


BMJ Open Catalonia Suicide Risk Code Epidemiology (CSRC-Epi) study: protocol for a population-representative nested case-control study of suicide attempts in Catalonia, Spain

Philippe Mortier ^{1,2}, Gemma Vilagut,^{1,2} Beatriz Puértolas Gracia,^{1,2} Ana De Inés Trujillo,^{1,3} Itxaso Alayo Bueno,^{1,2} Laura Ballester Coma,^{1,2,4} María Jesús Blasco Cubedo,^{1,2,5} Narcís Cardoner,^{6,7,8,9} Cristina Colls,¹⁰ Matilde Elices,^{8,11} Anna Garcia-Altes,^{2,10,12} Manel Gené Badia,¹³ Javier Gómez Sánchez,¹ Mario Martín Sánchez,¹⁴ Rosa Morros,^{15,16,17} Bibiana Prat Pubill,¹⁸ Ping Qin,¹⁹ Lars Mehlum,¹⁹ Ronald C Kessler,²⁰ Diego Palao,^{6,7,8,9} Víctor Pérez Sola,^{7,8,11,21} Jordi Alonso,^{1,2,5} On behalf of the CODIRISC Epidemiology Study Group

To cite: Mortier P, Vilagut G, Puértolas Gracia B, *et al*. Catalonia Suicide Risk Code Epidemiology (CSRC-Epi) study: protocol for a population-representative nested case-control study of suicide attempts in Catalonia, Spain. *BMJ Open* 2020;**10**:e037365. doi:10.1136/bmjopen-2020-037365

► Prepublication history for this paper is available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2020-037365>).

Received 30 January 2020
Revised 02 May 2020
Accepted 29 May 2020



© Author(s) (or their employer(s)) 2020. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to
Dr Philippe Mortier;
pmortier@imim.es

ABSTRACT

Introduction Suicide attempts represent an important public health burden. Centralised electronic health record (EHR) systems have high potential to provide suicide attempt surveillance, to inform public health action aimed at reducing risk for suicide attempt in the population, and to provide data-driven clinical decision support for suicide risk assessment across healthcare settings. To exploit this potential, we designed the Catalonia Suicide Risk Code Epidemiology (CSRC-Epi) study. Using centralised EHR data from the entire public healthcare system of Catalonia, Spain, the CSRC-Epi study aims to estimate reliable suicide attempt incidence rates, identify suicide attempt risk factors and develop validated suicide attempt risk prediction tools.

Methods and analysis The CSRC-Epi study is registry-based study, specifically, a two-stage exposure-enriched nested case-control study of suicide attempts during the period 2014–2019 in Catalonia, Spain. The primary study outcome consists of first and repeat attempts during the observation period. Cases will come from a case register linked to a suicide attempt surveillance programme, which offers in-depth psychiatric evaluations to all Catalan residents who present to clinical care with any suspected risk for suicide. Predictor variables will come from centralised EHR systems representing all relevant healthcare settings. The study's sampling frame will be constructed using population-representative administrative lists of Catalan residents. Inverse probability weights will restore representativeness of the original population. Analysis will include the calculation of age-standardised and sex-standardised suicide attempt incidence rates. Logistic regression will identify suicide attempt risk factors on the individual level (ie, relative risk) and the population level (ie, population attributable risk proportions). Machine learning techniques will be used to develop suicide attempt risk prediction tools.

Strengths and limitations of this study

- Suicide attempt cases (estimated n~6000) in the Catalonia Suicide Risk Code case registry are identified through in-depth psychiatric evaluation, which allows to carefully differentiate between suicidal and non-suicidal self-injurious behaviour.
- A wide range of predictor variables will be included, taken from centralised electronic health record data representing five clinical settings, that is, emergency care, primary care, outpatient mental healthcare, and general and psychiatric hospitalisations.
- A two-stage exposure-enriched nested case-control study combined with the use of inverse probability weights will enable efficient and population-representative estimations.
- An unknown proportion of suicide attempt cases do not contact healthcare services, and are therefore not included in this study.
- Limited information on history of suicide attempt before the study observation period will be available.

Ethics and dissemination This protocol is approved by the Parc de Salut Mar Clinical Research Ethics Committee (2017/7431/I). Dissemination will include peer-reviewed scientific publications, scientific reports for hospital and government authorities, and updated clinical guidelines.
Trial registration number NCT04235127.

INTRODUCTION

Suicide attempts constitute a major public health issue worldwide, despite the fact that prevention strategies have shown to be effective in reducing attempt rates.¹

Population-based surveys estimate the lifetime prevalence of suicide attempts among adults at 2.7% (range 0.5%–5.0%),² while a recent meta-analysis among children and adolescents found a pooled estimate of 6.0% (range 0.5%–34.1%) for suicide attempts in early life.³ Suicide attempts are related to subsequent suicide,⁴ which has a worldwide mortality rate estimated at 11.6 per 100 000 person-years, representing an annual loss of 34.6 million years of life.⁵ Apart from death by suicide, suicide attempts are also markers for subsequent persistent physical and mental health issues, repeat suicide attempt, psychiatric hospitalisations, impaired academic performance, unemployment, partner abuse victimisation and perpetration, having children removed by social services, loneliness, relationship difficulties, impaired social functioning and low life satisfaction.^{6–11}

Despite the considerable burden that suicide attempts represent in our society, there is a lack of reliable surveillance data on suicide attempts that could inform public health action.¹² This is in contrast with actual suicide rates, that are increasingly monitored in many countries worldwide.¹³ WHO advocates the use of centralised electronic health record (EHR) systems to develop national suicide attempt surveillance.¹⁴ However, currently used disease classification systems in EHR systems (eg, the International Classification of Diseases¹⁵) do not allow to distinguish between suicidal and non-suicidal intent of self-injurious behaviour.¹⁶ In addition, due to the often difficult ascertainment of self-injurious and suicidal intent, misclassification with regard to suicidal outcomes often occurs.¹⁷ Offering an in-depth psychiatric evaluation to each individual who presents to clinical care with suspected suicide risk, followed by a standardised registration of this clinical evaluation in a centralised EHR register, may therefore substantially improve the accuracy of public health surveillance of suicide attempts.

