



**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL**  
**FACULDADE DE MEDICINA**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM PSIQUIATRIA E CIÊNCIAS DO**  
**COMPORTAMENTO**

**DISSERTAÇÃO DE MESTRADO**

**MODELO PREDITIVO PARA TENTATIVA DE SUICÍDIO EM PACIENTES**  
**INTERNADOS POR USO DE CRACK: UMA ABORDAGEM POR APRENDIZADO**  
**DE MÁQUINA**

**VINÍCIUS SERAFINI ROGLIO**

**Orientador: Prof. Dr. Felix Henrique Paim Kessler**

**Co-orientador: Prof. Dr. Eduardo Nunes Borges**

Porto Alegre  
Junho de 2020



**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL**  
**FACULDADE DE MEDICINA**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM PSIQUIATRIA E CIÊNCIAS DO**  
**COMPORTAMENTO**

**DISSERTAÇÃO DE MESTRADO**

**MODELO PREDITIVO PARA TENTATIVA DE SUICÍDIO EM PACIENTES**  
**INTERNADOS POR USO DE CRACK: UMA ABORDAGEM POR APRENDIZADO**  
**DE MÁQUINA**

Dissertação apresentada ao Programa de Pós-Graduação em Psiquiatria e Ciências do Comportamento da Universidade Federal do Rio Grande do Sul como requisito parcial à obtenção do título de Mestre em Psiquiatria e Ciências do Comportamento.

Autor: Vinícius Serafini Roglio

Orientador: Felix Henrique Paim Kessler

Co-orientador: Eduardo Nunes Borges

Porto Alegre

Junho de 2020

### CIP - Catalogação na Publicação

Roglio, Vinícius  
Modelo preditivo para tentativa de suicídio em  
pacientes internados por uso de crack, uma abordagem  
por Aprendizado de Máquina / Vinícius Roglio. -- 2020.  
97 f.  
Orientador: Felix Henrique Paim Kessler.

Coorientador: Eduardo Nunes Borges.

Dissertação (Mestrado) -- Universidade Federal do  
Rio Grande do Sul, Faculdade de Medicina, Programa de  
Pós-Graduação em Psiquiatria e Ciências do  
Comportamento, Porto Alegre, BR-RS, 2020.

1. Tentativa de suicídio. 2. Aprendizado de  
Máquina. 3. Transtorno por uso de crack-cocaína. I.  
Henrique Paim Kessler, Felix, orient. II. Nunes  
Borges, Eduardo, coorient. III. Título.

**MODELO PREDITIVO PARA TENTATIVA DE SUICÍDIO EM PACIENTES  
INTERNADOS POR USO DE CRACK: UMA ABORDAGEM POR APRENDIZADO  
DE MÁQUINA**

Dissertação apresentada ao Programa de Pós-Graduação em Psiquiatria e Ciências do Comportamento da Universidade Federal do Rio Grande do Sul como requisito parcial à obtenção do título de Mestre em Psiquiatria e Ciências do Comportamento.

**Porto Alegre, 26 de junho de 2020**

A Comissão Examinadora, abaixo assinada, aprova a Dissertação elaborada por Vinícius Serafini Roglio, como requisito parcial para a obtenção do Grau de Mestre em Psiquiatria e Ciências do Comportamento.

**Comissão Examinadora:**

---

Prof. Dr. Felix Henrique Paim Kessler – Universidade Federal do Rio Grande do Sul  
Orientador

---

Prof. Dr. Maurício Kunz – Universidade Federal do Rio Grande do Sul

---

Prof.<sup>a</sup> Dr.<sup>a</sup> Stela Maris de Jesus Castro – Universidade Federal do Rio Grande do Sul

---

Dr.<sup>a</sup> Anne Orgler Sordi – Hospital de Clínicas de Porto Alegre

## AGRADECIMENTOS

Inicialmente agradeço a minha mãe, Maria Luiza, pelo imenso amor, incentivo e apoio que nunca me fizeram falta.

Aos meus colegas pesquisadores do Centro de Pesquisas em Álcool e Drogas (CPAD), Jaqueline, Diego, Felipe, Juliana, Fernando, Vanessa, Marcelo, Daiane e Carla pela intensa e contínua troca de conhecimentos, apoio e auxílio nas horas mais necessárias que permitiram a elaboração deste trabalho.

Ao meu orientador, Felix Kessler, pela incrível paciência e positividade em me guiar nessa jornada e pela confiança depositada.

Ao meu co-orientador, Eduardo Borges, pelas fantásticas aulas sobre aprendizado de máquina, metodologia e ciência nos intervalos e corredores do CPAD e sem o qual a minha experiência e aprendizado seriam muito menores.

Ao diretor e vice-diretora do CPAD, professores Flavio Pechansky e Lisia von Diemen, que sempre forneceram incentivo e os subsídios necessários para realização de um trabalho digno, integrativo e multidisciplinar.

Ao Programa de Pós-Graduação em Psiquiatria e Ciências do Comportamento da UFRGS e seu corpo docente, em especial ao professor Ives Cavalcante e à secretária do PPG-PSQ, Cláudia Grabinski, que prestaram auxílios e esclarecimentos com muita agilidade e eficiência.

Ao professor Rodrigo Grassi (PUC-RS) pela colaboração em partilhar do banco de dados de mulheres internadas por transtorno pro uso de crack.

Por fim, agradeço à UFRGS e ao Hospital de Clínicas de Porto Alegre (HCPA), pela qualidade do ensino, oportunidade de imersão em experiências de pesquisa, infraestrutura e pela excelência destas instituições públicas. E à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo incentivo financeiro neste período.

## RESUMO

O suicídio e os comportamentos do espectro suicida constituem um importante problema de saúde que levam a cerca de 800.000 mortes por ano mundialmente, segundo a Organização Mundial da Saúde. Na população de indivíduos com transtornos por uso de substâncias, principalmente de crack-cocaína, as taxas de suicídio e tentativa de suicídio se mostram muito elevadas em comparação à população geral, bem como também é observado a ocorrência de múltiplas comorbidades e condições socioeconômicas desfavoráveis, colocando essa população em situação de especial vulnerabilidade. Modelos teóricos de suicídio propõem explicações para o surgimento de comportamentos suicidas que integram fatores diáteses ou pré-motivacionais com fatores estressores ou precipitadores. Observa-se também uma grande sobreposição de fatores associados aos desfechos de suicídio e dependência química e possivelmente da existência de fatores específicos para esta população e de acordo com o gênero. Portanto, este trabalho objetiva investigar preditores de tentativa de suicídio em uma amostra de homens e mulheres internados por transtorno por uso de crack-cocaína e discutir a respeito dos elementos que podem contribuir para esse cenário. O artigo 1 apresenta um artigo original sobre a aplicação de duas metodologias analíticas de forma estratificada por gênero, uma descritiva e outra preditiva. A primeira sendo a regressão de Poisson com variância robusta a fim de estimar razões de prevalência e identificar fatores associados à tentativa de suicídio; e a segunda sendo o algoritmo Floresta Aleatória de Aprendizado de Máquina supervisionado, a fim de investigar preditores de tentativa de suicídio e avaliar a performance preditiva alcançada. O trabalho consiste de um estudo transversal que utiliza um banco de dados secundários mesclado de duas instituições especializadas no tratamento da dependência química de Porto Alegre/RS, contém uma amostra composta por homens e mulheres internados por transtorno por uso de crack e variáveis oriundas da aplicação de três instrumentos de pesquisa que coletaram informações sociodemográficas, psiquiátricas (transtornos e traumas) e sobre o padrão de uso de drogas. O artigo 2 apresenta um editorial sobre uso de drogas e suicídio a fim de incentivar futuros trabalhos e reforçar a necessidade de políticas públicas, estratégias de tratamento acessíveis e uso de tecnologias promissoras para promoção do bem estar social e redução da desigualdade.

## ABSTRACT

Suicide and suicide behaviors are an important health problem that lead to around 800,000 deaths per year worldwide, according to the World Health Organization. In the population of individuals with substance use disorders, especially crack-cocaine, the rates of suicide and attempted suicide are very high compared to the general population, as well as the occurrence of multiple comorbidities and unfavorable socioeconomic conditions, which place this population in a situation of special vulnerability. Theoretical models of suicide propose explanations for the emergence of suicidal behaviors that integrate diathesis or pre-motivational factors with stressors or precipitating factors. There is also a large overlap of factors associated with the outcomes of suicide and substance use disorder and possibly the existence of specific factors for this population and also gender-specific factors. Therefore, this study aims to investigate predictors of attempted suicide in a sample of men and women hospitalized for crack-cocaine disorder and discuss the elements that may contribute to this scenario. Article 1 presents an original article on the application of two analytical methodologies stratified by gender, a descriptive one and a predictive one. The first being the Poisson regression with robust variance in order to estimate prevalence ratios and to identify factors associated with suicide attempt; and the second being the supervised Machine Learning Random Forest algorithm, in order to investigate predictors of attempted suicide and evaluate the predictive performance achieved. The work consists of a cross-sectional study that uses a secondary database merged from two institutions specialized in the treatment of chemical dependence in Porto Alegre / RS, contains a sample composed of men and women hospitalized for crack use disorder and variables from the application of three research instruments that collected sociodemographic and psychiatric information (disorders and traumas) and the pattern of drug use. Article 2 presents an editorial on drug use and suicide in order to encourage future work and reinforce the need for public policies, accessible treatment strategies and the use of promising technologies to promote social welfare and reduce inequality.

## LISTA DE FIGURAS

Figura 1 – Taxa de óbitos por suicídio por 100.000 habitantes de indivíduos entre 10 e 79 anos - Brasil e regiões, 2000 a 2018.....	15
Figura 2 – Modelo de Estresse-Diátese para suicídio.....	20
Figura 3 – Exemplos de algoritmos e modelos de Aprendizado de Máquina por tipo de aprendizado e objetivo. ....	30
Figura 4 – Fluxo geral da aplicação de um algoritmo de aprendizado de máquina. ....	31
Figura 5 – Exemplo de estrutura de uma árvore de decisão. ....	34
Figura 6 – Exemplo de fluxo do processo de predição de um algoritmo de Floresta Aleatória. ....	35



## LISTA DE ABREVIATURAS E SIGLAS

Acc.	Acurácia
AM	Aprendizado de Máquina
AUC	<i>Area Under the Curve</i> (area abaixo da curva)
CEBRID	Centro Brasileiro de Informações sobre Drogas Psicotrópicas
GLM	<i>Generalized Linear Models</i> (Modelos Lineares Generalizados)
IA	Inteligência Artificial
IBGE	Instituto Brasileiro de Geografia e Estatística
IC	Intervalo de confiança
LENAD	Levantamento Nacional de Álcool e Drogas
ln	Logarítmo natural
LNUD	Levantamento sobre o Uso de Drogas
MIAI	Mortes intencionais auto-infligidas
NIR	<i>No Information Rate</i> (Taxa de não-informação)
OMS	Organização Mundial da Saúde
RC	Razão de chances
ROC	<i>Receiver Operating Characteristic (curve)</i>
RP	Razão de prevalência
RR	Risco relativo
Sens.	Sensibilidade
SIM	Sistema de Informações sobre Mortalidade
SPA	Substâncias psicoativas
Spec.	Especificidade
TUS	Transtorno por uso de substâncias
VD	Variável dependente
VI	Variável independente
VPN	Valor preditivo negativo
VPP	Valor preditivo positivo

## SUMÁRIO

<b>1.</b>	<b>INTRODUÇÃO.....</b>	<b>11</b>
1.1.	CRACK E COCAÍNA, EPIDEMIOLOGIA DO USO E PERFIL DO USUÁRIO .....	11
1.2.	EPIDEMIOLOGIA DO SUICÍDIO .....	14
1.3.	MODELOS TEÓRICOS DE SUICÍDIO.....	15
1.4.	SUICÍDIO, TRANSTORNOS PSIQUIÁTRICOS E POR USO DE CRACK .....	21
1.5.	MODELOS LINEARES GENERALIZADOS .....	23
1.5.1.	REGRESSÃO DE POISSON COM VARIÂNCIA ROBUSTA.....	26
1.6.	APRENDIZADO DE MÁQUINA .....	27
1.6.1.	APRENDIZADO DE MÁQUINA NA PESQUISA EM PSIQUIATRIA .....	32
1.6.2.	ALGORITMO FLORESTA ALEATÓRIA .....	33
<b>2.</b>	<b>OBJETIVOS .....</b>	<b>37</b>
2.1.	OBJETIVO PRINCIPAL.....	37
2.2.	OBJETIVOS SECUNDÁRIOS .....	37
<b>3.</b>	<b>ARTIGO 1 .....</b>	<b>38</b>
<b>4.</b>	<b>MATERIAL COMPLEMENTAR DO ARTIGO 1.....</b>	<b>73</b>
<b>5.</b>	<b>ARTIGO 2 .....</b>	<b>83</b>
<b>6.</b>	<b>CONSIDERAÇÕES FINAIS .....</b>	<b>87</b>
	<b>REFERÊNCIAS .....</b>	<b>91</b>

## 1 INTRODUÇÃO

Este capítulo apresenta o referencial teórico utilizado para elaboração deste trabalho e baseia-se na interdisciplinariedade entre as áreas de psiquiatria de adição, estatística, epidemiologia e computação. Os conteúdos apresentados fundamentam o estudo que permitiu a escrita das publicações realizadas como produtos desta dissertação. As seções 1.1 e 1.2 introduzem aspectos relacionados à população alvo e desfecho principal do trabalho, usuários de crack e tentativa de suicídio, respectivamente. A seção 1.3 apresenta uma revisão sobre as teorias do comportamento suicida. A seção 1.4 incorporam estes sob a perspectiva psiquiátrica. E as seções 1.5 e 1.6 introduzem as metodologias aplicadas. Os objetivos deste trabalho são apresentados no capítulo 2 e as produções científicas na sequência. Por fim, as considerações finais integram e concluem o trabalho.

### 1.1. CRACK E COCAÍNA, EPIDEMIOLOGIA DO USO E PERFIL DO USUÁRIO

A cocaína e o crack são substâncias psicoativas (SPA) de ação estimulante do sistema nervoso central e de alto poder dependógeno, agindo diretamente nos neurotransmissores químicos cerebrais dopaminérgicos, envolvidos nas respostas do corpo ao prazer. A substância se liga ao transportador de dopamina, bloqueando sua captura e mantendo-a por mais tempo nos espaços sinápticos. Após liberada, a dopamina continua estimulando o receptor, gerando uma super estimulação motora e sensorial, onde o usuário experimenta intensa euforia e prazer. Em aproximadamente 5 minutos após o uso, esse efeito cessa, fazendo da substância uma potente causadora de fissura e levando com que o processo de tolerância e dependência se instaure de forma muito rápida (Bastos and Bertoni, 2014).

A cocaína pode ser administrada por via nasal (aspirando a substância), oral ou endovenosa (injetada diluída em água). Enquanto o crack é um derivado da pasta base da cocaína misturada com bicarbonato de sódio, amônia e água, sendo administrado pela via pulmonar, através da combustão da substância. Esta é chamada de crack devido ao barulho que a pedra de crack emite ao ser queimada, que se assemelha ao ruído de fissuras sendo provocadas numa pedra.

No Brasil, um dos primeiros achados impactantes sobre o uso desta substância mostrou que os serviços ambulatoriais de São Paulo-SP observaram um crescimento na

procura por tratamento de problemas decorrentes do uso de crack de 17%, em 1990, para 67%, em 1994 (Dunn et al., 1996). Outros estudos pioneiros no país sobre perfil de consumo foram realizados no final da década de 1990, e em 2001 o Centro Brasileiro de Informações sobre Drogas Psicotrópicas (CEBRID) realizou o Primeiro Levantamento Domiciliar sobre o Uso de Drogas Psicotrópicas no Brasil, encontrando prevalência de uso na vida de cocaína e crack de 2,3% e 0,4%, respectivamente (Carlini et al., 2002). O estudo foi repetido em 2005, encontrando prevalências de 2,9% e 0,7% (Carlini, 2006).

Em 2012, o estudo realizou uma segunda fase de coletas e passou a ser chamado de Segundo Levantamento Nacional de Álcool e Drogas (LENAD), onde foram divulgadas prevalências de uso na vida nos valores de 3,7% e 1,3% (Laranjeira et al., 2014). O estudo também revelou que 27% dos usuários faziam uso todos os dias ou mais de 2 vezes por semana no último ano, levando o Brasil a representar 20% do consumo mundial de cocaína e crack e apresentando um aumento de 85% do uso de crack em 4 anos. Observa-se uma tendência crescente do consumo, possivelmente devida a mudanças referentes a diferentes formas de uso da cocaína (como merla e oxi), baixo custo e fácil acesso (Bastos and Bertoni, 2014). Entretanto, o Terceiro Levantamento sobre o Uso de Drogas (LNUD), realizado em 2015, divulgou prevalências do uso na vida de cocaína e crack no Brasil nos valores de 3,1% e 0,9%, respectivamente (Bastos et al., 2017). A tabela 1 apresenta o detalhamento destes dados.

**Tabela 1** - Estimativas e intervalos de 95% de confiança para prevalência do uso de cocaína e crack na vida, nos últimos 12 meses e nos últimos 30 dias por estudos nacionais - Brasil, 2001, 2005, 2012 e 2015.

	2001	2005	2012	2015
<b>Cocaína (%)</b>				
Uso na vida	2,3 (1,3-3,3)	2,9 (1,2-4,5)	3,7	3,1 (2,7-3,4)
Uso nos últimos 12 meses	-	0,7	1,7	0,9 (0,7-1,1)
Uso nos últimos 30 dias	-	0,4	-	0,3 (0,2-0,4)
<b>Crack (%)</b>				
Uso na vida	0,4 (0,1-0,9)	0,7	1,3	0,9 (0,7-1,1)
Uso nos últimos 12 meses	-	0,1	0,6	0,3 (0,2-0,4)

Uso nos últimos 30 dias	-	0,1	-	0,1 (0,04-0,2)
-------------------------	---	-----	---	----------------

“-” Indica estatísticas não divulgadas.

Fonte: 2001 e 2005 CEBRID, 2012 LENAD e 2017 LNUD.

O perfil do usuário de crack vem sendo descrito, em maioria relativa, como adultos jovens, homens, com baixa escolaridade, baixo poder aquisitivo, provenientes de uma frágil e desestruturada rede social e familiar, em situação de moradia instável e de alta vulnerabilidade social (Duailibi et al., 2008; Halpern et al., 2017; Horta et al., 2011). Portanto, uma população de difícil acesso, com importantes prejuízos nas esferas social e de saúde, envolvida com comportamentos violentos e sexuais de risco e complicações clínicas que amplificam sua marginalização (Bastos and Bertoni, 2014; Botti et al., 2014). Já o perfil do usuário de cocaína inalada é mais heterogêneo, visto que a substância é produzida por um processo de refinamento que requer mais pasta base, tornando-a mais cara e menos acessível às classes econômicas mais baixas.

Para além do agravo social, as comorbidades psiquiátricas mais comumente encontradas em usuários de crack são os transtornos de personalidade antissocial (Paim Kessler et al., 2012), transtorno de estresse pós-traumático, transtorno por uso de outras substâncias (poliusuários), transtorno de déficit de atenção e hiperatividade, risco de suicídio, transtorno psicótico (Narvaez et al., 2014), transtornos do humor (depressão e bipolaridade) e transtornos de ansiedade (agorafobia, fobia social e transtorno do pânico) (Paiva et al., 2017). A associação entre o consumo da droga e as comorbidades conduzem à piora dos sintomas, pouca adesão aos tratamentos e desfechos como recaídas precoces, recorrência de comportamentos de risco, sentimentos de desesperança e tendências suicidas, fatores que impactam diretamente na qualidade de vida e funcionalidade dos indivíduos (GRELLA et al., 2001; Narvaez et al., 2014; Paiva et al., 2017; Zubaran et al., 2010). O fato de o crack ser uma substância altamente deletéria, tanto fisicamente, com consequências circulatórias, cardíacas e neuronais (entre outras) do uso crônico; quanto psicologicamente, faz com que o usuário apresente um perfil social, clínico e psiquiátrico diferenciado (Guimarães et al., 2008; Toledo et al., 2017).

Por estas razões, o crack é atualmente considerado a droga ilícita que mais conduz a internações hospitalares e que mais demanda atendimento especializado no Brasil, onerando o sistema público de saúde (Santos Cruz et al., 2013; Siliquini et al., 2005). Porém, a complexidade do cenário que envolve o dependente impacta na adesão e retenção a recursos

terapêuticos, tornando-se um desafio para serviços de tratamento e gestores de políticas públicas (Faller et al., 2014; Lorenzoni et al., 2013).

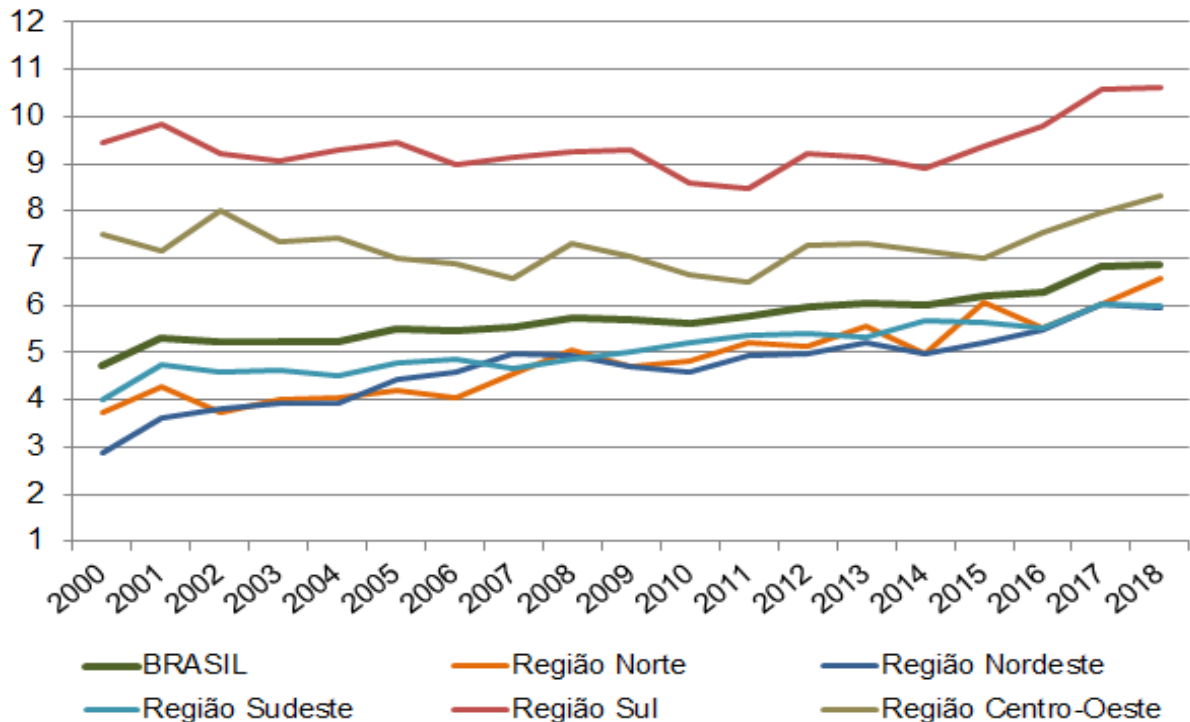
## 1.2. EPIDEMIOLOGIA DO SUICÍDIO

O suicídio é um problema de saúde pública complexo de dimensão global que atualmente representa 1,4% dos óbitos prematuros ocorridos em todo o planeta. De acordo com a OMS cerca de 800 mil mortes intencionais auto-infligidas (MIAI) são documentadas a cada ano, o que corresponde a um suicídio consumado a cada 40 segundos (WHO et al., 2012). Durante as últimas décadas o aumento das taxas de suicídio tem sido verificado em todo o mundo; estima-se que os suicídios são a segunda principal causa de mortalidade prematura em indivíduos de 15 a 29 anos (precedidos por acidentes de trânsito) e a terceira na população com idades entre 15 e 44 anos (WHO et al., 2012).

As taxas de suicídio são distintas entre os países, variando de acordo com a diversidade dos fatores culturais, sociais, econômicos e ambientais. Estima-se que 78% de todos os suicídios ocorram em países de baixa e média renda (Bachmann, 2018). No Brasil, a taxa MIAI entre 2010 e 2014 foi estimada em 5,23 / 100.000 habitantes (Dantas et al., 2017), taxa comparativamente baixa se comparada a de outros países, todavia, pondera-se que estes dados podem estar subestimados em decorrência de falhas no sistema de notificação (Scaini and Quevedo, 2018).

A figura 1 mostra a taxa de óbitos por 100.000 habitantes por lesões autoprovocadas intencionalmente entre 2000 e 2018 por indivíduos de 10 a 79 anos, calculado através dos dados do Sistema de Informações sobre Mortalidades do Sistema Único de Saúde (SIM/DATASUS; códigos CID-10 de X60 a X84) junto à projeção populacional brasileira realizada pelo Instituto Brasileiro de Geografia e Estatística (IBGE). O gráfico revela uma tendência crescente geral e a preponderância das maiores taxas para a região Sul do país ao longo de todo o período.

Figura 1 - Taxa de óbitos por suicídio por 100.000 habitantes de indivíduos entre 10 e 79 anos - Brasil e regiões, 2000 a 2018.



Fonte: Elaborado pelo autor com dados do SIM/DATASUS e IBGE.

### 1.3. MODELOS TEÓRICOS DE SUICÍDIO

O uso do termo suicídio foi pela primeira vez empregado pelo médico e filósofo Sir Thomas Browne em 1642, em seu livro “*Relegio Medici*”, sendo definido como o ato de intencionalmente terminar a própria vida (Karthick and Barwa, 2017). Ao longo de muitos anos de pesquisa, o conceito a respeito do suicídio adquiriu complexidade e expandiu para o estudo do fenômeno de forma mais ampla, que o particionou em principalmente três componentes: (1) ideação suicida, (2) tentativa de suicídio e (3) suicídio completado. Na literatura são encontrados outros termos, como suicidalidade (do inglês, *suicidality*) ou comportamentos suicidas. Por vezes, estes termos se referem ao fenômeno como um todo, incluindo os três componentes em um espectro e diferenciando-os em questões de descrição da fase, gravidade da intenção e tipo de intervenção necessária.

A ideação suicida pode ainda ser subcategorizada em ativa ou passiva. A ideação passiva consiste de pensamentos a respeito de morte ou do desejo de estar morto, sem

qualquer plano ou intenção de agir nesse sentido. E a ideação ativa refere-se a pensamentos sobre terminar a própria vida, incluindo a identificação de um método, planejamento ou ter a intenção de agir nesse sentido. A tentativa de suicídio consiste de um comportamento prejudicial ou destrutivo, autoinfligido, com pelo menos alguma intenção explícita ou inferida de morrer e com desfecho não fatal (Turecki and Brent, 2016). Por último, o suicídio completado refere-se ao fato do desfecho ter sido fatal.

