lee KCK, Cheng CPW, Law CW, Chan WC and Chang WC (2023) The prevalence and correlates of suicidal ideation in Chinese psychiatric patients during the fifth wave of COVID-19 in Hong Kong. *European Neuropsychopharmacology* 77, 4–11. doi:10.1016/j.euroneuro.2023.08.485
- Loewy RL, Pearson R, Vinogradov S, Bearden CE and Cannon TD (2011) Psychosis risk screening with the Prodromal Questionnaire – Brief Version (PQ-B). *Schizophrenia Research* 129, 42–46. doi:10.1016/j.schres.2011.03.029
- Lundberg S and Lee S-I (2017) A Unified Approach to Interpreting Model Predictions. In *Advances in Neural Information Processing Systems 30 (NIPS 2017)*. Red Hook, NY: Curran Associates, Inc, pp. 4765–4774. <https://arxiv.org/abs/1705.07874>.
- Mansfield AK, Keitner GI and Sheeran T (2019) The Brief Assessment of Family Functioning Scale (BAFFS): a three-item version of the General Functioning Scale of the Family Assessment Device. *Psychotherapy Research* 29, 824–831. doi:10.1080/10503307.2017.1422213
- Mediavilla R, Fernández-Jiménez E, Andreo J, Morán-Sánchez I, Muñoz-Sanjosé A, Moreno-Küstner B, Mascayano F, Ayuso-Mateos JL, Bravo-Ortiz M-F and Martínez-Alés G (2023) Association between perceived discrimination and mental health outcomes among health workers during the initial COVID-19 outbreak. *Spanish Journal of Psychiatry and Mental Health* 16, 221–224. doi:10.1016/j.rpsm.2021.06.001
- Milner A, Spittal MJ, Pirkis J and LaMontagne AD (2013) Suicide by occupation: systematic review and meta-analysis. *British Journal of Psychiatry* 203, 409–416. doi:10.1192/bjp.bp.113.128405
- MINDCOVID (2020) Impact on Mental Health and care needs associated with the COVID-19 pandemic: a comprehensive assessment in Spain. <https://www.mindcovid.org/>
- Miranda-Mendizabal A, Castellví P, Alayo I, Vilagut G, Blasco MJ, Torrent A, Ballester L, Almenara J, Lagares C, Roca M, Sesé A, Piqueras JA, Soto-Sanz V, Rodríguez-Marín J, Echeburúa E, Gabilondo A, Cebrià AI, Bruffaerts R, Auerbach RP, Mortier P, Kessler RC and Alonso J (2019) Gender commonalities and differences in risk and protective factors of suicidal thoughts and behaviors: a cross-sectional study of Spanish university students. *Depression and Anxiety* 36, 1102–1114. doi:10.1002/da.22960
- Miranda-Mendizabal A, Castellví P, Parés-Badell O, Alayo I, Almenara J, Alonso I, Blasco MJ, Cebrià A, Gabilondo A, Gili M, Lagares C, Piqueras JA, Rodríguez-Jiménez T, Rodríguez-Marín J, Roca M, Soto-Sanz V, Vilagut G and Alonso J (2019) Gender differences in suicidal behavior in adolescents and young adults: systematic review and meta-analysis of longitudinal studies. *International Journal of Public Health* 64, 265–283. doi:10.1007/s00038-018-1196-1
- Montesinos López OA, Montesinos López A and Crossa J (2022) Statistical Machine Learning Methods for Genomic Prediction [Internet]. Chapter 4, Overfitting, Model Tuning, and Evaluation of Prediction Performance, 14 January 2022. Cham: Springer. Available at: <https://www.ncbi.nlm.nih.gov/books/NBK583970/>.
