BMJ Open

The Catalonia Suicide Risk Code Epidemiology (CSRC-Epi) Study – a population-representative nested case-control study of suicide attempts in Catalonia, Spain.

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2020-037365
Article Type:	Protocol
Date Submitted by the Author:	30-Jan-2020
Complete List of Authors:	Mortier, Philippe ; Hospital del Mar Institute for Medical Research, Health Services Research Group Vilagut, Gemma; IMIM-Institut Hospital del Mar d'Investigacions Médiques, Puértolas Gracia, Beatriz; Hospital del Mar Institute for Medical Research, Health Services Research Group De Inés Trujillo, Ana; Hospital del Mar Institute for Medical Research, Health Services Research Group Alayo Bueno, Itxaso; Hospital del Mar Institute for Medical Research, Health Services Research Group Ballester Coma, Laura; Hospital del Mar Institute for Medical Research, Health Services Research Group Blasco Cubedo, María Jesús ; Hospital del Mar Institute for Medical Research, Health Services Research Group Cardoner, Narcís; Parc Taulí Sabadell, Hospital Universitari, Department of Mental Health Colls, Cristina; Agència de Qualitat i Avaluació Sanitàries de Catalunya, Catalan Health Department Elices, Matilde ; Hospital del Mar Institute for Medical Research, Neurosciences Research Programme Garcia-Altes, Anna; Agencia de Qualitat i Avaluació Sanitaries de Catalunya (AQuAS) Gené Badia, Manel; University of Barcelona Gómez Sánchez, Javier; Hospital del Mar Institute for Medical Research, Health Services Research Group Martín Sánchez, Mario; Agència de Salut Pública de Barcelona, Preventive Medicine and Public Health Training Unit PSMar-UPF-ASPB Morros, Rosa; Institut Universitari d'Investigació en Atenció Primària Jordi Gol (IDIAP Jordi Gol), Medicines Research Unit; Universitat Autònoma de Barcelona, Departament de Farmacologia i Terapèutica Prat Pubill, Bibiana; Ministry of Health, Catalan Government, Spain, Master Plan on Mental Health and Addictions Qin, Ping; National Centre for Suicide Research and Prevention, Institute of Clinical Medicine, University of Oslo Kessler, Ronald; Harvard Medical School, Department of Health Care Policy Palao, Diego; Fundació Parc Taulí-Institut Universitari UAB Pérez Sola, Víctor; Hospital del Mar Institute for Medical Research, Neurosciences Research Programme

	Alonso, Jordi; IMIM- Institut de Recerca Hospital del Mar,
Keywords:	Suicide & self-harm < PSYCHIATRY, EPIDEMIOLOGY, MENTAL HEALTH, PSYCHIATRY, PUBLIC HEALTH, STATISTICS & RESEARCH METHODS

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

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

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

The Catalonia Suicide Risk Code Epidemiology (CSRC-Epi) Study – 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[6,7,10,18], Cristina Colls[8], Matilde Elices[10,11], Anna García-Altés[2,8], Manel Gené Badia[12], Javier Gómez Sánchez[1], Mario Martín Sánchez[19], Rosa Morros Pedrós[13,14,15], Bibiana Prat Pubill[16], Ping Qin[17], Ronald C. Kessler[18], Diego Palao[6,7,11,20], Víctor Pérez[7,10,11,21], Jordi Alonso[1,2,5], on behalf of the CODIRISC Epidemiology Study Group*

*The ***CODIRISC Epidemiology Study Group*** are: Jordi Alonso, Itxaso Alayo Bueno, Laura Ballester Coma, Jordi Blanch, María Jesús Blasco Cubedo, Ana De Inés Trujillo, Maria Teresa Campillo Sanz, Narcís Cardoner, Anna Isabel Cebrià, Cristina Colls, Matilde Elices, Anna García-Altés, Ricard Gavaldà, 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.

AFFILIATIONS:

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

20. Institut d'Investigació i Innovació Parc Taulí (I3PT), Sabadell, Barcelona, Spain.
21. Institut de Neuropsiquiatria i Addiccions, Hospital del Mar, Barcelona, Spain.

CORRESPONDING AUTHOR

Philippe Mortier, M.D., M.Sc., Ph.D., Health Services Research Group, IMIM (Hospital del Mar Medical Research Institute), Dr. Aiguader, 88, 08003 Barcelona, Spain. pmortier@imim.es +34 933 16 07 43.

FUNDING STATEMENT:

This project was supported by ISCIII/FEDER PI17/00521, ISCIII/FEDER PI17/01205, and Generalitat de Catalunya (2017 SGR 452). The Catalonia Suicide Risk Code surveillance program is an initiative of the Mental Health and Addictions Plan of the Department of Health of the Catalan Government. Philippe Mortier has a Sara Borrell research contract awarded by the ISCIII (CD18/00049). Matilde Elices has a Juan de la Cierva research contract awarded by the ISCIII (FJCI-2017-31738). Víctor Pérez and Matilde Elices 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). Laura Ballester received funding by Ministerio de Educación, Cultura y Deporte (FPU15/05728). Diego Palao and Jordi Alonso received funding by ISCIII/FEDER PI17/01205.

COMPETING INTERESTS STATEMENT:

Diego Palao 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.

ABSTRACT:

Introduction: Suicide attempts (SA) represent an important public health burden. Centralized Electronic Health Record (EHR) systems have high yet underutilized potential to provide SA surveillance, to inform public health action aimed at reducing risk for SA 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 centralized EHR data from the entire public healthcare system of Catalonia, Spain, the CSRC-Epi study aims to estimate reliable SA incidence rates, identify SA risk factors, and develop validated SA 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 SA during the period 2014-2019 in Catalonia, Spain. The primary study outcome consists of first and repeat SA during the observation period. Cases will come from a case register linked to a SA surveillance program, 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 centralized 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-sex standardized SA incidence rates. Logistic regression will be used to identify SA risk factors on the individual-level (i.e., relative risk) and the population-level (i.e., population attributable risk proportions). Machine learning techniques will be used to develop SA risk prediction tools.

Ethics and dissemination: The protocol of this study has been 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.

Study registration number: NCT04235127.

Key words: Suicide Attempt; Nested Case-Control Studies; Registries; Electronic Health Records; Epidemiology; Incidence; Risk Factors; Risk Assessment; Supervised Machine Learning; Clinical Decision Support Systems

STRENGTHS AND LIMITATIONS OF THIS STUDY:

- SA cases (estimated n~6,000) in the CSRC 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 centralized EHR data representing five clinical settings, i.e., emergency care, primary care, outpatient mental healthcare, and general and psychiatric hospitalizations
- a two-stage exposure-enriched nested case-control study combined with the use of inverse probability weights will enable efficient and population-representative estimations

- SA risk prediction tools based on machine learning techniques will be developed to provide data-driven clinical decision support when assessing suicide attempt risk
- limited information on history of SA before the study observation period will be available

Word Count: 4,120.

