











Sex-Specific Risk Profiles for Suicide Among Persons with Substance Use Disorders in Denmark

Rachel Sayko Adams^{1,2} , Tammy Jiang³ , Anthony J. Rosellini⁴ , Erzsébet Horváth-Puhó⁵ , Amy E. Street^{6,7} , Katherine M. Keyes⁸ , Magdalena Cerdá⁹ , Timothy L. Lash¹⁰ , Henrik Toft Sørensen^{3,5}  & Jaimie L. Gradus^{3,5} 

Institute for Behavioral Health, Heller School for Social Policy and Management, Brandeis University, Waltham, MA, USA,¹ Rocky Mountain Mental Illness Research Education and Clinical Center, Veterans Health Administration, Aurora, CO, USA,² Department of Epidemiology, Boston University School of Public Health, Boston, MA, USA,³ Center for Anxiety and Related Disorders, Department of Psychological and Brain Sciences, Boston University, Boston, MA, USA,⁴ Department of Clinical Epidemiology, Aarhus University Hospital, Aarhus N, Denmark,⁵ National Center for PTSD, VA Boston Healthcare System, Boston, MA, USA,⁶ Department of Psychiatry, Boston University School of Medicine, Boston, MA, USA,⁷ Department of Epidemiology, Columbia University Mailman School of Public Health, New York, NY, USA,⁸ Center for Opioid Epidemiology and Policy, Department of Population Health, NYU Grossman School of Medicine, New York, NY, USA,⁹ and Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA, USA¹⁰

ABSTRACT

Background and Aims Persons with substance use disorders (SUDs) are at elevated risk of suicide death. We identified novel risk factors and interactions that predict suicide among men and women with SUD using machine learning. **Design** Case-cohort study. **Setting** Denmark. **Participants** The sample was restricted to persons with their first SUD diagnosis during 1995 to 2015. Cases were persons who died by suicide in Denmark during 1995 to 2015 ($n = 2774$) and the comparison subcohort was a 5% random sample of individuals in Denmark on 1 January 1995 ($n = 13\,179$). **Measurements** Suicide death was recorded in the Danish Cause of Death Registry. Predictors included social and demographic information, mental and physical health diagnoses, surgeries, medications, and poisonings. **Findings** Persons among the highest risk for suicide, as identified by the classification trees, were men prescribed antidepressants in the 4 years before suicide and had a poisoning diagnosis in the 4 years before suicide; and women who were 30+ years old and had a poisoning diagnosis 4 years before and 12 months before suicide. Among men with SUD, the random forest identified five variables that were most important in predicting suicide; reaction to severe stress and adjustment disorders, drugs used to treat addictive disorders, age 30+ years, antidepressant use, and poisoning in the 4 prior years. Among women with SUD, the random forest found that the most important predictors of suicide were prior poisonings and reaction to severe stress and adjustment disorders. Individuals in the top 5% of predicted risk accounted for 15% of all suicide deaths among men and 24% of all suicides among women. **Conclusions** In Denmark, prior poisoning and comorbid psychiatric disorders may be among the most important indicators of suicide risk among persons with substance use disorders, particularly among women.

Keywords Alcohol-related disorders, Denmark, machine learning, poisoning, substance use disorder, suicide.

Correspondence to: Rachel Sayko Adams, Institute for Behavioral Health, Heller School for Social Policy and Management, Brandeis University/ MS035, 415 South Street, Waltham, MA 02454-9110, USA. E-mail: radams@brandeis.edu

Submitted 1 September 2020; initial review completed 14 December 2020; final version accepted 10 February 2021

INTRODUCTION

Suicide is a global public health problem with nearly 800 000 suicide deaths annually [1]. Substance use disorders (SUDs) have been identified as a critical determinant of suicide risk [2–4]. SUDs are commonly comorbid with other psychiatric diagnoses such as depression, stress, and personality disorders, which independently confer suicide risk [5,6]. Suicide risk varies by type of SUD and sex [2,7], with

evidence of risk greater for men and strongest for alcohol- and opioid-related disorders [3,8,9]. Previous studies examined the association of SUD and suicide risk among predominantly male samples from the military or veterans [7,10,11]. Other studies explored the association between alcohol-related disorders and/or other SUDs with suicide risk [6,12,13], the association of nicotine dependence with or without other SUDs with suicide risk [14,15], with little consideration for how each SUD, or the presence of unique

combinations of SUDs, may increase risk [2]. One study of persons who died by suicide in Norway between 2009 and 2016 found that women were more likely to be diagnosed with sedative, hypnotic, or anxiolytic use disorder in the year before death compared to men, with no significant differences for prevalence of other types of SUD diagnoses by sex [16]. To date, information is limited concerning unique sex-specific risk factors for suicide death among the high-risk population of individuals with SUD.

Despite decades of research, suicide prediction has not improved significantly in clinical settings [17]. Suicide risk arises from multiple predisposing and concurrent risk factors [8], and conventional statistical methods used to identify patients at high risk have not been well suited to capture interactions among potentially correlated risk factors. Recently, novel supervised machine learning methods have allowed modeling of a broad constellation of predictors simultaneously to enhance suicide risk prediction [18]. Gradus *et al.* [19] were the first to use machine learning methods to identify sex-specific risk profiles for suicide among a complete civilian population. They found that SUD-related factors (e.g. alcohol-related disorders, prior poisoning) were among the most important predictors of sex-specific suicide risk in Denmark, and that ~20% of men and women who died by suicide had a SUD [19]. Risk factors for suicide may differ in the high-risk population of persons with SUD compared to the general population; however, this remains largely unexplored.

Efficacy of suicide assessment and prevention interventions delivered specifically to the SUD population depends on specific knowledge of suicide risk factors within this population, use of large samples, and an analytic approach capable of considering large numbers of potential predictor variables [20]. The study's goal was to use population-based, prospective Danish medical and social registry data and machine learning methods to identify sex-specific risk profiles for suicide among persons with SUD. To ensure accurate prediction of suicide death, the machine learning models were stratified by sex at birth, reflecting well-documented sex differences in the incidence of suicide and suicide risk factors [21–23]. Because individual SUDs were included as predictors, we could observe whether interactions among different SUDs improve risk prediction. To our knowledge, this is the first study to use machine learning techniques among the high-risk population of individuals with SUD to predict suicide death.

METHODS

Study sample

All residents of Denmark are provided universal medical coverage through a tax-funded healthcare system [24]. The source population was all persons born or legally residing in Denmark on 1 January 1995. The start of the study

period was chosen to correspond with, (i) the switch from International Classification of Diseases, Eighth Revision to International Classification of Diseases, Tenth Revision (ICD-10) in 1994; and (ii) all hospital outpatient clinic visits were reported to the Danish National Patient Registry starting in 1995 [25]. We used a case-cohort design efficient for examining rare outcomes [26]. Suicide cases were all individuals who had an incident SUD diagnosis and died by suicide between 1995 and 2015 in Denmark ($n = 2774$). The comparison subcohort was a 5% random sample of individuals living in Denmark on 1 January 1995 who had an incident SUD diagnosis during the study period ($n = 13\ 179$). We linked data across administrative and medical registries using Denmark's Civil Personal Register numbers; unique identifiers assigned to all Danish residents (see Supporting information Data S1 Appendix Table 1 for Danish registry information) [27–29].