Apart from surveillance, centralised EHR systems have high potential to be used in epidemiological studies on suicidal behaviour in the population.^{18–20} Indeed, it has been estimated that up to 92% of individuals that eventually die of suicide have some type of healthcare contact in the year prior to death,²¹ with rates ranging from 54% to 80% for primary care contacts,^{21 22} 31% to 66% for mental healthcare contacts,^{21 22} 24% to 60% for emergency department visits^{21 23–25} and 21% for psychiatric hospitalisations.²¹ Clinical data collected through centralised EHR systems could therefore provide new insight in the population distribution of suicide attempt risk factors, outline the different healthcare trajectories preceding an attempt and provide estimates of potential reductions in suicide attempt cases when designing prevention interventions. The need for nested case–control studies using EHR data to investigate suicidal behaviour has been recently highlighted.²⁶ Developments in statistical methods, including the use of inverse probability weighting in case–control studies²⁷ as well as two-stage (exposure-enriched) case–control designs,²⁸ now allow to

study rare events such as suicidal behaviour in an efficient and population-representative way.

From a clinical viewpoint, the use of centralised EHR data has high potential to provide data-driven clinical decision support when evaluating risk for future suicide attempts.²⁹ Such support is highly needed, as suicide attempts constitute a complex behavioural outcome, determined by highly multifactorial population processes, interdependencies and multilevel causality.³⁰ Clinicians are prone to heuristic-based decision making,³¹ that is, a rapid problem-solving approach, including subconscious cognitive shortcuts, linking a limited set of risk factors directly to suicide attempt potential. This leads to failure to detect real suicide potential, poor patient experience and ineffective clinical decision making. In recent years, a number of studies started implementing advanced analytical techniques on EHR data—including machine learning techniques—to model the complex additive and interactive effects between large numbers of predictor variables, and to improve the classification accuracy of data-driven suicide attempt risk prediction tools.³² When routinely implemented at the healthcare system level, such data-driven decision support tools could guide the adequate allocation of clinical resources, such as in-depth suicide risk assessments and tailored treatment interventions in multistage screening approaches.^{33 34}

Here, we present the protocol for the Catalonia Suicide Risk Code Epidemiology (CSRC-Epi) study, a large epidemiological study of suicide attempts occurring during the period 2014–2019 in Catalonia, Spain. The primary study outcome is suicide attempt among Catalan residents during the period 2014–2019. Secondary analyses will focus on actual suicide among those with previous suicide attempts. The CSRC study combines a nested case–control sampling design with the use of inverse probability weighting to enable the efficient analysis of a large amount of centralised EHR data. A unique aspect of the study is that data for the suicide attempt cases come from an especially designed suicide attempt surveillance protocol that stipulates that every Catalan resident presenting to clinical care with any suspected risk for suicide receives an in-depth psychiatric evaluation. These evaluations will allow us to differentiate carefully between suicidal and non-suicidal self-injurious behaviour in the study outcomes.

Study objectives

The CSRC-Epi study main objectives are:

- ▶ To provide reliable incidence measures for suicide attempts occurring in Catalonia during the period 2014–2019.
- ▶ To identify risk factor constellations for suicide attempts, both on the individual level (ie, the extent to which risk factors increase the risk for subsequent suicide attempt in an individual) and on the population level (ie, the proportion of the total cases of suicide attempt that are potentially attributable to risk factors).

- ▶ To develop suicide attempt risk prediction tools using machine learning techniques, and test their predictive accuracy.

METHODS AND ANALYSIS

Data sources

All data for this study will be obtained from the Health Evaluation and Quality Agency of Catalonia (AQuAS³⁵), a public entity attached to the Catalan Health Department. As from 2017, AQuAS manages the Public Data Analysis for Health Research and Innovation Programme (PADRIS³⁵) to provide researchers with access to large amounts of centralised EHR data, and to foster innovative health research.

Administrative data

A first source of data for the CSRC-Epi study will consist of six population-representative administrative lists of Catalan residents, one for each year in the 2014–2019 period. These lists constitute annual censuses of individuals with access to public healthcare, which, by law, includes every Catalan resident. These lists include socio-demographic variables (ie, sex, age, nationality, a range of socioeconomic indicators and healthcare catchment region) as well as associated small area geocode data (ie, data available on the healthcare catchment region level),³³ the dates of immigration and emigration in/out of Catalonia, the date of death, as well as a range of healthcare summary variables that are constructed to monitor healthcare needs in the Catalan population (ie, 12-month depression, 12-month complex mental disease and the number of 12-month healthcare contacts for each healthcare setting). These lists will be used to construct a sampling frame when conducting the nested case–control sampling, as further explained below.

CSRC case register

In 2014, the Catalan Health Department and the Catalan Health Service structurally implemented the CSRC surveillance programme³⁶ in the Catalan Public Healthcare system. The CSRC surveillance programme is a specifically designed suicide attempt surveillance protocol that stipulates that every Catalan resident presenting with any suspected risk for suicide in any public healthcare setting receives a face-to-face in-depth psychiatric evaluation at the nearest emergency department. This assessment includes differentiating suicide attempts from non-suicidal self-injurious behaviour, or from adverse mental health states without self-injurious intention. Each individual deemed at high risk for (repeat) suicide attempt is subsequently eligible for two brief follow-up interventions (ie, a mental healthcare visit within 10 days (within 72 hours when aged 17 or less), and a phone call after 30 days) to increase access to adequate mental healthcare use. Clinical data of all individuals that received a specialised assessment are registered centrally in the CSRC case register.

The CSRC case register subsequently includes all individuals with a suicide attempt during the observation period 2014–2019, including the exact date of event. For each event included in the SRC case register, a range of predictor variables for future suicidal behaviour are assessed: the Suicidal Scale of the Mini-International Neuropsychiatric Interview (MINI) 5.0.0,^{37 38} the type and lethality of the suicide attempt that warranted evaluation, presence and type of mental disorder, hopelessness, impulsivity, aggressiveness, altered state of conscience, use or dependence of alcohol, use or dependence of illicit drugs, serious somatic disease (including chronic diseases, chronic pain and disabilities), living status, presence of family or social support, social problems, stressful life events, access to lethal means and family history of suicide. These variables will be used as predictor variables in analyses predicting repetition of suicide attempt and suicide after a previous attempt.