INTRODUCTION

Suicide attempts (SA) constitute a major yet preventable [1] public health issue worldwide. Population-based surveys estimate the lifetime prevalence of SA among adults at 2.7% (range 0.5-5.0%) [2], while a recent meta-analysis among children and adolescents found a pooled lifetime estimate of 6.0% (range 0.5-34.1%) [3]. SA 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, SA are also predictive for persistent physical and mental health issues, repeat SA, psychiatric hospitalizations, impaired academic performance, unemployment, partner abuse victimization and perpetration, having children removed by social services, loneliness, relationship difficulties, impaired social functioning and low life satisfaction [6–11].

Despite this considerable societal impact, there is a lack of reliable surveillance data on SA that could inform public health action [12]. This is in contrast with actual suicide rates, that are increasingly monitored in many countries worldwide [13]. The WHO advocates the use of centralized Electronic Health Record (EHR) systems to develop national SA surveillance [14]. However, currently used disease classification systems in EHR systems (e.g., the *International Classification of Diseases* [ICD] [15]) 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 standardized registration of this clinical evaluation in a centralized EHR register, may therefore substantially improve the accuracy of public health surveillance of SA.

Apart from surveillance, centralized 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 54-80% for primary care contacts [21,22], 31-66% for mental healthcare contacts [21,22], 24-60% for emergency department visits [21,23–25], and 21% for psychiatric hospitalizations [21]. Clinical data collected through centralized EHR systems could therefore provide new insight in the population distribution of SA risk factors, outline the different healthcare trajectories preceding an attempt, and provide estimates of potential reductions in SA 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 studies [27] 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 centralized EHR data has high potential to provide data-driven clinical decision support when evaluating risk for future SA [29]. Such support is highly needed, as SA 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], i.e., a rapid problem-solving approach, including subconscious cognitive shortcuts, linking a limited set of risk factors directly to SA 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 SA risk prediction tools [32]. When routinely implemented at the healthcare system level, such data-driven decision support tools can guide the adequate allocation of clinical resources, such as in-depth suicide risk assessments and tailored treatment interventions in multi-stage screening approaches [33,34].

Here we present the protocol for the Catalonia Suicide Risk Code Epidemiology (CSRC-Epi) study, a large epidemiological study of SA occurring during the period 2014-2019 in Catalonia, Spain. The primary study outcome is SA among Catalan residents during the period 2014-2019. Secondary analyses will focus on actual suicide among those with previous SA. 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 centralized EHR data. A unique aspect of the study is that data for the SA cases come from an especially designed SA 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 include:

- to provide reliable incidence measures for SA occurring in Catalonia during the period 2014-2019.
- to identify risk factor constellations for SA, both on the individual-level (i.e., the extent to which risk factors increase the risk for subsequent SA in an individual) and on the population-level (i.e., the proportion of the total cases of SA that are potentially attributable to risk factors).
- to use machine learning techniques to develop clinically useful SA risk prediction tools that allow calculating personalized risk scores for future SA.

METHODS AND ANALYSIS

Data sources

All data for this study will be obtained from the Health Evaluation and Quality Agency of Catalonia (AQuAS [35]), a public entity attached to the Catalan Health Department. As from 2017, AQuAS manages the Public Data Analysis for Health Research and Innovation Program (PADRIS [35]) to provide researchers with access to large amounts of centralized 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 sociodemographic variables (i.e., sex, age, nationality, a range of socio-economic indicators, and healthcare catchment region) as well as associated small area geocode data (i.e., 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 (i.e., 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 program [36] in the Catalan Public Healthcare system. The CSRC surveillance program is a specifically designed SA 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 SA from non-suicidal self-injurious behaviour, or from adverse mental health states without self-injurious intention. Each individual deemed at high risk for (repeat) SA is subsequently eligible for two brief follow-up interventions (i.e., 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 specialized assessment are registered centrally in the CSRC case register.

The CSRC case register subsequently includes all individuals with SA 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 SA 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 SA, and suicide after a previous attempt.

Electronic Health Record (EHR) data

A third source of data will consist of centralized EHR registers, one for each of five clinical healthcare settings, i.e., emergency care visits, primary care visits, general hospitalizations, psychiatric hospitalizations, and outpatient mental health visits. These registers include a wide range of relevant predictor variables for suicidal behaviour, i.e.,

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 ICD-9-CM and ICD-10-CM disease classification system. The year of inception of the different registers is 2012 for emergency care visits and primary care visits, and 2008 for the other registers.

Pharmaceutical register

A fourth source of data will consist of a register containing all pharmaceutical drugs (i.e., over the counter as well as prescription drugs) and health-related products that have been delivered by officially recognized 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, i.e., 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 SA 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 International Classification of Disease system (9th or 10th revision).

[INSERT FIGURE 1 HERE]

Study Design

Figure 1 shows an overview of the CSRC-Epi study design. The CSRC-Epi study is a register-based study, i.e., a study that uses the exposure and outcome data from registries [43], 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 7.5-7.6 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 SA 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 [40]. We will therefore exclude cases and controls that have not reached age 10 by the end of the 2014-2019 period (i.e., ~10.9%), lowering the total expected target population to ~8.1 million.

Case selection

SA cases between 2014 and 2019 will be identified using the CSRC case register. Based on preliminary data exploration, we expect to include ~6,000 cases of clinically

confirmed SA by the end of 2019, of which ~8% (~480) will be repeat attempters. This substantially exceeds the number of SA cases included in previous register-based studies (median = 1,562, IQR = 1562-3250; [32]).

One of the main objectives of the CSRC program is to enable reliable surveillance of SA in the population, and to tackle underregistration and misclassification using regular EHR systems [16]. Nevertheless, failure to adhere to the CSRC program protocol may result in an unknown number of SA cases that remain undetected. Therefore, ICD disease classification codes in the five centralized EHR registers will be inspected to identify potentially missed cases of SA. For that purpose, a wide range of ICD codes related to SA (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 (i.e., intentional or accidental self-injurious behaviour) [16]. Outcome definition algorithms (i.e., predefined sets of ICD codes) have shown promising in increasing the accuracy of SA case detection using ICD codes [17]. Therefore, we intent to validate the range of ICD codes we identified against golden standard identification methods (i.e., manual review as well as text mining of clinical notes) to increase the accurate detection of potentially missed cases for our study.

[INSERT TABLE 1 HERE]

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-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 (i.e., ranging from ~17% for primary care visits to ~0.1% for general hospitalizations). 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 66 risk sets, one for each month in the 6-year observation period (i.e., June 2014 to December 2019). Within each risk set, a number of 30 age-sex matched controls will be randomly selected for each case (i.e., case or potentially missed case) without replacement (incidence density sampling or risk set sampling [41]). 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 SA and potential SA (see Data analytical plan), controls are selected for first SA within individuals (eligible controls including previous potential cases, but not previous SA cases), and if applicable, also for the first potential SA within individuals

(eligible controls including those without previous SA or potential SA 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, (i.e., 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 SA 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 SA assessed at the emergency department at time y and a last healthcare visit at primary care at time x, 30 age-sex matched controls will be selected among those individuals that have not committed a SA 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 SA during the period 2014-2019. We will both focus on first SA during the observation period, as well as on repetition of attempts, defined as SA among those with a previous SA during the observation period. As explained above, we will also identify potentially missed cases of SA, 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 focussing on clinical severity of SA (e.g., lethality and method of attempt). A secondary study outcome consists of suicide among those with a SA during 2014-2019. Cases of suicide will be identified through the mortality register, managed by the Catalan Department of Forensic Medicine (see above).