Outcome

The Danish Cause of Death Registry records age of death, manner of death (e.g. unintentional, suicide), place of death, and autopsy results [30]. We used this registry to ascertain suicide cases using ICD-10 codes X60–X84 [30]. In a separate independent expert review, 92% of deaths recorded as suicides were confirmed [31].

Substance use disorders

We obtained SUD diagnoses from two registries using two-digit ICD-10 codes: [32] the Danish Psychiatric Central Research Register records all psychiatric inpatient and outpatient diagnostic data [33], and the Danish National Patient Registry contains data on all inpatient hospitalizations in non-psychiatric hospitals, hospital outpatient visits, and emergency room visits [25]. SUDs were identified using the following ICD-10 codes in primary or secondary positions; alcohol-related disorders (F10), opioid-related disorders (F11), cannabis-related disorders (F12), sedative, hypnotic, or anxiolytic-related disorders (F13), cocaine-related disorders (F14), other stimulant-related disorders (F15), hallucinogen-related disorders (F16), nicotine dependence (F17), inhalant-related disorders (F18), and other psychoactive substance-related disorders (F19).

Other predictors

In addition to SUD diagnostic codes, we examined the following predictors of suicide; age, marital status, immigrant status, family suicidal behavior (i.e. parent or spouse), employment, income, mental health disorders, physical health disorders, surgeries, prescription drugs, psychological services, and poisonings (defined as adverse effects of and underdosing of drugs, medicaments, and biological

substances). We obtained data on age, marital status, immigrant status, and family suicidal behavior from the Danish Civil Registration System and Cause of Death Registry [28]. We obtained data on employment and income from the Integrated Database for Labor Market Research [34] and Income Statistics Register [35]. We used the Danish Psychiatric Central Research Register [33] and Danish National Patient Registry to obtain psychiatric disorder diagnoses using two-digit ICD-10 codes [25]. We also used the Danish National Patient Registry to obtain physical health diagnoses recorded using second-level ICD-10 groupings, and examined surgery procedure codes by body system. Prescription drug data were from the Danish National Prescription Registry [36,37], including dispensing date, product name, and level 3 Anatomical Therapeutic Classification codes (e.g. drugs used in addictive disorders [N07B]). We obtained data on psychological services (i.e. any encounter for a psychological service) from the Health Insurance Registry [38].

Statistical analysis

Consistent with other suicide-related machine learning studies [39–41], we dummy-coded variables to create time-varying predictors with intervals of 0–6, 0–12, 0–24, and 0–48 months before the date of death for suicide cases. To estimate the prevalence of each predictor during the person-time that gave rise to cases, we randomly selected a date for each member of the subcohort between the SUD diagnosis date and the end of follow-up and computed the prevalence of predictors 0–6, 0–12, 0–24, and 0–48 months before the selected date. Some predictors (e.g. sex, age, and immigrant status) were examined in their registry-based form without being dummy-coded as time-varying predictors. Employment and income were defined in the year before the date of death for suicide cases and in the year before the selected date between SUD diagnosis and end of follow-up for comparison subcohort members. Predictors from all time points were evaluated together.

We reduced the predictor set to mitigate the risk of overfitting, which arises when a model finds trivial patterns that are unique to a specific data set, but are not generalizable and fails to accurately predict events in external samples [42]. We performed data reduction separately for men and women using two steps: (i) removing rare predictors with fewer than 10 observations in any cell of a 2×2 contingency table of the predictor and suicide [43,44] and (ii) removing predictors with negligible main effect associations with suicide (unadjusted odds ratios between 0.9 and 1.1). The initial data set contained 2563 predictor variables. After data reduction, the final number of included predictors was 833 for men and 686 for women. Some SUD diagnoses were eliminated as predictors (e.g.

hallucinogen-related disorders for all time periods for men). Supporting information Data S1 Appendix Table 2 presents the predictors pre- and post-data reduction.

Given our dual interests in identifying novel predictors and interactions of suicide, we used classification trees and random forests, which are recursive partitioning methods that can automatically detect associations and interactions among predictors to optimize prediction accuracy and to provide metrics of predictor importance [45]. These methods were chosen over other machine learning methods (e.g. elastic net penalized regression) [46] to enhance interpretability of the variables used to the model and to visualize interactions. The classification tree model (CART) is a nonparametric method that builds a decision tree based on predictors and their combinations that result in the highest probability of differentiating cases from non-cases. We performed 10-fold cross-validation for the classification trees, setting the maximum tree depth and minimum number of observations in any node (terminal or parent) to five. To mitigate class imbalance, each individual tree was built using equal priors [47]. Risk of suicide was calculated for each identified combination of predictors. We used the R package rpart [48]. Although classification trees provide interpretable, visual representations of interactions that permit identification of specific combinations of risk factors associated with suicide risk, they are vulnerable to overfitting. Therefore, we used random forests that are less prone to overfitting to identify novel predictors of suicide among persons diagnosed with a SUD.

Random forests are another recursive partitioning method that ensembles a set of decision trees created using bootstrapped samples of the data. Each forest was built with 1000 trees. The number of variables randomly sampled at each split was 29 for men and 26 for women (i.e. square root of the total number of predictors for men and women; randomForest default). A minimum of 10 observations were needed to attempt a split. Given class imbalance, each tree was built using all suicide observations and an equally sized number of randomly selected non-suicide observations using the sampsize tuning parameter [49,50]. We used 2-fold cross-validation to generate individual-level random forests-predicted values. To evaluate the importance of each variable in predicting suicide, we inspected the mean decrease in accuracy values, which represents the reduction in accuracy if a predictor were randomly permuted [45]. Predictors that are more important for accurate suicide prediction will have a larger mean decrease in accuracy. We used the R package randomForest [51].

We evaluated prediction accuracy (i.e. discrimination) using receiver operating characteristics curve analysis conducted in 1000 bootstrap replicates to estimate the area under the curve (AUC) and 95% confidence intervals (CI)

[52]. The AUC represents the overall ability of the classifier to distinguish between subjects who died by suicide versus those who did not. We conducted analyses separately for men and women. Sex-stratified analyses were used instead of including sex as a predictor in analyses containing the entire sample (men and women combined). Including sex as a predictor would not necessarily reveal sex differences in the CART or random forest variable importance. Furthermore, the classification trees would only display different patterns of risk by sex at the point that sex is chosen as a splitting variable but not earlier in the tree. Analyses were conducted using SAS, version 9.4 (SAS Institute) [53] and R, version 3.5.2 [54]. The analysis was not pre-registered and the results should be considered exploratory.

RESULTS

Of the 2774 persons with SUD who were suicide cases, 1985 were men (72%) and 789 were women (28%; Table 1). Of the 13 179 persons with SUD in the comparison subcohort, 8223 were men (62%) and 4956 were women (38%). Age distributions in the suicide cases and comparison subcohorts were similar (mean [SD] for men: 48 [14] years vs. 50 [17] years; women: 52 [13] years vs. 51 [18] years). Persons who died by suicide were more frequently divorced, widowed, never married, or single than subcohort members. Suicide cases were less likely to be in the highest income quartile than the comparison subcohort; (men: 466 [24%] vs. 2232 [27%]; women: 191 [24%] vs. 1424 [29%]).