EHR data

A third source of data will consist of centralised EHR registers, one for each of five clinical healthcare settings, that is, emergency care, primary care, outpatient mental healthcare and general and psychiatric hospitalisations. These registers include a wide range of relevant predictor variables for suicidal behaviour, that is, history of self-injurious behaviours; all types of somatic conditions; neurodevelopmental, mental, behavioural, personality and substance use disorders; all types of medical procedures performed; and detailed information on the number and type of healthcare contacts. Diagnoses and procedures in the EHR data are coded using the International Classification of Diseases-9th revision-Clinical Modification (ICD-9-CM) and ICD-10-CM disease classification system. The year of inception of the different registers is 2012 for emergency care and primary care and 2008 for the other registers.

Pharmaceutical register

A fourth source of data will consist of a register containing all prescription drugs that have been delivered by officially recognised pharmacies, including the date of delivery. Note that this excludes prescribed medication that was not collected at the pharmacy. This register will provide an additional range of predictor variables for suicidal behaviour, that is, all prescriptions for psychopharmacological products, as well as prescriptions for a wide range of medication used to treat relevant somatic conditions or known to have psychotropic effects.

Mortality register

Suicide cases among those with a suicide attempt during the study observation period will be identified using data from the mortality register, managed by the Catalan Department of Forensic Medicine, which provides detailed data on causes of death using the ICD system (9th or 10th revision). State-of-the art forensic techniques, including psychological autopsy by a multidisciplinary team, is used

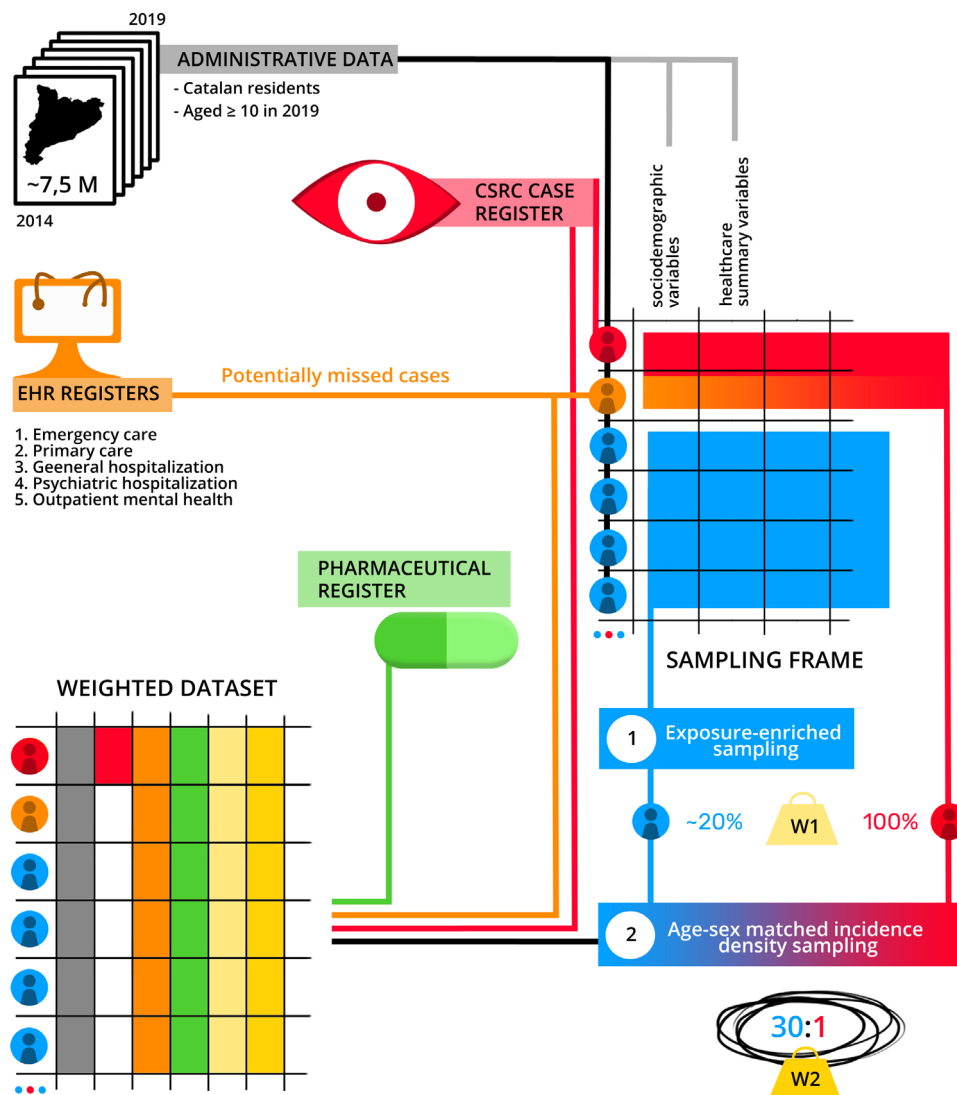


Figure 1 Overview of the CSRC-Epi study design. CSRC-Epi, Catalonia Suicide Risk Code Epidemiology; EHR, Electronic healthcare record; M, million; W, weight.

to determine death by suicide in the mortality register.³⁹ Nevertheless, forensic examination of suicidal intention is difficult, and misclassification may occur.

Study design

Figure 1 shows an overview of the CSRC-Epi study design. The CSRC-Epi study is a register-based study, that is, a study that uses the exposure and outcome data from registries,⁴⁰ which in turn, are representative for the target population. The target population consists of the dynamic cohort of all Catalan residents during the 6-year period, 2014–2019. As explained in detail below, we will conduct a two-stage nested case–control study within this dynamic cohort,^{41 42} which, in combination with the use of inverse probability weights, will allow us to construct a dataset representative for the original cohort of Catalan residents, and analyse the data accordingly.

Annual total population of Catalonia is between 7.5 and 7.6 million, with annual rates for immigration, emigration, birth and death being ~2.5%, ~1.9%, ~0.9% and

~0.8%, respectively.³⁹ Based on these figures, we expect a maximum of ~9.1 million individuals with Catalan residency status on at least one point in time over the 2014–2019 period. However, we expect suicide attempts in Catalonia to be extremely rare before the age of 10, in line with findings that self-injurious behaviour generally occurs as from the adolescent period.⁴³ We will, therefore, exclude cases and controls that have not reached age 10 by the end of the 2014–2019 period (ie, ~10.9%), lowering the total expected target population to ~8.1 million.