Muitas teorias a respeito do suicídio foram desenvolvidas cobrindo diversos aspectos, como biológicos, psicodinâmicos, cognitivos comportamentais e desenvolvimentais, na tentativa de explicar a etiologia do comportamento suicida. Muitos destes modelos apresentam sobreposições entre os fatores considerados e se complementam quanto aos fatores específicos da área de estudo que os originaram. Na sequência são apresentados resumidamente os principais modelos teóricos de suicídio que contribuíram para o estudo e execução deste trabalho.

O modelo de Estresse-Diátese é um dos modelos de suicídio mais antigos, tendo sido desenvolvido por Meehl, Bleuler, and Rosenthal na década de 1960 (Ingram and Luxton, 2005). Esse modelo classifica os preditores em distais e proximais e explica o comportamento suicida como um produto da interação entre dois componentes: (1) a presença de vulnerabilidades pré-existentes ou adquiridas (condicionadas ou aprendidas) e (2) a ocorrência de estressores precipitantes (gatilhos). O primeiro componente, chamado de diátese, consiste de um fator ou conjunto de fatores pré-disposicionais (distais) que provocam uma vulnerabilidade constitucional para desenvolver um distúrbio ou transtorno. Na época em que o modelo foi proposto, os fatores diáteses considerados eram predominantemente biológicos, porém, recentemente, o termo foi ampliado para incluir também predisposições cognitivas e sociais. O segundo componente, o estresse, pode ser representado por crises psicossociais ou outra experiência adversa e dolorosa que estabelece um evento proximal onde a tolerância ao estresse é ativada.

Segundo este modelo, os fatores relacionados à diátese podem condicionar ao impedimento de uma resposta emocionalmente regulada ao estresse e produzir o desejo de suicídio através da identificação de três elementos: (1) hipersensibilidade à sentimentos de derrota, (2) limitada habilidade de solução de problemas, gerando a percepção de que não há escapatória para o sofrimento e (3) desesperança quanto a um resgate da situação, gerando a percepção de que nada salvará o sujeito daquele sofrimento (Karthick and Barwa, 2017; Turecki and Brent, 2016). A combinação destes elementos produziria um sentimento de



aprisionamento e poderia levar à interpretação de que a única forma de cessar o sofrimento seria de terminar com a vida.

Outro modelo fortemente reconhecido é a teoria Interpessoal do suicídio, que é um modelo psicodinâmico desenvolvido por Joiner (2006) (Van Orden et al., 2010). A teoria propõe que o desejo de terminar a própria vida é causado pela presença simultânea de dois construtos interpessoais: (1) sentimento frustrado de não-pertencimento e (2) percepção de considerar-se um fardo (tradução livre do termo “*perceived burdensomeness*”). Sendo que ambos são acompanhados pela desesperança em relação a estes estados serem mutáveis, ou seja, uma convicção de que estes são constantes e imutáveis. Ainda, a teoria separa a capacidade do indivíduo de envolver-se em comportamentos suicidas do desejo de fazê-lo como um terceiro eixo focal, que diz respeito à capacidade adquirida de envolver-se em comportamentos suicidas. Segundo a teoria, a capacidade emerge da habituação à exposição repetida de experiências dolorosas e indutoras de medo, gerando processos opostos que poderiam gradualmente reduzir o medo da morte e fazer com que o desejo e a capacidade converjam em uma tentativa de suicídio.

Um outro modelo, inspirado na teoria Interpessoal do suicídio, é o modelo dos Três Passos proposto por Klonsky & May (2015). Este modelo propõe uma estrutura denominada “*ideação-à-ação*” (tradução livre da expressão “*ideation-to-action*”) e sugere que esta seja um guia geral para todos os modelos de suicídio, onde a ideação e a tentativa de suicídio são vistas como processos distintos e com diferentes explicações. O modelo é baseado em quatro elementos essenciais: (1) insuportável dor psicológica, (2) desesperança, (3) conexão com outros humanos e (4) capacidade de agir. De modo que estes elementos sejam capazes de modelar a progressão entre a ideação suicida e a tentativa de suicídio. Esse modelo parte do princípio que a combinação da presença de dor psicológica, independente da fonte, com a desesperança são os principais motivadores de uma ideação suicida. Em seguida, postula que a falta de conexão emocional com outros humanos serve de propulsor para a progressão à uma ideação suicida forte. E por último, sugere que a capacidade de agir determina a ocorrência de uma tentativa de suicídio.

O quarto e último modelo teórico de suicídio sumariado neste trabalho é o modelo Motivacional-Volitivo, proposto por O’Connor em 2011 e rerepresentado em 2018, sob o nome de modelo Motivacional-Volitivo Integrado do comportamento suicida (O’Connor and Kirtley, 2018). Este modelo integra os princípios e estruturas dos modelos citados anteriormente e é um dos modelos biopsicossociais mais abrangentes. O modelo propõe a identificação de três fases que descrevem a progressão do comportamento suicida: (1) pré-

motivacional, (2) motivacional e (3) volitiva. A fase pré-motivacional, inspirada do componente diátese do modelo original de Estresse-Diátese, descreve o contexto biopsicossocial em que a ideação e o comportamento suicidas são mais propensos a emergir. A fase motivacional descreve os fatores que levam ao surgimento da ideação suicida, pautados principalmente em aspectos cognitivos e comportamentais e que podem produzir a intenção de morte. E a terceira fase, a volitiva, descreve os fatores que conduzem a transição da ideação suicida para a tentativa de suicídio.

A evolução das teorias e modelos explicativos para o fenômeno do suicídio ocorreram principalmente no campo de estudo da psicologia e dão ênfase a diversos aspectos principais que foram adquirindo importância e unificaram alguns consensos atuais, como: (1) a influência de fatores diátese de origem biopsicossociais como vulnerabilizantes, (2) o papel do estresse e dos mecanismos cognitivos e psicodinâmicos de lidar com este, (3) a importância de sentimentos de desesperança, aprisionamento, dor e sofrimento, conexão e pertencimento, e (4) a capacidade de agir em relação ao envolvimento real em comportamentos suicidas. A tabela 2 apresenta um breve resumo das premissas de diversos autores ao longo das últimas décadas.

Tabela 2 - Resumo de premissas sobre o comportamento suicida propostas desde 1985.

<b>Autor(es) / ano</b>	<b>Premissa</b>
Shneidman 1985, 1993	A combinação de estresse, dor/sofrimento, e perturbação resultam em risco de suicídio. Necessidades psicológicas não atendidas causam dor mental torturante que leva ao suicídio como a única opção.
Schotte & Clunn 1987	Um déficit na capacidade interpessoal e cognitiva de solução de problemas aumenta o risco de comportamento suicida como uma reação ao estresse.
Plutchik, van Praag, & Conte 1989	Certos fatores de risco determinam o nível de agressão impulsiva que leva ao ato violento, enquanto a presença de outros fatores determina se a agressão é direcionada para dentro (internalizante) ou para fora (externalizante).
Baumeister 1990	O estado de autoconsciência distorcida quanto a dor insuportável causa prejuízo cognitiva e busca de um meio de escapar.

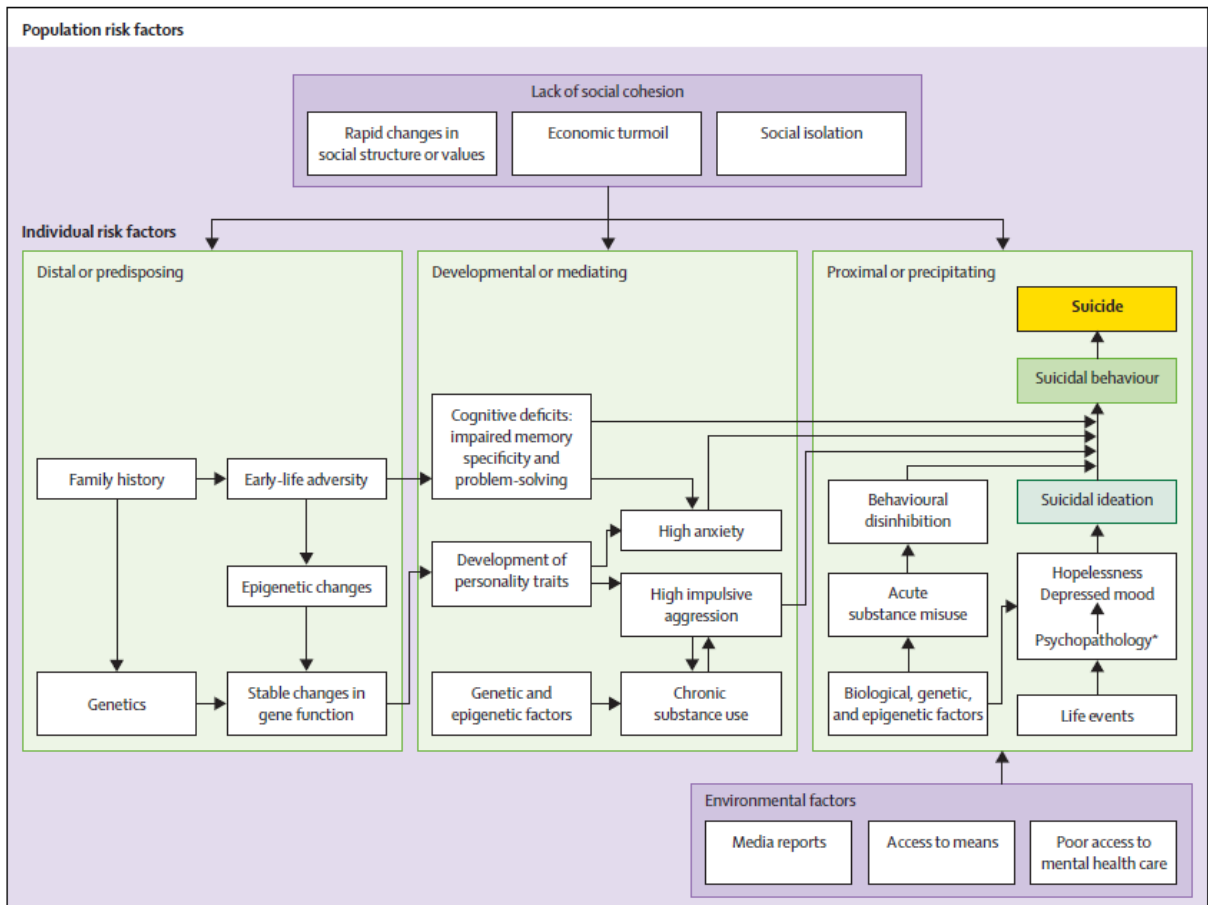
Beck, Brown, Berchick et. al. 1990	O esquema suicida composto por vieses de atenção, processamento de informações e memória prejudica a capacidade do indivíduo de recordar razões para viver ou ter esperança em relação à vida.
Williams 1997, 2001	O risco de suicídio aumenta quando os sentimentos de derrota e aprisionamento são altos e o potencial de resgate (como apoio social) é baixo.
Mann et. al. 1999	Modelo de estresse-diátese, onde o risco de suicídio é causado não apenas por transtorno psiquiátrico (estressor), mas também por uma diátese (tendência a experimentar mais ideação ou impulsividade suicida).
Rudd et. al. 2001, 2006	As características cognitivas, comportamentais, fisiológicas e afetivas do modo suicida predizem a vulnerabilidade aos gatilhos e a gravidade e duração das crises suicidas.
Joiner 2005	O desejo suicida é causado por altos níveis de <i>burdensomeness</i> e percepção frustrada de não-pertencimento. O comportamento suicida emerge quando a capacidade de agir é alta.
Brent & Mann 2006	A transmissão familiar da característica impulsivo-agressiva coloca os indivíduos em maior risco de comportamento suicida.
Johnson et. al 2008	Modelo de avaliação que propõe que o risco de suicídio é causado pela interação entre vieses nos sistemas de processamento de informações, esquemas e avaliação.
Wenzel e Beck 2008	Modelo diátese-estresse com três construtos principais: fatores de vulnerabilidade disposicional, processos cognitivos associados a distúrbios psiquiátricos e processos cognitivos associados a atos suicidas.
Williams et. al. 2008	Modelo de rede associativa, no qual a experiência de ideação ou comportamento suicida durante um episódio depressivo aumenta a probabilidade de ressurgir nos episódios subsequentes.
O'Connor 2011, 2018	Modelo de diátese-estresse, que especifica os componentes das fases pré-motivacional, motivacional (formação de idéias e intenções) e volitivas

	(ação comportamental) do comportamento suicida.
Klonsky e May 2014, 2015	Teoria dos Três Passos sugere a estrutura ideação-à-ação como guia geral para os modelos de suicídio e ressalta o papel da dor, desesperança, conexão e capacidade.

Fonte: Traduzido e adaptado de Barzilay e Apter (2014) e O'Connor e Nock (2014).

Ainda, verifica-se que o modelo de Estresse-Diátese forneceu elementos de base para a maioria das teorias, podendo ser dito como uma categoria abrangente de modelos de suicídio, onde se inclui, por exemplo, o modelo Motivacional-Volitivo Integrado e que se enquadra na estrutura de “ideação-à-ação”. Um exemplo moderno de modelo Estresse-Diátese é ilustrado na figura 2.

Figura 2 – Modelo de Estresse-Diátese para suicídio.



Fonte: Turecki e Brent (2016).

#### 1.4. SUICÍDIO, TRANSTORNOS PSIQUIÁTRICOS E POR USO DE CRACK

Apesar da etiologia do suicídio ser heterogênea, está bem descrito que a presença de transtornos psiquiátricos, sobretudo em fase sintomática, é o preditor de suicídio mais significativo (Arsenault-Lapierre et al., 2004; Bilsen, 2018; Hoertel et al., 2015). Um estudo prévio sinalizou que em pessoas com transtornos psiquiátricos a ocorrência de suicídios e de tentativas de suicídio é pelo menos 10 vezes superior aos índices evidenciados na população em geral. Além disso, estima-se que entre 60% e 98% de todos os suicídios completados sejam cometidos por indivíduos com transtornos psiquiátricos (Bachmann, 2018). A depressão, a psicose e os transtornos por uso de substâncias (TUS) são os fatores de risco mais citados na literatura, com um risco estimado entre 5 e 8% (Nordentoft, 2011). Porém os transtornos de ansiedade, personalidade, alimentares e relacionados a traumas, além de distúrbios mentais orgânicos, também contribuem para a ocorrência de MIAI (Bachmann, 2018).

Em consonância, uma recente metanálise sinalizou que em jovens, a presença de qualquer transtorno psiquiátrico aumenta a chance de tentativa de suicídio (RC = 3,6 IC<sub>95%</sub> 2,2-5,7) e de morte por suicídio (RC = 10,8 IC<sub>95%</sub> 4,7-25,0). Ressalta-se ainda, que a ocorrência de comorbidades psiquiátricas (mais de um transtorno concomitante) parece ser um agravador de risco cumulativo para o suicídio (Gili et al., 2019; Vijayakumar et al., 2011). Outro estudo de metanálise recente apontou que o risco relativo (RR) combinado de suicídio foi de 13,2 (IC<sub>95%</sub> 8,6-20,3) para transtornos psicóticos, 12,3 (IC<sub>95%</sub> 8,9-17,1) para transtornos do humor, 8,1 (IC<sub>95%</sub> 4,6-14,2) para transtornos de personalidade, 4,4 (IC<sub>95%</sub> 2,9-6,8) para transtornos por uso de substâncias e 4,1 (IC<sub>95%</sub> 2,4-6,9) para transtornos de ansiedade na população em geral. O RR global para esses transtornos mentais foi de 7,5 (IC<sub>95%</sub> 6,6-8,6). Por fim, os resultados demonstraram um risco de suicídio atribuível à população de transtornos mentais foi de até 21% (Too et al., 2019).

Outros estudos de metanálise realizados com grupos específicos de pacientes psiquiátricos também demonstram que se trata de um cenário preocupante. Em pacientes com transtorno depressivo maior a prevalência de tentativa de suicídio ao longo da vida foi de 31% (Dong et al., 2019); em esquizofrênicos a taxa exata estimada de suicídio foi de 579 / 100.000 pessoas / ano (Hor and Taylor, 2010); em bipolares as taxas são aproximadamente 20 a 30 vezes maiores que na população em geral (Plans et al., 2019). Em pacientes com qualquer transtornos de ansiedade (exceto transtorno obsessivo compulsivo) as taxas de ideação suicida

(RC = 2,9 IC<sub>95%</sub> 2,1-4,0), tentativa de suicídio (RC = 2,5 IC<sub>95%</sub> 1,9-3,1) e suicídios concluídos (RC = 3,3 IC<sub>95%</sub> 2,1-5,3) são maiores que na população sem estas condições (Kanwar et al., 2013). Observa-se que nestas metanálises citadas os transtornos por uso de substâncias são reconhecidos como variáveis preditoras ou confundidoras, o que sinaliza para uma lacuna importante de literatura a ser melhor explorada.

Observa-se que entre os transtornos psiquiátricos, os TUS estão entre os preditores de suicídio mais consolidados. Está bem descrito que o consumo de drogas é uma variável fortemente associada ao aumento do risco de suicídio na população geral (Conner et al., 2019; Kennedy et al., 2015; Poorolajal et al., 2016). Além disso, sabe-se que os TUS são amplamente comórbidos a outros transtornos (Gattamorta et al., 2017; Langås et al., 2011; Ross and Peselow, 2012). Assim, mesmo em subpopulações específicas de pacientes com transtornos mentais graves, como a esquizofrenia e o transtorno bipolar a predição de comportamentos suicidas tem sido associada ao uso de substâncias psicoativas (Fazel et al., 2019; Vijayakumar et al., 2011).

Apesar do uso de drogas ser um preditor bem definido de suicídio tanto na população geral quando em populações específicas, quando o recorte analítico se restringe à amostras de usuários de drogas e o consumo deixa de ser um fator de risco para ser a variável caracterizadora da amostra, observa-se que os preditores de suicídio não estão bem estabelecidos. Ainda, estima-se que o tipo de droga utilizado (devido às diferenças no mecanismo de ação de cada substância), a frequência do consumo, a quantidade da dose e a ocorrência de patologia dual, também podem interferir em diferentes intensidades na associação com os desfechos de ideação, tentativa e suicídio (Gili et al., 2019; Poorolajal et al., 2016).

No Brasil, o consumo de crack-cocaína configura um importante problema de saúde pública (Bastos and Bertoni, 2014; Laranjeira et al., 2014). Apesar dos índices de consumo do crack serem inferiores se comparados a outras drogas, observa-se que é a droga ilícita que mais conduz a internações em hospitais psiquiátricos (Duailibi et al., 2008) e que mais provoca demanda por atendimento (Siliquini et al., 2005) gerando custo expressivo para o sistema de saúde (Duailibi et al., 2008). Em usuários de crack os índices de ideação e tentativa de suicídio são especialmente elevados (von Diemen et al., 2014; Halpern et al., 2017; Narvaez et al., 2014). Um estudo de larga escala publicado recentemente verificou que enquanto na população geral a taxa de ideação suicida e de tentativas de suicídio autodeclaradas foram de 9,9 e 5,4%, respectivamente, em usuários de cocaínicos estas taxas

foram significativamente superiores, chegando a 40,0 e 20,8% respectivamente (Abdalla et al., 2019).

Investigações prévias conduzidas com diferentes amostras no Brasil demonstraram que comportamentos suicidas são particularmente frequentes em usuários de crack, ocorrendo em até 43% desta população, apesar disso, os fatores preditores do suicídio neste grupo não estão bem elucidados. Ressalta-se que se trata de uma população marcada por múltiplas vulnerabilidades psicossociais, altos índices de comorbidades clínicas e psiquiátricas e baixa adesão ao tratamento (Halpern et al., 2017; Paim Kessler et al., 2012; Silva et al., 2017), fatores que podem gerar um impacto significativo na ocorrência de MIAI. Cabe citar que as variáveis socioeconômicas identificadas por Halpern et. al. (2017) em usuários de crack são similares às citadas por Bachmann (2018) como mediadores de risco para suicídio: baixa renda, desemprego, baixa educação, baixo apoio social, moradia inadequada, superlotação, violência e outros.

## 1.5. MODELOS LINEARES GENERALIZADOS

Nas mais diversas áreas de pesquisa, incluindo a área da saúde, frequentemente se deseja estudar o comportamento de uma variável, dita como dependente, resposta ou desfecho, em relação a uma ou mais variáveis, ditas independentes, explicativas ou preditoras. Estas são responsáveis por explicar a variabilidade da variável dependente. Nestes casos, técnicas de modelagem são utilizadas, nas quais se incluem os modelos de regressão estatística. A proposta deste trabalho envolve a modelagem da variável “tentativa de suicídio”, tratada sempre como a variável dependente a ser modelada e de natureza categórica binária, ou seja, com apenas duas categorias de resposta possíveis, “sim” e “não”.

Segundo McCullagh e Nelder (1989), o modelo de regressão linear múltipla clássico foi sintetizado no início do século XIX pelos matemáticos Legendre, Gauss e Fisher, considerando a variável dependente com distribuição de probabilidade Normal (ou Gaussiana), o que implica que a variável que se deseja modelar seja de natureza quantitativa, por exemplo, número de pedras de crack ou então quantidade em gramas de crack utilizadas por dia normal de consumo. Entretanto, a suposição de normalidade (ainda amplamente utilizada para modelagem de dados atualmente) apresenta uma limitação para modelagem de variáveis que a violam, como por exemplo variáveis categóricas (onde as respostas são

categorias ao invés de métricas) ou quantitativas com distribuição assimétrica (onde há um desequilíbrio entre as probabilidades de ocorrência de valores pequenos e grandes). Outras suposições também se encontram frequentemente violadas em aplicações com dados reais, como a homocedasticidade, que é a suposição de que a variabilidade do resíduo do modelo de regressão, condicionado às variáveis independentes, é constante. Posteriormente, Nelder e Wedderburn (1972) expandiram a utilização de modelos de regressão às distribuições da família exponencial e propuseram uma teoria unificadora da modelagem estatística a qual denominaram Modelos Lineares Generalizados (do inglês, *Generalized Linear Models - GLM*). Dentre estas distribuições, se encontram, além da Normal, as distribuições exponencial, Beta, Gama, Binomial, Poisson, entre outras, e cujas médias são não-lineares num conjunto de parâmetros lineares.

Desta forma, a estrutura dos GLM é descrita em três componentes: o primeiro sendo o componente aleatório, que é a distribuição de probabilidade da variável dependente (aquela que se deseja modelar); o segundo sendo o componente linear (ou sistemático), que é o vetor de coeficientes associados às variáveis independentes; e o terceiro componente é a função de ligação, que mapeia o componente linear para o domínio da esperança do componente aleatório (Lee and Nelder, 1998). Assim, se o componente aleatório for uma variável normalmente distribuída e a função de ligação for a matriz identidade (o que significa ausência de uma função específica, restando apenas a igualdade entre os componentes aleatório e linear), temos o modelo de regressão linear clássico. Analogamente, se o componente aleatório seguir uma distribuição binomial (utilizado para variáveis categóricas) e a função de ligação for *logit*, temos o modelo de regressão logística clássico.

Na área da modelagem contemporânea podemos encontrar principalmente três tipos de modelagem, a descritiva, a explicativa e a preditiva, todas possíveis de serem alcançadas com aplicação de GLM, mas com focos em resultados diferentes da modelagem (Schooling and Jones, 2018; Shmueli, 2010). A modelagem descritiva tem por objetivo descrever a relação entre uma ou mais variáveis independentes (VI) e uma variável dependente (VD), através da estimação de uma estatística via aplicação de um modelo. Os GLM são os modelos mais utilizados para esse fim e a estatística de interesse é obtida pela interpretação do coeficiente de regressão produzido pelo modelo. Quando um modelo é ajustado com mais de uma VI, os coeficientes informam medidas de associação entre a VD e a respectiva VI de interesse controlado pelas demais VI presentes no modelo. Isso é de fundamental importância quando se deseja neutralizar o efeito que outras variáveis têm na VD ao estimar a associação entre esta e uma VI de interesse. Por exemplo: um investigador deseja estimar a associação entre



uso de drogas e tentativa de suicídio e possui uma amostra com participantes das mais diversas idades. Sabe-se que a idade também é uma variável associada ao suicídio, portanto, ele deseja estimar uma associação entre uso de drogas e suicídio de forma que a influência da idade não esteja diluída nessa associação. Para isso, na análise, o investigador identifica a variável “suicídio” como a VD e as variáveis “uso de drogas” e “idade” como VIs. Assim, o coeficiente da regressão referente a VI “uso de drogas” informará a associação entre uso de drogas e suicídio controlada por idade, ou seja, o efeito da idade no suicídio será captado no outro coeficiente, referente à VI “idade”, fazendo com que o primeiro apresente a associação desejada sem que a idade influencie nela. Portanto, quando um modelo é utilizado para estimar medidas de associação entre VI e VD via interpretação dos coeficientes, a modelagem utilizada se encaixa na abordagem de modelagem descritiva.

O segundo tipo de modelagem é a explicativa. Essa modelagem tem por objetivo investigar hipóteses de causalidade entre um ou mais fatores de risco (VI) e um desfecho (VD) através da aplicação de um modelo aos dados observados. Os modelos de regressão tradicionais da estatística são os mais utilizados para esse fim e o interesse principal é gerar interpretações a respeito de alterações no desfecho atribuíveis à variabilidade dos fatores de risco via interpretação dos coeficientes e fortemente endossado pela literatura. Portanto, modelos explicativos são baseados em potenciais fatores causais, utilizados para avaliar se fatores de risco particulares explicam a ocorrência ou curso de, por exemplo, uma doença. Esse tipo de modelagem é muito utilizado pela área de epidemiologia, sua estrutura costuma ser apresentada por um diagrama causal e sua forma mais adequada de investigação é por estudos longitudinais, como estudos de coorte ou ensaios clínicos. Esse tipo de modelagem, aliada ao delineamento do estudo, é a única que permite concluir o risco de determinado fator para um desfecho de forma a incorporar temporalidade. Existem outros conceitos importantes na modelagem explicativa, como confundimento, mediação e moderação da medida de efeito.

O terceiro tipo de modelagem é a preditiva. Essa modelagem tem por objetivo construir um modelo que capture os padrões oriundos de um grupo de preditores (variáveis independentes) a fim de projetar, ou prever, o comportamento de um desfecho (variável dependente). Essa modelagem não assume causalidade, portanto, as relações entre preditores e desfecho refletem associações ao invés de explicações (como na modelagem explicativa). O foco principal da modelagem preditiva é a avaliação da qualidade do ajuste do modelo aos dados reais, visando produzir estimativas precisas do desfecho usando dos padrões e associações capturados entre os preditores e o desfecho. Com base nisso, o modelo gera estimativas que podem ser comparadas aos dados reais do desfecho usados para ajustar o

modelo e então avaliar a diferença entre predito e observado, para concluir se o modelo de fato possui poder preditivo satisfatório.

Os GLM, portanto, são um conjunto de modelos estatísticos altamente flexíveis, com capacidade de estimar estatísticas de interesse, fornecer inferências baseadas em hipóteses e de modelar variáveis visando predição, explicação e/ou descrição. Sua difusão, compreensão e aplicação por pesquisadores de inúmeras áreas acadêmicas, principalmente naquelas não inseridas na esfera das ciências exatas, ocorrem de forma gradual.