Table 1 Characteristics of the suicide cases and members of the comparison subcohort, Denmark, 1995 to 2015.

Variable	Men		Women	
	Suicide cases (n = 1985)	Comparison subcohort (n = 8223)	Suicide cases (n = 789)	Comparison subcohort (n = 4956)
Age, mean (SD), y	48 (14)	50 (17)	52 (13)	51 (18)
Marital status (%)				
Married or registered partnership	472 (24)	2232 (27)	220 (28)	1469 (30)
Divorced/widowed/never married/single	1513 (76)	5963 (73)	569 (72)	3475 (70)
Immigrant (%)	70 (3.5)	330 (4.0)	14 (1.8)	168 (3.4)
Income quartile (%)				
<1	504 (25)	1919 (23)	197 (25)	1121 (23)
1 to <2	493 (25)	2031 (25)	199 (25)	1188 (24)
2 to <3	522 (26)	2041 (25)	202 (26)	1223 (25)
≥3	466 (24)	2232 (27)	191 (24)	1424 (29)

Table 2 Distribution of substance use disorders between 1 January 1995 and the index date among suicide cases and members of the comparison subcohort.

Variable (ICD-10 code)	Men		Women	
	Suicide cases (n = 1985)	Comparison subcohort (n = 8223)	Suicide cases (n = 789)	Comparison subcohort (n = 4956)
Alcohol-related disorders (F10)	1624 (82%)	6064 (74%)	610 (77%)	2784 (56%)
Opioid-related disorders (F11)	152 (7.7%)	259 (3.1%)	73 (9.3%)	210 (4.2%)
Cannabis-related disorders (F12)	237 (12%)	634 (7.7%)	50 (6.3%)	209 (4.2%)
Sedative, hypnotic, or anxiolytic-related disorders (F13)	142 (7.2%)	210 (2.6%)	170 (22%)	313 (6.3%)
Cocaine-related disorders (F14)	46 (2.3%)	100 (1.2%)	8 (1.0%)	26 (0.5%)
Other stimulant-related disorders (F15)	67 (3.4%)	158 (1.9%)	18 (2.3%)	45 (0.9%)
Hallucinogen-related disorders (F16)	–	19 (0.2%)	–	12 (0.2%)
Nicotine dependence (F17)	114 (5.7%)	1411 (17%)	50 (6.3%)	1636 (33%)
Other psychoactive substance-related disorders (F19)	286 (14%)	545 (6.6%)	126 (16%)	230 (4.6%)

We do not display values with small cell counts (≤5 observations) to protect study subjects' anonymity. There were no patients with ICD code F18 (inhalant-related disorders).

Alcohol-related disorder was the most common SUD among both cases and controls, but prevalence was higher among cases; (men: 1624 [82%] vs. 6064 [74%]; women: 610 [77%] vs. 2784 [56%]; Table 2). Among persons who died by suicide, the most common other SUDs were other psychoactive substance-related disorders (men: 14%; women: 16%), sedative, hypnotic, or anxiolytic-related disorders (men: 7.2%; women: 22%), cannabis-related disorders (men: 12.0%; women: 6.3%), and opioid-related disorders (men: 7.7%; women: 9.3%). Suicide cases were less likely to have nicotine dependence compared to the comparison subcohort: (men: 114 [5.7%] vs. 1411 [17%]; women: 50 [6.3%] vs. 1636 [33%]).

Classification trees

Men with the highest risk of suicide were those prescribed antipsychotics in the 4 years before their suicide and diagnosed with a brief psychotic disorder 2 years before the suicide, but who were not prescribed antidepressants or diagnosed with poisoning or schizophrenia ($n = 20$; suicide risk = 86%; Fig. 1). Men who were prescribed antidepressants and had a poisoning in the 4 years before suicide had a similar risk of suicide ($n = 533$; suicide risk = 85%).

Men who were prescribed drugs used to treat addictive disorders (e.g. buprenorphine), who were diagnosed with reaction to severe stress and with adjustment disorder in the previous 4 years ($n = 23$), but who were not prescribed antidepressants, antipsychotics, or anxiolytics, had an 84% suicide risk. Overall discrimination of cases from non-cases was good (AUC = 0.75, 95% CI = 0.73, 0.76) [55].

Among women with SUD, the highest risk of suicide occurred in those who were over age 30 years, had a poisoning diagnosis in the 4 years before the suicide, and a poisoning diagnosis in the 12 months before the suicide ($n = 345$, suicide risk = 92%; Fig. 2). Women with SUD under age 30 years who had a poisoning diagnosis in the previous 4 years and received a prescription for antipsychotics in the previous 6 months had the next highest risk ($n = 31$; suicide risk = 85%). This AUC indicates excellent discrimination (AUC = 0.86, 95% CI = 0.84, 0.88) [55].

Random forest

Among men, 76% and 74% of predictors had a mean decrease in accuracy values above zero in folds 1 and 2, respectively (combined mean = 4.6, SD = 3.9). Eleven predictors were among the top 30 most important