Case selection

Suicide attempt cases between 2014 and 2019 will be identified using the CSRC case register. Based on preliminary data exploration, we expect to include ~6000 cases of clinically confirmed suicide attempt by the end of 2019, of which ~8% (~480) will be repeat attempters. This substantially exceeds the number of suicide attempt cases included in previous register-based studies (median=1562, IQR=1562–3250³²).

Table 1 ICD codes to identify potentially missed cases of suicide attempt

	ICD-9-CM	Description	ICD-10-CM	Description
Suicide attempts and intentional self-injurious behaviour	E950*-E959*	Self-poisoning, hanging, strangulation, suffocation, fire arms, jumping, others	X71*-X83*	Drowning and submersion, firearms, explosive or thermal material, sharp or blunt objects, jumping from a high place, jumping or lying in front of a moving object, crashing of motor vehicle and other specified means
			T36*-T50* with a 5/6th character of 2	Drug poisoning (overdose)
			T51*-T65* with a 5/6th character of 2	Toxic effects of nonmedicinal substances
			T71* with a sixth character of 2	Asphyxiation, suffocation, strangulation
	T14.91	Suicide attempt		
Self-injurious behaviour of undetermined intent (intentional/accidental)	E980*-E989*	Self-poisoning, hanging, strangulation, suffocation, fire arms, jumping, others (undetermined intent)		
	994.7	Asphyxiation/strangulation	T71* with a sixth character of 4	Asphyxiation, suffocation, strangulation
	881*, 903.2, 903.3, 903.4	Open wound of elbow, forearm, and wrist; injury radial/ulnar vessels; injury palmar artery	S51.001-S51.009, S51.801-S51.809, S55.0-S55.199, S61.5-S61.509, S65.0-S65.199	Open wound of elbow/forearm/wrist, injury of ulnar/radial artery at forearm/wrist/arm level
	965*, 967*, 969*	Poisoning by analgesics, antipyretics, antirheumatics, sedatives and hypnotics	T36*-T50* with a 5/6th character of 4	Drug poisoning (overdose)
		T51*-T65* with a 5/6th character of 4	Toxic effects of non-medicinal substances	

ICD-9-CM, International Classification of Diseases-9th revision-Clinical Modification; ICD-10-CM, International Classification of Diseases-10th revision-Clinical Modification.

One of the main objectives of the CSRC programme is to enable reliable surveillance of suicide attempt in the population, and to tackle under-registration and misclassification using regular EHR systems.¹⁶ Nevertheless, failure to adhere to the CSRC programme protocol may result in an unknown number of suicide attempt cases that remain undetected. Therefore, ICD disease classification codes in the five centralised EHR registers will be inspected to identify potentially missed cases of suicide attempt. For that purpose, a wide range of ICD codes related to suicide attempt (see [table 1](#)) was identified through an extensive MEDLINE search, including a recent overview article with recommendations on the use of ICD codes for the surveillance of self-injurious behaviour.¹⁶ ICD codes are unable to determine suicidal intent, but do allow to identify subjects with intentional self-injurious behaviour and to differentiate them from subjects with self-injurious behaviour of undetermined intent (ie, intentional or accidental self-injurious behaviour).¹⁶ Outcome definition algorithms (ie, predefined sets of ICD codes) have shown promising in increasing the accuracy of suicide

attempt case detection using ICD codes.¹⁷ Therefore, we intent to validate the range of ICD codes we identified against golden standard identification methods (ie, manual review as well as text mining of clinical notes) to increase the accurate detection of potentially missed cases for our study.

Control selection

In a first stage, we will select a 20% stratified random sample of the 2014–2019 dynamic cohort members, using the six population-representative administrative lists of Catalan residents described above. Constructing this preliminary 20% subsample is necessary for two reasons: (1) we need to reduce the amount of data AQuAS will need to handle when conducting the age-matched and sex-matched incidence density sampling in the second stage; and (2) based on publicly available data, we estimate the probability of selecting controls with 12-month healthcare contacts for mental disorders to be relatively low (ie, ranging from ~17% for primary care visits to ~0.1% for general hospitalisations). Therefore, in order

to enrich the data for relevant exposure information, controls in the 20% subsample will be oversampled for number and specific types of healthcare use, using the past year healthcare summary variables available in the administrative lists. This will result in a higher number of controls with (mental) healthcare diagnoses eligible in the second stage.

In a second stage, we will create 67 risk sets, one for each month in the 6-year observation period (ie, June 2014 to December 2019). Within each risk set, a number of 30 age-matched and sex-matched controls will be randomly selected for each case (ie, case or potentially missed case) without replacement (incidence density sampling or risk set sampling⁴¹). Eligible controls will include future cases and controls will be allowed to be selected multiple times across risk sets. To allow for the joint and separate analyses of suicide attempt and potential suicide attempt (see the Data analytical plan section), controls are selected for first suicide attempt within individuals (eligible controls including previous potential cases, but not previous suicide attempt cases), and if applicable, also for the first potential suicide attempt within individuals (eligible controls including those without previous suicide attempt or potential suicide attempt only). After the final selection of controls, inverse probability weights will be constructed to restore population-representativeness of the original dynamic population cohort. Weights will be equal to 1 for cases and potentially missed cases, (ie, all are selected in both sampling stages); for controls, weights will reflect the selection probabilities at stage 1 (including the oversampling according to healthcare summary variables), as well as at stage 2 (including the age–sex matching and the total time at risk of each control during the observation period).^{42 44}

A specific objective of the SRC-Epi project is the construction of suicide attempt risk prediction tools by healthcare setting. For this purpose, a separate series of controls will be selected at the second sampling stage, this time matching by age, sex as well as type and timing of last healthcare contact. For example, for a suicide attempt assessed at the emergency department at time *y* and a last healthcare visit at primary care at time *x*, 30 age-matched and sex-matched controls will be selected among those individuals that have not committed a suicide attempt up to time *y*, restricting to those controls with primary healthcare visits around time *x*.