SA 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-sex standardized incidence rates, stratified by relevant sociodemographic variables (e.g., socio-economic status, healthcare region, etc.). As incidence rates do not inform of the distribution of cases over time, we will also estimate and visualize incidence proportions (or cumulative incidence) over time using the Kaplan Meier estimator to estimate one minus the survival function.

SA 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 SA during the observation period as the outcome variable, the weighted nested case-control dataset will be used. In order for the odds ratio to be a valid estimation of the hazard ratio (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 (e.g., the exact dates of

diagnoses or medical prescriptions) will be recoded by categorizing 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 Operator (LASSO [46]) method will be employed as to select a subset of predictor variables that predict the outcome best whilst maintaining a good model fit. For suicide and for repetition of SA as the outcome variable, the analysis will be very similar, but will now reduce to a cohort analysis of those individuals with first SA during 2014-2019 (which are all selected), and regression models will also include the clinical variables from the CRSC case register obtained though in-depth psychiatric evaluation.

Population-level effect sizes will be estimated by calculating bivariable and multivariable population-attributable risk proportions (PARP [47]), 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.

SA risk predictions

Risk prediction tools for (repetition of) SA will be constructed using machine learning techniques. A first series of algorithms will focus on estimating SA risk for different prediction windows, i.e., 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 SA after specific healthcare contacts using the separate series of controls selected in the second sampling stage (see above). Machine learning techniques will include elastic net penalized logistic regression [49], naïve Bayes classifiers [50], multivariate adaptive regression splines [51], Bayesian additive regression trees (BART) [52], random forests [53], gradient boosting [54], k-nearest neighbour algorithms [55], support vector machines [56], and artificial neural networks [57]. Stacking ensemble techniques (super learning) [58] will be implemented to further optimize 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 (i.e., thresholds set to delineate the top x% at highest risk [60]), 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 (i.e., the harmonic mean of recall and precision).

Study limitations

A first major limitation of the CSRC-Epi study is that any history of SA 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 program protocol, which assesses lifetime history of SA 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 program, SA cases may still go undetected. This will be countered by identifying potentially missed cases of SA in the EHR registers.

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 anonymized 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 SA risk factors constellations, on the need to improve SA 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 (co-design).

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.

FULL 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. doi:10.1186/s12888-019-2147-y

2 Nock MK, Borges G, Bromet EJ, *et al.* Cross-national prevalence and risk factors for suicidal ideation, plans and attempts. *Br J Psychiatry* Published Online First: 2008. doi:10.1192/bjp.bp.107.040113

3 Lim KS, 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**. doi:10.3390/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* Published Online First: 2016. doi:10.1017/S0033291715001804

5 Orpana HM, Marczak LB, Arora M, *et al.* Global, regional, and national burden of suicide mortality 1990 to 2016: Systematic analysis for the Global Burden of Disease Study 2016. *BMJ* Published Online First: 2019. doi:10.1136/bmj.l94

6 Goldman-Mellor SJ, Caspi A, Harrington HL, *et al.* Suicide attempt in young people a signal for long-term health care and social needs. *JAMA Psychiatry* Published Online First: 2014. doi:10.1001/jamapsychiatry.2013.2803

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-Threatening Behav* Published Online First: 2014. doi:10.1111/sltb.12077

8 Mortier P, Demyttenaere K, Auerbach RP, *et al.* The impact of lifetime suicidality on academic performance in college freshmen. *J Affect Disord* Published Online First: 2015. doi:10.1016/j.jad.2015.07.030

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* Published Online First: 2016. doi:10.1007/s10597-016-9987-4

10 Pajonk FG, Ruchholtz S, Waydhas C, *et al.* Long-term follow-up after severe suicide attempt by multiple blunt trauma. *Eur Psychiatry* 2005;**20**:115–20. doi:10.1016/j.eurpsy.2004.10.003

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. doi:10.1177/0004867416633298

12 Bachmann S. Epidemiology of suicide and the psychiatric perspective. Int. J. Environ. Res. Public Health. 2018. doi:10.3390/ijerph15071425

13 World Health Organisation. WHO | Suicide data. WHO Website. 2019.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. *World Heal Organ* 2016.

15 World Health Organization. The International Classification of Diseases. 2019.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.

17 Swain RS, Taylor LG, Braver ER, *et al.* A systematic review of validated suicide outcome classification in observational studies. *Int J Epidemiol* 2019;:1–14. doi:10.1093/ije/dyz038

18 Chute CG, Koo D. Public health, data standards, and vocabulary: Crucial infrastructure for reliable public health surveillance. *J Public Heal Manag Pract* 2002;**8**:11–7. doi:10.1097/00124784-200205000-00003

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. doi:10.1016/j.jbi.2007.07.001

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* Published Online First: 2016. doi:10.1146/annurev-publhealth-032315-021353

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* Published Online First: 2016. doi:10.1002/wps.20321

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. 2017. doi:10.1177/1403494817746274

23 Vasiliadis HM, 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. doi:10.1176/appi.ps.201400133

24 Morrison KB, Laing L. Adults' use of health services in the year before death by suicide in Alberta. *Heal Reports* 2011;**22**.

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. doi:10.1007/s11606-014-2767-3

26 Pirkis J, Nicholas A, Gunnell D. The case for case-control studies in the field of suicide prevention. *Epidemiol Psychiatr Sci* 2019;:2018–20. doi:10.1017/S2045796019000581

27 Kim RS. A new comparison of nested case–control and case–cohort designs and methods. *Eur J Epidemiol* Published Online First: 2014. doi:10.1007/s10654-014-9974-4

28 Huque MH, Carroll RJ, Diao N, *et al.* Exposure Enriched Case-Control (EECC) Design for the Assessment of Gene–Environment Interaction. *Genet Epidemiol* Published Online First: 2016. doi:10.1002/gepi.21986

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* Published Online First: 2017. doi:10.1002/wsbm.1378

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. doi:10.2105/AJPH.94.1.37

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* Published Online First: 2018. doi:10.2196/11135

32 Belsher BE, Smolenski DJ, Pruitt LD, *et al.* Prediction Models for Suicide Attempts and Deaths. *JAMA Psychiatry* 2019;**20910**:1–11. doi:10.1001/jamapsychiatry.2019.0174

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. doi:10.1007/978-3-030-03553-2_5

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* Published Online First: 2019. doi:10.1038/s41380-019-0531-0

35 Health Evaluation and Quality Agency of Catalonia (Aqua). Public Data Analysis for Health Research and Innovation Program (PADRIS). 2019.http://aquas.gencat.cat/en/ambits/analitica-dades/padris/ (accessed 1 Aug 2019).