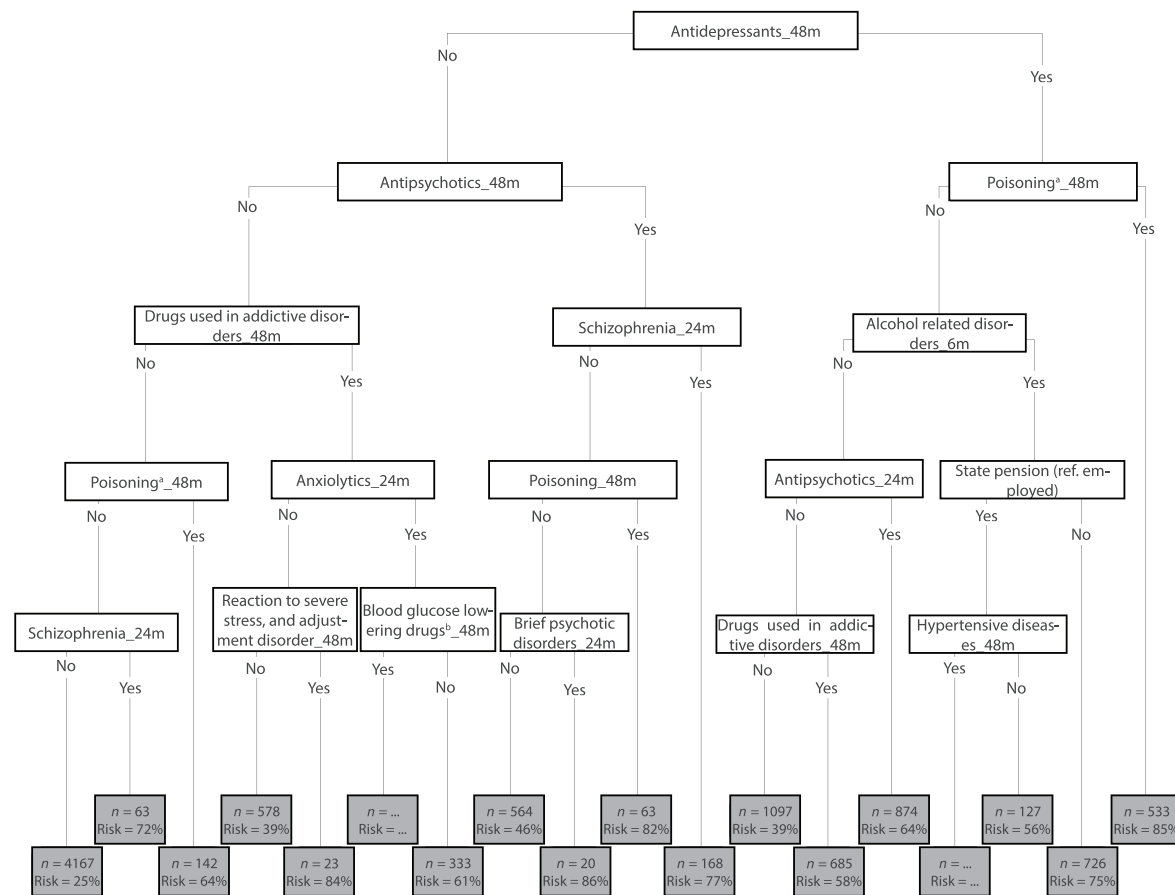


Figure 1 Classification tree depicting suicide predictors among men with substance use disorders in Denmark, 1995 – 2015

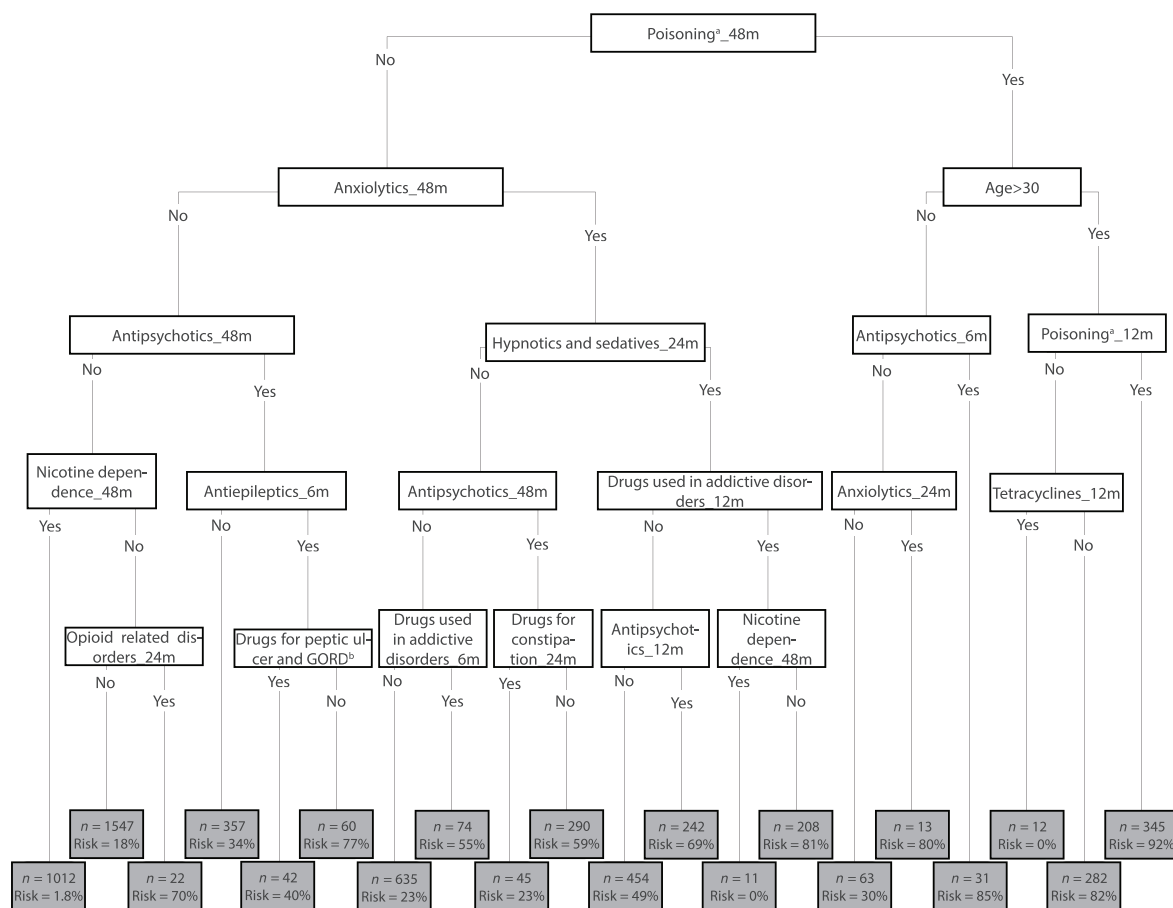


Figure 2 Classification tree depicting suicide predictors among women with substance use disorders in Denmark, 1995 – 2015

predictors in both folds (Fig. 3). The most important variables for predicting suicide included reaction to severe stress and adjustment disorders; drugs used to treat addictive disorders; age greater than 30 years; antidepressant use; and poisoning. The cross-validated AUC for the random forest was 0.77 (95% CI = 0.76, 0.78).

Among women, 77% to 78% (fold 1-fold 2) of predictors had a mean decrease in accuracy for values above zero, (i.e. model performance would have been compromised through their omission; combined mean = 3.0, SD = 2.7). Ten predictors were among the top 30 most important predictors in both folds, with most being poisoning and psychiatric disorders (Fig. 4). Poisoning across all time periods was the most important predictor of suicide. Reaction to severe stress and adjustment disorders; sedative, hypnotic, or anxiolytic-related disorders; use of anticholinergic agents; and other psychoactive substance-related disorders also emerged as important predictors. The cross-validated AUC for the random forest was 0.86 (95% CI = 0.85, 0.88).

Operating characteristics of high-risk thresholds

Cross-validated random forests-predicted probabilities were rank ordered and operating characteristics were computed

among individuals in the top quintile of the predicted risk distribution. Men in the top 5%, 10%, and 20% of predicted suicide risk accounted for 15%, 28%, and 48% of all suicide cases among men with SUD, respectively. Men in the bottom 95%, 90%, and 80% of predicted suicide risk accounted for 97%, 94%, and 87% of all men with SUD who did not die by suicide, respectively. Women in the top 5%, 10%, and 20% of predicted suicide risk accounted for 24%, 41%, and 66% of all suicide cases among women with SUD, respectively. Women in the bottom 95%, 90%, and 80% of predicted risk accounted for 98%, 95%, and 87% of all women with SUD who did not die by suicide, respectively.