### 1.5.1. REGRESSÃO DE POISSON COM VARIÂNCIA ROBUSTA

Em pesquisas epidemiológicas e da área da saúde em geral, é frequente o objetivo de estimar medidas de associação entre variáveis de exposição e desfecho (variáveis dependente e independente). Tais medidas são apropriadas de acordo com o delineamento do estudo utilizado. Para estudos de caso-controle a medida de associação mais apropriada é a razão de chances (RC), para estudos transversais é a razão de prevalências (RP) e para estudos de coorte é o risco relativo (ou razão de riscos, RR) (Fletcher and Fletcher, 2006). Vale ressaltar que a RP e o RR são medidas matematicamente idênticas, ambas são proporções, sendo que sua diferença se baseia em conceitos teóricos epidemiológicos. A RP é uma razão entre prevalências, sendo a prevalência uma proporção da presença de determinado desfecho na população ou amostra num recorte temporal fixado, enquanto o RR é uma razão de incidências, sendo a incidência uma proporção de novos casos de presença do desfecho dentro de um período de tempo fixado. Ou seja, os casos de presença do desfecho que já estavam estabelecidos no início do período de tempo de observação de um estudo são considerados no numerador para o cálculo da prevalência neste período, mas não são para o de incidência.

Entretanto, historicamente observa-se o uso da RC como estimativa ou aproximação da RP ou RR devido a ampla difusão do modelo de regressão logística, apropriado para modelagem de variáveis dependentes categóricas e que estima a RC pela exponencial dos coeficientes de regressão. Sendo a RC muitas vezes interpretada como uma RP ou RR. No entanto, esta aproximação só é válida para casos onde a proporção da presença do desfecho é rara (menor que 10%) em ambas as categorias da variável de exposição, caso contrário a RC superestima o valor da RP (ou RR). Como alternativa à popular regressão logística para

estimar RP (ou RR), é fortemente indicado o uso da regressão log-binomial ou então da regressão de poisson com variância robusta (Barros and Hirakata, 2003; Zou, 2004).

A regressão de Poisson é canonicamente apropriada para modelagem de variáveis quantitativas discretas positivas, mas pode ser usada para estimar RP ou RR para desfechos binários diretamente pelos coeficientes de regressão do modelo quando especificado com função de ligação logaritmo natural. Para definir o modelo, considere um desfecho binário  $Y$  (variável dependente, como tentativa de suicídio) onde a categoria de valor 0 representa “não tentou suicídio” e a de valor 1 representa “tentou suicídio”, considere também  $k$  preditores (variáveis independentes, como idade, transtornos psiquiátricos, traumas e variáveis sociodemográficas) definidos pelo vetor  $x = [x_1, x_2, \dots, x_k]$ . O modelo especifica que  $\ln(P(Y = 1 | x)) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k$ , onde  $\beta_0, \beta_1, \dots, \beta_k$  são os coeficientes de regressão do modelo. Se não houver termos de interação no modelo, a estimativa da RP (ou RR) é obtida pela exponencial dos coeficientes de regressão, sendo que cada medida de associação estimada é ajustada pelos demais preditores. Porém, a regressão de Poisson produz um intervalo de confiança menos preciso que o verdadeiro. Isso ocorre em consequência dos erros do modelo seguirem uma distribuição Poisson, que superestima os erros com resposta binária, cuja distribuição apropriada é a binomial. Entretanto, a precisão do intervalo de confiança pode ser corrigida com o modelo de Poisson com variância robusta, que utiliza um estimador robusto (estimador de Sanduíche) para as variâncias dos coeficientes de regressão (Nijem et al., 2009; Zou, 2004). Assim, considerando estudos transversais e de coorte, a regressão de poisson com variância robusta é indicada para estimar RP ou RR, respectivamente, diretamente através da exponencial dos coeficientes de regressão e sem prejuízo de imprecisão nas estimativas dos intervalos de confiança dos mesmos.

## 1.6. APRENDIZADO DE MÁQUINA

O Aprendizado de Máquina (AM) é um campo da pesquisa em Inteligência Artificial (IA), subcampo da engenharia e ciência da computação, que combina *expertises* como estatística multivariada, programação, otimização numérica e teoria da informação. O foco é ensinar o computador a reconhecer padrões através de exemplos e características encontradas nos dados para solução de problemas diversos (Lantz, 2015). As abordagens tradicionais para

programação se baseiam em regras que estabelecem como resolver um problema pela lógica, codificadas passo-a-passo. Enquanto as abordagens por AM podem ser definidas como uma tarefa automatizada, que lida com uma grande quantidade de dados para utilizar exemplos de como essa tarefa pode ser alcançada da maneira mais eficaz. Isso ocorre a partir do desenvolvimento de algoritmos que permitem aos computadores aprenderem através do processamento de dados empíricos, promovendo experiência e treinamento para se adaptarem aos diversos ambientes e desafios (Hu and Hao, 2012).

Em 1959, Arthur Samuel definiu AM como o "campo de estudo que dá aos computadores a habilidade de aprender sem serem explicitamente programados" (livre tradução) (Simon, 2013). Com o tempo, observa-se um aumento na ênfase da abordagem lógica indutiva (ao invés de simultaneamente indutiva e dedutiva como preconizada por Alan Turing na ampla área de IA), baseada em conhecimento empírico e voltada a abordar problemas solucionáveis de natureza prática. Assim, no início da década de 1990, o AM começa a sofrer uma surpreendente ascensão, impulsionada pelo expressivo crescimento do volume e variedade de dados disponíveis, bem como do aumento do poder de processamento computacional observado desde então. Esse fenômeno originou o conceito de *Big-Data* (em 1997) e popularizou os conceitos de ciência de dados (*Data Science*), mineração de dados (*Data Mining*), entre outros.

Muitos elementos do AM estão intimamente ligados, por vezes sobrepostos, à estatística computacional, uma disciplina que foca nas propriedades dos métodos estatísticos e sua complexidade computacional em fazer previsões através do uso de computadores. Ela tem fortes laços com a otimização matemática, que produz métodos, teoria e domínios de aplicação para este campo. O AM é usado em uma variedade de tarefas computacionais onde desenvolver algoritmos explícitos é impraticável e onde o objetivo é muitas vezes uma previsão completa, cuja máquina é adaptada para atender a meta de previsão (van der Laan and Rose, 2011). Nesse sentido, os métodos de modelagem por AM podem ser distinguidos em dois paradigmas da modelagem de dados, os modelos algorítmicos e os modelos probabilísticos (Breiman, 2001a). Este último incorpora os modelos tradicionais da área da estatística, como os GLM e alguns autores passaram a utilizar o nome Aprendizado Estatístico para este campo combinado (James et al., 2013).

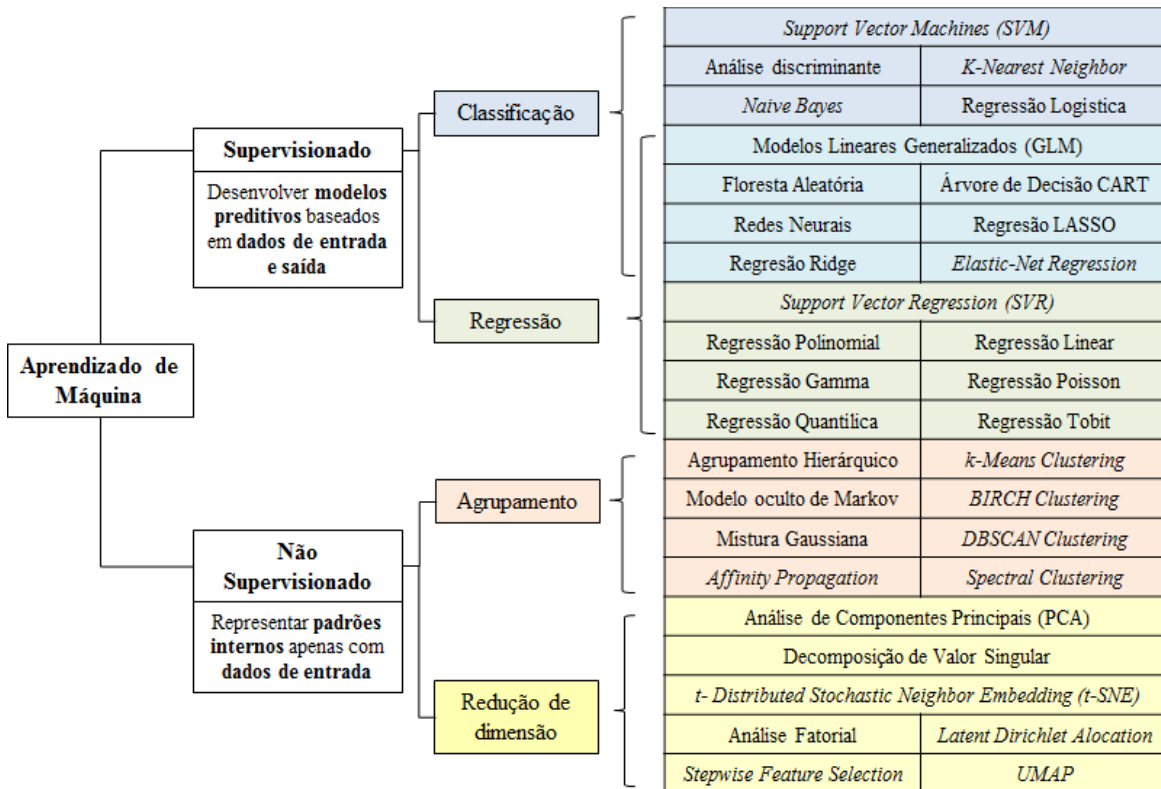
Os modelos de AM se subdividem principalmente em três tipos, os de aprendizagem supervisionada, não supervisionada e os de aprendizagem por reforço. Os de aprendizagem supervisionada são aqueles que permitem a criação de modelos preditivos e que atendem problemas de regressão e classificação. Neste tipo de aprendizagem, exemplos de dados de

entrada e de saídas desejadas são apresentadas ao modelo, que objetiva aprender uma regra geral que mapeia as entradas para as saídas. A saída pode ser uma variável quantitativa ou categórica, sendo que seus valores são chamados de rótulos pela literatura da área. É chamado de supervisionado pois as classes de saída desejadas são informadas, visando que o computador as reproduza através de padrões identificados entre estas e os demais dados de entrada. A filtragem de mensagens de spam nos aplicativos de e-mail é um exemplo de classificação, onde as entradas são mensagens de e-mail rotuladas nas classes "spam" e "não spam" (Drucker et al., 1999). Os modelos de regressão logística e de regressão linear também são exemplos de aplicação deste tipo de aprendizado.

Os algoritmos de aprendizagem não supervisionada são aqueles que utilizam apenas dados de entrada, sem rótulos desejados, com o objetivo de identificar similaridades, reduzir dimensão e definir agrupamentos (Hu and Hao, 2012; Lantz, 2015), como a análise de cluster e a análise de componente principais (técnicas já consolidadas pela estatística multivariada). Por último, os algoritmos de aprendizagem por reforço lidam com a exposição de um sistema a situações dinâmicas que geram penalizações ou recompensas, de acordo com a proximidade atingida do objetivo, fazendo com que o modelo calibre iterativamente suas respostas a novos cenários. Para cada tipo de aprendizagem há uma diversidade de modelos que se adequam às diferentes naturezas dos dados e objetivos de modelagem. A figura 3 apresenta algumas das diversas técnicas de AM supervisionadas e não supervisionadas.

A utilização de um método de AM supervisionado segue algumas etapas gerenciais, iniciando da mesma forma que todos os procedimentos estatísticos tradicionais, pela aquisição dos dados, seguido pela limpeza e preparação do banco de dados para análises. Após, um algoritmo de AM é escolhido conforme as especificações de natureza das variáveis e objetivo da análise para então ser aplicado, testado, diagnosticado e aprimorado. Para isso, o banco de dados é dividido em duas partições aleatoriamente, processo chamado de *hold-out*. O processo de aplicação é chamado de aprendizagem ou treinamento, pois é quando o algoritmo utiliza os dados empíricos para entender o modelo de relações entre as variáveis. A etapa de teste é quando o modelo treinado é aplicado na partição menor, ou seja, em uma amostra independente que não fora utilizada no processo de aprendizagem de máquina que ajustou o modelo. Alguns autores sugerem a proporção de 80% para treino e 20% para teste (James et al., 2013).

Figura 3 - Exemplos de algoritmos e modelos de Aprendizado de Máquina por tipo de aprendizado e objetivo.



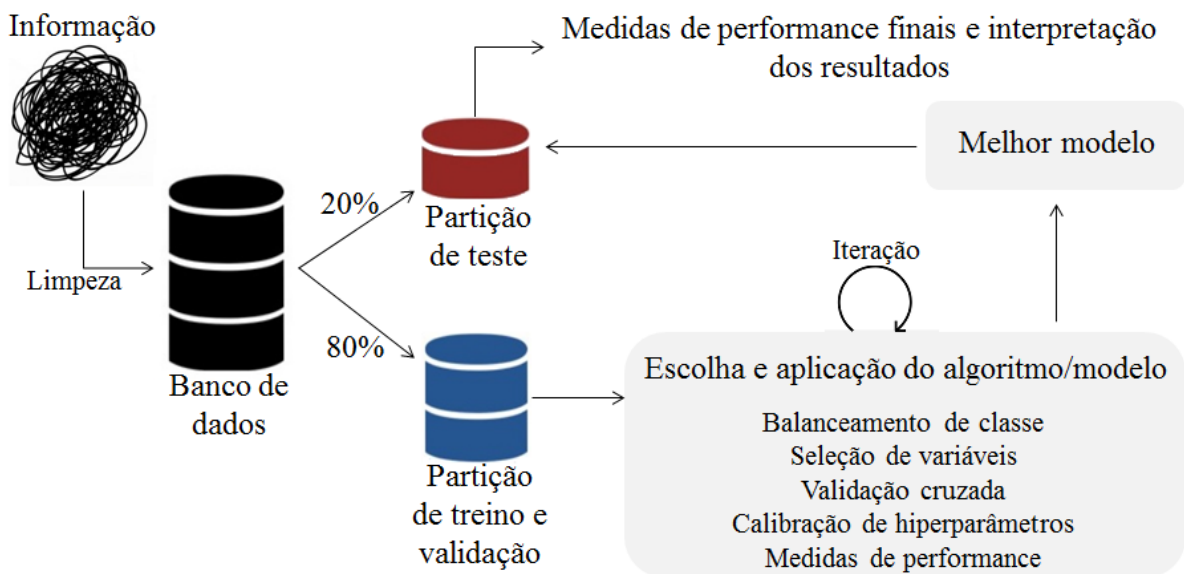
Fonte: Elaborado pelo autor, adaptado de múltiplas fontes online.

Na partição de teste, ocorre a avaliação do poder preditivo do modelo. Para modelos supervisionados classificatórios onde o desfecho é binário, isso é realizado através de medidas de qualidade de ajuste como a sensibilidade, especificidade, valor preditivo positivo (VPP) e valor preditivo negativo (VPN). As denominações destas medidas na área da saúde são as mesmas empregadas nos testes diagnósticos (nova técnica *versus* confirmatório), sendo que na área da computação o termo precisão (do inglês, *precision*) é utilizado em equivalência para ambas, sensibilidade (precisão da classe A) e especificidade (precisão da classe B). O termo revocação (do inglês, *recall*) é utilizado em equivalência para ambos VPP e VPN. Outras métricas importantes também muito utilizadas em ambas as áreas são a acurácia, acurácia balanceada e a área abaixo da curva ROC (*receiver operating characteristic curve*), sendo esta última a mais popular.

A área de AM também incorporou aplicações de técnicas de reamostragem para verificação de consistência e adequabilidade de modelos a dados empíricos, baseando-se no conceito tradicional de *bootstrap* e expandindo-o para uma validação cruzada (do inglês, *k-*

*fold cross-validation*). Essa técnica particiona o banco destinado à aprendizagem em  $k$  partições (*folds*) aleatoriamente. O algoritmo é então treinado em cada combinação de  $k-1$  partições e testado na restante, gerando assim uma medida de variabilidade que permite avaliar a consistência do modelo antes de testá-lo na partição final de teste. A técnica conhecida como *leave-one-out* é um caso particular da validação cruzada em que  $k=n$ . O principal benefício e objetivo dessa técnica é evitar o sobreajuste (*over-fitting*), que ocorre quando um modelo estuda os dados de treinamento a tal ponto que influencia negativamente a predição do modelo em novos dados. Um modelo sobreajustado apresenta medidas de desempenho ótimas na partição de treino e pobres na partição de teste (James et al., 2013). A figura 4 ilustra o processo geral de aplicação de um algoritmo de AM supervisionado.

Figura 4 - Fluxo geral da aplicação de um algoritmo de aprendizado de máquina.



Fonte: Elaborado pelo autor, adaptado de múltiplas fontes.

Como já mencionado, as aplicações de técnicas de aprendizado de máquina expandiram em consoância com o aumento expressivo de geração e armazenamento de dados. Estas técnicas foram propostas de modo que se tenha disponível um volume de dados suficiente para aplicação dos diversos métodos de repartição e reamostragem, como a validação cruzada, sem prejuízos na precisão dos resultados. Isso implica em uma importante limitação destes métodos frente a amostras pequenas, principalmente para os modelos preditivos.

### 1.6.1. APRENDIZADO DE MÁQUINA NA PESQUISA EM PSIQUIATRIA

O campo das adições é uma ampla área da medicina, que envolve domínios multifatoriais e subjetivos e onde tradicionalmente se observa uma tendência de abordagem estatística através de GLM, devido à sua popularidade e adequação às hipóteses de pesquisa. Neste contexto o objetivo é ajustar um modelo parcimonioso que gere estimativas e interpretações que respondam aos objetivos, o que pode ser falho devido à supressão de relações não lineares e possivelmente resultando em insuficiente capacidade preditiva. Para propiciar uma perspectiva analítica diferente deste cenário, algumas aplicações de AM emergiram no campo.

O uso de AM no campo da saúde mental iniciou-se na década de 1980 e atualmente apresenta ferramentas indispensáveis para uso em diferentes segmentos (Kononenko, 2001). Embora ainda incipientes, pesquisas na área utilizam diferentes metodologias para classificação, aumento da acurácia diagnóstica, identificação do perfil de pacientes, exploração e validação de indicadores do campo (Ahn et al., 2016; Ahn and Vassileva, 2016; Pariyadath et al., 2014). O tratamento para transtorno por uso de substâncias (TUS) é uma das áreas que pode se beneficiar de evidências que conduzam indivíduos a receber abordagens mais adequadas às suas características e demandas específicas, possivelmente auxiliando na formulação de planos terapêuticos (Shatte et al., 2019). A inclusão de abordagens de AM junto à experiência clínica aumenta a chance de compreensão dos complexos fatores subjacentes aos desfechos positivos em programas de tratamento para adição (Connor et al., 2007; Gowin et al., 2015). Assim, o AM se propõe como uma possibilidade viável de compreender as inter-relações de diferentes fatores em um plano analítico complexo, o que pode resultar em informações ainda não exploradas e úteis para a prática clínica.

A literatura científica acerca das aplicações de AM no campo da psiquiatria de adição recentemente deu seus primeiros passos. Encontram-se aplicações para predição de transtornos por uso de cocaína através de medidas de impulsividade (Ahn et al., 2016); identificação de marcadores de dependência química em heroína e metanfetamina através de características de personalidade, demográficas, psiquiátricas e neurocognitivas (Ahn and Vassileva, 2016); previsão de tempo de abstinência por meio da gravidade de dependência em álcool, fissura, qualidade de vida, medidas psicométricas e demográficas (Connor et al., 2007); entre outros estudos. Um notável fato sobre estes estudos revela a necessidade de explorar cada vez mais essa metodologia: resultados variando entre alta e excelente acurácia



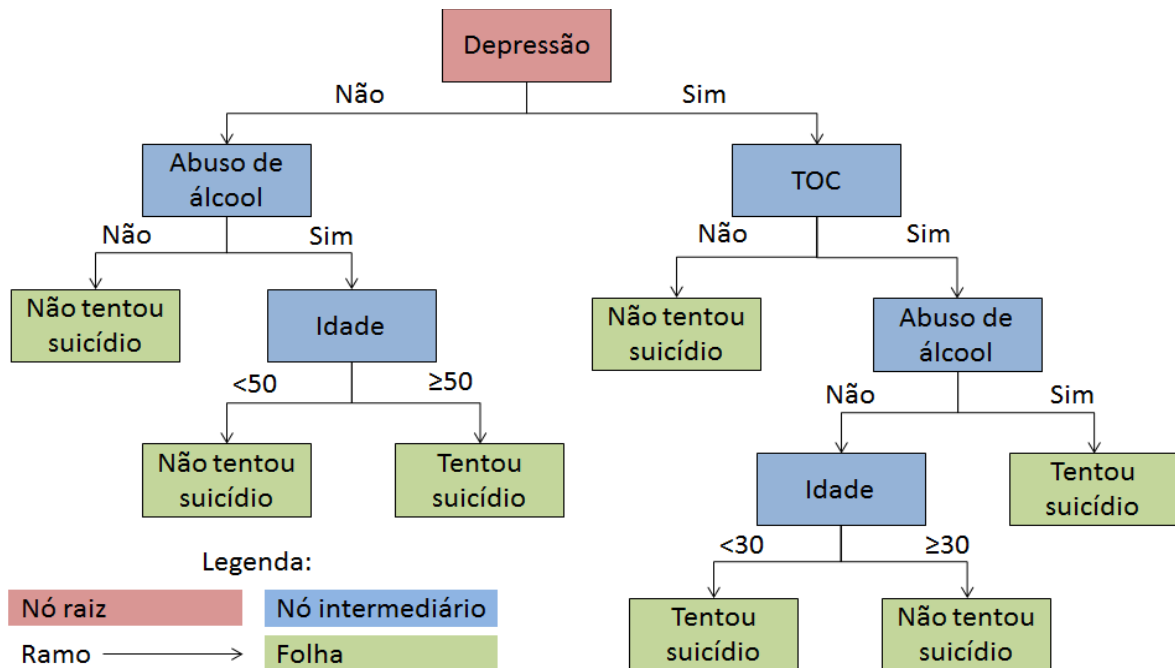
pela análise da área abaixo da curva ROC embora utilizando amostras relativamente pequenas.

Por fim, dois estudos merecem expressivo reconhecimento, o primeiro sendo uma revisão teórica de diversas aplicabilidades computacionais na área da psiquiatria, trazendo ricas discussões sobre dimensionalidade de constructos, abordagens para bancos de dados, classificação de diagnósticos, predição de resposta a tratamento e outros importantes aspectos que unem AM e psiquiatria (Huys et al., 2016). O segundo estudo avaliou uma grande amostra de indivíduos utilizando um banco de dados populacional para comparar 17 modelos preditivos de AM, utilizando 28 preditores entre características sociodemográficas, tipo de tratamento, drogas de preferência e problemas de saúde mental. O melhor dos modelos analisados alcançou AUC de 0,82 para predição de sucesso no tratamento de transtornos por uso de substância (Acion et al., 2017) e mostrou uma potencialidade preditiva acima da alcançada pelos modelos estatísticos tradicionais com termos de interação de segunda ordem.

### 1.6.2. ALGORITMO FLORESTA ALEATÓRIA

O algoritmo Floresta Aleatória (do inglês, *Random Forest*), proposto por Breiman (2001), é um método de aprendizado de máquina supervisionado da classe de algoritmos baseados em árvores e é construído pela combinação de diversas árvores de decisão (do inglês, *decision trees*). Uma árvore de decisão é um algoritmo supervisionado de modelagem preditiva por si só e consiste no encadeamento de regras de decisão (ramos) que tracejam caminhos desde um nó raiz até uma folha. Os nós consistem das variáveis preditoras, os ramos consistem das possíveis respostas a elas e a folha consiste da resposta predita à variável que se pretende fazer a predição (variável dependente ou desfecho segundo a nomenclatura estatística). Assim, uma árvore de decisão é construída fracionando um problema complexo de decisão (ou predição) em diversas decisões mais simples (Safavian and Landgrebe, 1991) e hierarquicamente estruturadas, podendo lidar com relações não lineares e interações complexas. Esse método foi apresentado por Breiman, Friedman, Olshen e Stone (Breiman et al., 1984) e pode ser usado para predição de variáveis categóricas ou quantitativas, sendo chamado de árvore de classificação ou árvore de regressão, respectivamente. A figura 5 ilustra um exemplo dessa estrutura.

Figura 5 - Exemplo de estrutura de uma árvore de decisão.



Fonte: Elaborado pelo autor com dados fictícios.

A construção de uma árvore de decisão inicia na definição da variável preditora do nó raiz e, subsequentemente, dos nós intermediários, até que todas as variáveis predictoras sejam utilizadas. A definição da variável em cada nó é realizada através do cálculo de uma métrica que quantifica a heterogeneidade da variável desfecho entre cada ramo do nó. As métricas mais populares são o índice de impureza de Gini e a entropia. A variável preditora cujo os ramos maximizam a heterogeneidade do desfecho, consequentemente maximizando a homogeneidade do desfecho dentro de cada ramo, é escolhida para ser o nó. Portanto, o nó raiz é interpretado como a variável preditora mais importante da árvore.

Árvores de decisão são consideradas modelos preditivos não paramétricos, pois o método não assume qualquer suposição quanto à distribuição de probabilidade das variáveis utilizadas para fornecer um resultado válido. A determinação de uma regra de decisão em um ramo que parte de uma variável quantitativa, por exemplo, é feita testando-se diversas combinações de divisões iterativamente, até encontrar a configuração que maximiza a métrica utilizada. Isso torna o método mais robusto e resistente à dados discrepantes.

No entanto, uma única árvore de decisão é considerada um modelo preditivo pobre. Embora possua baixo viés preditivo, geralmente apresenta alta variabilidade e baixa acurácia, sendo ineficiente para predição. Este problema ocorre também com outros algoritmos, chamados de aprendizes fracos (do inglês, *weak learners*), e pode ser completamente superado com o uso de técnicas de reamostragem. O processo é similar à validação cruzada

mencionada anteriormente, mas utilizado dentro do procedimento de ajuste inicial de um modelo em particular, a ideia é gerar diversas subamostras com os dados de treino aleatoriamente, ajustar uma árvore de decisão para cada subamostra exceto uma, submeter cada elemento desta à predição por cada uma das diversas árvores, e finalmente, combinar as predições para cada elemento através de média (para árvores de regressão) ou votação pela maioria (para árvores de classificação) para se obter uma predição final para cada elemento. A

Figura 6 - Exemplo de fluxo do processo de predição de um algoritmo de Floresta Aleatória.