36 Barcelona: Servei Català de la Salut. Atenció a les persones en risc de suïcidi: codi risc de suïcidi (CRS). 2015.https://scientiasalut.gencat.cat/handle/11351/1654?locale-attribute=es

37 Sheehan D V., Lecrubier Y, Sheehan KH, *et al.* The Mini-International Neuropsychiatric Interview (M.I.N.I.): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. In: *Journal of Clinical Psychiatry*. 1998.

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* Published Online First: 2012. doi:10.1080/13811118.2013.722052

39 Government of Catalonia (Generalitat de Catalunya). Statistical Institute of Catalonia (Institut d'Estadística de Catalunya - IDESCAT). https://www.idescat.cat/

40 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* Published Online First: 2017. doi:10.1007/s10802-016-0214-0

41 Rothman KJ, Greenland S, Associate TLL. Modern Epidemiology, 3rd Edition. *Hastings Cent Rep* Published Online First: 2014. doi:10.1002/hast.292

42 Borgan Ø, Breslow NE, Chatterjee N, *et al. Handbook of Statistical Methods for*

Case-Control Studies. 2018. doi:10.1201/9781315154084

43 Mathes T, Pieper D. Study design classification of registry-based studies in systematic reviews. *J Clin Epidemiol* 2018;**93**:84–7. doi:10.1016/j.jclinepi.2017.09.016

44 Kim RS, Kaplan RC. Analysis of secondary outcomes in nested case-control study designs. *Stat Med* Published Online First: 2014. doi:10.1002/sim.6231

45 Kim RS. Analysis of Nested Case-Control Study Designs: Revisiting the Inverse Probability Weighting Method. *Commun Stat Appl Methods* 2013;**20**:455–66. doi:10.5351/csam.2013.20.6.455

46 Tibshirani R. Regression Shrinkage and Selection Via the Lasso. *J R Stat Soc Ser B* Published Online First: 1996. doi:10.1111/j.2517-6161.1996.tb02080.x

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-Threatening Behav* Published Online First: 2009. doi:10.1521/suli.2009.39.5.548

48 Walsh CG, Ribeiro JD, Franklin JC. Predicting Risk of Suicide Attempts Over Time Through Machine Learning. *Clin Psychol Sci* Published Online First: 2017. doi:10.1177/2167702617691560

49 Zou H, Hastie T. Regularization and variable selection via the elastic net. *J R Stat Soc Ser B Stat Methodol* Published Online First: 2005. doi:10.1111/j.1467-9868.2005.00503.x

50 Rish I. An empirical study of the naive Bayes classifier. *IJCAI 2001 Work Empir methods Artif Intell* Published Online First: 2001. doi:10.1039/b104835j

51 Friedman JH. Multivariate Adaptive Regression Splines. *Ann Stat* Published Online First: 1991. doi:10.1214/aos/1176347963

52 Chipman HA, George EI, Mcculloch RE. BART: BAYESIAN ADDITIVE REGRESSION TREES BART: BAYESIAN ADDITIVE REGRESSION TREES12. *Source Ann Appl Stat Ann Appl Stat* Published Online First: 2010. doi:10.1214/09-AOAS285

53 Breiman L. Random forests. *Mach Learn* Published Online First: 2001. doi:10.1023/A:1010933404324

54 Friedman JH. Greedy function approximation: A gradient boosting machine. *Ann Stat* Published Online First: 2001. doi:10.2307/2699986

55 Cover TM, Hart PE. Nearest Neighbor Pattern Classification. *IEEE Trans Inf Theory* Published Online First: 1967. doi:10.1109/TIT.1967.1053964

56 Cortes C, Vapnik V. Support-Vector Networks. *Mach Learn* Published Online First: 1995. doi:10.1023/A:1022627411411

57 Ayat S, Farahani HA, Aghamohamadi M, *et al.* A comparison of artificial neural networks learning algorithms in predicting tendency for suicide. *Neural Comput Appl* Published Online First: 2013. doi:10.1007/s00521-012-1086-z

58 Van Der Laan MJ, Polley EC, Hubbard AE. Super learner. *Stat Appl Genet Mol*

Biol Published Online First: 2007. doi:10.2202/1544-6115.1309

59 Krautenbacher N, Theis FJ, Fuchs C. Correcting Classifiers for Sample Selection Bias in Two-Phase Case-Control Studies. *Comput Math Methods Med* Published Online First: 2017. doi:10.1155/2017/7847531

60 Kessler RC, Stein MB, Petukhova M V., *et al.* Predicting suicides after outpatient mental health visits in the Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS). *Mol Psychiatry* Published Online First: 2017. doi:10.1038/mp.2016.110

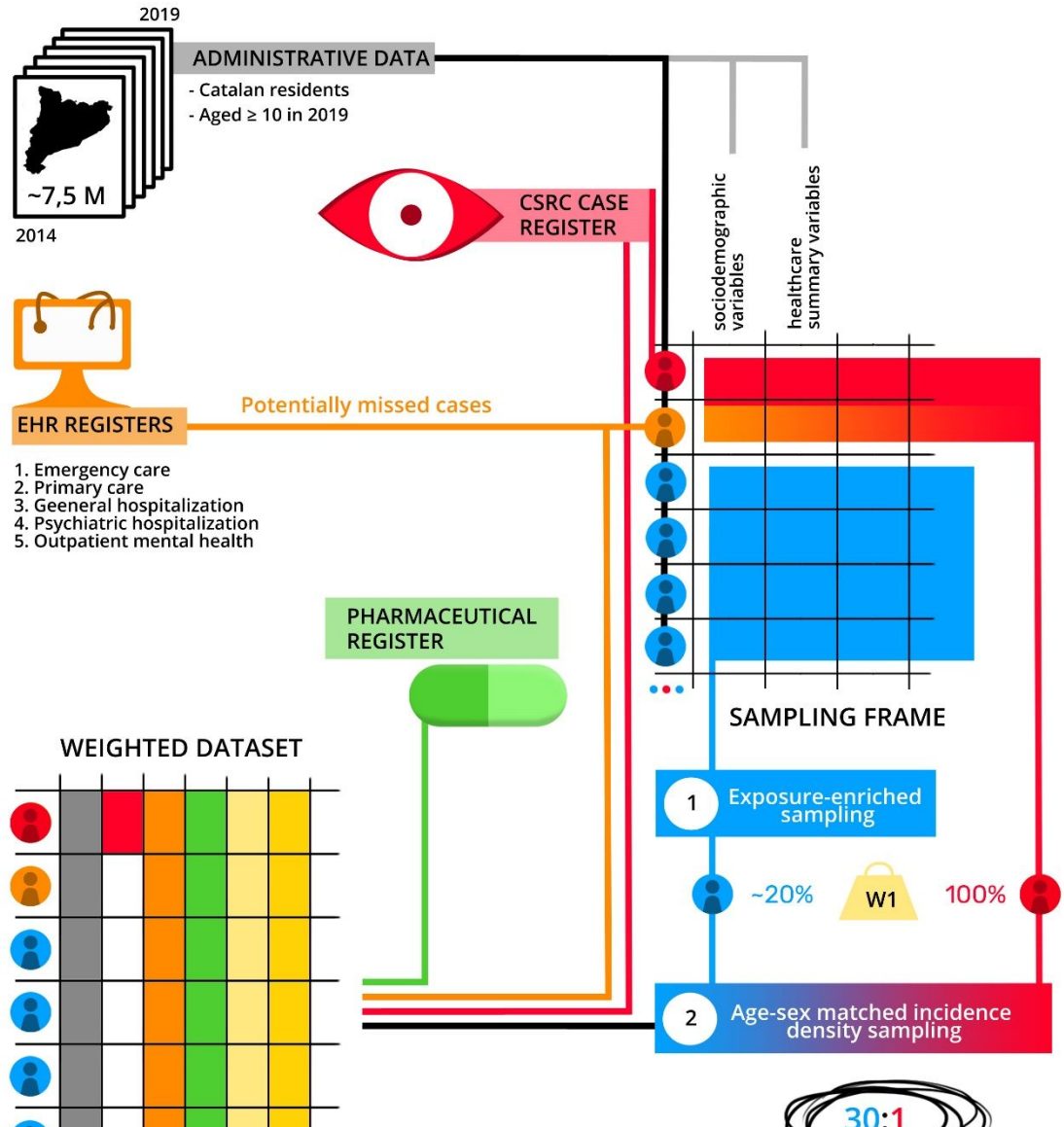
61 Universitat Pompeu Fabra (UPF), Barcelona S. UPF Open Access Publishing. 2019.https://guiesbibtic.upf.edu/acces-obert/eng