DISCUSSION

Earlier studies have identified persons with SUD at increased risk for death by suicide, with evidence of suicide risk varying by type of SUD and by sex [2–4,7,8]. Our analytic sample of persons with SUD who died by suicide in Denmark during 1995 to 2015 was drawn from the larger population-based sample reported in Gradus *et al.* [19] We found that in addition to previously documented risk factors for suicide related to psychiatric conditions, drugs used

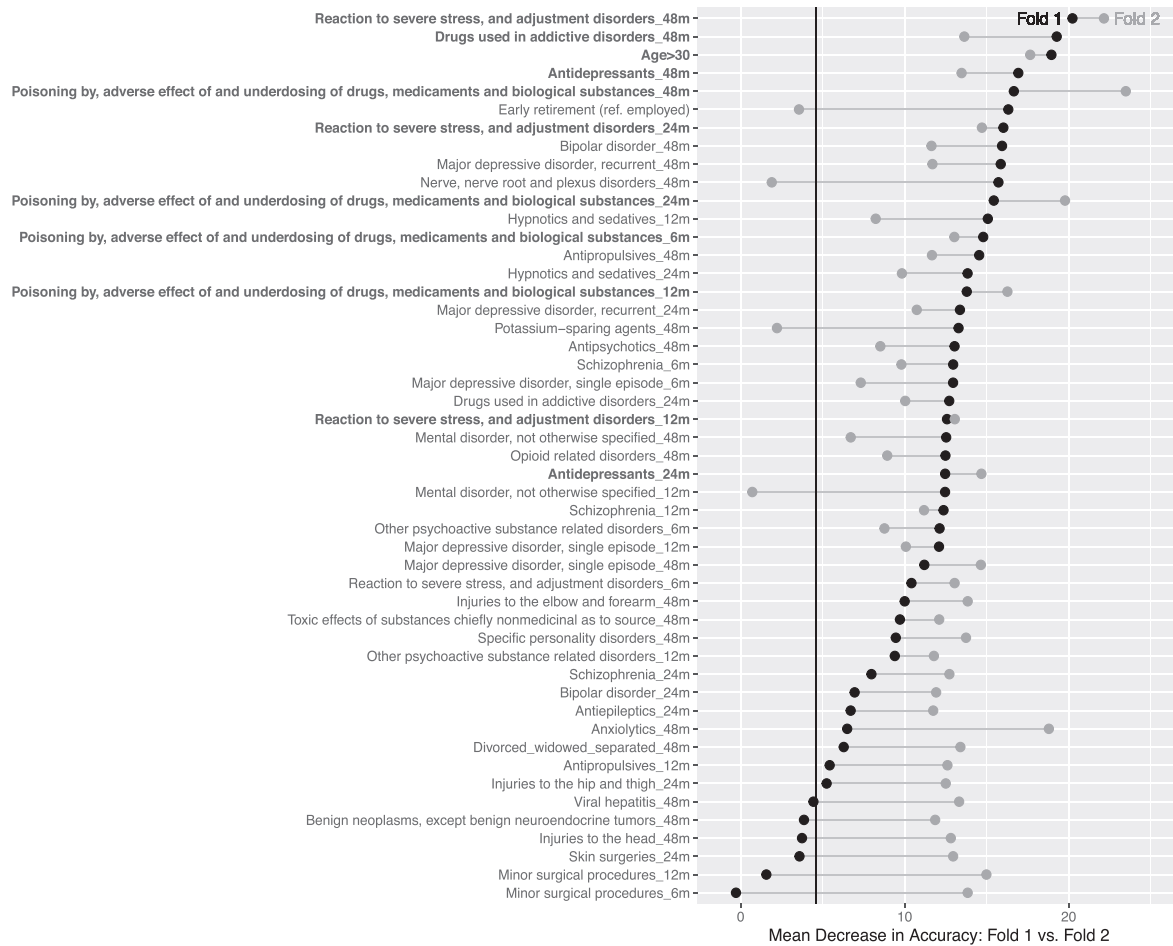


Figure 3 Variable importance of suicide predictors among men with substance use disorders in Denmark from split sample cross-validation, 1995 – 2015

to treat addictive disorders were a particularly important risk factor for men with SUD. Evidence-based medication treatments are available for treatment of addiction to opioids, alcohol, and nicotine [56], which may suggest that these treatments are a marker for underlying SUDs that are especially predictive of suicide risk. More research is needed to examine if narrower classifications of medication treatments (e.g. buprenorphine vs. methadone) are more predictive of suicide risk. For women, the most important predictors of suicide were poisoning in all four time periods, indicating that incidents of poisoning are an acute marker of future suicide risk among women with SUD—a risk that may persist for many years after the poisoning with clinical implications for the emergency department where many persons present during an overdose [57].

For men with SUD, the most important predictors were similar to those found in the parent study [19] of all Danish men who died by suicide. However, important physical conditions were unique to the SUD sample (e.g. injuries to elbow/forearm). Importantly, we were able to predict a larger proportion of suicide deaths among women with SUD compared to men. For women with SUD, the most

important predictors were related to prior poisoning. In comparison, in the parent study of all Danish women who died by suicide, there was more heterogeneity in the leading predictors for suicide for women, including antipsychotic use, prior suicide attempt, and psychiatric disorders [19]. Our findings point toward previous poisonings and psychiatric diagnoses as particularly important targets for future research and suicide prevention among women with SUD diagnoses. As well, sex-specific suicide risk assessment within the population with SUD may be warranted.

As discussed, we found that prior poisoning was one of the most important predictors of suicide risk among persons with SUD, particularly among women. The definition of this code set (T36-T50) is poisoning by, adverse effects of, and underdosing of drugs, medicaments, and biological substances. Therefore, this broad risk indicator may include non-fatal drug overdoses, inclusive of accidental, intentional self-harm, assault, or undetermined intent, or poisoning because of adverse effects or underdosing. Poisoning due to adverse effects or underdosing may be less important to suicide risk prediction; although the codes