- Mortier P, Vilagut G, Alayo I, Ferrer M, Amigo F, Aragonès E, Aragón-Peña A, Asunsolo Del Barco A, Campos M, Espuga M, González-Pinto A, Haro JM, López Fresneña N, Martínez de Salazar A, Molina JD, Ortí-Lucas RM, Parellada M, Pelayo-Terán JM, Pérez-Gómez B, Pérez-Zapata A, Pijoan JI, Plana N, Polentinos-Castro E, Portillo-Van Diest A, Puig MT, Rius C, Sanz F, Serra C, Urreta-Barallobre I, Kessler RC, Bruffaerts R, Vieta E, Pérez-Solà V, Alonso J, Alonso J, Alayo I, Alonso M, Álvarez M, Amann B, Amigo FF, Anmella G, Aragón A, Aragonés N, Aragonès E, Arizón AI, Asunsolo A, Ayora A, Ballester L, Barbas P, Basora J, Bereciartua E, Ignasi Bolibar IB, Bonfill X, Cotillas A, Cuartero A, de Paz C, Cura ID, Jesus Del Yerro M, Diaz D, Domingo JL, Empananza JI, Espallargues M, Espuga M, Estevan P, Fernandez MI, Fernandez T, Ferrer M, Ferreres Y, Fico G, Forjaz MJ, Barranco RG, Garcia Torrecillas Garcia-ribera JM, Garrido A, Gil E, Gomez M, Gomez J, Pinto AG, Haro JM, Hernando M, Insigna MG, Iriberrí M, Jimenez N, Jimenez X, Larrauri A, Leon F, Lopez-Fresneña N, Lopez C, Lopez-Atanes Juan Antonio Lopez-Rodríguez M, Lopez-Cortacans G, Marcos A, Martin J, Martin V, Martinez-Cortés M, Martinez-Martinez R, Martinez de Salazar AD, Martinez I, Marzola M, Mata N, Molina JM, de Dios Molina J, Molinero E, Mortier P, Muñoz C, Murrú A, Olmedo J, Ortí RM, Padrós R, Pallejà M, Parra R, Pascual J, Pelayo JM, Pla R, Plana N, Aznar CB, Gomez BP, Zapata AP, Pijoan JI, Polentinos E, Puertolas B, Puig MT, Quílez A, Quintana MJ, Quiroga A, Rentero D, Rey C, Rius C, Rodríguez-Blázquez C, Rojas MJ, Romero Y, Rubio G, Rumayor M, Ruiz P, Saenz M, Sanchez J, Sanchez-Arcilla I, Sanz F, Serra C, Serra-Sutton V, Serrano M, Sola S, Solera S, Soto M, Tarrago A, Tolosa N, Vazquez M, Viciola M, Vieta E, Vilagut G, Yago S, Yañez J, Zapico Y, Zorilla LM, Zorrilla I, Zurbano SL and Perez-Solà V (2022) Four-month incidence of suicidal thoughts and behaviors among healthcare workers after the first wave of the Spain COVID-19 pandemic. *Journal of Psychiatric Research* 149, 10–17. doi:10.1016/j.jpsychires.2022.02.009
- Mortier P, Vilagut G, Ferrer M, Alayo I, Bruffaerts R, Cristóbal-Narváez P, Del Cura-González I, Domènech-Abella J, Féllez-Nobrega M, Olaya B, Pijoan JI, Vieta E, Pérez-Solà V, Kessler RC, Haro JM, Alonso J, Alonso M, Álvarez-Villalba M, Amann B, Amigo FF, Anmella G, Aragón A, Aragonés N, Aragonès E, Arizón AI, Asunsolo A, Ayora A, Ballester L, Barbas P, Basora J, Bereciartua E, Bravo I, Bolibar I, Bonfill X, Cotillas-Rodero A, Cuartero A, De Paz C, Del Yerro MJ, De Vocht J, Diaz D, Domingo JL, Empananza JI, Espallargues M, Espuga M, Estevan-Burdeus P, Fernández MI, Fernández T, Ferreres Y, Fico G, Forjaz MJ, García-Barranco R, García-Ribera C, García-Torrecillas JM, Garrido-Barral A, Gil E, Giola-Insigna M, Gómez M, Gómez J, González-Pinto A, Hernando M, Iriberrí M, Jansen L, Jansen L, Jiménez X, Larrauri A, León-Vázquez F, López-Atanes M, López-Fresneña N, López-Rodríguez C, López-Rodríguez JA, López-Cortacans G, Marcos A,