Table 1. International Classification of Diseases Codes to identify potentially missed cases of suicide attempt.

	ICD-9-CM	Description	ICD-10-CM	Description
suicide attempts and intentional self-injurious behavior	**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 6th character of 2	asphyxiation, suffocation, strangulation
			T14.91	suicide attempt
self-injurious behavior 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 6th 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 nonmedicinal substances

Note: ICD = International Classification of Diseases; CM = Clinical Modification.

Figure 1. Overview of the CSRC-Epi study design



Note: M = million; EHR = Electronic Healthcare Record; W = weight.

STROBE 2007 (v4) Statement—Checklist of items that should be included in reports of *cohort studies*

Section/Topic	Item #	Recommendation	Reported on page #
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	4
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	5
Objectives	3	State specific objectives, including any prespecified hypotheses	6
Methods			
Study design	4	Present key elements of study design early in the paper	6,8
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	6-8
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up	8,9
		(*b*) For matched studies, give matching criteria and number of exposed and unexposed	9,10
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	6-8
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	6-8
Bias	9	Describe any efforts to address potential sources of bias	9,10
Study size	10	Explain how the study size was arrived at	8
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	10,11
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	10,11
		(*b*) Describe any methods used to examine subgroups and interactions	10,11
		(*c*) Explain how missing data were addressed	10,11
		(*d*) If applicable, explain how loss to follow-up was addressed	NA
		(*e*) Describe any sensitivity analyses	NA
Results			

Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	8-10
		(b) Give reasons for non-participation at each stage	NA
		(c) Consider use of a flow diagram	Figure 1
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	8
		(b) Indicate number of participants with missing data for each variable of interest	NA
		(c) Summarise follow-up time (eg, average and total amount)	NA
Outcome data	15*	Report numbers of outcome events or summary measures over time	NA
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	NA
		(*b*) Report category boundaries when continuous variables were categorized	NA
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	NA
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	NA
Discussion			
Key results	18	Summarise key results with reference to study objectives	NA
Limitations			
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	12
Generalisability	21	Discuss the generalisability (external validity) of the study results	12
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	3

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

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

BMJ Open

The Catalonia Suicide Risk Code Epidemiology (CSRC-Epi) Study – protocol for a population-representative nested case-control study of suicide attempts in Catalonia, Spain.

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2020-037365.R1
Article Type:	Protocol
Date Submitted by the Author:	02-May-2020

Complete List of Authors:	Mortier, Philippe ; Hospital del Mar Institute for Medical Research, Health Services Research Group Vilagut, Gemma; IMIM-Institut Hospital del Mar d'Investigacions Médiques, Puértolas Gracia, Beatriz; Hospital del Mar Institute for Medical Research, Health Services Research Group De Inés Trujillo, Ana; Hospital del Mar Institute for Medical Research, Health Services Research Group Alayo Bueno, Itxaso; Hospital del Mar Institute for Medical Research, Health Services Research Group Ballester Coma, Laura; Hospital del Mar Institute for Medical Research, Health Services Research Group Blasco Cubedo, María Jesús ; Hospital del Mar Institute for Medical Research, Health Services Research Group Cardoner, Narcís; Parc Taulí Sabadell, Hospital Universitari, Department of Mental Health Colls, Cristina; Agència de Qualitat i Avaluació Sanitàries de Catalunya, Catalan Health Department Elices, Matilde ; Hospital del Mar Institute for Medical Research, Neurosciences Research Programme Garcia-Altes, Anna; Agencia de Qualitat i Avaluació Sanitaries de Catalunya (AQuAS) Gené Badia, Manel; University of Barcelona Gómez Sánchez, Javier; Hospital del Mar Institute for Medical Research, Health Services Research Group Martín Sánchez, Mario; Agència de Salut Pública de Barcelona, Preventive Medicine and Public Health Training Unit PSMar-UPF-ASPB Morros, Rosa; Institut Universitari d'Investigació en Atenció Primària Jordi Gol (IDIAP Jordi Gol), Medicines Research Unit; Universitat Autònoma de Barcelona, Departament de Farmacologia i Terapèutica Prat Pubill, Bibiana; Ministry of Health, Catalan Government, Spain, Master Plan on Mental Health and Addictions Qin, Ping; National Centre for Suicide Research and Prevention, Institute of Clinical Medicine, University of Oslo Kessler, Ronald; Harvard Medical School, Department of Health Care Policy Palao, Diego; Fundació Parc Taulí-Institut Universitari UAB Pérez Sola, Víctor; Hospital del Mar Institute for Medical Research, Neurosciences Research Programme

	Alonso, Jordi; IMIM- Institut de Recerca Hospital del Mar,
Primary Subject Heading:	Mental health
Secondary Subject Heading:	Epidemiology, Public health, Research methods
Keywords:	Suicide & self-harm < PSYCHIATRY, EPIDEMIOLOGY, MENTAL HEALTH, PSYCHIATRY, PUBLIC HEALTH, STATISTICS & RESEARCH METHODS

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

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

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

The 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[6,7,10,18], Cristina Colls[8], Matilde Elices[10,11], Anna García-Altés[2,8], Manel Gené Badia[12], Javier Gómez Sánchez[1], Mario Martín Sánchez[19], Rosa Morros Pedrós[13,14,15], Bibiana Prat Pubill[16], Ping Qin[17], Ronald C. Kessler[18], Diego Palao[6,7,11,20], Víctor Pérez[7,10,11,21], Jordi Alonso[1,2,5], on behalf of the CODIRISC Epidemiology Study Group*

*The ***CODIRISC Epidemiology Study Group*** are: Jordi Alonso, Itxaso Alayo Bueno, Laura Ballester Coma, Jordi Blanch, María Jesús Blasco Cubedo, Ana De Inés Trujillo, Maria Teresa Campillo Sanz, Narcís Cardoner, Anna Isabel Cebrià, Cristina Colls, Matilde Elices, Anna García-Altés, Ricard Gavaldà, 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.