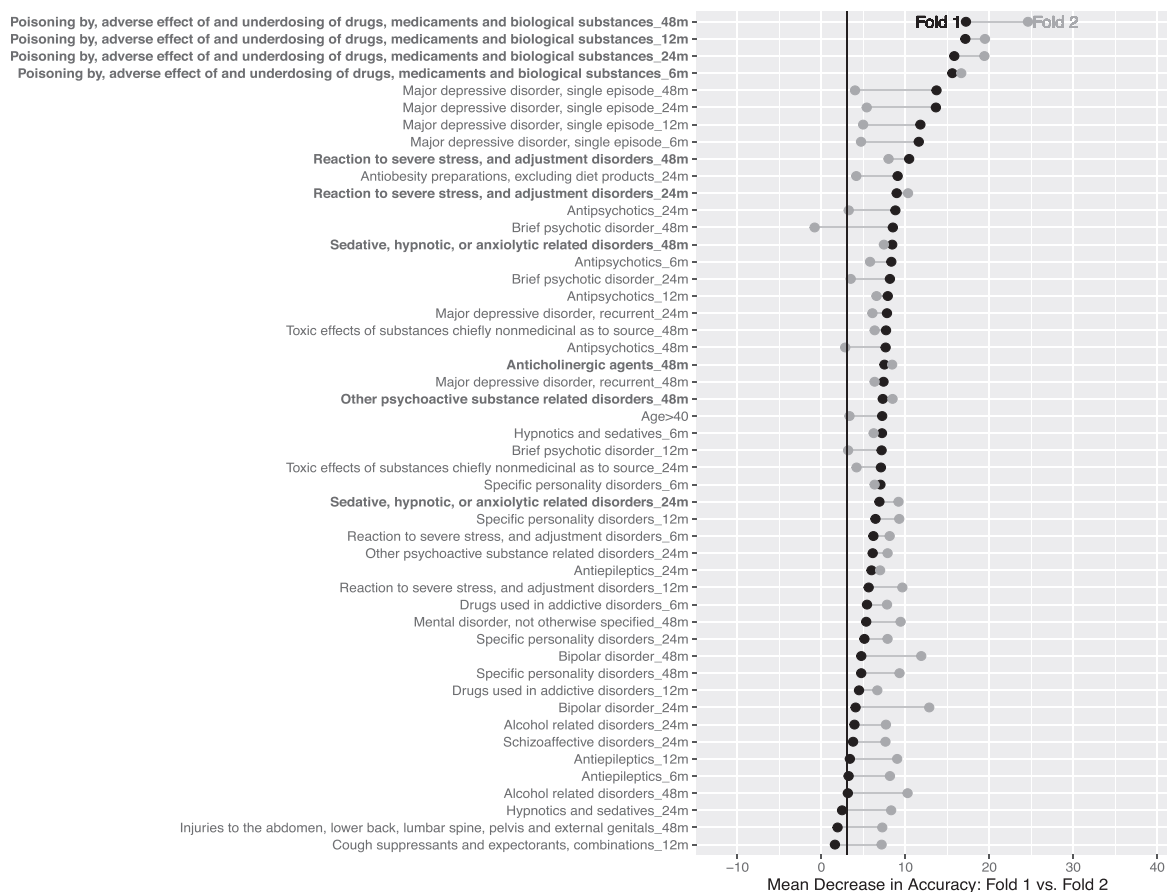


Figure 4 Variable importance of suicide predictors among women with substance use disorders in Denmark from split sample cross-validation, 1995-2015

were collapsed in the models and this could not be examined. In addition to non-fatal accidental overdose, the poisoning indicator may have captured prior suicide attempts by poisoning. It is noteworthy that in the parent study [19], suicide attempt was an important predictor for both men and women, whereas in our study prior suicide attempt was eliminated during data reduction because of small sample cells. It is possible that in the subgroup with SUD, poisoning may be a common method of suicide attempt. Yet, in people with SUD, suicide attempts may be more likely to be coded as poisoning if a substance is used for which they have a known SUD (e.g. alcohol), because of additional uncertainty around intent. Therefore, prior poisoning may reflect a suicide attempt for some individuals in our sample. In a study of people who died by suicide in Norway between 2009 and 2016 who had contact with substance use treatment services within the year before their death, women were more likely than men to die by poisoning as a method of suicide (AOR 1.81, 95% CI = 1.09–3.02) [16]. More research is needed to disentangle how indicators of suicide risk behavior may be present within non-fatal poisoning [58–60], particularly among persons with documented SUD. Our findings suggest that prior poisoning for any reason may be among the

strongest signals of suicide risk among persons with SUD [57], and targeted post-poisoning intervention strategies should be developed for this population, particularly for women.

We did not find evidence that combinations of SUDs had high importance for suicide risk. A partial explanation may be that most persons who died by suicide had alcohol-related disorders. A meta-analysis concluded that over one-third of suicide deaths were preceded by acute alcohol use [61], a proportion that may be even greater among persons with a SUD who died by suicide. These findings underscore the importance of addressing at-risk alcohol use as a focus of suicide prevention efforts [62] for both the general population and the high-risk population of persons with SUD. Sedative, hypnotic, or anxiolytic-related disorders and other psychoactive substance-related disorders were identified as important SUD-specific risk factors, albeit demonstrating lower accuracy regarding suicide risk prediction. This may highlight important future research avenues. Contrary to previous studies [14,15], our study revealed in the CART results for women, and the random forest results, that there was no evidence that nicotine dependence meaningfully predicted increased suicide risk.

Limitations

Our study should be interpreted in light of several limitations. There may be potential measurement error of suicide deaths such that suicides are under-recorded and may be misrecorded as deaths of undetermined intent, accidents, or ill-defined and unknown cause of mortality [31]. A validation study of suicide deaths in Denmark found that 92% of recorded suicide deaths were confirmed after expert review [31]. Furthermore, given the rarity of suicide, low specificity of suicide classification will lead to greater bias than low sensitivity [63]. Specificity of suicide classification is believed to be high because the majority of persons who did not die by suicide are correctly classified as such. Therefore, we expect any misclassification of suicide deaths to have a limited biasing effect in our study. Although Danish registries collect high-quality data with long-term follow-up, like other population-based epidemiologic suicide risk studies, included predictors may be subject to misclassification, although concerns about biased results due this misclassification is assuaged for reasons outlined above. Further, our ability to include predictors was limited by data availability. We lacked information on quality of life in the days and hours before a suicide. Because we lacked information characterizing suicide method, we are unable to provide information about predictors of suicide method (e.g. suicide by poisoning) among persons with SUD, which may be particularly informative for safety interventions regarding lethal means. Future research should consider using more discreet categories of ICD-10-CM codes in this broad code set (T36-T50) to better understand risk specific to non-fatal drug overdoses, inclusive of accidental, intentional self-harm, assault, or undetermined intent, or poisoning because of adverse effects or underdosing. Our findings should not be interpreted as causal effects because our models are non-causal and cannot clarify the direction or magnitude of associations. It is unclear if these results are generalizable to other populations with SUD outside of Denmark, where there are differences in the distribution of SUDs, access and approaches to substance use treatment, and stigma associated with SUD. For instance, although Denmark has seen a dramatic increase in prescription opioid use and ranks fifth in global opioid consumption [64], other countries such as the United States have experienced greater morbidity and mortality associated with the opioid epidemic [65].

CONCLUSIONS

This is the first study, to our knowledge, to use machine learning techniques with a population-based sample of individuals with SUD to enhance sex-specific suicide risk prediction. Among the high-risk subgroup of persons with SUD who died by suicide in Denmark, we found that prior

poisoning was a critical indicator of suicide risk, particularly among women, and that suicide risk may continue for many years after the poisoning. Therefore, poisoning among persons with SUD may be an important signal for clinicians to initiate more proactive and ongoing assessment and for researchers to pilot and test new interventions to reduce future death by suicide. Our findings are consistent with prior studies, which suggest that psychiatric disorders confer additional risk for suicide among the high-risk subgroup of persons with SUD. Future studies are needed to replicate these analyses to inform suicide risk prediction among persons with SUD and to investigate whether SUD type is associated with suicide method (e.g. poisoning) to inform lethal means safety interventions.

Declaration of interests

None.

Acknowledgements

This work was supported by National Institute of Mental Health (NIMH) grant R01MH109507 (PI: J.L.G.), and grant R248-2017-521 from the Lundbeck Foundation (PI: H.T.S.). The funding sources had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication. The authors do not have any conflicts of interest to disclose. E.H.-P. and H.T.S. contributed to the acquisition of data. E.H.-P., A.J.R., and J.L.G. contributed to data analysis and all authors take responsibility for data interpretation.

Author contributions

Rachel Adams: Conceptualization; investigation. **Tammy Jiang:** Investigation; methodology; project administration; visualization. **Anthony Rosellini:** Conceptualization; investigation; methodology; validation; visualization. **Erzsébet Horváth-Puhó:** Conceptualization; data curation; formal analysis; funding acquisition; investigation; methodology; supervision; validation; visualization. **Amy Street:** Conceptualization; funding acquisition; investigation; methodology. **Katherine Keyes:** Investigation. **Magdalena Cerda:** Investigation. **Timothy Lash:** Conceptualization; funding acquisition; investigation; methodology. **Henrik Sørensen:** Conceptualization; data curation; funding acquisition; investigation; methodology; software; validation; visualization. **Jaimie Gradus:** Conceptualization; data curation; formal analysis; funding acquisition; investigation; methodology; project administration; resources; software; supervision; validation; visualization.