Data analytical plan

The primary outcome of this study is suicide attempt during the period 2014–2019. We will both focus on first suicide attempt during the observation period, as well as on repetition of attempts, defined as suicide attempts among those with a previous suicide attempt during the observation period. As explained above, we will also identify potentially missed cases of suicide attempt, using the ICD codes identified in the literature (see [table 1](#)). Cases and potentially missed cases will be considered both as joint as well as separate outcomes in the analyses.

In addition, we will conduct separate analyses focusing on clinical severity of suicide attempts (eg, lethality and method of attempt). A secondary study outcome consists of suicide among those with a suicide attempt during 2014–2019. Cases of suicide will be identified through the mortality register, managed by the Catalan Department of Forensic Medicine (see Mortality register).

Suicide attempt occurrence

Population-representative annual incidence rates will be estimated by dividing annual number of cases by total annual sums of person-years at risk, multiplied by 100 000, using the weighted nested case–control dataset. We will calculate both crude as well as age-standardised and sex-standardised incidence rates, stratified by relevant sociodemographic variables (eg, socioeconomic status, healthcare region, etc). As incidence rates do not inform of the distribution of cases over time, we will also estimate and visualise incidence proportions (or cumulative incidence) over time using the Kaplan-Meier estimator to estimate one minus the survival function.

Suicide attempt risk factor associations

To estimate the individual-level associations between predictor variables and outcome variable, (conditional) bivariable and multivariable binomial logistic regression will be applied. For a first suicide attempt during the observation period as the outcome variable, the weighted nested case–control dataset will be used. In order for the OR to be a valid estimation of the HR (and hence, relative risk) multiple inputs for those controls selected multiple times and for those cases also selected as controls will be included in the analyses.⁴⁵ Time-varying predictor variables (eg, the exact dates of diagnoses or medical prescriptions) will be recoded by categorising time-to-event into discrete time intervals. Relevant time-to-event cut-offs for these intervals will be identified by examining the changes in odds across time-to-event for a high number of short time intervals. Given the high amount of predictor variables under study, the Least Absolute Shrinkage and Selection Operator⁴⁶ method will be employed as to select a subset of predictor variables that predict the outcome best while maintaining a good model fit. For suicide and for repetition of suicide attempt as the outcome variable, the analysis will be very similar, but will now reduce to a cohort analysis of those individuals with first suicide attempt during 2014–2019 (which are all selected), and regression models will also include the clinical variables from the CRSC case register obtained through in-depth psychiatric evaluation.

Population-level effect sizes will be estimated by calculating bivariable and multivariable population-attributable risk proportions (PARP⁴⁷), based on summary measures of individual predicted probabilities obtained from the logistic regression models described above, comparing original models versus models in which regression coefficients of the predictor variable(s) under study are set to zero. PARP subsequently provide an estimate of the

proportion of cases that could potentially be prevented if certain risk factor(s) in the population were to be eliminated, assuming a causal relationship between risk factor and outcome variable. PARP estimates are of high value for policy makers involved in suicide prevention, as they provide insight in the population-level impact of risk factor(s) with regard to suicide risk. Specifically, PARP take into account that high-prevalence risk factors carrying low individual risk may be equally or even more important to consider than low-prevalence risk factors carrying high risk for the affected individuals. Taking into account such knowledge is important, as it is the combination of both individual-level and population-level interventions that has shown to be successful in reducing adverse outcomes with complex multicausal aetiologies.³⁰

Suicide attempt risk predictions

Risk prediction tools for (repetition of) suicide attempt will be constructed using machine learning techniques. A first series of algorithms will focus on estimating suicide attempt risk for different prediction windows, that is, censoring data of both cases and controls at different time points relative to the time of event.⁴⁸ A second series of algorithms will predict suicide attempt after specific healthcare contacts using the separate series of controls selected in the second sampling stage (see Control selection). Machine learning techniques will include elastic net penalised logistic regression,⁴⁹ naïve Bayes classifiers,⁵⁰ multivariate adaptive regression splines,⁵¹ Bayesian additive regression trees,⁵² random forests,⁵³ gradient boosting,⁵⁴ k-nearest neighbour algorithms,⁵⁵ support vector machines⁵⁶ and artificial neural networks.⁵⁷ Stacking ensemble techniques (super learning)⁵⁸ will be implemented to further optimise prediction accuracy, by using the predictions from the above mentioned algorithms (the base learners) as input to train new models (meta learners). To avoid model overfit, model development and tuning will be conducted on a training dataset using k-fold cross-validation, and model predictive accuracy will be evaluated in a separate validation dataset using a recalibrated algorithm. Sample selection bias introduced by the two-stage nested case-control design will be addressed using appropriately corrected classifiers.⁵⁹ Model predictive accuracy will be evaluated by the area under the receiver operating characteristic curve, as well as accuracy measures calculated for different thresholds of the continuous predicted probability (ie, thresholds set to delineate the top x% at highest risk⁶⁰), including positive predictive values (precision, or the probability that a predicted case is actually a true case), sensitivity values (recall, or the proportion of true cases that has been predicted correctly), and F_1 -scores (ie, the harmonic mean of recall and precision).

Study limitations

A first major limitation of the CSRC-Epi study is that any history of suicide attempt before the study observation period is unknown for both cases and controls. However,

indirect information will be available, consisting of (1) the Suicidal Scale of the MINI (item 6) used in the CSRC programme protocol, which assesses lifetime history of suicide attempt among the cases. The timing of these previous attempts will be unknown, and this information will be based on patients' self-reported information and (2) the EHR registers, which include episodes of self-harm before 2014, among both cases and controls. It will, however, not be possible to determine the suicidal intent of these episodes, and EHR register data are only available as from 2008. A second main limitation of the CSRC-Epi study is that, although the entire Catalan population has free access to public healthcare, about 20% of the population opts for private coverage or uses both public and private healthcare systems. This limits the population-representativeness of the EHR data to an unknown extent. As mentioned above, a third limitation is that, due to variable adherence to the CSRC surveillance programme, suicide attempt cases may still go undetected. This will be countered by identifying potentially missed cases of suicide attempt in the EHR registers. Related to this, it should be acknowledged that an unknown proportion of suicide attempt cases do not contact healthcare services, and are therefore not included in this study. This limitation is inherent to studies using EHR data, and points to the need on complementing the knowledge that can be gained from registry-based studies with findings from general population epidemiological survey research.