AFFILIATIONS:

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

19. Preventive Medicine and Public Health Training Unit PSMar-UPF-ASPB, Parc de Salut Mar, Agència de Salut Pública de Barcelona, Pompeu Fabra University, Barcelona, Spain
20. Institut d'Investigació i Innovació Parc Taulí (I3PT), Sabadell, Barcelona, Spain.
21. Institut de Neuropsiquiatria i Addiccions, Hospital del Mar, Barcelona, Spain.

CORRESPONDING AUTHOR

Philippe Mortier, M.D., M.Sc., Ph.D., Health Services Research Group, IMIM (Hospital del Mar Medical Research Institute), Dr. Aiguader, 88, 08003 Barcelona, Spain. pmortier@imim.es +34 933 16 07 43.

ABSTRACT:

Introduction: Suicide attempts represent an important public health burden. Centralized 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 centralized 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 program, 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 centralized 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-sex standardized suicide attempt incidence rates. Logistic regression will identify suicide attempt risk factors on the individual-level (i.e., relative risk) and the population-level (i.e., population attributable risk proportions). Machine learning techniques will be used to develop suicide attempt risk prediction tools.

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.

Study registration number: NCT04235127.

Key words: Suicide Attempt; Nested Case-Control Studies; Registries; Electronic Health Records; Epidemiology; Incidence; Risk Factors; Risk Assessment; Supervised Machine Learning; Clinical Decision Support Systems

STRENGTHS AND LIMITATIONS OF THIS STUDY:

- suicide attempt cases (estimated n~6,000) in the CSRC 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 centralized EHR data representing five clinical settings, i.e., emergency care, primary care, outpatient mental healthcare, and general and psychiatric hospitalizations
- 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

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 hospitalizations, impaired academic performance, unemployment, partner abuse victimization 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]. The WHO advocates the use of centralized Electronic Health Record (EHR) systems to develop national suicide attempt surveillance [14]. However, currently used disease classification systems in EHR systems (e.g., the *International Classification of Diseases* [ICD] [15]) 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 standardized registration of this clinical evaluation in a centralized EHR register, may therefore substantially improve the accuracy of public health surveillance of suicide attempts.

Apart from surveillance, centralized 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 54-80% for primary care contacts [21,22], 31-66% for mental healthcare contacts [21,22], 24-60% for emergency department visits [21,23–25], and 21% for psychiatric hospitalizations [21]. Clinical data collected through centralized 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 studies [27] 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 centralized 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], i.e., 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 multi-stage 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 centralized 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 (i.e., the extent to which risk factors increase the risk for subsequent suicide attempt in an individual) and on the population-level (i.e., 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 (AQuAS [35]), a public entity attached to the Catalan Health Department. As from 2017, AQuAS manages the Public Data Analysis for Health Research and

Innovation Program (PADRIS [35]) to provide researchers with access to large amounts of centralized 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 sociodemographic variables (i.e., sex, age, nationality, a range of socio-economic indicators, and healthcare catchment region) as well as associated small area geocode data (i.e., 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 (i.e., 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 program [36] in the Catalan Public Healthcare system. The CSRC surveillance program 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 (i.e., 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 specialized 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.

Electronic Health Record (EHR) data

A third source of data will consist of centralized EHR registers, one for each of five clinical healthcare settings, i.e., emergency care, primary care, outpatient mental healthcare, and general and psychiatric hospitalizations. These registers include a wide range of relevant predictor variables for suicidal behaviour, i.e., 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 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 recognized 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, i.e., 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 International Classification of Disease system (9th or 10th revision). State-of-the art forensic techniques, including psychological autopsy by a multidisciplinary team, is used to determine death by suicide in the mortality register [39]. Nevertheless, forensic examination of suicidal intention is difficult, and misclassification may occur.

[INSERT FIGURE 1 HERE]

Study Design

Figure 1 shows an overview of the CSRC-Epi study design. The CSRC-Epi study is a register-based study, i.e., 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 7.5-7.6 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 (i.e., ~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 ~6,000 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 = 1,562, IQR = 1562-3250; [32]).

One of the main objectives of the CSRC program is to enable reliable surveillance of suicide attempt in the population, and to tackle underregistration and misclassification using regular EHR systems [16]. Nevertheless, failure to adhere to the CSRC program protocol may result in an unknown number of suicide attempt cases that remain undetected. Therefore, ICD disease classification codes in the five centralized 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 (i.e., intentional or accidental self-injurious behaviour) [16]. Outcome definition algorithms (i.e., 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 (i.e., manual review as well as text mining of clinical notes) to increase the accurate detection of potentially missed cases for our study.

[INSERT TABLE 1 HERE]

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-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 (i.e., ranging from ~17% for primary care visits to ~0.1% for general hospitalizations). 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 (i.e., June 2014 to December 2019). Within each risk set, a number of 30 age-sex matched controls will be randomly selected for each case (i.e., case or

potentially missed case) without replacement (incidence density sampling or risk set sampling [41]). 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 Data analytical plan), 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, (i.e., 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-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 focussing on clinical severity of suicide attempts (e.g., 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 above).

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-sex standardized incidence rates, stratified by relevant sociodemographic variables (e.g., socio-economic status, healthcare region, etc.). As incidence rates do not inform of the distribution of cases over time, we will also estimate and visualize 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 odds ratio to be a valid estimation of the hazard ratio (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 (e.g., the exact dates of diagnoses or medical prescriptions) will be recoded by categorizing 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 Operator (LASSO [46]) method will be employed as to select a subset of predictor variables that predict the outcome best whilst 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 though in-depth psychiatric evaluation.

Population-level effect sizes will be estimated by calculating bivariable and multivariable population-attributable risk proportions (PARP [47]), 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- 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, i.e., 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 above). Machine learning techniques will include elastic net penalized logistic regression [49], naïve Bayes classifiers [50], multivariate adaptive regression splines [51], Bayesian additive regression trees (BART) [52], random forests [53], gradient boosting [54], k-nearest neighbour algorithms [55], support vector machines [56], and artificial neural networks

[57]. Stacking ensemble techniques (super learning) [58] will be implemented to further optimize 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 (i.e., thresholds set to delineate the top x% at highest risk [60]), 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 (i.e., 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 program 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 program, 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 epidemiologic 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 anonymized 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 (co-design).

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.

CONTRIBUTORSHIP STATEMENT

Initial draft of the protocol: PM, GV, BPG, IAB, and JA. 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, AGA, MGB, JGS, MMS, RMP, BPP, PQ, RCK, DP, VP, and JA. All authors read and approved the final manuscript.

COMPETING INTERESTS:

Diego Palao 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.