References

- World Health Organization. Suicide data [Internet]. World Health Organization. [cited 2019 Aug 20]. Available from: http://www.who.int/mental_health/prevention/suicide/suicideprevent/en/
- Lynch F. L., Peterson E. L., Lu C. Y., Hu Y., Rossom R. C., Waitzfelder B. E., et al. Substance use disorders and risk of suicide in a general US population: a case control study. *Addict Sci Clin Pract* 2020; **15**: 14.
- Wilcox H. C., Conner K. R., Caine E. D. Association of alcohol and drug use disorders and completed suicide: an empirical review of cohort studies. *Drug Alcohol Depend* 2004 Dec 7; **76**: S11–S19.
- Poorolajal J., Haghtalab T., Farhadi M., Darvishi N. Substance use disorder and risk of suicidal ideation, suicide attempt and suicide death: a meta-analysis. *J Public Health (Oxf)* 2016; **38**: e282–e291.
- Windfuhr K., Kapur N. Suicide and mental illness: a clinical review of 15 years findings from the UK National Confidential Inquiry into suicide. *Br Med Bull* 2011; **100**: 101–21.
- Østergaard M. L. D., Nordentoft M., Hjorthøj C. Associations between substance use disorders and suicide or suicide attempts in people with mental illness: a Danish nation-wide, prospective, register-based study of patients diagnosed with schizophrenia, bipolar disorder, unipolar depression or personality disorder. *Addiction* 2017 Jul; **112**: 1250–9.
- Ilgen M. A., Bohnert A., Ignacio I. Psychiatric diagnoses and risk of suicide in veterans. *Arch Gen Psychiatry* 2010; **67**: 1152–8.
- Esang M., Ahmed S. A. Closer look at substance use and suicide. *Am J Psychiatry Resid J* 2018; **13**: 6–8.
- World Health Organization. Global Health Estimates 2016: Deaths by Cause, Age, Sex, by Country and by Region, 2000–2016. Geneva; 2018.
- Bohnert K. M., Ilgen M. A., Louzon S., McCarthy J. F., Katz I. R. Substance use disorders and the risk of suicide mortality among men and women in the US veterans health administration. *Addiction* 2017 Jul; **112**: 1193–201.
- Bohnert K. M., Ilgen M. A., McCarthy J. F., Ignacio R. V., Blow F. C., Katz I. R. Tobacco use disorder and the risk of suicide mortality. *Addiction* 2014 Jan; **109**: 155–62.
- Flenborg-Madsen T., Knop J., Mortensen E. L., Becker U., Sher L., Grønbaek M. Alcohol use disorders increase the risk of completed suicide—irrespective of other psychiatric disorders. A longitudinal cohort study. *Psychiatry Res* 2009 May 15; **167**: 123–30.
- Edwards A., Ohlsson H., Sundquist J., Sundquist K., Kendler K. S. Alcohol use disorder and risk of suicide in a Swedish population-based cohort. *Am J Psychiatry* 2020; **177**: 627–34.
- Yaworski D., Robinson J., Sareen J., Bolton J. M. The relation between nicotine dependence and suicide attempts in the general population. *Can J Psychiatry* 2011 Mar 1; **56**: 161–70.
- Kessler R., Borges G., Sampson N. A., Miller M., Nock M. K. The association between smoking and subsequent suicide-related outcomes in the National Comorbidity Survey panel sample - PubMed. *Mol Psychiatry* 2008; **14**: 1132–42.
- Myhre M. Ø., Kildahl A. T., Walby F. A. Suicide after contact with substance misuse services: a national registry study. *BJPsych Open* 2020 May 7; **6**: e45.
- Franklin J. C., Ribeiro J. D., Fox K. R., Bentley K. H., Kleiman E. M., Huang X., et al. Risk factors for suicidal thoughts and behaviors: a meta-analysis of 50 years of research. *Psychol Bull* 2017; **143**: 187–232.
- Jiang T., Gradus J. L., Rosellini A. J. Supervised machine learning: a brief primer. *Behav Ther* [Internet] 2020 May 16 [cited 2020 Jul 23]. Available from: **51**: 675–87; <http://www.sciencedirect.com/science/article/pii/S0005789420300678>
- Gradus J. L., Rosellini A. J., Horváth-Puhó E., Street A. E., Galatzer-Levy I. R., Jiang T., et al. Prediction of sex-specific suicide risk using machine learning and single-payer health care registry data from Denmark. *JAMA Psychiat* 2020; **77**: 25–34.
- Fazel S., O'Reilly L. Machine learning for suicide research-can it improve risk factor identification? *JAMA Psychiat* [Internet] 2019 Oct 23 [cited 2019 Nov 14]. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/31642876>
- Weissman M. M., Bland R. C., Canino G. J., Greenwald S., Hwu H. G., Joyce P. R., et al. Prevalence of suicide ideation and suicide attempts in nine countries. *Psychol Med* 1999 Jan; **29**: 9–17.
- Morthorst B., Soegaard B., Nordentoft M., Erlangsen A. Incidence rates of deliberate self-harm in Denmark 1994–2011. *Crisis* 2016; **37**: 256–64.
- Miranda-Mendizabal A., Castellví P., Parés-Badell O., Alayo I., Almenara J., Alonso I., et al. Gender differences in suicidal behavior in adolescents and young adults: systematic review and meta-analysis of longitudinal studies. *Int J Public Health* 2019 Mar; **64**: 265–83.
- Schmidt M., Schmidt S. A. J., Adelborg K., Sundbøll J., Laugesen K., Ehrenstein V., et al. The Danish health care system and epidemiological research: from health care contacts to database records. *Clin Epidemiol* 2019; **11**: 563–91.
- Schmidt M., Schmidt S. A. J., Sandegaard J. L., Ehrenstein V., Pedersen L., Sørensen H. T. The Danish National Patient Registry: a review of content, data quality, and research potential. *Clin Epidemiol* 2015; **7**: 449–90.
- Barlow W. E., Ichikawa L., Rosner D., Izumi S. Analysis of case-cohort designs. *J Clin Epidemiol* 1999 Dec; **52**: 1165–72.
- Pedersen C. B., Gotzsche H., Møller J. O., Mortensen P. B. The Danish civil registration system. A cohort of eight million persons. *Dan Med Bull* 2006 Nov; **53**: 441–9.
- Pedersen C. B. The Danish civil registration system. *Scand J Public Health* 2011 Jul; **39**: 22–5.
- Schmidt M., Pedersen L., Sørensen H. T. The Danish civil registration system as a tool in epidemiology. *Eur J Epidemiol* 2014 Aug; **29**: 541–9.
- Helweg-Larsen K. The Danish register of causes of death. *Scand J Public Health* 2011 Jul; **39**: 26–9.
- Tøllefsen I. M., Helweg-Larsen K., Thiblin I., Hem E., Kastrup M. C., Nyberg U., et al. Are suicide deaths under-reported? Nationwide re-evaluations of 1800 deaths in Scandinavia. *BMJ Open* 2015 Nov 1; **5**: e009120.
- Free 2021 ICD-10-CM Codes [Internet]. [cited 2021 Jan 7]. Available from: <https://www.icd10data.com/ICD10CM/Codes>
- Mors O., Perto G. P., Mortensen P. B. The Danish psychiatric central research register. *Scand J Public Health* 2011 Jul; **39**: 54–7.
- Timmermans B. The Danish Integrated Database for Labor Market Research: Towards Demystification for the English Speaking Audience.
- Baadsgaard M., Quitzau J. Danish registers on personal income and transfer payments. *Scand J Public Health* 2011 Jul; **39**: 103–5.