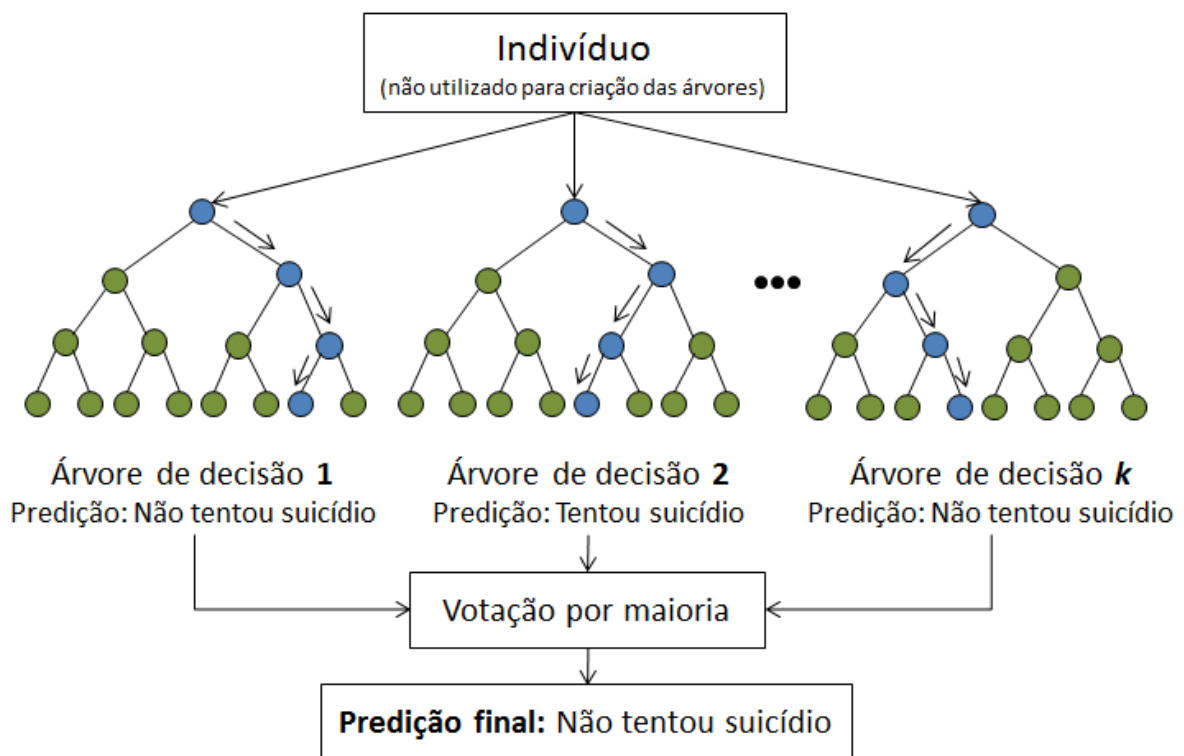


figura 6 ilustra um exemplo de fluxo para o procedimento.

Fonte: Traduzido e adaptado de William Koehrsen (<https://mudium.com> acessado em 20/04/2020).

Esse processo reduz a variabilidade da função de predição e aumenta substancialmente sua acurácia. Quando o processo de reamostragem é feito selecionando subamostras com igual probabilidade, o método de predição é chamado de *bagging* (referente à *bootstrapping aggregating*). Quando o processo de reamostragem utiliza da predição realizada com uma subamostra para atribuir pesos aos dados conforme suas contribuições na predição para que a próxima subamostra seja gerada com probabilidades desiguais, o método é chamado de *boosting*.

A combinação de algoritmos para produção de um único modelo preditivo insere este modelo composto numa classe de modelos chamada de *ensemble models*. *Bagging* e *boosting* são técnicas comuns para criação de *ensemble models*. Entretanto, é comum encontrarmos no banco de dados variáveis mais fortemente associadas com o desfecho e, portanto, estas seriam atribuídas aos nós raiz ou nós de alta hierarquia em praticamente todas as árvores de decisão, fazendo com que as árvores sejam muito similares e os resultados das predições das diversas árvores sejam correlacionados. A correlação entre as árvores implica na existência de certo grau de redundância na informação fornecida pelas predições e limita os benefícios do uso da média (ou votação pela maioria) destas como medida agregadora, além de não reduzir a variância da função de predição.

Para contornar esse problema, uma terceira técnica de reamostragem foi proposta, onde o processo é quase idêntico ao *bagging*, com a exceção de que em cada nó das árvores ajustadas às subamostras há uma limitação para uso de um subgrupo aleatório de variáveis preditoras ao invés de dispor de todas elas na definição do nó. A aleatorização é feita sem reposição e a técnica impede que preditores fortemente associados com o desfecho influenciem uma grande parcela das predições produzidas, reduzindo a correlação entre as árvores e ainda impondo o uso de todos os preditores disponíveis. Por esse motivo, problemas como a multicolinearidade, que prejudica os modelos clássicos de regressão, não ameaçam este método. Para o caso de modelos baseados em árvores, esse método mostrou superioridade em relação às outras técnicas de reamostragem e é chamado de Floresta Aleatória.

Esse algoritmo, assim como muitos outros algoritmos preditivos, possuem variáveis de configuração que controlam o próprio processo de treinamento e agindo como parâmetros de performance, estas são chamadas de hiperparâmetros. Por exemplo, o modelo algorítmico Floresta Aleatória possui dois hiperparâmetros principais: (1) número de árvores de decisão combinadas e (2) número de variáveis preditoras para avaliação em cada nó (Hastie et al., 2009; James et al., 2013). Uma Floresta Aleatória pode, por exemplo, estabilizar seu poder preditivo em 500 árvores de decisão, não precisando gerar mais do que isso. Os hiperparâmetros podem ser calibrados na fase de treinamento de modo que se maximize o poder preditivo do modelo sem aumentar o custo de processamento computacional.

## 2 OBJETIVOS

Sob a perspectiva do embasamento teórico apresentado nas seções anteriores e visando aplicação de duas técnicas analíticas, sendo um modelo estatístico (regressão de Poisson com variância robusta) e um algoritmo de Aprendizado de Máquina supervisionado (Floresta Aleatória), apresentam-se a seguir os objetivos almejados por esse trabalho.

### 2.1 OBJETIVO PRINCIPAL

Desenvolver um modelo preditivo de tentativa de suicídio, assim como investigar fatores associados à tentativa de suicídio em pacientes com transtorno por uso de cocaína-crack internados, estratificando por gênero.

### 2.2. OBJETIVOS SECUNDÁRIOS

- Estimar prevalência de tentativa de suicídio, traumas na infância e comorbidades psiquiátricas em homens e mulheres dependentes de cocaína-crack internados;
- Identificar preditores de suicídio específicos da população de homens e mulheres dependentes de cocaína-crack internados;
- Estimar razões de prevalência para fatores associados à tentativa de suicídio em dependentes de cocaína-crack internados, estratificado por gênero;
- Discutir sobre o uso de diferentes tecnologias que propiciem novas análises no campo da psiquiatria de adição.

A seguir, são apresentadas as duas publicações realizadas na elaboração deste trabalho a fim de alcançar os objetivos citados. O artigo 1 refere-se a um artigo original publicado no periódico PLOS ONE que explora preditores e fatores associados à tentativa de suicídio em homens e mulheres internados com transtorno por uso de crack. E o artigo 2 refere-se a um editorial sobre uso de drogas e suicídio publicado no periódico *Brazilian Journal of Psychiatry*.

### 3. ARTIGO 1

PLOS ONE – Original Manuscript

## **Prediction of attempted suicide in men and women with crack cocaine use disorder in Brazil**

Vinícius Serafini Roglio<sup>1,2</sup>, Eduardo Nunes Borges<sup>1,3</sup>, Francisco Diego Rabelo-da-Ponte<sup>1,2,4</sup>, Felipe Ornell<sup>1,2</sup>, Juliana Nichterwitz Scherer<sup>1</sup>, Jaqueline Bohrer Schuch<sup>1,2</sup>, Ives Cavalcante Passos<sup>4</sup>, Breno Sanvicente-Vieira<sup>5</sup>, Rodrigo Grassi-Oliveira<sup>5</sup>, Lisia von Diemen<sup>1,2</sup>, Flavio Pechansky<sup>1,2</sup>, Felix Henrique Paim Kessler<sup>1,2</sup>

**1** Center for Drug and Alcohol Research, Hospital de Clínicas de Porto Alegre, Universidade Federal do Rio Grande do Sul, Brazil. **2** Graduate Program in Psychiatry and Behavioral Sciences, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil. **3** Center for Computational Sciences, Universidade Federal do Rio Grande, Brazil. **4** Molecular Psychiatry Laboratory, Hospital de Clínicas de Porto Alegre, Universidade Federal do Rio Grande do Sul, Brazil. **5** Developmental Cognitive Neuroscience Lab, Pontifical Catholic University of Rio Grande do Sul, Brazil.

Corresponding author: Eduardo Nunes Borges ([eduardoborges@furg.br](mailto:eduardoborges@furg.br))

OPEN ACCESS

<https://doi.org/10.1371/journal.pone.0232242>

Editor: Vincenzo De Luca, University of Toronto, CANADA

Received: July 23, 2019

Accepted: April 10, 2020

Published: May 4, 2020

Copyright: © 2020 Roglio et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Abstract**

**Background:** Suicide is a severe health problem, with high rates in individuals with addiction. Considering the lack of studies exploring suicide predictors in this population, we aimed to investigate factors associated with attempted suicide in inpatients diagnosed with cocaine use disorder using two analytical approaches.

**Methods:** This is a cross-sectional study using a secondary database with 247 men and 442 women hospitalized for cocaine use disorder. Clinical assessment included the Addiction Severity Index, the Childhood Trauma Questionnaire, and the Structured Clinical Interview for the Diagnostic and Statistical Manual of Mental Disorders, totalling 58 variables. Descriptive Poisson regression and predictive Random Forest algorithm were used complementarily to estimate prevalence ratios and to build prediction models, respectively. All analyses were stratified by gender.

**Results:** The prevalence of attempted suicide was 34% for men and 50% for women. In both genders, depression ( $PR_M=1.56$ ,  $PR_W=1.27$ ) and hallucinations ( $PR_M=1.80$ ,  $PR_W=1.39$ ) were factors associated with attempted suicide. Other specific factors were found for men and women, such as childhood trauma, aggression, and drug use severity. The men's predictive model had prediction statistics of  $AUC=0.68$ ,  $Acc.=0.66$ ,  $Sens.=0.82$ ,  $Spec.=0.50$ ,  $PPV=0.47$  and  $NPV=0.84$ . This model identified several variables as important predictors, mainly related to drug use severity. The women's model had higher predictive power ( $AUC=0.73$  and all other statistics were equal to 0.71) and was parsimonious.

**Conclusions:** Our findings indicate that attempted suicide is associated with depression, hallucinations and childhood trauma in both genders. Also, it suggests that severity of drug use may be a moderator between predictors and suicide among men, while psychiatric issues shown to be more important for women.

**Keywords:** Suicide; attempted suicide; suicidal behavior; crack-cocaine; smoked cocaine; snorted cocaine; drug use disorder.

## Introduction

Suicide is a complex global health problem that currently accounts for 1.4% of premature deaths worldwide. According to the World Health Organization, a total of 800,000 suicides are documented each year, which corresponds to one completed suicide every 40 seconds. Unlike many diseases and risk behaviors, the mortality rate has not declined over the past two decades [1]. For instance, suicide is among the top three causes of death of people aged 15-44 [2]. Suicide prevalence rates differ by country, varying according to the heterogeneity of cultural, social, economic, and environmental factors [3]. Furthermore, each of the components of suicidal behavior, such as ideation, attempt, and mortality, also have different epidemiology. [4]. The etiology of suicide is heterogeneous, although the literature shows that previous suicide attempts and the presence of psychiatric disorders are the most established predictors of suicide [5–8]. Within this category, substance use disorders (SUD), such as cocaine and alcohol use disorders, are among the most significant predictors of suicide [9].

In Brazil, crack-cocaine use and suicidal behavior are public health problems [10–12]. A recently large-scale study with a representative sample of Brazilians found rates of suicidal attempts and deaths in the general population of 9.9 and 5.4%, respectively. In crack-cocaine users these rates were significantly higher, 40.0 and 20.8%, respectively [13]. Previous studies have shown that 47% of crack-cocaine users had a current suicide risk [14], and a prevalence of suicidal behaviors of 30% in crack-cocaine addicts, in Brazil. The subject of suicide has been the focus of studies in psychiatry in the last decades, but the understanding about this behavior remains insufficient [15]. Moreover, the predictive factors in this vulnerable population were not well explored yet. These individuals present multiple psychosocial vulnerabilities, high rates of clinical and psychiatric comorbidities, and low adherence to treatment [16–19], which may have a significant impact on suicidal behavior. Furthermore,



crack and cocaine are the illicit drugs that most lead to demand for detoxification treatment in psychiatric facilities [20,21], generating considerable cost to the public health system.

Gender-related aspects also seem to influence or mediate the drug use profile and the severity of crack-cocaine dependence [22,23] and may be potentially associated with suicide [3,24–26]. Although some predictors of suicide are common to both genders, such as prior mental disorders, substance abuse, and early exposure to violence, some other potential risk factors appear to be gender-specific [26]. Female-specific risk factors are related to physical and sexual abuse, eating disorders, bipolar disorder, depressive symptoms, interpersonal problems and previous abortion. Male-specific risk factors are often related to disruptive behavior, hopelessness, family or friend history of suicide, mental disorders due to alcohol and drug abuse, externalizing disorders, homelessness, and access to means [27,28]. Considering the differences in suicidal behavior between men and women, targeted and tailored prevention and intervention strategies are recommended [24]. However, only a few studies have evaluated gender differences related to suicide, especially among cocaine users, which is still a gap to be explored in the field of psychiatry.

Traditional statistical methods have provided valuable insights into addiction psychiatry. However, few papers have explored the complex relationship between suicide and its predictors in men and women with cocaine use disorder [29]. The use of different analytical methods, such as machine learning, to complement each other, may be an essential step towards better understanding the phenomenon of suicide, bringing advances to the field [30] and providing a foundation for changes to treatment protocols.

This study is driven by the hypothesis that the characteristics of subjects who attempt suicide differ between men and women with cocaine use disorder, who have distinct psychiatric and substance use profiles. We aim to investigate associated factors and predictors of suicide attempts in cocaine use disorder inpatients, through descriptive and predictive

models using secondary databases obtained from retrospective cross-sectional studies, and stratifying by gender.

## **Methods**

Analytical procedures were organized in distinct phases according to the Knowledge Discovery in Databases (KDD) methodology [31]. First, we presented information about study participants and data selection. Then, we detailed preprocessing and data transformation, including creation of new variables. Finally, we presented all statistical analyses and data mining, including configuration of the machine learning algorithms, solutions to the problem of class imbalance, evaluation measures, and the software packages used in the implementation.

### ***Participants and assessments***

Our data sources are secondary databases populated with data from 2012 to 2018 from two inpatient units specialized in addiction treatment in Porto Alegre, Southern Brazil. This study was approved by the Hospital de Clínicas de Porto Alegre (HCPA) Research Ethics Committee (number 180166). The inclusion criteria were as follows: (1) fulfilled DSM-IV criteria for cocaine use disorder; (2) reporting crack-cocaine as the preferred drug in case of multiple drug use; (3) 18 or more years of age; (4) absence of cognitive deficits that impair the reliability of the answers; and (5) voluntarily acceptance of all research procedures via written informed consent.

Some clinical variables were extracted from the hospital medical records, but most data were collected using three instruments: the sixth version of the Addiction Severity Index (ASI-6) [32,33]; the Structured Clinical Interview for DSM-IV axis I disorders (SCID-I) [34]; and the Childhood Trauma Questionnaire (CTQ) [35,36]. Trained researchers assessed all

participants by administration of the Brazilian versions of clinical and semi-structured interviews. Details of the data collection procedure are described elsewhere [37,38]. The initial sample comprised 863 individuals who met the inclusion criteria. We pre-selected a set of variables related to suicide according to recent literature reviews [39,40].

### ***Preprocessing and data transformation***

The data sample selected in the previous phase was preprocessed in order to clean and standardize variable types, formats, and content. We then created new variables based on transformations and combinations of the original ones. Supplementary material Text S1 presents detailed information on the method for creating these variables. We included sociodemographic variables; tuberculosis, hepatitis and HIV infection, and number of other chronic diseases; life proportion since first treatment for alcohol or drugs; years and life proportion of regular use of alcohol, cannabis, cocaine (snorted and smoked), and tobacco; age at first use of each substance; number of distinct psychoactive substances used for more than 50 days during lifetime; alcohol and drug craving and withdrawal symptoms at the time of hospitalization; social support comprising partner, adult relatives and/or close friends; close relationship(s) with other drug user(s); difficulty in controlling aggressiveness, talking about feelings, and enjoying leisure time; physical abuse (childhood and lifetime); sexual abuse (childhood and lifetime); dichotomized CTQ scores; idealization of upbringing, based on the statements that are not used to compute trauma scores in the CTQ; SCID-I diagnoses grouped and dichotomized as psychotic, bipolar, depression, obsessive-compulsive, post-traumatic stress, alcohol use, eating and anxiety disorders. The outcome of attempted suicide is binary with answers: “no suicide attempts in lifetime” and “at least one suicide attempt (under drug effect or not).”

We also removed noisy instances and individuals with sparse information. We excluded 42 subjects because the outcome was missing. Several other participants were removed because the instruments were not sufficiently complete: 38 for the ASI-6, 80 for the SCID-I, and 34 for the CTQ. We then dropped all variables with more than 7% missing values. This threshold was chosen based on analysis of the missing data pattern presented in Fig S1. At this point, we observed that only 4 out of 57 variables had more than 3% of data missing. The final sample comprised 57 variables on 669 inpatients (247 men and 422 women), representing our study population.

### ***Data mining and statistical analysis***

In this study, we adopted two distinct approaches to analyze data. We first performed a Generalized Linear Model (GLM) with Poisson distribution, log link, and robust estimation of variance to calculate attempted suicide prevalence ratios (PR), to investigate risk factors. PR was chosen as the ideal statistic rather than the odds ratio because it is considered more adequate for our study design, and interpretation is more intuitive [41]. With this approach, we aimed to descriptively explore variables associated with attempted suicide using a partitioned deviance statistical method. This analysis was performed with the whole sample (still stratifying by gender). Associations between suicide attempt and all variables were investigated through PR estimations controlled for age and ethnicity. Then, all statistically significant variables, along with age, were used as independent variables in a descriptive multiple GLM. Finally, stepwise selection with backward elimination was applied until all main effect p-values were  $< 0.2$  (except for age, used to control).

### ***Machine learning approach***

The second approach was a supervised machine learning algorithm called Random Forest (a type of predictive modelling), which requires a data preparation. The dataset was randomly split into distinct partitions, 80% for training/validating and 20% for testing. We ran the Multivariate Imputation by Chained Equations (MICE) algorithm [42] to impute missing training data, setting the number of iterations equal to 10, and used the Random Forest imputations method for all variables. A detailed description of data imputation is presented in the supplementary material and illustrated in Fig S2.

We applied the Recursive Feature Elimination (RFE) algorithm [43] to reduce the complexity of the predictive model, by retaining only the variables that most predict attempted suicide. This algorithm starts by fitting the model to all predictors. At each iteration, the features are ranked using their importance to prediction. The least relevant feature is removed, and the model is refitted. The set of features with the best performance is determined, and they are used to learn the final model.

Random Forest [44] is an ensemble method for supervised learning. It grows hundreds of different decision trees and combines them as a single classifier using majority voting. Each tree grows using a bootstrap of the training set and random search of best splits, parameterized by the number of candidate predictors. We used repeated cross-validation [45] for all learned models. In this schema, the original training dataset is randomly partitioned into  $k$  equal size samples, where each one is used as the validation data, with the remaining samples used as training data. This procedure is repeated  $n$  times yielding several random partitions of the original training dataset. The results are averaged to produce a single prediction quality estimation. The advantage of this method is that all examples are used for both training and validation in several distinct models. We used  $k=10$  and  $n=5$  as resampling method for RFE and for tuning the Random Forest hyperparameter. We varied the randomly selected predictors using a grid search incrementing its value from 1 to  $1.5\sqrt{sf}$ , where  $sf$  is

the number of selected features. We used the area under the Receiver Operating Characteristic (ROC) curve (AUC) [46] to evaluate prediction quality. The number of candidate predictors with the best AUC was selected to fit the model with all training data.

#### *Class imbalance problem*

When the frequency of suicide attempts had a significant negative impact on model fitting, we dealt with the class imbalance problem by subsampling the training data using three distinct strategies: down-sampling that randomly samples the majority class to be the same size as the minority; up-sampling that randomly samples with replacement the minority class to be the same size as the majority; and SMOTE [47], which is a hybrid technique that down-samples the majority class and synthesizes new data points in the minority.

#### *Assessment of model performance*

Final predictive models were evaluated using the following quality measures: AUC, Balanced Accuracy, Sensitivity, Specificity, Positive Predictive Value (PPV), and Negative Predictive Value (NPV). The classification threshold was defined as the point that maximized the sum of Sensitivity and Specificity. Additionally, the result of a statistical test of difference of proportions upon the null hypothesis of equality between No Information Rate (NIR) and Accuracy was reported. NIR is the best guess given no information beyond the overall distribution of the outcome, i.e., it is the prevalence of the majority category.

#### ***Implementation***

Data curation and GLM were implemented using IBM SPSS syntax because the original databases were stored in SPSS native format. All the procedures and analyses of remaining phases of the KDD process were implemented in R scripts using RStudio IDE and

the packages VIM (visualize missing data), mice (imputation), DMwR (SMOTE sampling), caret (training), randomForest (predictive algorithm), and pROC (plot ROC curves).

We decided to stratify all analyses by gender because suicide is a phenomenon with signatures that differ between genders in terms of the lethality of the method used, psychosocial risk factors and prevalence of psychiatric disorders [24,48].

## **Results**

### ***Descriptive approach***

Table 1 presents the sample's sociodemographic profile and the prevalence rates of attempted suicide within those variables, stratified by gender and controlled by age and ethnicity. The overall prevalence of attempted suicide was 44%, 34% for men and 50% for women. For men, there was a significant association between age and suicide attempt, showing a 3% increase in prevalence of attempted suicide for each incremental year of age. No other direct relationship was found between the outcome and any other sociodemographic variables.

Similar analyses were performed for all social and family-related variables (Table S1), and all psychiatric, clinical and drug-related variables (Table S2). For men, 18 variables had statistically significant prevalence ratios, and another set of 18 variables was detected in the women's data. Eleven were significant in common to both genders, including hallucinations not under the effect of drug or abstinence, alcohol withdrawal symptoms, childhood physical abuse, drug withdrawal symptoms, being a victim of a violent crime, depressive disorder, childhood emotional abuse, and psychiatric hospitalization not related to alcohol or drug use, among a few others. The three highest PR out of the seven variables significant for men only were for physical abuse (PR=2.14, p=.009), childhood physical neglect (PR=1.74, p=.001), and having a close relationship with other drug dependent people (PR=1.73, p=.010).

Similarly, the highest PR for women were for difficulty controlling aggression (PR=1.50, p=.001), childhood sexual abuse (PR=1.36, p=.003), and chronic respiratory disease (PR=1.30, p=.006).

**Table 1.** Sociodemographic variables and prevalence ratios (PR) for attempted suicide, stratified by gender.

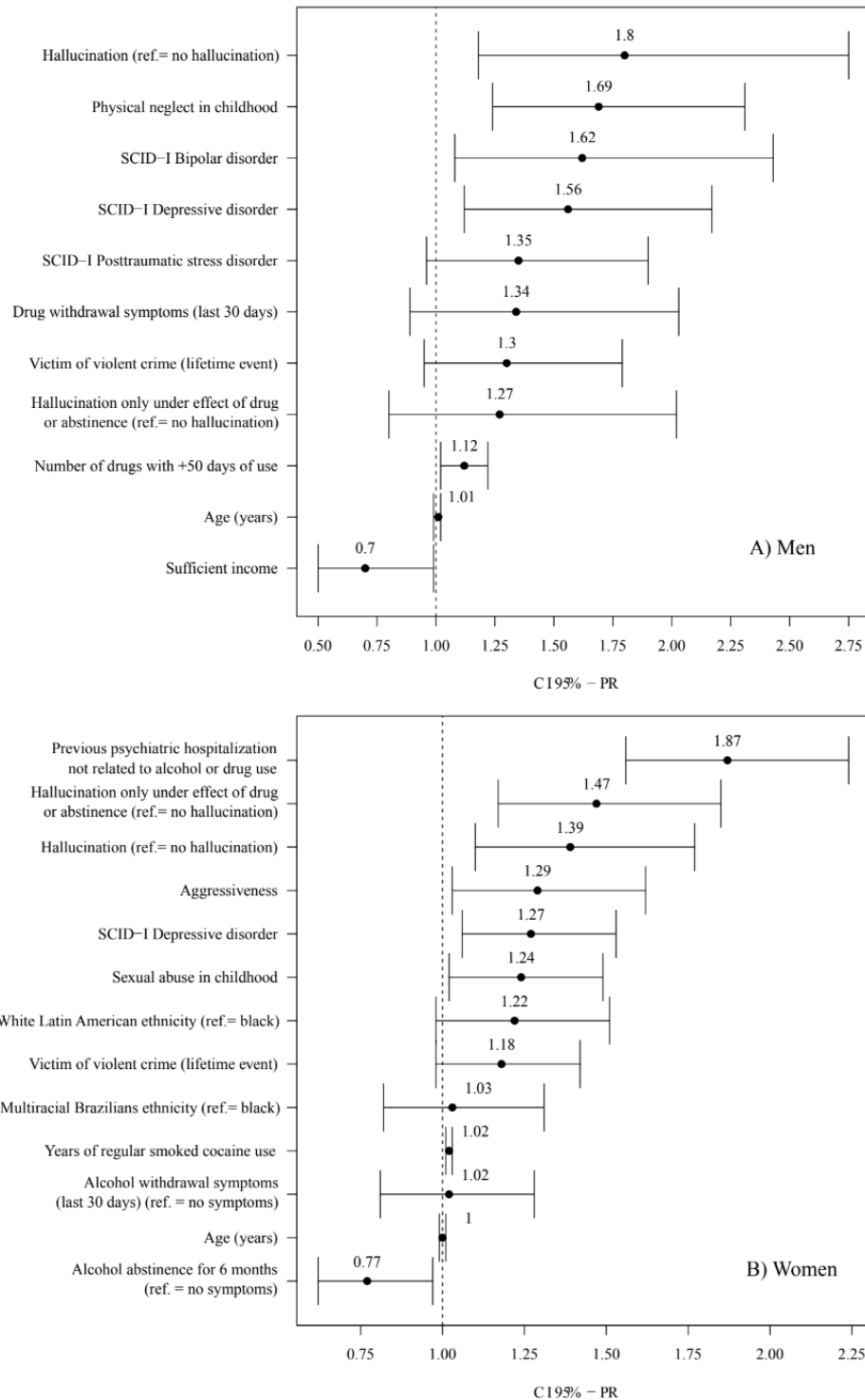
	Gender*		Attempted suicide**					
	Men	Women	Men	PR	p	Women	PR	p
<b>Age (years) <sup>1</sup></b>	34.3 ± 8.5	31.4 ± 8.7	-	1.02	0.03	-	1.00	0.551
<b>Ethnicity <sup>2</sup></b>								
Black	50 (20.5)	172 (40.8)	19 (38.0)	ref.	-	77 (44.8)	ref.	-
Multiracial Brazilians	54 (22.1)	98 (23.2)	14 (25.9)	0.72	0.252	53 (54.1)	1.20	0.158
White American Latin	140 (57.4)	152 (36.0)	52 (37.1)	0.95	0.812	81 (53.3)	1.18	0.153
<b>Education <sup>2</sup></b>								
None	58 (23.6)	147 (35.9)	20 (34.5)	ref.	-	77 (52.4)	ref.	-
Basic Education	109 (44.3)	153 (37.4)	39 (35.8)	1.15	0.523	66 (43.1)	0.82	0.111
Secondary Education	79 (32.1)	109 (26.7)	27 (34.2)	1.06	0.808	60 (55.0)	1.04	0.768
<b>Occupational status <sup>2</sup></b>								
Unemployed	97 (39.6)	203 (48.1)	39 (40.2)	ref.	-	100 (49.3)	ref.	-
Informal job	62 (25.3)	134 (31.8)	23 (37.1)	0.89	0.590	66 (49.3)	1.00	0.529
Employed	86 (35.1)	85 (20.1)	24 (27.9)	0.76	0.200	45 (52.9)	1.09	0.465
<b>Marital status <sup>2</sup></b>								
Married	74 (30.5)	137 (35.7)	29 (39.2)	ref.	-	66 (48.2)	ref.	-
Non-married	169 (69.5)	247 (64.3)	55 (32.5)	0.86	0.421	125 (50.6)	1.05	0.674

\*Summary of variables in the line within gender by <sup>1</sup>mean ± standard deviation or <sup>2</sup>frequency (%).