- Martín J, Martín V, Martínez-Cortés M, Martínez-Martínez R, De Salazar ADM, Martínez I, Marzola M, Mata N, Molina JM, Molina JD, Molinero E, Muñoz-Ruipérez C, Murru A, Navarro L, Olmedo-Galindo J, Ortí-Lucas RM, Padrós R, Pallejà M, Parra R, Pascual J, Pelayo-Terán JM, Pla R, Plana N, Pérez-Aznar C, Pérez-Gómez B, Pérez-Zapata A, Polentinos-Castro E, Puértolas B, Puig MT, Quílez Á, Quintana MJ, Quiroga A, Rentero D, Rey C, Rius C, Rodríguez-Blázquez C, Rojas-Giraldo MJ, Romero-Barzola Y, Rubio G, Ruiz P, Rumayor M, Sáenz M, Sánchez J, Sánchez-Arcilla I, Sanz F, Serra C, Serra-Sutton V, Serrano M, Solá S, Solera S, Soto M, Tarragó A, Tolosa N, Vázquez M, Viciola M, Voorspoels W, Yago-González S, Yáñez-Sánchez J, Zapico Y, Zorita LM, Zorrilla I and Zurbano SL (2021a) Thirty-day suicidal thoughts and behaviours in the Spanish adult general population during the first wave of the Spain COVID-19 pandemic. *Epidemiology and Psychiatric Sciences*, 30, e19. doi:10.1017/S2045796021000093
- Mortier P, Vilagut G, Ferrer M, Serra C, Molina JD, López-Fresneña N, Puig T, Pelayo-Terán JM, Pijoan JI, Emparanza JI, Espuga M, Plana N, González-Pinto A, Ortí-Lucas RM, Salazar AM, Rius C, Aragonès E, Cura-González I, Aragón-Peña A, Campos M, Parellada M, Pérez-Zapata A, Forjaz MJ, Sanz F, Haro JM, Vieta E, Pérez-Solá V, Kessler RC, Bruffaerts R and Alonso J (2021b) Thirty-day suicidal thoughts and behaviors among hospital workers during the first wave of the Spain COVID-19 outbreak. *Depression and Anxiety* 38, 528–544. doi:10.1002/da.23129
- Mortier P, Vilagut G, García-Mieres H, Alayo I, Ferrer M, Amigo F, Aragonès E, Aragón-Peña A, Asúnsolo Del Barco Á, Campos M, Espuga M, González-Pinto A, Haro JM, López Fresneña N, Martínez de Salazar AD, Molina JD, Ortí-Lucas RM, Parellada M, Pelayo-Terán JM, Pérez-Gómez B, Pérez-Zapata A, Pijoan JI, Plana N, Polentinos-Castro E, Portillo-Van Diest A, Puig T, Rius C, Sanz F, Serra C, Urreta-Barallobre I, Kessler RC, Bruffaerts R, Vieta E, Pérez-Solá V and Alonso J (2024) Health service and psychotropic medication use for mental health conditions among healthcare workers active during the Spain Covid-19 Pandemic – a prospective cohort study using web-based surveys. *Psychiatry Research* 334, 115800. doi:10.1016/j.psychres.2024.115800
- Murata S, Rezeppa T, Thoma B, Marengo L, Krancevich K, Chiyka E, Hayes B, Goodfriend E, Deal M, Zhong Y, Brummit B, Courty T, Riston S, Brent DA and Melhem NM (2021) The psychiatric sequelae of the COVID-19 pandemic in adolescents, adults, and health care workers. *Depression and Anxiety* 38, 233–246. doi:10.1002/da.23120
- Navarro MC, Ouellet-Morin I, Geoffroy M-C, Boivin M, Tremblay RE, Côté SM and Orri M (2021) Machine learning assessment of early life factors predicting suicide attempt in adolescence or young adulthood. *JAMA Network Open* 4, e211450. doi:10.1001/jamanetworkopen.2021.1450
- Nock MK, Millner AJ, Ross EL, Kennedy CJ, Al-Suwaidi M, Barak-Corren Y, Castro VM, Castro-Ramirez F, Lauricella T, Murman N, Petukhova M, Bird SA, Reis B, Smoller JW and Kessler RC (2022) Prediction of suicide attempts using clinician assessment, patient self-report, and electronic health records. *JAMA Netw Open* 5, e2144373. doi:10.1001/jamanetworkopen.2021.44373
- Nock MK, Stein MB, Heeringa SG, Ursano RJ, Colpe LJ, Fullerton CS, Hwang I, Naifeh JA, Sampson NA, Schoenbaum M, Zaslavsky AM and Kessler RC (2014) Prevalence and correlates of suicidal behavior among soldiers. *JAMA Psychiatry* 71, 514. doi:10.1001/jamapsychiatry.2014.30
- Nordin N, Zainol Z, Mohd Noor MH and Chan LF (2023) An explainable predictive model for suicide attempt risk using an ensemble learning and Shapley Additive Explanations (SHAP) approach. *Asian J Psychiatry* 79, 103316. doi:10.1016/j.ajp.2022.103316
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Müller A, Nothman J, Louppe G, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M and Duchesnay É. 2012. Scikit-learn: machine Learning in Python.