Patient and public involvement

No patients involved.

ETHICS AND DISSEMINATION

The protocol of this study has been approved by the Parc de Salut Mar Clinical Research Ethics Committee (CEIC protocol 2017/7431/I). The study is in line with the principles established in the Declaration of Helsinki, with the Charter of Fundamental Rights of the European Union (2000/C 364/01), and the European Convention on Human Rights. All data for this study come from the PADRIS programme, and will involve processing of completely anonymised EHR data. For record linkage activities, the Spanish Order SAS/3470/2009 for data obtained in observational studies is followed. This is also in line with the Data Protection Directive of the European Union (Directive 95/46/EC).

We aim to create awareness of the proposed action in the general public, by providing comprehensive information on the need for detecting new suicide attempt risk factors constellations, on the need to improve suicide attempt risk estimation, and on the ongoing exploration of clinical decision support for the improved assessment of suicide risk. The following communication measures will be taken: (1) the design of a website providing clear and balanced information on the project; (2) balanced newspaper articles and interviews to the press and (3) providing all healthcare settings with patient folders on

the project, providing a clear and balanced summary of the project.

Communication with patients and with next of kin will be in lay terms only. Feasible formats are internet websites and patient forums, patient folders and carefully planned releases to the press. We will provide clear and balanced information on the project, acknowledging the unique experience of each patient and stressing our final aim of improving (not replacing) human clinical practice and care. Patient groups will also be involved in the design of the overall communication strategy (codesign).

Targeted expert audiences will consist of those involved in suicide research, as well as in general psychiatry, psychiatric epidemiology and translational psychiatry. Scientific publications will be sent to peer-reviewed journals through open-access publishing, and added to the Pompeu Fabra University's repository of open access articles.⁶¹ Further dissemination of results will be through scientific conferences and workshops. Hospital and government authorities involved with mental healthcare will be informed of the study results through scientific reports, which will present series of suicide intervention prevention frameworks. In addition, we will provide the Spanish and Catalan Department of Health with updated clinical guidelines for the assessment of suicide risk. Clinicians with mental healthcare expertise as well as emergency department clinicians and general practitioners will be informed of these recommendations through the project's website and the professional associations' websites, periodicals and meetings.

Author affiliations

- ¹Health Services Research Group, IMIM (Hospital del Mar Medical Research Institute), Barcelona, Spain
- ²CIBER Epidemiología y Salud Pública (CIBERESP), Madrid, Spain
- ³Department of Social Psychology, Autonomous University of Barcelona (UAB), Cerdanyola del Vallès, Barcelona, Spain
- ⁴Department of Psychology, University of Girona (UdG), Girona, Spain
- ⁵Department of Health & Experimental Sciences, Pompeu Fabra University (UPF), Barcelona, Spain
- ⁶Depression and Anxiety Program, Department of Mental Health, Parc Taulí Sabadell, Hospital Universitari, Sabadell, Spain
- ⁷Department of Psychiatry and Legal Medicine, Universitat Autònoma de Barcelona (UAB), Cerdanyola Del Vallès, Barcelona, Spain
- ⁸Centro de Investigación en Red de Salud Mental, CIBERSAM, Madrid, Spain
- ⁹Institut d'Investigació i Innovació Parc Taulí (I3PT), Sabadell, Barcelona, Spain
- ¹⁰Agència de Qualitat i Avaluació Sanitàries de Catalunya - Health Evaluation and Quality Agency of Catalonia (AQuAS), Catalan Health Department, Barcelona, Spain
- ¹¹Neurosciences Research Programme, IMIM (Hospital del Mar Medical Research Institute), Barcelona, Spain
- ¹²Institut d'Investigació Biomèdica (IIB Sant Pau), Barcelona, Spain
- ¹³Legal Medicine Unit, Faculty of Medicine, University of Barcelona, Barcelona, Spain
- ¹⁴Preventive Medicine and Public Health Training Unit PSMar-UPF-ASP, Parc de Salut Mar, Agència de Salut Pública de Barcelona, Pompeu Fabra University, Barcelona, Spain
- ¹⁵Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol), Barcelona, Spain
- ¹⁶Departament de Farmacologia, de Terapèutica i de Toxicologia, Universitat Autònoma de Barcelona, Barcelona, Spain
- ¹⁷Institut Català de la Salut (ICS), Metropolitana Nord, Barcelona, Spain
- ¹⁸Master Plan on Mental Health and Addictions, Ministry of Health, Catalan Government, Barcelona, Spain

¹⁹National Centre for Suicide Research and Prevention, Institute of Clinical Medicine, University of Oslo, Oslo, Norway

²⁰Department of Health Care Policy, Harvard Medical School, Boston, MA, USA

²¹Institut de Neuropsiquiatria i Addiccions, Hospital del Mar, Barcelona, Spain

Collaborators The *CODIRISC Epidemiology Study Group* are: Jordi Alonso, Itxaso Alayo Bueno, Laura Ballester Coma, Jordi Blanch, María Jesús Blasco Cubedo, Ana De Inés Trujillo, María Teresa Campillo Sanz, Narcís Cardoner, Anna Isabel Cebrià, Cristina Colls, Matilde Elices, Anna García-Altés, Ricard Gavalda, Manel Gené Badia, Javier Gómez Sánchez, Ronald C. Kessler, Lars Mehlum, Cristina Molina Parilla, Rosa Morros Pedrós, Philippe Mortier, Jordi Ortiz, Diego Palao, Rosa Maria Pérez Pérez, Víctor Pérez, Maria J. Portella, Bibiana Prat Pubill, Beatriz Puértolas Gracia, Raquel Suárez Pérez, Ping Qin, and Gemma Vilagut.

Contributors Initial draft of the protocol: PM, GV, BPG, IAB and JA. Critical review of the protocol: LM, PQ, and RCK. Initial draft of the manuscript: PM, GV, BPG, IAB, and JA. Critical review of the manuscript: PM, GV, BPG, ADIT, IAB, LBC, MJBC, NC, CC, ME, AG-A, MGB, JGS, MMS, RM, BPP, PQ, RCK, DP, VPS and JA. All authors read and approved the final manuscript.