\*\*Summary of attempted suicide (yes) within rows and PR controlled by age and ethnicity.



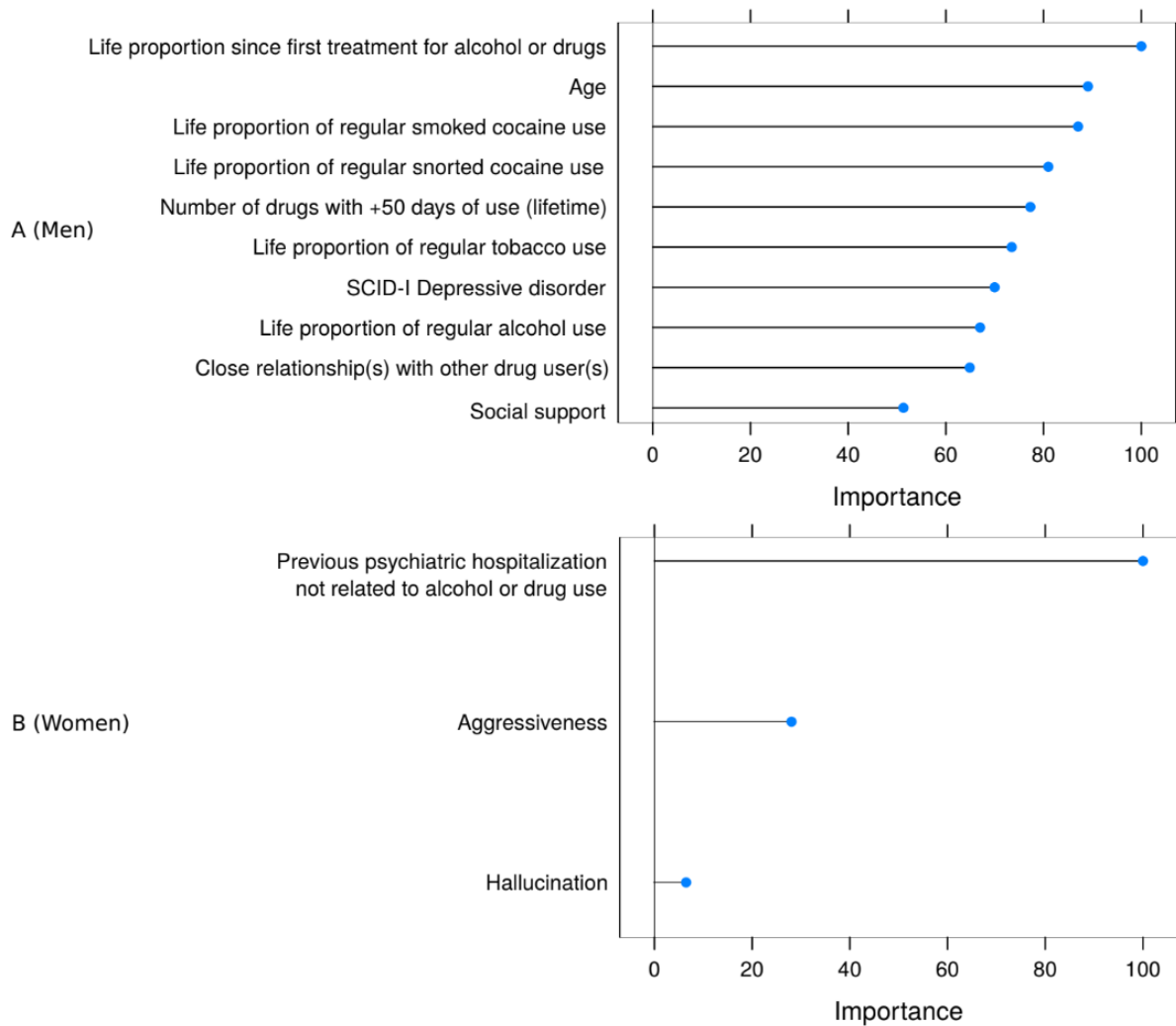
Fig 1 presents the final multiple GLMs, applying backward elimination to the respective 18 independent variables for each gender. Because of some missing data, the men's PRs were calculated with  $n=246$ , where 6 out of 11 variables were statistically significant. For women, with  $n=402$ , 8 out of 13 variables were statistically significant. The main results of this statistical approach include diagnosis of depression ( $PR_M=1.56$ ,  $p=.008$ ;  $PR_W=1.27$ ,  $p=.011$ ) and report of hallucinations throughout life ( $PR_M=1.80$ ,  $p=.006$ ;  $PR_W=1.39$ ,  $p=.006$ ), both of which remained significant factors associated with attempted suicide in men and women. The specific risk factor for each gender can be observed in the figure.



**Fig 1. Prevalence ratios and 95% confidence intervals of lifetime suicide attempt for the variables in the multiple Poisson regression models among men (A; n=246) and women (B; n=402) crack-cocaine inpatient users.** The remaining variables in the last step of the backward elimination are shown. Variables with no indication of reference category (ref.) are binary (*yes/no*), and their reference category is *no*.

### *Predictive approach*

For the machine learning approach, the RFE algorithm reduced the set of training features (predictors) from 57 to 27 for men and to only three for women. Fig 2 presents the ROC curves on the test data of the Random Forest fitted models for (A) men and (B) women. The AUCs were better than random in both cases ( $AUC_M=0.680$  and  $AUC_W=0.734$ ).



**Fig 2. ROC curves for the predictive models of attempted suicide plotted on the test data for each Random Forest learned model - men (A; n=50) and women (B; n=90).** The dot indicates the sensibility and specificity at which the distance between the curve and the transversal line is the largest.

Table 2 presents the performance of the Random Forest models with the unseen test dataset. For women, some of the measures returned the same value, of 0.714, because the confusion matrix was symmetrical.

**Table 2.** Evaluation of the Random Forest Models.

Gender	AUC	Sens.	Spec.	PPV	NPV	Bal. Acc.	NIR	H0: Acc. > NIR p-value
Men	0.680	0.823	0.500	0.467	0.842	0.662	0.653	0.765
Women	0.734	0.714	0.714	0.714	0.714	0.714	0.500	<0.001

AUC = Area under the receiver operating characteristic curve.

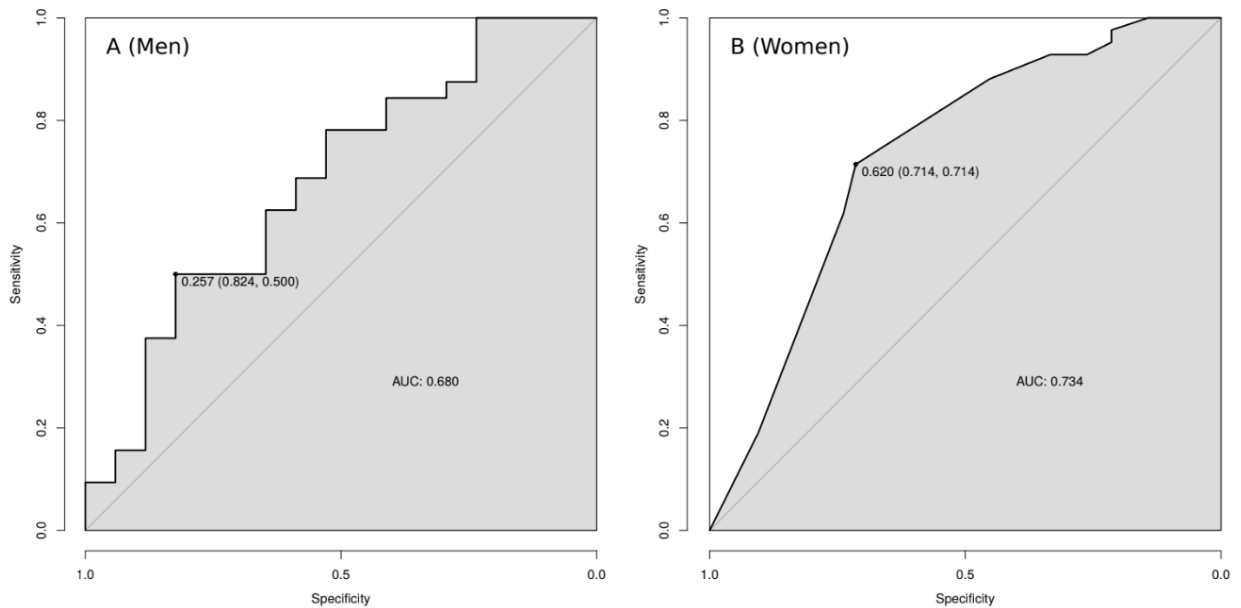
Sens. = Sensitivity. Spec. = Specificity. Bal. Acc. = Balanced Accuracy.

PPV = Positive Predictive Value. NPP = Negative Predictive Value.

NIR = No Information Rate.

According to the statistical test, the accuracy of the women's model was significantly higher than NIR. Additionally, the balance between the evaluation metrics is further evidence toward good model performance. However, the accuracy of the men's model was not significantly higher than NIR, which would imply that it does not predict attempted suicide better than using only the majority outcome category. It is advisable to interpret this result with caution, because the test does not consider all the statistics necessary to fully evaluate the model in order to ultimately conclude that it is not useful. Note that if all test instances were predicted as *no attempted suicide*, according to NIR, then sensitivity would be zero. The metrics for the category of interest should therefore be evaluated (Sens.<sub>M</sub>=0.823 and PPV<sub>M</sub>=0.467).

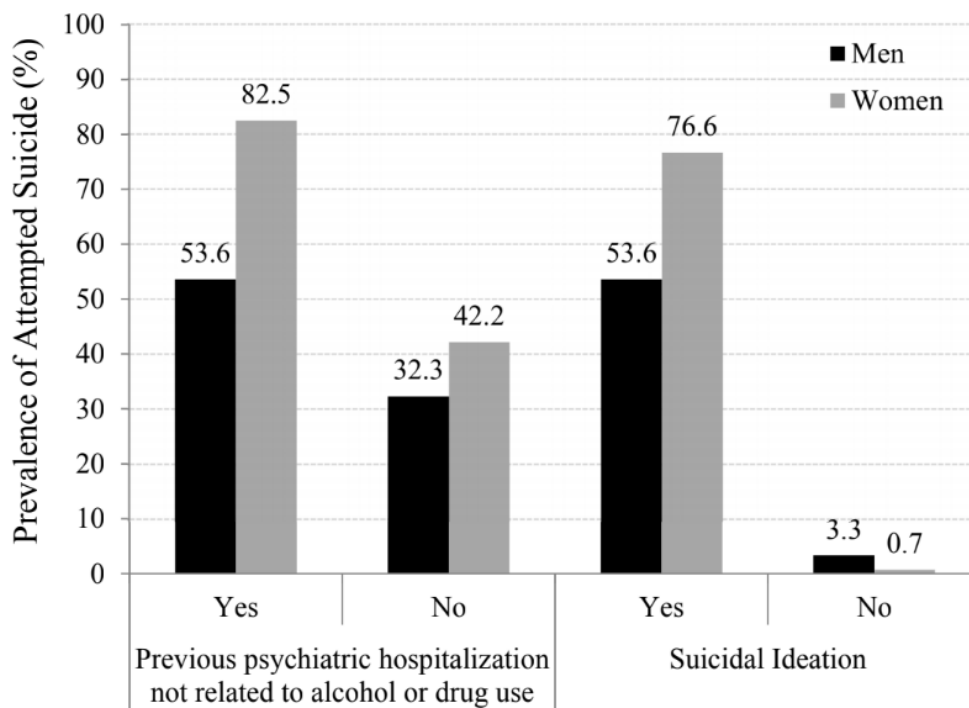
Fig 3 shows the ten most important variables for the men's model and the three variables in the women's model. In this approach, variables concerning an amount of time over lifetime, such as years of regular substance use, were divided by age to create the features of life proportion of regular use, detailed in the supplementary material Text S1.



**Fig 3. Variables with greater importance for the prediction of attempted suicide by the Random Forest algorithm according to gender.** The top ten variables out of the 27 for men (A) and the three variables for women (B) are shown.

We tested the inclusion of suicidal ideation as a predictor, with the result that all predictive statistics improved drastically. To illustrate the difference between the use of suicidal ideation and, for instance, previous psychiatric hospitalization not related to alcohol or drugs, Fig 4 shows the prevalence of attempted suicide between the combined categories of gender and those variables. Overall, the prevalence of attempted suicide was greater for those who presented a previous psychiatric hospitalization (not related to alcohol or drug) than for those who did not, and the difference was greater for women than for men ( $PR_M=1.55$ ,  $p=0.031$ ;  $PR_W=1.91$ ,  $p<0.001$ ). However, the prevalence of attempted suicide for those who presented suicidal ideation was very much higher than for those who did not ( $PR_M=15.8$ ;

PR<sub>w</sub>=112.6, both  $p < 0.001$ ), almost restricting the possible occurrence of attempted suicide to the presence of suicidal ideation. Both scenarios were expected, but in the case of the second, we conclude that it maintains a very proximal relationship with suicide, overlapping important distal factors and overfitting prediction without providing new knowledge. Therefore, it was included in neither approach.



**Fig 4. Prevalence of attempted suicide among men and women within the categories of suicidal ideation and previous psychiatric hospitalization (not related to alcohol or drug use).**

## Discussion

To the best of our knowledge, this is the first study that conducts in-depth analyses of associated factors and prediction of attempted suicide in an inpatient sample of men and women with cocaine use disorder. The main findings of this study provide evidence to support the conclusion that men and women have different factors associated with attempted suicide,

endorsing our decision to stratify by gender, and showing that complementary analytical approaches can provide additional insights and improve hypothesis generation. In women, our findings were more promising, with both analytical approaches showing agreement between risk factors associated with attempted suicide. In men the results were less consistent, where some variables were associated with the outcome in the descriptive model, although the predictive approach showed low accuracy.

Descriptive, explanatory, and predictive models are similar in practice but serve different analytical goals [49]. We presented factors associated with attempted suicide through a descriptive model approach that estimates prevalence ratios, and complementarily, a predictive model approach that explores those factors as predictors in a machine learning algorithm. The use of different approaches can help us to better understand the relationship structure between predictors and outcome. Although GLM can be used to explore such relationships using full factorial models, this strategy often suffers from poor statistical power when the sample size is not planned. As a result, it is very common to investigate linear relationships with main effect coefficients. In contrast, the Random Forest algorithm is a nonparametric method that relies on bootstrapping strategy to explore nonlinear relationships. Its weakness lies in the possibility of random results when applied to small samples, with which GLM often performs well. In this sense, using both approaches add robustness for confirmatory aspect and hypothesis generation. More than that, for some complex clinical questions, such as suicidal behavior, a linear thinking as conceptualized by the risk factors era with descriptive analyses may not solve them. Moreover, the accuracy of all the input data to predict the outcome is as important as the estimation of the impact of a specific predictor.

The descriptive analytical approach indicated that attempted suicide by individuals with cocaine use disorder is associated with some common factors among men and women. Among these factors, we highlight the presence of certain psychiatric symptoms and disorders

(e.g. depression), abuse of psychoactive substances, psychiatric hospitalizations, and childhood maltreatment.

Of these associated factors, childhood trauma appears to play a critical role. A large cohort study showed that childhood trauma increased by 2 to 5 times the risk of suicide attempt across the lifespan [50]. More than that, several studies have demonstrated that childhood trauma is highly associated with suicide attempts in several psychiatric disorders, such as bipolar disorder, depression, schizophrenia [51], and drug addiction [1,52,53]. A recent meta-analysis of longitudinal studies reported higher rates of childhood maltreatment in individuals with SUD compared to the general population [54]. In our study, physical and sexual abuses during childhood were associated with suicide attempt in both genders. An association between physical, sexual and emotional abuse and attempted suicide was also seen in different populations, such as subjects with bipolar disorders [55], prisoners [56], and women in custody [57], and was also related to increased risk of psychiatric disorder across lifespan [58].

Other factors we identified in association with attempted suicide among men and women were major depressive disorders and hallucinations. Mood disorders are commonly accompanied by SUD, leading to worse course of illness and several functional and cognitive impairments [59,60]. Studies suggest that mood disorders and SUD share common genetic predispositions and environmental triggers that may be related to the chronic trajectory and severity of disorders [61]. It is well established that psychopathology is a precipitating risk factor for suicide since 90% of individuals who died from suicide had some type of psychiatric disorder [1]. Furthermore, it is plausible that hallucinations and psychotic experiences may lead individuals with SUD to attempt suicide. A cross-sectional study, for example, demonstrated that auditory verbal hallucinations increased the odds of suicide attempt in adolescents with suicidal ideation regardless of gender [62,63].



By the other hand, there were also different associated factors with attempted suicide among men and women. In men, exclusive factors included bipolar disorder, insufficient income and number of drugs used for more than 50 days over lifetime. Evidence indicates that substance abuse expresses a higher severity of bipolar disorder and drug use behavior may be a mechanism of emotional regulation typical of this disorder [64,65]. Concurrently, other authors indicate that impulsivity or coping skills may mediate drug use [66–68]. Furthermore, our findings are corroborated by other studies which found evidence that subjects in precarious employment and with low incomes had a higher risk of suicide [69,70]. A recent systematic review found that worse economic status and unemployment has been associated with suicidal behavior [71]. Lack of economic resources seems to increase the burden of the preexisting clinical condition and it is common for users of a highly addictive drug to spend a significant part of their income on acquiring that drug [72–74]. This cycle most probably prolongs the burden related to economic vulnerability as long as drug use addictive behavior remains.

Looking at the Random Forest model of male crack-cocaine users, the most important variables are related to drug addiction severity and poor social-family support. In this context, drug severity related factors could act as moderators of the relationship between other variables and attempted suicide, as well as themselves being predictors. This fact also reflects a more complex model given the homogeneous importance assigned to many variables. The most important predictor for men was life proportion since the first treatment for alcohol or drugs, which is consistent with previous studies [75]. In a national survey in the United States, the number of substances had greater importance for predicting suicide attempts than the type of substances used [76].

For women, exclusive variables in the descriptive approach included difficulty in controlling aggression, presence of abstinence symptoms, and regular years of snorted cocaine

use. Interestingly, a study with a community sample of adults revealed that high levels of withdrawal symptoms increased the risk of suicide in women, but not in men [77]. Alcohol-withdrawal symptoms could be considered as a stressor factor since drug withdrawal leads to emotional and physical aversive states. These long-term effects might work as unpredictable and uncontrollable stressors, leading to a process of stress sensitization and enhancement of the endogenous aversive state due to drug withdrawal [78,79]. Thus, these subjects may not have good coping skills for facing the severity of withdrawal symptoms, leading to suicide attempts [77,80].

Our findings also associated the number of years of regular smoked-cocaine use and previous psychiatric hospitalization not related to alcohol or drug use with suicide in women. Both variables are understood as hallmarks of illness chronicity in psychiatric conditions [81]. In the neuroprogression model, the number of psychiatric episodes, number of hospitalizations, substance use, and childhood trauma contribute to impaired response to treatment and cognitive and psychosocial functioning and to higher rates of suicide attempt [82,83]. We also found that aggression was significantly associated with attempted suicide in women. This factor could have an overlap with suicidal behavior, interpersonal aggression, and self-harm as a spectral event, in which some subjects may engage in one of these different types of violent behavior [84,85]. Interestingly, it is suggested that aggression and impulsivity are diathesis factors in suicidal behavior [86,87].

Moreover, the women's predictive model was much easier to interpret, had higher predictive power, and was closer to a parsimonious model, which involves the minimum possible number of parameters to be estimated while thoroughly explaining the behavior of the outcome. Here, previous psychiatric hospitalization not related to alcohol or drugs, aggression, and hallucinations were the most important variables for predicting attempted

suicide, drawing attention to other psychiatric events in this subpopulation and corroborating descriptive results for this group.

Regarding treatment, both suicidal behavior and addiction involve several problematic aspects in the individual's life, one being aggravated by the other, so intensive psychotherapy is highly recommended in order to rearrange the lives of these people as a whole. As our sample has shown, this population has a high prevalence of attempted suicide and is already at risk of suicide, so approaches focused on suicide prevention are needed early on. A randomized controlled trial of individuals who attempted suicide showed that cognitive therapy was effective in preventing suicide reattempt and depression [88]. However, there is no focus on gender-related psychiatric differences in several studies about suicide despite having gender distinctness. Hence, it urges to develop tailored strategies focused on suicide risk factors according to gender as implementing specific mental health policies among men and women, eliminating health care access barriers, and restricting access to lethal means [89]. Following our specific findings on gender differences, psychosocial treatments should be targeted coping strategies and social skills due to the less social support common in men. While in women, it should be tailored on preventing psychiatric comorbidities and encouraging anger management. In relation to pharmacological treatment, lithium was effective to prevent suicide in subjects with mood disorder as shown in a meta-analysis of randomized clinical trials [90]. But for its administration in cocaine addicts, further studies with this population are needed to support this treatment alternative.

### **Limitations**

Data were based on retrospective self-report of attempted suicide, and subjects may underreport data and have biased recall, especially in a SUD sample. In the predictive approach, the model for men did not achieve satisfactory overall classification accuracy. This

result could be caused by many factors, such as a weak explanation of attempted suicide by the set of variables used or, and more plausible, due to the small sample size in terms of the needs for effective fitting models based on supervised machine learning. These circumstances reflect insufficient information for high precision. However, this is nevertheless one of the largest clinical samples of cocaine use disorder inpatients. Also, suicide attempts are more prevalent in women than men, which could be explained by the fact that men tend to use more lethal methods, leading to greater suicide mortality than women [68]. The prevalence difference translates into a class imbalance problem [91], which, although it was addressed in model training, still leaves the test data unbalanced and the overall classification accuracy less meaningful for evaluating the model than class-specific measures, such as sensitivity and PPV. We suggest that further studies should focus on predictive models of both suicide attempt and death in order to achieve more robust findings. Furthermore, the study design does not allow for causal conclusions between predictors and suicide but explores the relationship between the variables. These limitations are due to the use of secondary databases, because the factors associated with attempted suicide captured by the predictive models were limited to the variables available in the data collection instruments. A measure of impulsivity, for instance, is a highly desirable variable for predictive suicide studies. It could help to explain suicide attempts that occurred without suicidal ideation. Factors that aggravate vulnerabilities and stress, such as quality of life and sexual orientation, among many others, would also be desirable.

## **Conclusions**

Our findings indicate that attempted suicide is associated with depression and hallucinations in both genders, as well as previous hospitalization for mental issues - related to psychiatry for women and indirectly related to substance abuse for men. This scenario

suggests that the severity of drug use may be a moderator between predictors and suicide among men. Our study also corroborates the influence of childhood trauma on suicide attempts, with different types of trauma being related to risk factors for both men and women. It is noteworthy that both of our analytical approaches showed excellent consistency with respect to the variables associated with suicide among women.

We hypothesize that highly accurate predictive models would support important clinical decisions such as suicide prediction, selection of treatment options, and prognosis orientations. The present study advances the field of machine learning in mental disorders and presents a model for prediction of attempted suicide in patients with cocaine use disorder. Future longitudinal studies should replicate these findings using larger samples from multiple centers and pursue a data fusion approach, combining clinical data with other biological measures, such as genetics, to build more accurate tools to predict suicide attempts in subjects with cocaine use disorder.

### **Acknowledgments**

We would like to thank the data collection teams and their supervisors. We are also grateful to the hospitals' care staff and especially to the participants, who, although undergoing treatment, agreed to cooperate.

### **References**

1. Turecki G, Brent DA. Suicide and suicidal behaviour. *Lancet*. 2016;387: 1227–1239. doi:10.1016/S0140-6736(15)00234-2
2. World Health Organization. Prevention of suicidal behaviours: a task for all.
3. Bachmann S. Epidemiology of Suicide and the Psychiatric Perspective. *Int J Environ Res Public Health*. 2018;15: 1425. doi:10.3390/ijerph15071425

4. Mościcki EK. Epidemiology of Suicidal Behavior. *Suicide Life-Threatening Behav.* 1995;25: 22–35. doi:10.1111/J.1943-278X.1995.TB00390.X
5. Arsenault-Lapierre G, Kim C, Turecki G. Psychiatric diagnoses in 3275 suicides: a meta-analysis. *BMC Psychiatry.* 2004;4: 37. doi:10.1186/1471-244X-4-37
6. Runeson B, Tidemalm D, Dahlin M, Lichtenstein P, Langstrom N. Method of attempted suicide as predictor of subsequent successful suicide: national long term cohort study. *BMJ.* 2010;341: c3222–c3222. doi:10.1136/bmj.c3222
7. Hoertel N, Franco S, Wall MM, Oquendo MA, Kerridge BT, Limosin F, et al. Mental disorders and risk of suicide attempt: a national prospective study. *Mol Psychiatry.* 2015;20: 718–726. doi:10.1038/mp.2015.19
8. Bilsen J. Suicide and Youth: Risk Factors. *Front Psychiatry.* 2018;9. doi:10.3389/fpsy.2018.00540
9. Borges G, Walters EE, Kessler RC. Associations of Substance Use, Abuse, and Dependence with Subsequent Suicidal Behavior. *Am J Epidemiol.* 2000;151: 781–789. doi:10.1093/oxfordjournals.aje.a010278
10. Bastos FI, Bertoni N. PESQUISA NACIONAL SOBRE O USO DE CRACK. Quem são os usuários de crack e/ou similares do Brasil? Quantos são nas capitais brasileiras? ICICT/FIOCRUZ, editor. Rio de Janeiro; 2014. Available: <https://www.arca.fiocruz.br/handle/icict/10019>
11. Bastos FIPM. 3rd National survey on drug use by the brazilian population. ICICT/FIOCRUZ, editor. Rio de Janeiro; 2017. Available: <https://www.arca.fiocruz.br/handle/icict/34614>
12. Laranjeira R, Madruga CS, Pinsky I, Caetano R, Ribeiro M, Mitsuhiro S. II Levantamento Nacional de Álcool e Drogas-Consumo de Álcool no Brasil: Tendências entre 2006/2012. São Paulo Inpad. 2013.