FUNDING:

This project was supported by ISCIII/FEDER PI17/00521, ISCIII/FEDER PI17/01205, and Generalitat de Catalunya (2017 SGR 452). The Catalonia Suicide Risk Code surveillance program is an initiative of the Mental Health and Addictions Plan of the Department of Health of the Catalan Government. Philippe Mortier has a Sara Borrell research contract awarded by the ISCIII (CD18/00049). Matilde Elices has a Juan de la Cierva research contract awarded by the ISCIII (FJCI-2017-31738). Víctor Pérez and Matilde Elices 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). Beatriz Puértolas Gracia and Ana De Inés Trujillo received funding from ISCIII FI18/00012 and FPU2017-06447, respectively. Laura Ballester received funding by Ministerio de Educación, Cultura y Deporte (FPU15/05728). Diego Palao and Jordi Alonso received funding by ISCIII/FEDER PI17/01205.

DATA SHARING STATEMENT

All electronic healthcare record data related to the CSRC-Epi study are available upon request at the Health Evaluation and Quality Agency of Catalonia (AQUAS), a public entity attached to the Catalan Health Department, as part of the Public Data Analysis for Health Research and Innovation Program (PADRIS).

FULL 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. doi:10.1186/s12888-019-2147-y

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. doi:10.1192/bjp.bp.107.040113

3 Lim KS, 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**. doi:10.3390/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. doi:10.1017/S0033291715001804

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**:l94. doi:10.1136/bmj.l94

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. doi:10.1001/jamapsychiatry.2013.2803

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-Threatening Behav* 2014;**44**:497–509. doi:10.1111/sltb.12077

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. doi:10.1016/j.jad.2015.07.030

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. doi:10.1007/s10597-016-9987-4

10 Pajonk FG, Ruchholtz S, Waydhas C, *et al.* Long-term follow-up after severe suicide attempt by multiple blunt trauma. *Eur Psychiatry* 2005;**20**:115–20. doi:10.1016/j.eurpsy.2004.10.003

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. doi:10.1177/0004867416633298

12 Bachmann S. Epidemiology of Suicide and the Psychiatric Perspective. *Int J Environ Res Public Health* 2018;**15**:1425. doi:10.3390/ijerph15071425

13 World Health Organisation. WHO | Suicide data. WHO Website. 2019.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. *World Heal Organ* 2016.

15 World Health Organization. The International Classification of Diseases. 2019.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.

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. doi:10.1093/ije/dyz038

18 Chute CG, Koo D. Public health, data standards, and vocabulary: Crucial infrastructure for reliable public health surveillance. *J Public Heal Manag Pract* 2002;**8**:11–7. doi:10.1097/00124784-200205000-00003

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. doi:10.1016/j.jbi.2007.07.001

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. doi:10.1146/annurev-publhealth-032315-021353

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. doi:10.1002/wps.20321

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. doi:10.1177/1403494817746274

23 Vasiliadis HM, 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. doi:10.1176/appi.ps.201400133

24 Morrison KB, Laing L. Adults' use of health services in the year before death by suicide in Alberta. *Heal Reports* 2011;**22**.

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. doi:10.1007/s11606-014-2767-3

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. doi:10.1017/S2045796019000581

27 Kim RS. A new comparison of nested case–control and case–cohort designs and methods. *Eur J Epidemiol* 2015;**30**:197–207. doi:10.1007/s10654-014-9974-4

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. doi:10.1002/gepi.21986

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. doi:10.1002/wsbm.1378

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. doi:10.2105/ajph.94.1.37

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. doi:10.2196/11135

32 Belsher BE, Smolenski DJ, Pruitt LD, *et al.* Prediction Models for Suicide Attempts and Deaths. *JAMA Psychiatry* 2019;**20910**:1–11. doi:10.1001/jamapsychiatry.2019.0174

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. doi:10.1007/978-3-030-03553-2_5

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. doi:10.1038/s41380-019-0531-0

35 Health Evaluation and Quality Agency of Catalonia (Aqua). Public Data Analysis for Health Research and Innovation Program (PADRIS). 2019.http://aquas.gencat.cat/en/ambits/analitica-dades/padris/ (accessed 1 Aug 2019).

36 Barcelona: Servei Català de la Salut. Atenció a les persones en risc de suïcidi: codi risc de suïcidi (CRS). 2015.https://scientiasalut.gencat.cat/handle/11351/1654?locale-attribute=es

37 Lecrubier Y, Sheehan D, Weiller E, *et al.* The Mini International Neuropsychiatric Interview (MINI). A short diagnostic structured interview: reliability and validity according to the CIDI. *Eur Psychiatry* 1997;**12**:224–31. doi:10.1016/S0924-9338(97)83296-8

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. doi:10.1080/13811118.2013.722052

39 Martin-Fumadó C, Gómez-Durán EL. Suicide investigation: Psychological autopsy. *Spanish J Leg Med* 2017;**43**:135–7. doi:10.1016/j.remle.2017.11.001

40 Mathes T, Pieper D. Study design classification of registry-based studies in systematic reviews. *J Clin Epidemiol* 2018;**93**:84–7. doi:10.1016/j.jclinepi.2017.09.016

41 Rothman KJ, Greenland S, Associate TLL. Modern Epidemiology, 3rd Edition. *Hastings Cent Rep* Published Online First: 2014. doi:10.1002/hast.292

42 Borgan Ø, Breslow NE, Chatterjee N, *et al. Handbook of Statistical Methods for Case-Control Studies*. 2018. doi:10.1201/9781315154084

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. doi:10.1007/s10802-016-0214-0

44 Kim RS, Kaplan RC. Analysis of secondary outcomes in nested case-control study designs. *Stat Med* 2014;**33**:4215–26. doi:10.1002/sim.6231

45 Kim RS. Analysis of Nested Case-Control Study Designs: Revisiting the Inverse Probability Weighting Method. *Commun Stat Appl Methods* 2013;**20**:455–66. doi:10.5351/csam.2013.20.6.455

46 Tibshirani R. Regression Shrinkage and Selection Via the Lasso. *J R Stat Soc Ser B* 1996;**58**:267–88. doi:10.1111/j.2517-6161.1996.tb02080.x

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-Threatening Behav* 2009;**39**:548–57. doi:10.1521/suli.2009.39.5.548

48 Walsh CG, Ribeiro JD, Franklin JC. Predicting Risk of Suicide Attempts Over Time Through Machine Learning. *Clin Psychol Sci* 2017;**5**:457–69. doi:10.1177/2167702617691560

49 Zou H, Hastie T. Regularization and variable selection via the elastic net. *J R Stat Soc Ser B (Statistical Methodol* 2005;**67**:301–20. doi:10.1111/j.1467-9868.2005.00503.x

50 Rish I. An empirical study of the naive Bayes classifier. *IJCAI 2001 Work Empir methods Artif Intell* Published Online First: 2001. doi:10.1039/b104835j

51 Friedman JH. Multivariate Adaptive Regression Splines. *Ann Stat* 1991;**19**:1–67. doi:10.1214/aos/1176347963

52 Chipman HA, George EI, McCulloch RE. BART: Bayesian additive regression trees. *Ann Appl Stat* 2010;**4**:266–98. doi:10.1214/09-AOAS285