Funding This project was supported by ISCIII/FEDER P17/00521, ISCIII/FEDER P17/01205, and Generalitat de Catalunya (2017 SGR 452). The Catalonia Suicide Risk Code surveillance programme is an initiative of the Mental Health and Addictions Plan of the Department of Health of the Catalan Government. PM has a Sara Borrell research contract awarded by the ISCIII (CD18/00049). ME has a Juan de la Cierva research contract awarded by the ISCIII (FJCI-2017-31738). VPS and ME want to thank unrestricted research funding from Secretaria d'Universitats i Recerca del Departament d'Economia i Coneixement (2017 SGR 134 to 'Mental Health Research Group'), and Generalitat de Catalunya (Government of Catalonia). BPG and ADIT received funding from ISCIII F18/00012 and FPU2017-06447, respectively. LBC received funding by Ministerio de Educación, Cultura y Deporte (FPU15/05728). DP and JA received funding by ISCIII/FEDER P17/01205.

Competing interests DP has received grants and also served as consultant or advisor for Angelini, Janssen, Lundbeck and Servier. The other authors have no competing interests to declare.

Patient and public involvement Patients and/or the public were involved in the design, or conduct, or reporting, or dissemination plans of this research. Refer to the Methods section for further details.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD

Philippe Mortier <http://orcid.org/0000-0003-2113-6241>

REFERENCES

- 1 Lewitzka U, Sauer C, Bauer M, *et al*. Are national suicide prevention programs effective? A comparison of 4 verum and 4 control countries over 30 years. *BMC Psychiatry* 2019;19:1–10.
- 2 Nock MK, Borges G, Bromet EJ, *et al*. Cross-National prevalence and risk factors for suicidal ideation, plans and attempts. *Br J Psychiatry* 2008;192:98–105.
- 3 Lim K-S, Wong CH, McIntyre RS, *et al*. Global lifetime and 12-month prevalence of suicidal behavior, deliberate self-harm and non-suicidal self-injury in children and adolescents between 1989 and 2018: a meta-analysis. *Int J Environ Res Public Health* 2019;16:ijerph16224581.
- 4 Ribeiro JD, Franklin JC, Fox KR, *et al*. Self-injurious thoughts and behaviors as risk factors for future suicide ideation, attempts, and death: a meta-analysis of longitudinal studies. *Psychol Med* 2016;46:225–36.
- 5 Naghavi M. Global burden of disease self-harm Collaborators. global, regional, and national burden of suicide mortality 1990 to 2016: systematic analysis for the global burden of disease study 2016. *BMJ* 2019;364:194.