- Pellegrini L, Maietti E, Rucci P, Burato S, Menchetti M, Berardi D, Maina G, Fineberg NA and Albert U (2021) Suicidality in patients with obsessive-compulsive and related disorders (OCRDs): a meta-analysis. *Comprehensive Psychiatry* 108, 152246. doi:10.1016/j.comppsy.2021.152246
- Petrie K, Crawford J, Baker STE, Dean K, Robinson J, Veness BG, Randall J, McGorry P, Christensen H and Harvey SB (2019) Interventions to reduce symptoms of common mental disorders and suicidal ideation in physicians: a systematic review and meta-analysis. *The Lancet Psychiatry* 6, 225–234. doi:10.1016/S2215-0366(18)30509-1
- Portillo-Van Diest A, Vilagut G, Alayo I, Ferrer M, Amigo F, Amann BL, Aragón-Peña A, Aragonès E, Asúnsolo Del Barco Á, Campos M, Del Cura-González I, Espuga M, González-Pinto A, Haro JM, Larrauri A, López-Fresneña N, Martínez de Salazar A, Molina JD, Ortí-Lucas RM, Parellada M, Pelayo-Terán JM, Pérez-Zapata A, Pijoan JI, Plana N, Puig T, Rius C, Rodríguez-Blázquez C, Sanz F, Serra C, Urreta-Barallobre I, Kessler RC, Bruffaerts R, Vieta E, Pérez-Solá V, Alonso J and Mortier P (2023) Traumatic stress symptoms among Spanish healthcare workers during the COVID-19 pandemic: a prospective study. *Epidemiology and Psychiatric Sciences* 32, e50. doi:10.1017/S2045796023000628
- Posner K, Brown GK, Stanley B, Brent DA, Yershova KV, Oquendo MA, Currier GW, Melvin GA, Greenhill L, Shen S and Mann JJ (2011) The Columbia–Suicide Severity Rating Scale: initial validity and internal consistency findings from three multisite studies with adolescents and adults. *American Journal of Psychiatry* 168, 1266–1277. doi:10.1176/appi.ajp.2011.10111704
- Rezvani S and Wang X (2023) A broad review on class imbalance learning techniques. *Applied Soft Comput* 143, 110415. doi:10.1016/j.asoc.2023.110415
- Ribeiro JD, Franklin JC, Fox KR, Bentley KH, Kleiman EM, Chang BP and Nock MK (2016) Self-injurious thoughts and behaviors as risk factors for future suicide ideation, attempts, and death: a meta-analysis of longitudinal studies. *Psychological Medicine* 46, 225–236. doi:10.1017/S0033291715001804
- Rogoža D, Strumila R, Klivickaitė E, Diržius E and Čėnaitė N (2021) Depressive symptoms, help-seeking, and barriers to mental healthcare among healthcare professionals in Lithuania. *Acta Medica Lituanica* 28, 59–76. doi:10.15388/Amed.2020.28.1.3
- Sahimi HMS, Mohd Daud TI, Chan LF, Shah SA, Rahman FHA and Nik Jaafar NR (2021) Depression and suicidal ideation in a sample of Malaysian healthcare workers: a preliminary study during the COVID-19 pandemic. *Frontiers in Psychiatry*, 12, 658174. doi:10.3389/fpsy.2021.658174
- Saito T and Rehmsmeier M (2015) The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS One* 10, e0118432. doi:10.1371/journal.pone.0118432
- Sangha O, Stucki G, Liang MH, Fossel AH and Katz JN (2003) The self-administered comorbidity questionnaire: a new method to assess comorbidity for clinical and health services research. *Arthritis Care Res (Hoboken)* 49, 156–163. doi:10.1002/art.10993
- Schafer KM, Kennedy G, Gallyer A and Resnik P (2021) A direct comparison of theory-driven and machine learning prediction of suicide: a meta-analysis. *PLoS One* 16, e0249833. doi:10.1371/journal.pone.0249833
- Schrijvers DL, Bollen J and Sabbe BGC (2012) The gender paradox in suicidal behavior and its impact on the suicidal process. *Journal of Affective Disorders* 138, 19–26. doi:10.1016/j.jad.2011.03.050
- Seki T, Kawazoe Y and Ohe K (2021) Machine learning-based prediction of in-hospital mortality using admission laboratory data: a retrospective, single-site study using electronic health record data. *PLoS One* 16, e0246640. doi:10.1371/journal.pone.0246640
- Sheehan DV, Harnett-Sheehan K and Raj BA (1996) The measurement of disability. *International Clinical Psychopharmacology* 11, 89–95. doi:10.1097/00004850-199606003-00015