13. Abdalla RR, Miguel AC, Brietzke E, Caetano R, Laranjeira R, Madruga CS. Suicidal behavior among substance users: data from the Second Brazilian National Alcohol and Drug Survey (II BNADS). *Brazilian J Psychiatry*. 2019;41: 437–440. doi:10.1590/1516-4446-2018-0054
14. Paim Kessler FH, Barbosa Terra M, Faller S, Ravy Stolf A, Carolina Peuker A, Benzano D, et al. Crack Users Show High Rates of Antisocial Personality Disorder, Engagement in Illegal Activities and Other Psychosocial Problems. *Am J Addict*. 2012;21: 370–380. doi:10.1111/j.1521-0391.2012.00245.x
15. Gradus JL, Rosellini AJ, Horváth-Puhó E, Street AE, Galatzer-Levy I, Jiang T, et al. Prediction of Sex-Specific Suicide Risk Using Machine Learning and Single-Payer Health Care Registry Data From Denmark. *JAMA Psychiatry*. 2019. doi:10.1001/jamapsychiatry.2019.2905
16. Paim Kessler FH, Barbosa Terra M, Faller S, Ravy Stolf A, Carolina Peuker A, Benzano D, et al. Crack Users Show High Rates of Antisocial Personality Disorder, Engagement in Illegal Activities and Other Psychosocial Problems. *Am J Addict*. 2012;21: 370–380. doi:10.1111/j.1521-0391.2012.00245.x
17. Narvaez JCM, Jansen K, Pinheiro RT, Kapczinski F, Silva RA, Pechansky F, et al. Psychiatric and substance-use comorbidities associated with lifetime crack cocaine use in young adults in the general population. *Compr Psychiatry*. 2014;55: 1369–1376. doi:10.1016/j.comppsy.2014.04.021
18. Silva DC da, Ávila AC de, Yates MB, Cazassa MJ, Dias FB, Souza MH de, et al. Sintomas psiquiátricos e características sociodemográficas associados à tentativa de suicídio de usuários de cocaína e crack em tratamento. *J Bras Psiquiatr*. 2017;66: 89–95. doi:10.1590/0047-2085000000155
19. Halpern SC, Scherer JN, Roglio V, Faller S, Sordi A, Ornell F, et al. Vulnerabilidades

- clínicas e sociais em usuários de crack de acordo com a situação de moradia: um estudo multicêntrico de seis capitais brasileiras. *Cad Saude Publica*. 2017;33. doi:10.1590/0102-311x00037517
20. Duailibi LB, Ribeiro M, Laranjeira R. Profile of cocaine and crack users in Brazil. *Cad Saude Publica*. 2008;24: s545–s557. doi:10.1590/S0102-311X2008001600007
  21. Siliquini R, Morra A, Versino E, Renga G. Recreational drug consumers: who seeks treatment? *Eur J Public Health*. 2005;15: 580–586. doi:10.1093/eurpub/cki031
  22. Lejuez CW, Bornovalova MA, Reynolds EK, Daughters SB, Curtin JJ. Risk factors in the relationship between gender and crack/cocaine. *Exp Clin Psychopharmacol*. 2007;15: 165–175. doi:10.1037/1064-1297.15.2.165
  23. Pope SK, Falck RS, Carlson RG, Leukefeld C, Booth BM. Characteristics of Rural Crack and Powder Cocaine Use: Gender and Other Correlates. *Am J Drug Alcohol Abuse*. 2011;37: 491–496. doi:10.3109/00952990.2011.600380
  24. Freeman A, Mergl R, Kohls E, Székely A, Gusmao R, Arensman E, et al. A cross-national study on gender differences in suicide intent. *BMC Psychiatry*. 2017;17: 234. doi:10.1186/s12888-017-1398-8
  25. Ibrahim N, Amit N, Che Din N, Ong HC. Gender differences and psychological factors associated with suicidal ideation among youth in Malaysia. *Psychol Res Behav Manag*. 2017;Volume 10: 129–135. doi:10.2147/PRBM.S125176
  26. 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;64: 265–283. doi:10.1007/s00038-018-1196-1
  27. Stefanello S, Cais CF da S, Mauro MLF, Freitas GVS de, Botega NJ. Gender differences in suicide attempts: preliminary results of the multisite intervention study



- on suicidal behavior (SUPRE-MISS) from Campinas, Brazil. *Rev Bras Psiquiatr.* 2007;30: 139–143. doi:10.1590/S1516-44462006005000063
28. 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;64: 265–283. doi:10.1007/s00038-018-1196-1
29. Karila L, Petit A, Lowenstein W, Reynaud M. Diagnosis and Consequences of Cocaine Addiction. *Curr Med Chem.* 2012;19: 5612–5618. doi:10.2174/092986712803988839
30. Acion L, Kelmansky D, van der Laan M, Sahker E, Jones D, Arndt S, et al. Use of a machine learning framework to predict substance use disorder treatment success. *PLoS One.* 2017;12: e0175383. doi:10.1371/journal.pone.0175383
31. Fayyad U, Piatetsky-Shapiro G, Smyth P. The KDD process for extracting useful knowledge from volumes of data. *Commun ACM.* 1996;39: 27–34. doi:10.1145/240455.240464
32. McLELLAN AT, LUBORSKY L, WOODY GE, O'BRIEN CP. An Improved Diagnostic Evaluation Instrument for Substance Abuse Patients. *J Nerv Ment Dis.* 1980;168: 26–33. doi:10.1097/00005053-198001000-00006
33. Kessler F, Faller S, Souza-Formigoni MLO, Cruz MS, Brasiliano S, Stolf AR, et al. Avaliação multidimensional do usuário de drogas e a Escala de Gravidade de Dependência. *Rev Psiquiatr do Rio Gd do Sul.* 2010;32: 48–56. doi:10.1590/S0101-81082010000200005
34. Del-Ben CM, Vilela JAA, Crippa JA de S, Hallak JEC, Labate CM, Zuardi AW. Confiabilidade da “Entrevista Clínica Estruturada para o DSM-IV - Versão Clínica” traduzida para o português. *Rev Bras Psiquiatr.* 2001;23: 156–159. doi:10.1590/S1516-44462001000300008

35. Bernstein DP, Fink L, Handeisman L, Foote J, Lovejoy M, Wenzel K, et al. Initial reliability and validity of a new retrospective measure of child abuse and neglect. *Am J Psychiatry*. 1994;151: 1132–1136. doi:10.1176/ajp.151.8.1132
36. Grassi-Oliveira R, Stein LM, Pezzi JC. Tradução e validação de conteúdo da versão em português do Childhood Trauma Questionnaire. *Rev Saude Publica*. 2006;40: 249–255. doi:10.1590/S0034-89102006000200010
37. Sanvicente-Vieira B, Rovaris DL, Ornell F, Sordi A, Rothmann LM, Niederauer JPO, et al. Sex-based differences in multidimensional clinical assessments of early-abstinence crack cocaine users. Santana GL, editor. *PLoS One*. 2019;14: e0218334. doi:10.1371/journal.pone.0218334
38. Scherer JN, Silvello D, Volpato VL, Roglio VS, Fara L, Ornell F, et al. Predictive factors associated with driving under the influence among Brazilian drug-using drivers. *Accid Anal Prev*. 2019;123: 256–262. doi:10.1016/j.aap.2018.12.003
39. Yuodelis-Flores C, Ries RK. Addiction and suicide: A review. *Am J Addict*. 2015;24: 98–104. doi:10.1111/ajad.12185
40. Franklin JC, Ribeiro JD, Fox KR, Bentley KH, Kleiman EM, 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. doi:10.1037/bul0000084
41. Barros AJ, Hirakata VN. Alternatives for logistic regression in cross-sectional studies: an empirical comparison of models that directly estimate the prevalence ratio. *BMC Med Res Methodol*. 2003;3: 21. doi:10.1186/1471-2288-3-21
42. Buuren S van, Groothuis-Oudshoorn K. mice : Multivariate Imputation by Chained Equations in R. *J Stat Softw*. 2011;45. doi:10.18637/jss.v045.i03
43. Chandrashekar G, Sahin F. A survey on feature selection methods. *Comput Electr Eng*. 2014;40: 16–28. doi:10.1016/j.compeleceng.2013.11.024

44. Breiman L. Random Forests. *Mach Learn.* 2001;45: 5–32. doi:10.1023/A:1010933404324
45. Kim J-H. Estimating classification error rate: Repeated cross-validation, repeated hold-out and bootstrap. *Comput Stat Data Anal.* 2009;53: 3735–3745. doi:10.1016/j.csda.2009.04.009
46. Jin Huang, Ling CX. Using AUC and accuracy in evaluating learning algorithms. *IEEE Trans Knowl Data Eng.* 2005;17: 299–310. doi:10.1109/TKDE.2005.50
47. Chawla N V., Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic Minority Over-sampling Technique. *J Artif Intell Res.* 2002;16: 321–357. doi:10.1613/jair.953
48. Mergl R, Koburger N, Heinrichs K, Székely A, Tóth MD, Coyne J, et al. What Are Reasons for the Large Gender Differences in the Lethality of Suicidal Acts? An Epidemiological Analysis in Four European Countries. Niederkrotenthaler T, editor. *PLoS One.* 2015;10: e0129062. doi:10.1371/journal.pone.0129062
49. Shmueli G. To Explain or to Predict? *Stat Sci.* 2010;25: 289–310. doi:10.1214/10-STS330
50. Dube SR, Anda RF, Felitti VJ, Chapman DP, Williamson DF, Giles WH. Childhood Abuse, Household Dysfunction, and the Risk of Attempted Suicide Throughout the Life Span. *JAMA.* 2001;286: 3089. doi:10.1001/jama.286.24.3089
51. Mohammadzadeh A, Azadi S, King S, Khosravani V, Sharifi Bastan F. Childhood trauma and the likelihood of increased suicidal risk in schizophrenia. *Psychiatry Res.* 2019;275: 100–107. doi:10.1016/j.psychres.2019.03.023
52. Garami J, Valikhani A, Parkes D, Haber P, Mahlberg J, Misiak B, et al. Examining Perceived Stress, Childhood Trauma and Interpersonal Trauma in Individuals With Drug Addiction. *Psychol Rep.* 2019;122: 433–450. doi:10.1177/0033294118764918
53. Norman RE, Byambaa M, De R, Butchart A, Scott J, Vos T. The Long-Term Health

- Consequences of Child Physical Abuse, Emotional Abuse, and Neglect: A Systematic Review and Meta-Analysis. Tomlinson M, editor. *PLoS Med.* 2012;9: e1001349. doi:10.1371/journal.pmed.1001349
54. Halpern SC, Schuch FB, Scherer JN, Sordi AO, Pachado M, Dalbosco C, et al. Child Maltreatment and Illicit Substance Abuse: A Systematic Review and Meta-Analysis of Longitudinal Studies. *Child Abus Rev.* 2018;27: 344–360. doi:10.1002/car.2534
55. Adigüzel V, Özdemir N, Şahin ŞK. Childhood traumas in euthymic bipolar disorder patients in Eastern Turkey and its relations with suicide risk and aggression. *Nord J Psychiatry.* 2019;73: 490–496. doi:10.1080/08039488.2019.1655589
56. Angelakis I, Austin JL, Gooding P. Childhood maltreatment and suicide attempts in prisoners: a systematic meta-analytic review. *Psychol Med.* 2019; 1–10. doi:10.1017/S0033291719002848
57. DeCou CR, Lynch SM, DeHart DD, Belknap J. Evaluating the association between childhood sexual abuse and attempted suicide across the lifespan: Findings from a nationwide study of women in jail. *Psychol Serv.* 2016;13: 254–260. doi:10.1037/ser0000096
58. Chen LP, Murad MH, Paras ML, Colbenson KM, Sattler AL, Goranson EN, et al. Sexual Abuse and Lifetime Diagnosis of Psychiatric Disorders: Systematic Review and Meta-analysis. *Mayo Clin Proc.* 2010;85: 618–629. doi:10.4065/mcp.2009.0583
59. Cardoso T de A, Mondin TC, Souza LD de M, da Silva RA, Magalhães PVS, Kapczinski F, et al. Functioning in bipolar disorder with substance abuse/dependence in a community sample of young adults. *J Affect Disord.* 2015;187: 179–182. doi:10.1016/j.jad.2015.08.046
60. Daigre C, Grau-López L, Rodríguez-Cintas L, Ros-Cucurull E, Sorribes-Puertas M, Esculies O, et al. The role of dual diagnosis in health-related quality of life among

- treatment-seeking patients in Spain. *Qual Life Res.* 2017;26: 3201–3209. doi:10.1007/s11136-017-1668-4
61. Post RM, Kalivas P. Bipolar disorder and substance misuse: pathological and therapeutic implications of their comorbidity and cross-sensitisation. *Br J Psychiatry.* 2013;202: 172–176. doi:10.1192/bjp.bp.112.116855
  62. Fujita J, Takahashi Y, Nishida A, Okumura Y, Ando S, Kawano M, et al. Auditory verbal hallucinations increase the risk for suicide attempts in adolescents with suicidal ideation. *Schizophr Res.* 2015;168: 209–212. doi:10.1016/j.schres.2015.07.028
  63. Ishii T, Hashimoto E, Ukai W, Kakutani Y, Sasaki R, Saito T. Characteristics of Attempted Suicide by Patients with Schizophrenia Compared with Those with Mood Disorders: A Case-Controlled Study in Northern Japan. Clelland JD, editor. *PLoS One.* 2014;9: e96272. doi:10.1371/journal.pone.0096272
  64. Arias F, Szerman N, Vega P, Mesías B, Basurte I, Rentero D. Trastorno bipolar y trastorno por uso de sustancias. Estudio Madrid sobre prevalencia de patología dual. *Adicciones.* 2016;29: 186. doi:10.20882/adicciones.782
  65. Wilens TE, Martelon M, Anderson JP, Shelley-Abrahamson R, Biederman J. Difficulties in emotional regulation and substance use disorders: A controlled family study of bipolar adolescents. *Drug Alcohol Depend.* 2013;132: 114–121. doi:10.1016/j.drugalcdep.2013.01.015
  66. Maser JD, Akiskal HS, Schettler P, Scheftner W, Mueller T, Endicott J, et al. Can Temperament Identify Affectively Ill Patients Who Engage in Lethal or Near-Lethal Suicidal Behavior? A 14-Year Prospective Study. *Suicide Life-Threatening Behav.* 2002;32: 10–32. doi:10.1521/suli.32.1.10.22183
  67. Watkins HB, Meyer TD. Is there an empirical link between impulsivity and suicidality in bipolar disorders? A review of the current literature and the potential psychological

- implications of the relationship. *Bipolar Disord.* 2013;15: 542–558.  
doi:10.1111/bdi.12090
68. Beyer JL, Weisler RH. Suicide Behaviors in Bipolar Disorder. *Psychiatr Clin North Am.* 2016;39: 111–123. doi:10.1016/j.psc.2015.09.002
69. Mäki N, Martikainen P. A register-based study on excess suicide mortality among unemployed men and women during different levels of unemployment in Finland. *J Epidemiol Community Health.* 2012;66: 302 LP – 307. doi:10.1136/jech.2009.105908
70. Min KB, Park SG, Hwang SH, Min JY. Precarious employment and the risk of suicidal ideation and suicide attempts. *Prev Med (Baltim).* 2015;71: 72–76. doi:10.1016/j.ypmed.2014.12.017
71. Iemmi V, Bantjes J, Coast E, Channer K, Leone T, McDaid D, et al. Suicide and poverty in low-income and middle-income countries: a systematic review. *The Lancet Psychiatry Elsevier*; Aug 1, 2016 pp. 774–783. doi:10.1016/S2215-0366(16)30066-9
72. Greenwald MK, Steinmiller CL. Cocaine behavioral economics: from the naturalistic environment to the controlled laboratory setting. *Drug Alcohol Depend.* 2014;141: 27–33. doi:10.1016/j.drugalcdep.2014.04.028
73. North SC, Pollio ED. Financing Cocaine Use in a Homeless Population. *Behavioral Sciences* . 2017. doi:10.3390/bs7040074
74. Yang JC, Roman-Urrestarazu A, Brayne C. Binge alcohol and substance use across birth cohorts and the global financial crisis in the United States. *PLoS One.* 2018;13: e0199741. Available: <https://doi.org/10.1371/journal.pone.0199741>
75. Ilgen MA, Harris AHS, Moos RH, Tiet QQ. Predictors of a Suicide Attempt One Year After Entry Into Substance Use Disorder Treatment. *Alcohol Clin Exp Res.* 2007;31: 635–642. doi:10.1111/j.1530-0277.2007.00348.x
76. Borges G, Walters EE, Kessler RC. Associations of Substance Use, Abuse, and

- Dependence with Subsequent Suicidal Behavior. *Am J Epidemiol.* 2000;151: 781–789. doi:10.1093/oxfordjournals.aje.a010278
77. Campos RC, Holden RR, Costa F, Oliveira AR, Abreu M, Fresca N. The moderating effect of gender on the relationship between coping and suicide risk in a Portuguese community sample of adults. *J Ment Health.* 2017;26: 66–73. doi:10.1080/09638237.2016.1222066
78. Harkness KL, Hayden EP, Lopez-Duran NL. Stress sensitivity and stress sensitization in psychopathology: an introduction to the special section. *J Abnorm Psychol.* 2015;124: 1–3. doi:10.1037/abn0000041
79. Chartoff EH, Carlezon WAJ. Drug withdrawal conceptualized as a stressor. *Behav Pharmacol.* 2014;25: 473–492. doi:10.1097/FBP.0000000000000080
80. Brådvik L, Berglund M. A Suicide Peak after Weekends and Holidays in Patients with Alcohol Dependence. *Suicide Life-Threatening Behav.* 2003;33: 186–191. doi:10.1521/suli.33.2.186.22773
81. Forty L, Ulanova A, Jones L, Jones I, Gordon-Smith K, Fraser C, et al. Comorbid medical illness in bipolar disorder. *Br J Psychiatry.* 2018/01/02. 2014;205: 465–472. doi:DOI: 10.1192/bjp.bp.114.152249
82. Fusar-Poli P, Díaz-Caneja CM, Patel R, Valmaggia L, Byrne M, Garety P, et al. Services for people at high risk improve outcomes in patients with first episode psychosis. *Acta Psychiatr Scand.* 2016;133: 76–85. doi:10.1111/acps.12480
83. Hauser M, Galling B, Correll CU. Suicidal ideation and suicide attempts in children and adolescents with bipolar disorder: a systematic review of prevalence and incidence rates, correlates, and targeted interventions. *Bipolar Disord.* 2013;15: 507–523. doi:10.1111/bdi.12094
84. Harford TC, Yi H, Grant BF. Associations between childhood abuse and interpersonal

- aggression and suicide attempt among U.S. adults in a national study. *Child Abuse Negl.* 2014;38: 1389–1398. doi:10.1016/j.chiabu.2014.02.011
85. McMahon K, Hoertel N, Olfson M, Wall M, Wang S, Blanco C. Childhood maltreatment and impulsivity as predictors of interpersonal violence, self-injury and suicide attempts: A national study. *Psychiatry Res.* 2018;269: 386–393. doi:10.1016/j.psychres.2018.08.059
86. Tsujii N, Mikawa W, Tsujimoto E, Adachi T, Niwa A, Ono H, et al. Reduced left precentral regional responses in patients with major depressive disorder and history of suicide attempts. *PLoS One.* 2017;12: e0175249. Available: <https://doi.org/10.1371/journal.pone.0175249>
87. Mann JJ. Neurobiology of suicidal behaviour. *Nat Rev Neurosci.* 2003;4: 819–828. doi:10.1038/nrn1220
88. Brown GK, Ten Have T, Henriques GR, Xie SX, Hollander JE, Beck AT. Cognitive Therapy for the Prevention of Suicide Attempts. *JAMA.* 2005;294: 563. doi:10.1001/jama.294.5.563
89. Cibis A, Mergl R, Bramesfeld A, Althaus D, Niklewski G, Schmidtke A, et al. Preference of lethal methods is not the only cause for higher suicide rates in males. *J Affect Disord.* 2012;136: 9–16. doi:10.1016/j.jad.2011.08.032
90. Cipriani A, Pretty H, Hawton K, Geddes JR. Lithium in the Prevention of Suicidal Behavior and All-Cause Mortality in Patients With Mood Disorders: A Systematic Review of Randomized Trials. *Am J Psychiatry.* 2005;162: 1805–1819. doi:10.1176/appi.ajp.162.10.1805
91. Japkowicz N. The Class Imbalance Problem: Significance and Strategies. Proceedings of the 2000 International Conference on Artificial Intelligence (IC-AI'2000): Special Track on Inductive Learning. Las Vegas, Nevada; 2000.



#### 4. MATERIAL COMPLEMENTAR DO ARTIGO 1

**Text S1.** Detailed information on the methods for the machine learning approach.

##### **Methods**

###### *Participants and assessments*

Data were collected using three instruments: the sixth version of the Addiction Severity Index (ASI-6); the Structured Clinical Interview for DSM-IV axis I disorders (SCID-I); and the Childhood Trauma Questionnaire (CTQ).

ASI-6 was developed by Thomas McLellan [25] and validated in Brazil by Kessler et al. [26]. It is a structured interview lasting approximately 45 minutes and addresses the impact of substance use in the last 30 days, last six months, and over lifetime. The instrument comprises 243 questions regarding seven working areas: Medical, Employment and Support, Alcohol use, Drug use, Legal aspects, and Socio-familiar and Psychiatric aspects. The instrument also contains a section of sociodemographic information.

The Clinical Interview for Axis I Disorders of the Diagnostic and Statistical Manual of Mental Disorders IV (SCID-I) was adapted for Brazil by Del-Ben et al. [27]. This instrument consists of a semi-structured interview lasting approximately 60 minutes and provides presence or absence of diagnoses of current and past psychiatric disorders of Axis I according to criteria of the fourth version of the Diagnostic and Statistical Manual of Mental Disorders. It is divided into six modules: mood episodes, psychotic episodes, psychotic disorders, mood disorders, substance use disorders, anxiety, and other disorders.

The Childhood Trauma Questionnaire (CTQ) was developed by Bernstein et al. [28] and validated in Brazil by Grassi-Oliveira et al. [29]. This instrument can be self-applied in 8 minutes and graduates the frequency of 28 assertions related to childhood situations on a five-point Likert scale, investigating five traumatic components: childhood physical abuse (PA), emotional abuse (EA), sexual abuse (SA), physical neglect (FN), and emotional neglect (EN).

### ***Preprocessing and data transformation***

The new variables created based on changes and combinations of the original ones are:

1. Number of chronic diseases. ASI6's second section consists of medical issues. Questions M3-M14 explore by self-report high pressure (M3), diabetes (M4), heart disease (M5), stroke or ischemia (M6), epilepsy or convulsions (M7), cancer (M8), HIV or AIDS (M9), tuberculosis (M10), hepatitis (M11), cirrhosis or other chronic liver disease (M12), chronic kidney disease (M13) and chronic respiratory problem (M14) with the question "Have you ever been told by a doctor or healthcare provider that you had any of the following physical or medical conditions"? This variable was created using the sum of those, except for M10, M11, and M14, that were used individually; and also M9, that was replaced by anti-HIV test. So, this is a discrete variable ranging from 0 to 8.
2. Life proportion since the first treatment for alcohol or drugs. In the ASI6's section about Alcohol and Drugs, D03 question asks, "How old were you the first time you entered alcohol or drug abuse treatment"? To create this variable, we divided this answer by age, so it is a continuous variable ranging from 0 to 1.
3. Life proportion of regular use of alcohol, marijuana, snorted and smoked cocaine, and tobacco. Here, we have five variables, calculated similarly to the previous ones. Questions D08 and D09 ask, "How many years in your life have you drunk alcohol on

a regular basis, 3+ days per week?” and “How many years in your life have you drunk at least (5-men, 4-women) drinks per day on a regular basis, 3+ days per week?”, respectively. We used the maximum of those two divided by age to create the life proportion of regular use of alcohol. Variables D25B, D27B, and D28B ask the same as D08 but related to marijuana, snorted cocaine, and smoked cocaine instead of alcohol, respectively. D57 refers to years of daily use of tobacco. All of those divided by age create the life proportion of regular use of the respective substance, also continuous variables ranging from 0 to 1.

4. Number of psychoactive substances used for more than 50 days (lifetime). Also in the Alcohol and Drugs section of ASI6, the questions D25C through D33C ask “Have you used [substance] on 50 or more days in your life?” for the following substances: marijuana (D25C), sedatives (D26C), snorted cocaine (D27C), smoked cocaine (D28C), stimulants (D29C), hallucinogens (D30C), heroin (D31C), other opioids (D32C) and inhalants (D33C). The question D10 asks, “Have you drunk at least (5-men, 4-women) drinks in a day in 50 or more days in your life”? Finally, we dichotomize D57 (0 = no tobacco; 1 through higher = yes) and sum all of those in this new variable, creating a discrete variable ranging from 1 to 11.
5. Social support network. Another section of ASI6 has questions about Family and Social. Questions F9A, F9B, and F9C ask, “If you need help, can you count on: (A/B/C)?”, where person A is partner, B is adult relatives and C is close friends (one binary answer for each). Also, question F12 asks, “Aside from your partner, other adult relatives and close friends, are there any people you keep in touch with that you can count on if you really need help”? The sum of these four answers creates this variable, which is a discrete variable ranging from 0 to 4.

6. Close relationship(s) with other drug user(s). Similarly to the previous variable, the questions F8A, F8B, and F8C ask, “Do/does your A/B/C have a current problem with alcohol or use drugs?”, where person A is partner, B is adult relatives, and C is close friends (one binary answer for each). The sum of these three answers creates this variable, which is a discrete variable ranging from 0 to 3.
7. Difficulty in controlling aggressiveness. ASI6’s P13A and P14A questions ask, “Have you ever (since 18) had difficulty controlling your temper, or urges to hit or harm someone?” and “Have you ever (since 18) pushed, hit, thrown things at, or used weapons against someone?”, respectively. This is a binary variable (0=no; 1=yes) created by the answer “yes” for at least one of the questions.
8. Dichotomized CTQ scores. Each trauma is composed of a sum of 5 Likert statements about the respective trauma construct. There is a specific threshold for each trauma being considered present, as follows: PN>9, AP>9, EN>14, EA>12, and SA>7.
9. Physical abuse in childhood. According to CTQ’s application instructions, the questions must be answered concerning life before 16 years to be considered in childhood. Following this threshold, we combined ASI6’s questions F23 and F29, which ask “Have you ever been physically assaulted/abused by someone you knew?” and “Have you ever been the victim of a violent crime like being mugged, assaulted?” with their subsequent questions, F24 and F30. They ask, “How old were you when this first happened”? We merge these data considering under 16 years old with CTQ’s dichotomized physical abuse to create this new binary variable (0=no; 1=yes).
10. Sexual abuse in childhood. This variable was created similarly to the previous one, combining CTQ’s dichotomized sexual abuse variable and ASI6’s F26 and F27 questions, which ask “Have you ever been sexually assaulted/abused by someone you

knew?” and age of the first occurrence, respectively. This combination results in a binary variable.