- 6 Goldman-Mellor SJ, Caspi A, Harrington H, *et al.* Suicide attempt in young people: a signal for long-term health care and social needs. *JAMA Psychiatry* 2014;71:119–27.
- 7 Joffe BI, Van Lieshout RJ, Duncan L, *et al.* Suicidal ideation and behavior in adolescents aged 12–16 years: a 17-year follow-up. *Suicide Life Threat Behav* 2014;44:497–509.
- 8 Mortier P, Demyttenaere K, Auerbach RP, *et al.* The impact of lifetime suicidality on academic performance in college freshmen. *J Affect Disord* 2015;186:254–60.
- 9 De Luca SM, Franklin C, Yueqi Y, *et al.* The relationship between suicide ideation, behavioral health, and College academic performance. *Community Ment Health J* 2016;52:534–40.
- 10 Pajonk F-G, Ruchholtz S, Waydhas C, *et al.* Long-Term follow-up after severe suicide attempt by multiple blunt trauma. *Eur psychiatr* 2005;20:115–20.
- 11 Stanford S, Jones MP, Loxton DJ. Understanding women who self-harm: predictors and long-term outcomes in a longitudinal community sample. *Aust N Z J Psychiatry* 2017;51:151–60.
- 12 Bachmann S. Epidemiology of suicide and the psychiatric perspective. *Int J Environ Res Public Health* 2018;15:1425.
- 13 World Health Organisation. Who | suicide data. who website, 2019. Available: https://www.who.int/mental_health/prevention/suicide/suicideprevent/en/
- 14 WHO. *Practice manual for establishing and maintaining surveillance systems for suicide attempts and self-harm.* WHO, 2016.
- 15 World Health Organization. The International classification of diseases, 2019. Available: <https://www.who.int/classifications/icd/en/> [Accessed 31 Jul 2019].
- 16 Hedegaard H, Schoenbaum M, Claassen C, *et al.* Issues in developing a surveillance case definition for nonfatal suicide attempt and intentional self-harm using International classification of diseases, tenth revision, clinical modification (ICD-10-CM) coded data. *Natl Health Stat Report* 2018;1–19.
- 17 Swain RS, Taylor LG, Braver ER, *et al.* A systematic review of validated suicide outcome classification in observational studies. *Int J Epidemiol* 2019;48:1636–49.
- 18 Chute CG, Koo D. Public health, data Standards, and vocabulary: crucial infrastructure for reliable public health surveillance. *J Public Health Manag Pract* 2002;8:11–17.
- 19 Kukafka R, Ancker JS, Chan C, *et al.* Redesigning electronic health record systems to support public health. *J Biomed Inform* 2007;40:398–409.
- 20 Casey JA, Schwartz BS, Stewart WF, *et al.* Using electronic health records for population health research: a review of methods and applications. *Annu Rev Public Health* 2016;37:61–81.
- 21 Schaffer A, Sinyor M, Kurdyak P, *et al.* Population-Based analysis of health care contacts among suicide decedents: identifying opportunities for more targeted suicide prevention strategies. *World Psychiatry* 2016;15:135–45.
- 22 Stene-Larsen K, Reneflot A. Contact with primary and mental health care prior to suicide: a systematic review of the literature from 2000 to 2017. *Scand J Public Health* 2019;47:9–17.
- 23 Vasiladiadis H-M, Ngamini-Ngui A, Lesage A. Factors associated with suicide in the month following contact with different types of health services in Quebec. *Psychiatr Serv* 2015;66:121–6.
- 24 Morrison KB, Laing L. *Adults' use of health services in the year before death by suicide in Alberta.* *Heal Reports.* 22, 2011.
- 25 Ahmedani BK, Simon GE, Stewart C, *et al.* Health care contacts in the year before suicide death. *J Gen Intern Med* 2014;29:870–7.
- 26 Pirkis J, Nicholas A, Gunnell D. The case for case-control studies in the field of suicide prevention. *Epidemiol Psychiatr Sci* 2020;29:e62.
- 27 Kim RS. A new comparison of nested case-control and case-cohort designs and methods. *Eur J Epidemiol* 2015;30:197–207.
- 28 Huque MH, Carroll RJ, Diao N, *et al.* Exposure enriched case-control (EECC) design for the assessment of gene-environment interaction. *Genet Epidemiol* 2016;40:570–8.
- 29 Sitapati A, Kim H, Berkovich B, *et al.* Integrated precision medicine: the role of electronic health records in delivering personalized treatment. *Wiley Interdiscip Rev Syst Biol Med* 2017;9:e1378.
- 30 Knox KL, Conwell Y, Caine ED. If suicide is a public health problem, what are we doing to prevent it? *Am J Public Health* 2004;94:37–45.
- 31 Horrocks M, Michail M, Aubeeluck A, *et al.* An electronic clinical decision support system for the assessment and management of suicidality in primary care: protocol for a mixed-methods study. *JMIR Res Protoc* 2018;7:e11135.
- 32 Belsher BE, Smolenski DJ, Pruitt LD, *et al.* Prediction models for suicide attempts and deaths. *JAMA Psychiatry* 2019;76:642–11.
- 33 Kessler RC, Bernecker S, Bossarte RM, *et al.* The role of big data analytics in predicting suicide. In: *Personalized psychiatry-Big data analytics in mental health*, 2019.
- 34 Kessler RC, Bossarte RM, Luedtke A, *et al.* Suicide prediction models: a critical review of recent research with recommendations for the way forward. *Mol Psychiatry* 2020;25:168–79.
- 35 Health Evaluation and Quality Agency of Catalonia (Aqua). Public data analysis for health research and innovation program (PADRIS), 2019. Available: <http://aquas.gencat.cat/en/ambits/analitica-dades/padris/> [Accessed 1 Aug 2019].
- 36 Barcelona: Servei Catal de la Salut. Atenció a les persones en RISC de suïcidi: codi RISC de suïcidi (CRS), 2015. Available: <https://scientiasalut.gencat.cat/handle/11351/1654?locale-attribute=es>
- 37 Lecrubier Y, Sheehan DV, Weiller E, *et al.* The mini international neuropsychiatric interview (mini). A short diagnostic structured interview: reliability and validity according to the CIDI. *Eur psychiatr* 1997;12:224–31.
- 38 Roaldset JO, Linaker OM, Bjørkly S. Predictive validity of the mini suicidal scale for self-harm in acute psychiatry: a prospective study of the first year after discharge. *Arch Suicide Res* 2012;16:287–302.
- 39 Martín-Fumadó C, Gómez-Durán EL. Suicide investigation: psychological autopsy. *Span J Leg Med* 2017;43:135–7.
- 40 Mathes T, Pieper D. Study design classification of registry-based studies in systematic reviews. *J Clin Epidemiol* 2018;93:84–7.
- 41 Rothman KJ, Greenland S, Associate TLL. *Modern epidemiology.* 3rd Edition. Hastings Cent Rep, 2014.
- 42 Ø B, Breslow NE, Chatterjee N, *et al.* *Handbook of statistical methods for case-control studies*, 2018.
- 43 Glenn CR, Lanzillo EC, Esposito EC, *et al.* Examining the course of suicidal and Nonsuicidal Self-Injurious thoughts and behaviors in outpatient and inpatient adolescents. *J Abnorm Child Psychol* 2017;45:971–83.
- 44 Kim RS, Kaplan RC. Analysis of secondary outcomes in nested case-control study designs. *Stat Med* 2014;33:4215–26.
- 45 Kim RS. Analysis of nested case-control study designs: revisiting the inverse probability weighting method. *Commun Stat Appl Methods* 2013;20:455–66.
- 46 Tibshirani R. Regression shrinkage and selection via the LASSO. *J R STAT SOC B* 1996;58:267–88.
- 47 Krysinska K, Martin G. The struggle to prevent and evaluate: application of population attributable risk and preventive fraction to suicide prevention research. *Suicide Life Threat Behav* 2009;39:548–57.
- 48 Walsh CG, Ribeiro JD, Franklin JC. Predicting risk of suicide attempts over time through machine learning. *Clin Psychol Sci* 2017;5:457–69.
- 49 Zou H, Hastie T. Regularization and variable selection via the elastic net. *J Royal Statistical Soc B* 2005;67:301–20.
- 50 Rish I. *An empirical study of the naive Bayes classifier.* IJCAI workshop on empirical methods in artificial intelligence, 2001.
- 51 Friedman JH. Multivariate adaptive regression splines. *Ann. Statist.* 1991;19:1–67.
- 52 Chipman HA, George EI, McCulloch RE. Bart: Bayesian additive regression trees. *Ann Appl Stat* 2010;4:266–98.
- 53 Breiman L. Random forests. *Mach Learn* 2001;45:5–32.
- 54 Friedman JH. Greedy function approximation: a gradient boosting machine. *Ann Stat* 2001;29:1189–232.
- 55 Cover T, Hart P. Nearest neighbor pattern classification. *IEEE Trans Inform Theory* 1967;13:21–7.
- 56 Cortes C, Vapnik V. Support-vector networks. *Mach Learn* 1995;20:273–97.
- 57 Ayat S, Farahani HA, Aghamohamadi M, *et al.* A comparison of artificial neural networks learning algorithms in predicting tendency for suicide. *Neural Comput & Applic* 2013;23:1381–6.
- 58 van der Laan MJ, Polley EC, Hubbard AE. Super learner. *Stat Appl Genet Mol Biol* 2007;6:Article25.
- 59 Krautenbacher N, Theis FJ, Fuchs C. Correcting classifiers for sample selection bias in two-phase case-control studies. *Comput Math Methods Med* 2017;2017:1–18.
- 60 Kessler RC, Stein MB, Petukhova MV, *et al.* Predicting suicides after outpatient mental health visits in the army study to assess risk and resilience in Servicemembers (army STARRS). *Mol Psychiatry* 2017;22:544–51.
- 61 Universitat Pompeu Fabra (UPF), Barcelona S. Upf open access publishing, 2019. Available: <https://guiesbibtic.upf.edu/acces-obert/eng>