- Sohn MN, Dimitropoulos G, Ramirez A, McPherson C, Anderson A, Munir A, Patten SB, McGirr A and Devoe DJ (2023) Non-suicidal self-injury, suicidal thoughts and behaviors in individuals with an eating disorder relative to healthy and psychiatric controls: a systematic review and meta-analysis. *International Journal of Eating Disorders* 56, 501–515. doi:10.1002/eat.23880
- Somé NH, Noormohammadpour P and Lange S (2024) The use of machine learning on administrative and survey data to predict suicidal thoughts and behaviors: a systematic review. *Frontiers in Psychiatry* 15, 1291362. doi:10.3389/fpsy.2024.1291362

- Spitzer RL, Kroenke K, Williams JBW and Löwe B** (2006) A brief measure for assessing generalized anxiety disorder. *Archives of Internal Medicine* **166**, 1092. doi:[10.1001/archinte.166.10.1092](https://doi.org/10.1001/archinte.166.10.1092)
- van Buuren S** (2018). *Flexible Imputation of Missing Data*, (2nd edn.) Chapman and Hall/CRC. doi:[10.1201/9780429492259](https://doi.org/10.1201/9780429492259)
- White IR, Royston P and Wood AM** (2011) Multiple imputation using chained equations: issues and guidance for practice. *Statistics in Medicine* **30**, 377–399. doi:[10.1002/sim.4067](https://doi.org/10.1002/sim.4067)
- Wood AM, White IR and Royston P** (2008) How should variable selection be performed with multiply imputed data? *Statistics in Medicine* **27**, 3227–3246. doi:[10.1002/sim.3177](https://doi.org/10.1002/sim.3177)
- World Health Organization** (2024) <https://www.who.int/news-room/factsheets/detail/suicide> [World Health Organization. Suicide data.].
- Xiaoming X, Ming A, Su H, Wo W, Jianmei C, Qi Z, Hua H, Xuemei L, Lixia W, Jun C, Lei S, Zhen L, Lian D, Jing L, Handan Y, Haitang Q, Xiaoting H, Xiaorong C, Ran C, Qinghua L, Xinyu Z, Jian T, Jing T, Guanghua J, Zhiqin H, Nkundimana B and Li K** (2020) The psychological status of 8817 hospital workers during COVID-19 epidemic: a cross-sectional study in Chongqing. *Journal of Affective Disorders* **276**, 555–561. doi:[10.1016/j.jad.2020.07.092](https://doi.org/10.1016/j.jad.2020.07.092)
- Xu X, Wang W, Chen J, Ai M, Shi L, Wang L, Hong S, Zhang Q, Hu H, Li X, Cao J, Lv Z, Du L, Li J, Yang H, He X, Chen X, Chen R, Luo Q, Zhou X, Tan J, Tu J, Jiang G, Han Z and Kuang L** (2021) Suicidal and self-harm ideation among Chinese hospital staff during the COVID-19 pandemic: prevalence and correlates. *Psychiatry Research* **296**, 113654. doi:[10.1016/j.psychres.2020.113654](https://doi.org/10.1016/j.psychres.2020.113654)
- Yarkoni T and Westfall J** (2017) Choosing prediction over explanation in psychology: lessons from machine learning. *Perspectives on Psychological Science* **12**, 1100–1122. doi:[10.1177/1745691617693393](https://doi.org/10.1177/1745691617693393)
- Zhang Y, Wang J, Xiong X, Jian Q, Zhang L, Xiang M, Zhou B and Zou Z** (2022) Suicidality in patients with primary diagnosis of panic disorder: a single-rate meta-analysis and systematic review. *Journal of Affective Disorders* **300**, 27–33. doi:[10.1016/j.jad.2021.12.075](https://doi.org/10.1016/j.jad.2021.12.075)
- Zhao Y and Long Q** (2017) Variable selection in the presence of missing data: imputation-based methods. *WIREs Computational Statistics* **9**, e1402. doi:[10.1002/wics.1402](https://doi.org/10.1002/wics.1402)
- Zhou Y, Wang W, Sun Y, Qian W, Liu Z, Wang R, Qi L, Yang J, Song X, Zhou X, Zeng L, Liu T, Li Z and Zhang X** (2020) The prevalence and risk factors of psychological disturbances of frontline medical staff in China under the COVID-19 epidemic: workload should be concerned. *Journal of Affective Disorders* **277**, 510–514. doi:[10.1016/j.jad.2020.08.059](https://doi.org/10.1016/j.jad.2020.08.059)
- Zuromski KL, Ustun B, Hwang I, Keane TM, Marx BP, Stein MB, Ursano RJ and Kessler RC** (2019) Developing an optimal short-form of the PTSD Checklist for DSM-5 (PCL-5). *Depression and Anxiety* **36**, 790–800. doi:[10.1002/da.22942](https://doi.org/10.1002/da.22942)