36. Kildemoes H. W., Sørensen H. T., Hallas J. The Danish National Prescription Registry. *Scand J Public Health* 2011 Jul; **39**: 38–41.
37. Pottegård A., Schmidt S. A. J., Wallach-Kildemoes H., Sørensen H. T., Hallas J., Schmidt M. Data resource profile: the Danish National Prescription Registry. *Int J Epidemiol* 201701; **46**: 798798–NaN–798f.
38. Andersen J. S., Olivarius N. D. E., Krasnik A. The Danish National Health Service Register. *Scand J Public Health* 2011 Jul; **39**: 34–7.
39. McCarthy J. F., Bossarte R. M., Katz I. R., Thompson C., Kemp J., Hannemann C. M., et al. Predictive modeling and concentration of the risk of suicide: implications for preventive interventions in the US Department of veterans affairs. *Am J Public Health* 2015 Sep; **105**: 1935–42.
40. Kessler R. C., Hwang I., Hoffmire C. A., McCarthy J. F., Petukhova M. V., Rosellini A. J., et al. Developing a practical suicide risk prediction model for targeting high-risk patients in the veterans health administration. *Int J Methods Psychiatr Res* [Internet] 2017 Sep [cited 2018 Nov 3]. **26**. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5614864/>
41. Kessler R. C., Warner L. C. H., Ivany L. C., Petukhova M. V., Rose S., Bromet E. J., et al. Predicting U.S. Army suicides after hospitalizations with psychiatric diagnoses in the Army study to assess risk and resilience in Servicemembers (Army STARRS). *JAMA Psychiat* 2015 Jan 1; **72**: 49–57.
42. Hawkins D. M. The problem of overfitting. *J Chem Inf Comput Sci* 2004 Jan 1; **44**: 1–12.
43. Kessler R. C., Warner C. H., Ivany C., Petukhova M. V., Rose S., Bromet E. J., et al. Predicting suicides after psychiatric hospitalization in US Army soldiers: the Army study to assess risk and resilience in Servicemembers (Army STARRS). *JAMA Psychiat* 2015 Jan 1; **72**: 49–57.
44. Rosellini A. J., Monahan J., Street A. E., Hill E. D., Petukhova M., Reis B. Y., et al. Using administrative data to identify U.S. Army soldiers at high-risk of perpetrating minor violent crimes. *J Psychiatr Res* 2017; **84**: 128–36.
45. Strobl C., Malley J., Tutz G. An introduction to recursive partitioning: rationale, application and characteristics of classification and regression trees, bagging and random forests. *Psychol Methods* 2009 Dec; **14**: 323–48.
46. Zou H., Hastie T. Regularization and variable selection via the elastic net. *J R Stat Soc Series B Stat Methodology* 2005; **67**: 301–20.
47. Kuhn M, Johnson K. Remedies for Severe Class Imbalance. In: Kuhn M, Johnson K, editors. *Applied Predictive Modeling [Internet]*. New York, NY: Springer New York; 2013 [cited 2019 Apr 18]. p. 419–43. Available from: https://doi.org/10.1007/978-1-4614-6849-3_16
48. Therneau T, Atkinson B, port BR (producer of the initial R, maintainer 1999-2017). rpart: Recursive Partitioning and Regression Trees [Internet]. 2019 [cited 2019 Apr 18]. Available from: <https://CRAN.R-project.org/package=rpart>
49. Chen C., Breiman L. *Using Random Forest to Learn Imbalanced Data*. Berkeley: University of California; 2004 Jan 1.
50. Huang B. F. F., Boutros P. C. The parameter sensitivity of random forests. *BMC Bioinformatics* 2016 Sep 1; **17**: 331.
51. Liaw A, Wiener M. Classification and Regression by randomForest. R News;
52. Robin X., Turck N., Hainard A., Tiberti N., Lisacek F., Sanchez J.-C., et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 2011 Mar 17; **12**: 77.
53. SAS Institute Inc. SAS/GRAPH 9.4. Cary, NC; 2013.
54. R Development Core Team. R: A language and environment for statistical computing [Internet]. Vienna, Austria; 2017. Available from: <https://www.R-project.org/>
55. Hosmer D., Lemeshow S. *Applied Logistic Regression, 3rd edn*. Hoboken, New Jersey: John Wiley & Sons; 2013 528 p.
56. National Institute on Drug Abuse. Addiction Medications [Internet]. National Institute on Drug Abuse. 2020 [cited 2020 Jul 24]. Available from: <https://www.drugabuse.gov/publications/principles-adolescent-substance-use-disorder-treatment-research-based-guide/evidence-based-approaches-to-treating-adolescent-substance-use-disorders/addiction-medications>
57. Goldman-Mellor S., Olfson M., Lidon-Moyano C., Schoenbaum M. Mortality following nonfatal opioid and sedative/hypnotic drug overdose. *Am J Prev Med* 2020 Jul 1; **59**: 59–67.
58. Rockett I. R., Hobbs G., De Leo D., Stack S., Frost J. L., Ducatman A. M., et al. Suicide and unintentional poisoning mortality trends in the United States, 1987-2006: two unrelated phenomena? *BMC Public Health* 2010 Nov 17; **10**: 705.
59. Bohnert A. S. B., Roeder K., Ilgen M. Unintentional overdose and suicide among substance users: a review of overlap and risk factors. *Drug Alcohol Depend* 2010; **110**: 183–92.
60. Bohnert A. S. B., Ilgen M. A. Understanding links among opioid use, overdose, and suicide. *N Engl J Med* 201903; **380**: 71–9.
61. Cherpitel C. J., Borges G. L. G., Wilcox H. C. Acute alcohol use and suicidal behavior: a review of the literature. *Alcohol Clin Exp Res* 2004 May; **28**: 18S–28S.
62. Conner K. R., Huguet N., Caetano R., Giesbrecht N., McFarland B. H., Nolte K. B., et al. Acute use of alcohol and methods of suicide in a US National Sample. *Am J Public Health* 2014 Jan; **104**: 171–8.
63. Jurek A. M., Greenland S., Maldonado G. How far from non-differential does exposure or disease misclassification have to be to bias measures of association away from the null? *Int J Epidemiol* 2008 Apr; **37**: 382–5.
64. Nissen S., Pottegård A., Ryg J. Trends of opioid utilisation in Denmark: a Nationwide study. *Drugs Real World Outcomes* 2019; **6**: 155–64.
65. Meyer A., LeClair C., McDonald J. V. Prescription opioid prescribing in Western Europe and the United States. *R I Med J (2013)* 2020 Mar 2; **103**: 45–8.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Appendix Table S1 Danish registry information.

Appendix Table S2 Variables included in the suicide deaths analyses among men and women.