11. Physical abuse in a lifetime. This variable is binary and was created by the answers “yes” for at least one of ASI6’s F24 and F29 questions or CTQ’s dichotomized childhood physical abuse.
12. Sexual abuse in a lifetime. Similarly to the previous, this variable uses only ASI’s F26 question together with CTQ’s dichotomized childhood sexual abuse.
13. Idealization of upbringing. A numerical variable that sums the answers of the CTQ statements 10, 16 and 22, which are not used to compute any trauma score and is used to assess issues related to social desirability or a tendency to deny negative experiences during childhood.
14. SICD-I diagnoses. They were grouped and dichotomized as psychotic, bipolar, depression, obsessive-compulsive, post-traumatic stress, alcohol use, eating and anxiety disorders.

### ***Data mining and statistical analysis***

#### *Machine learning approach*

Figure S2 presents the density plots of numeric variables comparing the original (blue) and the imputed (red) distribution in training data for (A) men and (B) women. This algorithm creates multiple replacement values for multivariate missing data based on Fully Conditional Specification, where a separate model imputes each incomplete variable. After that, we compute the average for numeric variables, median for ordinal categorical, and mode for nominal ones. These calculated values were used to impute the test data. That is, for each variable, all test instances were imputed with the same value. The chosen imputation strategy

prevents any test data from being used to learn the classification model. Furthermore, it considers each test instance alone without using any statistics on the test set.

**Table S1.** Prevalence ratios of attempted suicide for social and family-related variables, stratified by gender, controlled by age and ethnicity.

	Gender*		Attempted Suicide**					
	Men	Women	Men	PR	p	Women	PR	p
<b>Sufficient income</b>	119 (48.2)	159 (38.5)	31 (26.1)	0.61	0.007	83 (52.2)	1.07	0.532
<b>Have children</b>	165 (66.8)	364 (87.3)	59 (35.8)	0.95	0.792	176 (48.4)	0.79	0.064
<b>Have an authentic/genuine friendship</b>	141 (57.8)	200 (47.6)	50 (35.5)	1.04	0.812	103 (51.5)	1.08	0.420
<b>Social Support</b>								
..Have no support from partner, close friends or adult relatives	7 (2.8)	26 (6.2)	2 (28.6)	ref.	-	11 (42.3)	ref.	-
..One person supports him/her	66 (26.7)	121 (28.7)	25 (37.9)	1.45	0.535	55 (45.5)	1.07	0.786
..Two people support him/her	93 (37.7)	156 (37)	28 (30.1)	1.13	0.838	83 (53.2)	1.28	0.303
..Three people support him/her	59 (23.9)	90 (21.3)	24 (40.7)	1.58	0.446	49 (54.4)	1.29	0.312
..Four people support him/her	22 (8.9)	29 (6.9)	7 (31.8)	1.15	0.831	13 (44.8)	1.09	0.779
<b>Close relationship(s) with other drug user(s)</b>								
..Partner, friends and adult relatives are not drug dependent	115 (46.6)	160 (37.9)	31 (27.2)	ref.	-	72 (45.9)	ref.	-
..One of them is drug dependent	82 (33.2)	180 (42.7)	30 (36.6)	1.36	0.142	91 (50.6)	1.10	0.394
..Two or three of them are drug dependent	50 (20.2)	82 (19.4)	24 (48.0)	1.73	0.010	45 (54.9)	1.18	0.211
<b>Homeless (lifetime event)</b>	113 (48.9)	185 (54.3)	46 (40.7)	1.39	0.066	94 (50.8)	1.00	0.937
<b>Arrested (lifetime event)</b>	97 (39.6)	93 (22.1)	32 (33.0)	0.94	0.713	41 (44.1)	0.85	0.202
<b>Difficulty in talking about your feelings</b>	133 (54.1)	277 (67.1)	52 (39.1)	1.31	0.135	141 (50.9)	1.17	0.248
<b>Difficulty in enjoying your leisure time</b>	167 (68.4)	257 (62.2)	65 (38.9)	1.40	0.115	132 (51.4)	1.08	0.460
<b>Difficulty in controlling aggressiveness</b>	151 (61.6)	280 (66.7)	56 (37.1)	1.19	0.354	156 (55.7)	1.50	0.001

<b>Victim of violent crime (lifetime event)</b>	100 (41.0)	135 (32.3)	46 (46.0)	1.83	<0.001	80 (59.3)	1.28	0.010
<b>Saw someone being killed</b>	195 (78.9)	248 (59.2)	73 (37.4)	1.59	0.076	125 (50.4)	1.02	0.822
<b>Childhood physical neglect</b>	84 (34.0)	148 (35.1)	41 (48.8)	1.74	0.001	76 (51.4)	1.03	0.752
<b>Childhood emotional neglect</b>	51 (20.6)	131 (31.0)	24 (47.1)	1.47	0.051	70 (53.4)	1.08	0.497
<b>Childhood emotional abuse</b>	103 (41.7)	183 (43.4)	46 (44.7)	1.61	0.006	102 (55.7)	1.21	0.048
<b>Physical abuse in life</b>	187 (75.7)	349 (82.7)	74 (39.6)	2.14	0.009	183 (52.4)	1.35	0.062
<b>Childhood physical abuse</b>	156 (63.2)	255 (60.4)	67 (42.9)	2.17	0.001	141 (55.3)	1.31	0.012
<b>Sexual abuse in life</b>	59 (23.9)	239 (56.6)	27 (45.8)	1.41	0.052	131 (54.8)	1.24	0.041
<b>Childhood sexual abuse</b>	59 (23.9)	209 (49.5)	27 (45.8)	1.41	0.052	121 (57.9)	1.36	0.003

\*Summary of variables in the line within gender by frequency (%).

\*\*Summary of attempted suicide (yes) within rows and PR controlled by age and ethnicity. Ref.= reference category (no for binary variables).

**Table S2.** Prevalence ratios of attempted suicide for psychiatric, clinical and drug-related variables, stratified by gender, controlled by age and ethnicity.

	Gender*		Attempted Suicide**					
	Men	Women	Men	PR	p	Women	PR	p
<b>HIV</b> <sup>2</sup>	23 (9.3)	78 (18.5)	7 (30.4)	0.77	0.428	39 (50.0)	0.99	0.955
<b>Tuberculosis</b> <sup>2</sup>	17 (6.9)	25 (5.9)	6 (35.3)	0.93	0.837	12 (48.0)	0.95	0.801
<b>Hepatitis</b> <sup>2</sup>	31 (12.6)	45 (10.8)	15 (48.4)	1.30	0.230	24 (53.3)	1.08	0.609
<b>Chronic respiratory disease</b> <sup>2</sup>	41 (16.6)	155 (36.8)	18 (43.9)	1.34	0.138	91 (58.7)	1.30	0.006
<b>Age at first alcohol use</b> <sup>1</sup>	14.3 ± 2.8	15.4 ± 4.5	-	0.96	0.216	-	1.01	0.336
<b>Years of regular alcohol use</b> <sup>3</sup>	2 [0-11]	0 [0-4]	-	1.02	0.042	-	1.01	0.204
<b>Alcohol withdrawal symptoms (last 30 days)</b> <sup>2</sup>								
.. No symptoms	36 (14.8)	62 (14.9)	33 (29.5)	ref.	-	106 (52.5)	ref.	-
.. Abstinent for 6 months	112 (45.9)	202 (48.7)	28 (29.2)	0.91	0.657	61 (40.4)	0.76	0.024
.. Yes	96 (39.3)	151 (36.4)	24 (66.7)	2.15	<0.001	39 (62.9)	1.21	0.101
<b>Alcohol craving (last 30 days)</b> <sup>2</sup>	56 (23.0)	87 (20.6)	26 (46.4)	1.38	0.090	53 (60.9)	1.28	0.017
<b>Age at first cannabis use</b> <sup>1</sup>	15.2 ± 4.4	15.4 ± 4.5	-	1.00	0.842	-	1.00	0.957
<b>Years of regular cannabis use</b> <sup>3</sup>	7 [2-13.5]	4 [0-12]	-	1.00	0.681	-	1.00	0.783
<b>Age at first snorted cocaine use</b> <sup>1</sup>	18.1 ± 4.6	18.5 ± 5.9	-	0.99	0.862	-	1.00	0.849

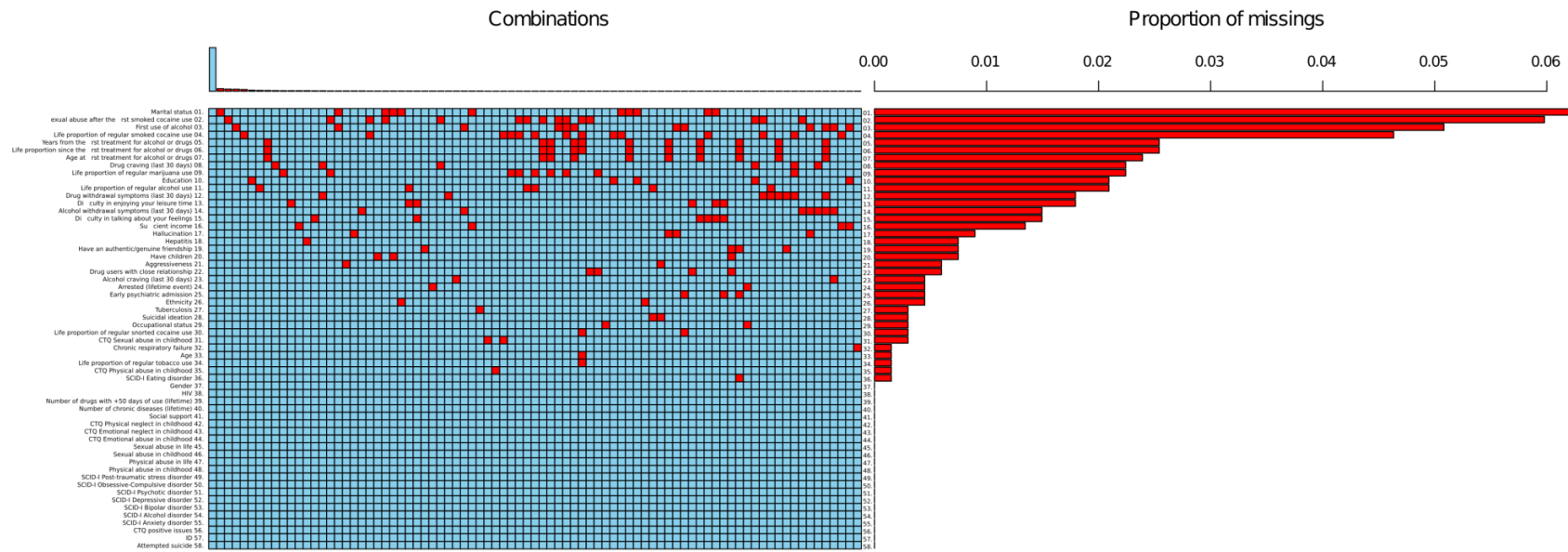
<b>Years of regular snorted cocaine use<sup>3</sup></b>	5.5 [2-13]	4 [0-10]	-	1.01	0.420	-	1.02	0.026
<b>Age at first smoked cocaine use<sup>1</sup></b>	24.6 ± 8.1	19.1 ± 7.2	-	1.00	0.953	-	1.00	0.656
<b>Years of regular smoked cocaine use<sup>3</sup></b>	6 [3-11]	5 [2-10]	-	1.00	0.899	-	0.99	0.334
<b>Drug withdrawal symptoms (last 30 days)<sup>2</sup></b>	146 (61.1)	302 (72.2)	63 (43.2)	1.99	0.002	161 (53.3)	1.31	0.032
<b>Drug craving (last 30 days)<sup>2</sup></b>	161 (66.5)	301 (73.1)	61 (37.9)	1.24	0.274	161 (53.5)	1.26	0.066
<b>Age at first tobacco use<sup>1</sup></b>	14.2 ± 3.7	13.7 ± 4	-	1.01	0.639	-	0.99	0.622
<b>Years of daily tobacco use<sup>3</sup></b>	14 [4-22]	15 [8-21]	-	1.02	0.064	-	0.99	0.692
<b>Number of psychoactive drugs used for more than 50 days in life<sup>1</sup></b>	4.5 ± 1.5	4.2 ± 1.3	-	1.22	<0.001	-	1.09	0.019
<b>Age at first treatment for alcohol and/or drugs<sup>1</sup></b>	28.1 ± 8.7	24.5 ± 8.3	-	0.99	0.655	-	1.00	0.940
<b>Previous psychiatric hospitalization not related to alcohol or drug use<sup>2</sup></b>	28 (11.4)	80 (19.0)	15 (53.6)	1.55	0.031	66 (82.5)	1.93	<0.001
<b>Hallucination (lifetime event)<sup>2</sup></b>								
..No	116 (47.0)	202 (48.6)	24 (20.7)	ref.	-	74 (36.6)	ref.	-
..Yes	54 (21.9)	89 (21.4)	29 (53.7)	2.53	<0.001	60 (67.4)	1.85	<0.001
..Only under effect of drug or abstinence	77 (31.2)	125 (30.0)	33 (42.9)	1.97	0.002	72 (57.6)	1.56	<0.001
<b>Felt yourself incapacitated due to psychological or psychiatric issues (last 30 days)<sup>2</sup></b>	86 (44.1)	168 (45.0)	36 (41.9)	1.18	0.366	96 (57.1)	1.12	0.257
<b>Suicidal Ideation</b>	155 (63.0)	263 (64.8)	83 (53.6)	15.8	<0.001	209 (76.6)	112.6	<0.001
<b>SCID-I Psychotic disorder<sup>2</sup></b>	40 (16.2)	33 (7.8)	18 (45.0)	1.24	0.274	20 (60.6)	1.22	0.182
<b>SCID-I Depressive disorder<sup>2</sup></b>	67 (27.1)	131 (31.0)	36 (53.7)	1.83	<0.001	78 (59.5)	1.30	0.006
<b>SCID-I Bipolar disorder<sup>2</sup></b>	15 (6.1)	42 (10.0)	9 (60.0)	1.69	0.020	26 (61.9)	1.24	0.119
<b>SCID-I Alcohol use disorder<sup>2</sup></b>	159 (64.4)	245 (58.1)	64 (40.3)	1.53	0.046	133 (54.3)	1.24	0.044
<b>SCID-I Eating disorder<sup>2</sup></b>	7 (2.8)	10 (2.4)	1 (14.3)	0.46	0.418	5 (50.0)	0.97	0.932
<b>SCID-I Anxiety disorder<sup>2</sup></b>	121 (49.0)	262 (62.1)	48 (39.7)	1.30	0.133	138 (52.7)	1.15	0.176
<b>SCID-I Post-traumatic stress disorder<sup>2</sup></b>	26 (10.5)	111 (26.3)	16 (61.5)	1.71	0.008	59 (53.2)	1.10	0.389
<b>SCID-I Obsessive-Compulsive disorder<sup>2</sup></b>	16 (6.5)	78 (18.5)	9 (56.3)	1.55	0.073	40 (51.3)	1.02	0.887

\*Summary of variables in the line within gender by <sup>1</sup>mean ± standard deviation, <sup>2</sup>frequency (%) or <sup>3</sup>median [quartile1-quartile3].

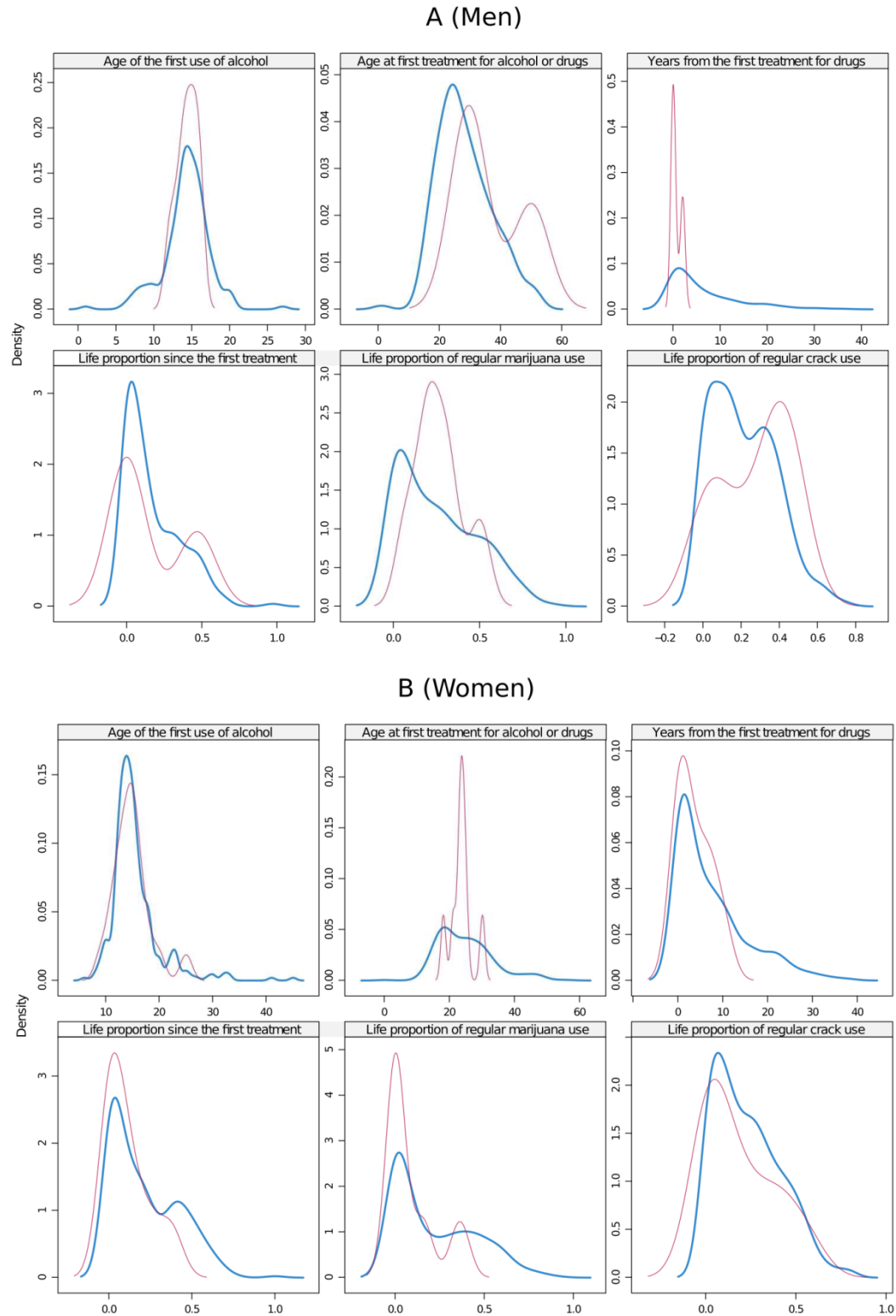
\*\*Summary of attempted suicide (*yes*) within rows and PR controlled by age and ethnicity. Ref.= reference category (*no* for binary variables).



**Fig S1. Missing data pattern of the selected sample.** Each line represents a variable, each column a participant, and the red dots mark the missing data along with the histogram of their frequency. Only 4 out of 57 variables have more than 3% of data missing.



**Fig S2. Comparison of the distributions of original (blue) and inputted (red) training data by MICE method for (A) men and (B) women.**



## 5. ARTIGO 2

**BJP – Brazilian Journal of Psychiatry**

Braz J Psychiatry. 2019 Sep-Oct;41(5):000-000  
doi:10.1590/1516-4446-2019-0633  
Brazilian Psychiatric Association

EDITORIAL

### **Drugs and suicidal behavior: a call for positive, broad and preventive interventions**

Vinícius S. Roglio, Felix H.P. Kessler  
Center for Drug and Alcohol Research, Hospital de Clínicas de Porto Alegre,  
Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil.

Drug misuse and suicidal behavior are among the most concerning and unclear issues in public health and psychiatry worldwide. Over the years, healthcare professionals have been facing much stigma to evaluate those who seek aid for these issues, but there is still a long way to go concerning prevention, accessibility, and treatment quality. The World Health Organization (WHO) has stated that reducing the suicide mortality rate is a global imperative, considering that it accounts for 800,000 annual deaths.<sup>1</sup>

In this issue, Abdalla et al.<sup>2</sup> provide high-quality evidence that corroborates the association between drug use and suicidal behavior in a representative Brazilian sample, showing a prevalence of suicide attempts ranging from 12.4 to 20.8% in alcohol, cannabis, and cocaine users, vs. 4.7% for non-users. Statistics on suicidal behavior in users with more severe dependence are impressive. Data from 690 male patients hospitalized for substance use disorder (SUD; alcohol and cocaine/crack use) in southern Brazil show a lifetime prevalence of suicidal ideation and suicide attempt of 59.8% and 35%, respectively. Moreover, the prevalence of suicide attempt in the 30 days before seeking treatment was 12%.

Although the cross-sectional designs of the aforementioned studies preclude any inference as to causality, there is a well-established wide overlap of biological, environmental, and psychosocial risk factors for substance use and suicide. In this sense, the data from these Brazilian studies corroborate the global literature in finding that the relationship between drug use and suicidal behavior seems to be reciprocal and multidirectional. These are complex phenomena, and a causal model to explain the pathways involved would be hard to build.

The proposed theoretical model of stress-diathesis is the most comprehensive and explains suicidal behavior as the interaction between acquired vulnerabilities (conditioned and/or learned) and triggering stressors.<sup>3</sup> These include genetic factors, trauma, a history of abuse or neglect, personality traits, psychiatric disorders, socioeconomic problems, discrimination, emotional imbalance, rejection, feelings of failure or helplessness, and chronic pain, among others; the same factors are also associated with abuse of psychoactive substances.

It is estimated that the prevalence of psychiatric disorders in individuals who die by suicide is up to 90%.<sup>3</sup> With the exception of schizophrenia, mood, and personality disorders, SUD is the best-documented predictor of suicidal behavior, and it is frequently comorbid with others. Systematic reviews and meta-analysis reinforce the augmented risk of suicidal behavior in individuals with SUD, especially use of alcohol, opioids, intravenous drugs, and chronic cannabis use.<sup>4-7</sup> Interestingly, the risk factors seem to be the same as in the general population, although effect sizes are larger, suggesting more intense associations with suicide in the presence of substance abuse, which thus seems to act as an enhancer of suicide risk. In the context of low- and middle-income countries (LMICs), a systematic review including 108 studies revealed a consistent pattern of direct associations among all substances and all suicidal behaviors.<sup>8</sup>

An extensive meta-analysis of 365 studies shows that predictive modeling of suicide has stagnated; this area of research awaits new insights and is in urgent need of methodological innovation.<sup>9</sup> Since it is clear that drug misuse is not only a key element in aggravating suicide risk but also a modifiable factor, prevention strategies must be implemented even before the whole puzzle has been solved. These interventions must not be directed only to individuals at risk, but rather should be planned widely and at the community level. First, the focus should shift to intensifying efforts to reduce poverty and mitigate the consequences of socioeconomic and environmental stressors. The WHO highlights the need to implement a system for evaluating and monitoring substance use and for conducting research to deepen knowledge in this area, fostering public policies which encourage education, sport, leisure, and culture, especially in LMICs. Government plans and public policies need to integrate preventive and diagnostic strategies with treatment services, the latter focusing on evidence-based brief interventions, which may include training in problem-solving skills and emotional selfregulation for target populations. Public policies to halt the advance of licit drugs and the cannabis industry are also relevant, especially among young individuals.<sup>8</sup>

Telephone- and internet-based interventions are becoming increasingly popular in this setting, but the evidence for their cost-effectiveness is still low; additional studies on these interventions are warranted. Novel technologies using artificial intelligence – such as mobile applications (apps) for symptom monitoring and to facilitate psychotherapeutic approaches – may be useful, especially for highrisk populations such as psychiatric patients, children and adolescents, older adults, and minorities. Interventions that leverage big data through social media should also be explored in the next decade. Lastly, it is important to observe new approaches that aim to decrease stress levels and improve well-being through mindfulness techniques and positive psychology approaches focused on emotions – such as gratitude and forgiveness – and training in optimism and transcendence (e.g., spirituality). In the future, these strategies may represent an inexpensive, user-friendly intervention that can be deployed to large numbers of individuals, possibly even through virtual reality, to promote higher resilience levels and prevent tragic outcomes.<sup>10-12</sup>

## **Disclosure**

The authors report no conflicts of interest.

## **References**

1. World Health Organization (WHO). Preventing suicide: a global imperative. Geneva: WHO; 2014.
2. Abdalla RR, Miguel AC, Brietzke E, Caetano R, Laranjeira R, Madruga CS. Suicidal behavior among substance users: data from the second Brazilian national alcohol and drug survey (II BNADS). *Braz J Psychiatry*. 2019;41:XX-XX.
3. Turecki G, Brent DA. Suicide and suicidal behaviour. *Lancet*. 2016;387:1227-39.
4. 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:e282-91.
5. Yuodelis-Flores C, Ries RK. Addiction and suicide: a review. *Am J Addict*. 2015;24:98-104.

6. Wilcox HC, Conner KR, Caine ED. Association of alcohol and drug use disorders and completed suicide: an empirical review of cohort studies. *Drug Alcohol Depend.* 2004;76 Suppl:S11-9.
7. Borges G, Bagge CL, Orozco R. A literature review and metaanalyses of cannabis use and suicidality. *J Affect Disord.* 2016; 195:63-74.
8. Breet E, Goldstone D, Bantjes J. Substance use and suicidal ideation and behaviour in low- and middle-income countries: a systematic review. *BMC Public Health.* 2018;18:549.
9. Franklin JC, Ribeiro JD, Fox KR, Bentley KH, Kleiman EM, Huang X, et al. Risk factors for suicidal thoughts and behaviors: a metaanalysis of 50 years of research. *Psychol Bull.* 2017;143:187-232.
10. Bolier L, Haverman M, Westerhof GJ, Riper H, Smit F, Bohlmeijer E. Positive psychology interventions: a meta-analysis of randomized controlled studies. *BMC Public Health.* 2013;13:119.
11. Hensley ML, Liu SZ, Siegel SJ. The effects of virtual reality on mental wellness: a literature review. *Ment Health Fam Med.* 2019; 14:811-8.
12. Krentzman AR. Review of the application of positive psychology to substance use, addiction, and recovery research. *Psychol Addict Behav.* 2013;27:151-65.

## 6. CONSIDERAÇÕES FINAIS

No artigo 1 deste trabalho apresentamos uma investigação de preditores de tentativa de suicídio por meio de um algoritmo de aprendizado de máquina e também de fatores associados a esse desfecho por meio da estimação de razões de prevalências, bem como uma discussão sobre a concordância destes em uma amostra de pacientes com transtorno por uso de crack internados. Para isso, se fez uso de variáveis sociodemográficas, clínicas, transtornos psiquiátricos, padrão do uso de drogas e traumas na infância e na vida adulta oriundas de um banco de dado secundário mesclado de duas instituições especializadas em tratamento para dependência química de Porto Alegre-RS. Enquanto no artigo 2, fornecemos uma breve discussão de incentivo à estudos que busquem aprofundar o conhecimento a respeito da interação entre uso de drogas e comportamentos suicidas, bem como reforçamos a necessidade de políticas públicas abrangentes e estratégias de tratamento acessíveis que promovam o bem estar psicológico e reduzam a desigualdade social. Para esse fim, enfatizamos o papel da elaboração de novas abordagens e do uso de tecnologias promissoras que possibilitem sua difusão na sociedade.