53 Breiman L. Random forests. *Mach Learn* 2001;**45**:5–32. doi:10.1023/A:1010933404324

54 Friedman JH. Greedy function approximation: A gradient boosting machine. *Ann Stat* 2001;**29**:1189–232. doi:10.2307/2699986

55 Cover T, Hart P. Nearest neighbor pattern classification. *IEEE Trans Inf Theory* 1967;**13**:21–7. doi:10.1109/TIT.1967.1053964

56 Cortes C, Vapnik V. Support-Vector Networks. *Mach Learn* 1995;**20**:273–97. doi:10.1023/A:1022627411411

57 Ayat S, Farahani HA, Aghamohamadi M, *et al.* A comparison of artificial neural networks learning algorithms in predicting tendency for suicide. *Neural Comput Appl* 2013;**23**:1381–6. doi:10.1007/s00521-012-1086-z

58 van der Laan MJ, Polley EC, Hubbard AE. Super Learner. *Stat Appl Genet Mol Biol* 2007;**6**. doi:10.2202/1544-6115.1309

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. doi:10.1155/2017/7847531

60 Kessler RC, Stein MB, Petukhova M V., *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. doi:10.1038/mp.2016.110

61 Universitat Pompeu Fabra (UPF), Barcelona S. UPF Open Access Publishing. 2019.https://guiesbibtic.upf.edu/acces-obert/eng

Table 1. International Classification of Diseases Codes to identify potentially missed cases of suicide attempt.

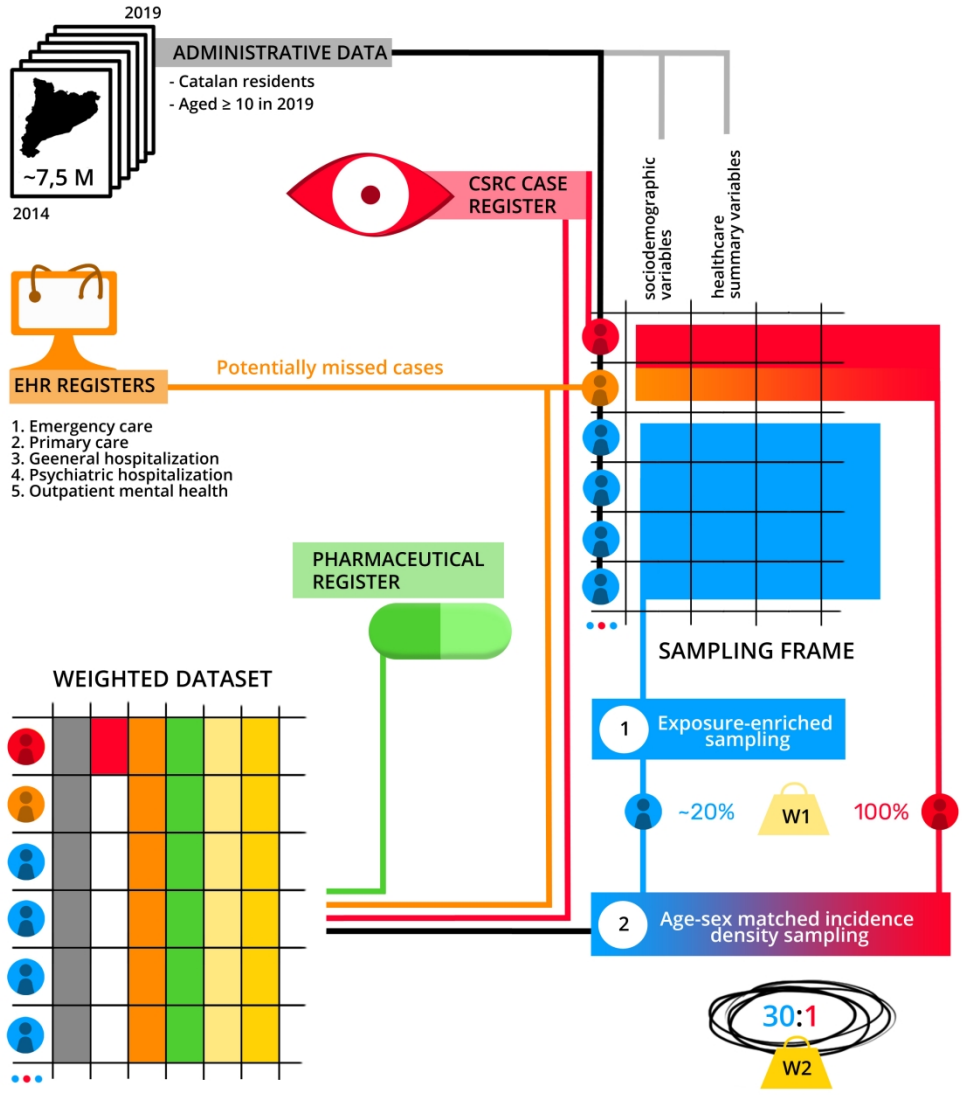
	ICD-9-CM	Description	ICD-10-CM	Description
suicide attempts and intentional self-injurious behavior	**E950*-E959***	self-poisoning, hanging, strangulation, suffocation, fire arms, jumping, others	**X71*–X83***	drowning and submersion, firearms, explosive or thermal material, sharp or blunt objects, jumping 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 6th character of 2	asphyxiation, suffocation, strangulation
			T14.91	suicide attempt
self-injurious behavior 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 6th 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 nonmedicinal substances

Note: ICD = International Classification of Diseases; CM = Clinical Modification.

Figure 1. Overview of the CSRC-Epi study design

[INSERT IMAGE]

Note: M = million; EHR = Electronic Healthcare Record; W = weight.



218x249mm (300 x 300 DPI)

STROBE 2007 (v4) Statement—Checklist of items that should be included in reports of *cohort studies*

Section/Topic	Item #	Recommendation	Reported on page #
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	4
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	5
Objectives	3	State specific objectives, including any prespecified hypotheses	6
Methods			
Study design	4	Present key elements of study design early in the paper	6,8
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	6-8
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up	8,9
		(*b*) For matched studies, give matching criteria and number of exposed and unexposed	9,10
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	6-8
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	6-8
Bias	9	Describe any efforts to address potential sources of bias	9,10
Study size	10	Explain how the study size was arrived at	8
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	10,11
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	10,11
		(*b*) Describe any methods used to examine subgroups and interactions	10,11
		(*c*) Explain how missing data were addressed	10,11
		(*d*) If applicable, explain how loss to follow-up was addressed	NA
		(*e*) Describe any sensitivity analyses	NA
Results			

Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	8-10
		(b) Give reasons for non-participation at each stage	NA
		(c) Consider use of a flow diagram	Figure 1
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	8
		(b) Indicate number of participants with missing data for each variable of interest	NA
		(c) Summarise follow-up time (eg, average and total amount)	NA
Outcome data	15*	Report numbers of outcome events or summary measures over time	NA
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	NA
		(*b*) Report category boundaries when continuous variables were categorized	NA
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	NA
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	NA
Discussion			
Key results	18	Summarise key results with reference to study objectives	NA
Limitations			
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	12
Generalisability	21	Discuss the generalisability (external validity) of the study results	12
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	3

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

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