A motivação para este trabalho se deu na importância do tema e na observação de uma lacuna na literatura a respeito de preditores de suicídio em usuários de crack, bem como da oportunidade de unir duas técnicas analíticas oriundas de áreas diferentes. Embora os resultados não tenham demonstrado preditores específicos para esta população que já não tenham sido evidenciados por outros trabalhos com a população geral, eles corroboram na consolidação de alguns fatores psicossociais associados a comportamentos suicidas em homens e mulheres usuários de crack, assim como estimulam o uso de novas técnicas de modelagem em psiquiatria.

Em linha com teorias atuais, a etiologia do suicídio e da dependência de crack possuem uma sobreposição no que se refere à vulnerabilidades, principalmente de ordem psicológica e suas influências no comportamento, enfatizando que a dependência de crack pode exercer um forte papel de moderador na relação entre diversos fatores vulnerabilizantes e comportamento suicida. Isso desperta importante preocupação com a população de usuários de crack, o que se confirmou nas altas prevalências de tentativa de suicídio encontradas por homens e mulheres neste trabalho.

Embora os modelos de aprendizado de máquina supervisionados sejam chamados de preditivos, é importante reforçar que este estudo utilizou um banco de dados oriundo de um estudo transversal e, portanto, é incapaz de gerar conclusões a respeito de causalidade entre os preditores e a tentativa de suicídio. Restringe-se à predição baseada em dados sem cronologia precisa em relação à tentativa de suicídio e aplicação a dados ainda não observados ao invés de uma previsão sobre futuros comportamentos suicidas. Epidemiologicamente, inferências causais são investigadas por delineamentos de pesquisa longitudinais, ou, em nível mais fraco de evidência, pela extensa discussão a respeito de critérios de causalidade, como temporalidade, força da associação, consistência, gradiente, plausibilidade biológica, entre outros. Entretanto, para desfechos do espectro do comportamento suicida, observa-se grande dificuldade em delinear tais estudos, assim como uma complicação ética importante, o que justifica a aplicação de um modelo preditivo ao invés de explicativo.

Segundo os modelos de suicídio mais atuais (estresse-diátese, biopsicossocial e motivacional-volitivo), que combinam uma série de fatores estressores e vulnerabilizantes ao longo da vida com fatores precipitantes do comportamento suicida, é compreensível a dificuldade em reproduzi-los através de modelos matemáticos. O volume e complexidade dos fatores psicossociais desenvolvidos ao longo da vida são diversos e os fatores precipitantes apresentam desafios a respeito de sua adequada aferição para pesquisa. Devido à limitada disponibilidade de variáveis oriundas dos instrumentos, este trabalho apresentou um foco maior nos fatores vulnerabilizantes distais à tentativa de suicídio, como transtornos psiquiátricos, traumas, características sociodemográficas e de uso de drogas. Desta forma, os resultados reproduziram a importância da depressão e do trauma na infância pela abordagem descritiva, assim como apresentaram comportamentos mais internalizantes por parte das mulheres (como internações psiquiátricas prévias) e externalizantes por parte dos homens (como uso crônico de mais de uma droga) em associação à tentativa de suicídio pela abordagem preditiva. Resultados, estes, que já eram esperados. O trabalho se beneficiaria da presença de outras variáveis preditores de suicídio já estipuladas pela literatura, como orientação sexual, impulsividade, qualidade de vida, comportamento autolesivo e principalmente, de um bloco de questões focadas em fatores estressantes junto à data da tentativa de suicídio reportada para servir de informações mais proximais e possivelmente precipitantes desta.

Um resultado indireto deste trabalho e que merece destaque, à parte dos objetivos, foi observado na etapa de pré-processamento dos dados e limpeza do banco, onde se observou que cerca de dois terço dos homens foram excluídos da análise devido a dados faltantes,



enquanto para mulheres essa parcela foi de menos de um terço. Isso se deve à dificuldade de retenção dos homens ao programa de tratamento, fazendo com que muitos deles solicitem alta hospitalar precoce e impedindo que a equipe de coleta finalize as aplicações dos instrumentos de pesquisa.

A respeito da qualidade dos dados disponíveis por coletas via instrumentos clínicos e de pesquisa da área psiquiátrica para aplicação de técnicas de aprendizado de máquina, observa-se que a área ainda carece de robustez da informação. Muitas variáveis se propõem a mensurar informações subjetivas e partem de definições que variam no grau de flexibilidade por parte do aplicador do instrumento, assim como também é infrequente o satisfatório entendimento por parte do entrevistado. A qualidade dos resultados da aplicação de técnicas de aprendizado de máquina, assim como de qualquer metodologia estatística é fortemente dependente da padronização de significado que o dado fornece. Para variáveis latentes, por exemplo, é sugerido que a área continue investindo no aprimoramento de seus instrumentos de coleta e mensuração, com a utilização de métodos como a Teoria da Resposta ao Item (TRI). Essa perspectiva vai de encontro à proposta contida na quinta versão do Manual Diagnóstico e Estatístico de Transtornos Mentais, que visa tratar os transtornos psiquiátricos como um espectro, em contraste à abordagem por diagnósticos categóricos. Não sugerindo que estes sejam abolidos, pois também apresentam benefícios, porém mais à prática clínica do que à pesquisa.

Embora os modelos de aprendizado de máquina tenham sofrido sua ascensão em aplicações comerciais com grandes bancos de dados (*big-data*), eles fornecem um arsenal analítico capaz de revolucionar diversas áreas de pesquisa e devem ser explorados para tal. Sob essa ótica, o tamanho amostral deste trabalho pode ser considerado relativamente pequeno (visando geração de resultado aceitável por AM), embora se tenha utilizado uma das maiores amostras de pacientes internados por transtorno por uso de crack do Brasil. Desta forma, é possível classificar a utilização de uma abordagem ainda não convencional para pesquisa em psiquiatria, como o aprendizado de máquina supervisionado, em um banco de dados secundário com múltiplas variáveis como uma forma adequada de reutilizar a informação e convidar esta discussão metodológica para a área da psiquiatria de adição.

Por fim, sugere-se que futuros trabalhos que venham a focar em modelagem preditiva considerem uma abordagem por modelos competitivos. Essa abordagem propõe que o analista faça uso de diversos modelos probabilísticos e algorítmicos diferentes, combinando resultados e comparando suas performances preditivas, visando o ajuste de um modelo cuja estrutura melhor se adeque às características dos dados em uso. Para trabalhos que pretendem focar na

identificação de preditores ou fatores associados ao desfecho, sem necessariamente propor um modelo “produto” que de fato faça predição para novos dados, é possível iterar o ajuste de um modelo, digamos, mil vezes, e apresentar a proporção de vezes que cada preditor foi observado dentre os dez mais importantes. Essa abordagem contorna o problema de um possível resultado aleatório oriundo de uma amostra considerada pequena, agregando robustez e consistência ao identificar um grupo de preditores. Infelizmente, a idealização destas estratégias foram pensadas ao longo do período de estudo sobre o tema e concebidas como viáveis ao final deste, não sendo aplicadas neste trabalho devido a limitação de tempo. Pretendemos dar seguimento nesta direção em futuros trabalhos, também se beneficiando de mais coletas realizadas no período para utilização de uma amostra maior.

## REFERÊNCIAS

- Abdalla RR, Miguel AC, Brietzke E, Caetano R, Laranjeira R, Madruga CS. Suicidal behavior among substance users: data from the Second Brazilian National Alcohol and Drug Survey (II BNADS). *Brazilian J Psychiatry* 2019;41:437–40. <https://doi.org/10.1590/1516-4446-2018-0054>.
- Acion L, Kelmansky D, van der Laan M, Sahker E, Jones D, Arndt S. Use of a machine learning framework to predict substance use disorder treatment success. *PLoS One* 2017;12:e0175383. <https://doi.org/10.1371/journal.pone.0175383>.
- Ahn W-Y, Ramesh D, Moeller FG, Vassileva J. Utility of Machine-Learning Approaches to Identify Behavioral Markers for Substance Use Disorders: Impulsivity Dimensions as Predictors of Current Cocaine Dependence. *Front Psychiatry* 2016;7. <https://doi.org/10.3389/fpsy.2016.00034>.
- Ahn W-Y, Vassileva J. Machine-learning identifies substance-specific behavioral markers for opiate and stimulant dependence. *Drug Alcohol Depend* 2016;161:247–57. <https://doi.org/10.1016/j.drugalcdep.2016.02.008>.
- Arsenault-Lapierre G, Kim C, Turecki G. Psychiatric diagnoses in 3275 suicides: a meta-analysis. *BMC Psychiatry* 2004;4:37. <https://doi.org/10.1186/1471-244X-4-37>.
- Bachmann S. Epidemiology of Suicide and the Psychiatric Perspective. *Int J Environ Res Public Health* 2018a;15:1425. <https://doi.org/10.3390/ijerph15071425>.
- Bachmann S. Epidemiology of suicide and the psychiatric perspective. *Int J Environ Res Public Health* 2018b;15:1425. <https://doi.org/10.3390/ijerph15071425>.
- Barros AJ, Hirakata VN. Alternatives for logistic regression in cross-sectional studies: an empirical comparison of models that directly estimate the prevalence ratio. *BMC Med Res Methodol* 2003;3:21. <https://doi.org/10.1186/1471-2288-3-21>.
- Bastos FI, Bertoni N. PESQUISA NACIONAL SOBRE O USO DE CRACK. Quem são os usuários de crack e/ou similares do Brasil? Quantos são nas capitais brasileiras? Rio de Janeiro: 2014.
- Bastos FIPM, de Vasconcellos MTL, De Boni RB, dos Reis NB, Coutinho CF de S. III Levantamento Nacional sobre o Uso de Drogas pela População Brasileira. São Paulo: 2017.
- Bilsen J. Suicide and Youth: Risk Factors. *Front Psychiatry* 2018;9.

<https://doi.org/10.3389/fpsy.2018.00540>.

- Botti NCL, Machado JS de A, Tameirão FV. Perfil sociodemográfico e padrão do uso de crack entre usuários em tratamento no Centro de Atenção Psicossocial . *Estud e Pesqui Em Psicol* 2014;14:290–303.
- Breiman L. Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author). *Stat Sci* 2001a;16:199–231. <https://doi.org/10.1214/ss/1009213726>.
- Breiman L. Random Forests. *Mach Learn* 2001b;45:5–32. <https://doi.org/10.1023/A:1010933404324>.
- Breiman L, Friedman J, Stone CJ, Olshen RA. *Classification and Regression Trees*. Taylor & Francis; 1984.
- Carlini E., Galduróz JCF, Noto AR, Nappo SA. I Levantamento Domiciliar sobre o Uso de Drogas Psicotrópicas no Brasil: Estudo Envolvendo as 107 Maiores Cidades do País. São Paulo: 2002.
- Carlini EA. II Levantamento Domiciliar Sobre o Uso de Drogas Psicotrópicas no Brasil: Estudo Envolvendo as 108 Maiores Cidades do País. São Paulo: 2006.
- Conner KR, Bridge JA, Davidson DJ, Pilcher C, Brent DA. Metaanalysis of Mood and Substance Use Disorders in Proximal Risk for Suicide Deaths. *Suicide Life-Threatening Behav* 2019;49:278–92. <https://doi.org/10.1111/sltb.12422>.
- Connor JP, Symons M, Feeney GFX, Young RM, Wiles J. The Application of Machine Learning Techniques as an Adjunct to Clinical Decision Making in Alcohol Dependence Treatment. *Subst Use Misuse* 2007;42:2193–206. <https://doi.org/10.1080/10826080701658125>.
- Dantas AP, Azevedo UN de, Nunes AD, Amador AE, Marques M V., Barbosa IR. Analysis of suicide mortality in Brazil: spatial distribution and socioeconomic context. *Rev Bras Psiquiatr* 2017;40:12–8. <https://doi.org/10.1590/1516-4446-2017-2241>.
- von Diemen L, Kapczinski F, Sordi AO, de Magalhães Narvaez JC, Guimarães LSP, Kessler FHP, et al. Increase in brain-derived neurotrophic factor expression in early crack cocaine withdrawal. *Int J Neuropsychopharmacol* 2014;17:33–40. <https://doi.org/10.1017/S146114571300103X>.
- Dong M, Zeng L-N, Lu L, Li X-H, Ungvari GS, Ng CH, et al. Prevalence of suicide attempt in individuals with major depressive disorder: a meta-analysis of observational surveys. *Psychol Med* 2019;49:1691–704. <https://doi.org/10.1017/S0033291718002301>.
- Drucker H, Donghui Wu, Vapnik VN. Support vector machines for spam categorization. *IEEE Trans Neural Networks* 1999;10:1048–54. <https://doi.org/10.1109/72.788645>.

- Duailibi LB, Ribeiro M, Laranjeira R. Profile of cocaine and crack users in Brazil. *Cad Saude Publica* 2008;24:s545–57. <https://doi.org/10.1590/S0102-311X2008001600007>.
- Dunn J, Laranjeira RR, Da Silveira DX, Formigoni MOSL, Ferri CP. Crack Cocaine: An Increase in Use among Patients Attending Clinics in São Paulo: 1990-1993. *Subst Use Misuse* 1996;31:519–27. <https://doi.org/10.3109/10826089609045824>.
- Faller S, Peuker AC, Sordi A, Stolf A, Souza-Formigoni ML, Cruz MS, et al. Who seeks public treatment for substance abuse in Brazil? Results of a multicenter study involving four Brazilian state capitals. *Trends Psychiatry Psychother* 2014;36:193–202. <https://doi.org/10.1590/2237-6089-2014-0040>.
- Fazel S, Wolf A, Larsson H, Mallett S, Fanshawe TR. The prediction of suicide in severe mental illness: development and validation of a clinical prediction rule (OxMIS). *Transl Psychiatry* 2019;9:98. <https://doi.org/10.1038/s41398-019-0428-3>.
- Fletcher SW, Fletcher RH. *Epidemiologia Clínica - Elementos Essenciais*. 4th ed. Porto Alegre: Saraiva; 2006.
- Gattamorta KA, Mena MP, Ainsley JB, Santisteban DA. The Comorbidity of Psychiatric and Substance Use Disorders Among Hispanic Adolescents. *J Dual Diagn* 2017;13:254–63. <https://doi.org/10.1080/15504263.2017.1343965>.
- Gili M, Castellví P, Vives M, de la Torre-Luque A, Almenara J, Blasco MJ, et al. Mental disorders as risk factors for suicidal behavior in young people: A meta-analysis and systematic review of longitudinal studies. *J Affect Disord* 2019;245:152–62. <https://doi.org/10.1016/j.jad.2018.10.115>.
- Gowin JL, Ball TM, Wittmann M, Tapert SF, Paulus MP. Individualized relapse prediction: Personality measures and striatal and insular activity during reward-processing robustly predict relapse. *Drug Alcohol Depend* 2015;152:93–101. <https://doi.org/10.1016/j.drugalcdep.2015.04.018>.
- GRELLA CE, HSER Y-I, JOSHI V, ROUNDS-BRYANT J. Drug Treatment Outcomes for Adolescents with Comorbid Mental and Substance Use Disorders. *J Nerv Ment Dis* 2001;189:384–92. <https://doi.org/10.1097/00005053-200106000-00006>.
- Guimarães CF, Santos DVV dos, Freitas RC de, Araujo RB. Perfil do usuário de crack e fatores relacionados à criminalidade em unidade de internação para desintoxicação no Hospital Psiquiátrico São Pedro de Porto Alegre (RS). *Rev Psiquiatr Do Rio Gd Do Sul* 2008;30:101–8. <https://doi.org/10.1590/S0101-81082008000300005>.
- Halpern SC, Scherer JN, Roglio V, Faller S, Sordi A, Ornell F, et al. Vulnerabilidades clínicas e sociais em usuários de crack de acordo com a situação de moradia: um estudo

- multicêntrico de seis capitais brasileiras. *Cad Saude Publica* 2017;33. <https://doi.org/10.1590/0102-311x00037517>.
- Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning*. 2nd ed. New York, NY: Springer New York; 2009. <https://doi.org/10.1007/b94608>.
- Hoertel N, Franco S, Wall MM, Oquendo MA, Kerridge BT, Limosin F, et al. Mental disorders and risk of suicide attempt: a national prospective study. *Mol Psychiatry* 2015;20:718–26. <https://doi.org/10.1038/mp.2015.19>.
- Hor K, Taylor M. Review: Suicide and schizophrenia: a systematic review of rates and risk factors. *J Psychopharmacol* 2010;24:81–90. <https://doi.org/10.1177/1359786810385490>.
- Horta RL, Horta BL, Rosset AP, Horta CL. Perfil dos usuários de crack que buscam atendimento em Centros de Atenção Psicossocial. *Cad Saude Publica* 2011;27:2263–70. <https://doi.org/10.1590/S0102-311X2011001100019>.
- Hu F, Hao Q. *Intelligent sensor networks: the integration of sensor networks, signal processing and machine learning*. CRC Press; 2012.
- Huys QJM, Maia T V, Frank MJ. Computational psychiatry as a bridge from neuroscience to clinical applications. *Nat Neurosci* 2016;19:404–13. <https://doi.org/10.1038/nn.4238>.
- Ingram RE, Luxton DD. *Vulnerability-Stress Models*. Dev. Psychopathol. A Vulnerability-Stress Perspect., 2455 Teller Road, Thousand Oaks California 91320 United States: SAGE Publications, Inc.; 2005, p. 32–46. <https://doi.org/10.4135/9781452231655.n2>.
- James G, Witten D, Hastie T, Tibshirani R. *An Introduction to Statistical Learning*. vol. 103. New York, NY: Springer New York; 2013. <https://doi.org/10.1007/978-1-4614-7138-7>.
- Kanwar A, Malik S, Prokop LJ, Sim LA, Feldstein D, Wang Z, et al. THE ASSOCIATION BETWEEN ANXIETY DISORDERS AND SUICIDAL BEHAVIORS: A SYSTEMATIC REVIEW AND META-ANALYSIS. *Depress Anxiety* 2013;n/a-n/a. <https://doi.org/10.1002/da.22074>.
- Karthick S, Barwa S. A review on theoretical models of suicide. *Int J Adv Sci Res* 2017;3:101. <https://doi.org/10.7439/ijasr.v3i9.4382>.
- Kennedy MC, Marshall BDL, Hayashi K, Nguyen P, Wood E, Kerr T. Heavy alcohol use and suicidal behavior among people who use illicit drugs: A cohort study. *Drug Alcohol Depend* 2015;151:272–7. <https://doi.org/10.1016/j.drugalcdep.2015.03.006>.
- Klonsky ED, May AM. The Three-Step Theory (3ST): A New Theory of Suicide Rooted in the “Ideation-to-Action” Framework. *Int J Cogn Ther* 2015;8:114–29. <https://doi.org/10.1521/ijct.2015.8.2.114>.

- Kononenko I. Machine learning for medical diagnosis: history, state of the art and perspective. *Artif Intell Med* 2001;23:89–109. [https://doi.org/10.1016/S0933-3657\(01\)00077-X](https://doi.org/10.1016/S0933-3657(01)00077-X).
- van der Laan MJ, Rose S. *Targeted Learning*. New York, NY: Springer New York; 2011. <https://doi.org/10.1007/978-1-4419-9782-1>.
- Langås A-M, Malt UF, Opjordsmoen S. Comorbid mental disorders in substance users from a single catchment area - a clinical study. *BMC Psychiatry* 2011;11:25. <https://doi.org/10.1186/1471-244X-11-25>.
- Lantz B. *Machine Learning with R*. 2nd ed. Packt Publishing Ltd.; 2015.
- Laranjeira R, Madruga CS, Pinsky I, Caetano R, Ribeiro M, Mitsuhiro S. II Levantamento Nacional de Álcool e Drogas (LENAD). São Paulo: 2014.
- Lee Y, Nelder JA. Generalized linear models for the analysis of quality-improvement experiments. *Can J Stat* 1998;26:95–105. <https://doi.org/10.2307/3315676>.
- Lorenzoni V, Curzio O, Karakachoff M, Saponaro A, Sanza M, Mariani F, et al. The effects of the macro-environment on treatment retention for problem cocaine users. *Int J Drug Policy* 2013;24:52–9. <https://doi.org/10.1016/j.drugpo.2012.07.001>.
- McCullagh P, Nelder JA. *Generalized Linear Models*. 2nd ed. New York: Chapman & Hall/CRC Taylor & Francis Group; 1989.
- Narvaez JCM, Jansen K, Pinheiro RT, Kapczynski F, Silva RA, Pechansky F, et al. Psychiatric and substance-use comorbidities associated with lifetime crack cocaine use in young adults in the general population. *Compr Psychiatry* 2014;55:1369–76. <https://doi.org/10.1016/j.comppsy.2014.04.021>.
- Nelder JA, Wedderburn RWM. Generalized Linear Models. *J R Stat Soc Ser A* 1972;135:370. <https://doi.org/10.2307/2344614>.
- Nijem K, Kristensen P, Al-Khatib A, Bjertness E. Application of different statistical methods to estimate relative risk for self-reported health complaints among shoe factory workers exposed to organic solvents and plastic compounds. *Nor Epidemiol* 2009;15. <https://doi.org/10.5324/nje.v15i1.233>.
- Nordentoft M. Absolute Risk of Suicide After First Hospital Contact in Mental Disorder. *Arch Gen Psychiatry* 2011;68:1058. <https://doi.org/10.1001/archgenpsychiatry.2011.113>.
- O'Connor RC, Kirtley OJ. The integrated motivational–volitional model of suicidal behaviour. *Philos Trans R Soc B Biol Sci* 2018;373:20170268. <https://doi.org/10.1098/rstb.2017.0268>.

- Van Orden KA, Witte TK, Cukrowicz KC, Braithwaite SR, Selby EA, Joiner TE. The interpersonal theory of suicide. *Psychol Rev* 2010;117:575–600. <https://doi.org/10.1037/a0018697>.
- Paim Kessler FH, Barbosa Terra M, Faller S, Ravy Stolf A, Carolina Peuker A, Benzano D, et al. Crack Users Show High Rates of Antisocial Personality Disorder, Engagement in Illegal Activities and Other Psychosocial Problems. *Am J Addict* 2012;21:370–80. <https://doi.org/10.1111/j.1521-0391.2012.00245.x>.
- Paiva CB, Ferreira IB, Bosa VL, Narvaez JC de M. Depression, anxiety, hopelessness and quality of life in users of cocaine/crack in outpatient treatment. *Trends Psychiatry Psychother* 2017;39:34–42. <https://doi.org/10.1590/2237-6089-2015-0065>.
- Pariyadath V, Stein EA, Ross TJ. Machine learning classification of resting state functional connectivity predicts smoking status. *Front Hum Neurosci* 2014;8. <https://doi.org/10.3389/fnhum.2014.00425>.
- Plans L, Barrot C, Nieto E, Rios J, Schulze TG, Papiol S, et al. Association between completed suicide and bipolar disorder: A systematic review of the literature. *J Affect Disord* 2019;242:111–22. <https://doi.org/10.1016/j.jad.2018.08.054>.
- 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 (Bangkok)* 2016;38:e282–91. <https://doi.org/10.1093/pubmed/fdv148>.
- Ross S, Peselow E. Co-Occurring Psychotic and Addictive Disorders. *Clin Neuropharmacol* 2012;35:235–43. <https://doi.org/10.1097/WNF.0b013e318261e193>.
- Safavian SR, Landgrebe D. A survey of decision tree classifier methodology. *IEEE Trans Syst Man Cybern* 1991;21:660–74. <https://doi.org/10.1109/21.97458>.
- Santos Cruz M, Andrade T, Bastos FI, Leal E, Bertoni N, Lipman L, et al. Patterns, determinants and barriers of health and social service utilization among young urban crack users in Brazil. *BMC Health Serv Res* 2013;13:536. <https://doi.org/10.1186/1472-6963-13-536>.
- Scaini G, Quevedo J. Suicide rates in the United States continue to rise. Are rates in Brazil underestimated? *Rev Bras Psiquiatr* 2018;40:347–8. <https://doi.org/10.1590/1516-4446-2018-4004>.
- Schooling CM, Jones HE. Clarifying questions about “risk factors”: predictors versus explanation. *Emerg Themes Epidemiol* 2018;15:10. <https://doi.org/10.1186/s12982-018-0080-z>.
- Shatte ABR, Hutchinson DM, Teague SJ. Machine learning in mental health: a scoping



- review of methods and applications. *Psychol Med* 2019;49:1426–48. <https://doi.org/10.1017/S0033291719000151>.
- Shmueli G. To Explain or to Predict? *Stat Sci* 2010;25:289–310. <https://doi.org/10.1214/10-STS330>.
- Siliquini R, Morra A, Versino E, Renga G. Recreational drug consumers: who seeks treatment? *Eur J Public Health* 2005;15:580–6. <https://doi.org/10.1093/eurpub/cki031>.
- Silva DC da, Ávila AC de, Yates MB, Cazassa MJ, Dias FB, Souza MH de, et al. Sintomas psiquiátricos e características sociodemográficas associados à tentativa de suicídio de usuários de cocaína e crack em tratamento. *J Bras Psiquiatr* 2017;66:89–95. <https://doi.org/10.1590/0047-2085000000155>.
- Simon P. *Too Big to Ignore: The Business Case for Big Data*. New Jersey: JOHN WILEY & SONS INC; 2013.
- Toledo L, Góngora A, Bastos FIPM. À margem: uso de crack, desvio, criminalização e exclusão social – uma revisão narrativa. *Cien Saude Colet* 2017;22:31–42. <https://doi.org/10.1590/1413-81232017221.02852016>.
- Too LS, Spittal MJ, Bugeja L, Reifels L, Butterworth P, Pirkis J. The association between mental disorders and suicide: A systematic review and meta-analysis of record linkage studies. *J Affect Disord* 2019;259:302–13. <https://doi.org/10.1016/j.jad.2019.08.054>.
- Turecki G, Brent DA. Suicide and suicidal behaviour. *Lancet* 2016;387:1227–39. [https://doi.org/10.1016/S0140-6736\(15\)00234-2](https://doi.org/10.1016/S0140-6736(15)00234-2).
- Vijayakumar L, Kumar MS, Vijayakumar V. Substance use and suicide. *Curr Opin Psychiatry* 2011;24:197–202. <https://doi.org/10.1097/YCO.0b013e3283459242>.
- WHO, Federation R, OMS/WHO. SUPRE Prevention of suicidal behaviours: a task for all 2012:1–2.
- Zou G. A Modified Poisson Regression Approach to Prospective Studies with Binary Data. *Am J Epidemiol* 2004;159:702–6. <https://doi.org/10.1093/aje/kwh090>.
- Zubaran C, Foresti K, Thorell MR, Franceschini P, Homero W. Depressive Symptoms in Crack and Inhalant Users in Southern Brazil. *J Ethn Subst Abuse* 2010;9:221–36. <https://doi.org/10.1080/15332640.2010.501626>.