- 1 Full title: Development and validation of the Behavioral Health Acuity Risk model: a predictive model for
- 2 suicide prevention through clinical interventions
- 3 Short title: Predictive model for suicide prevention
- 4
- 5 Varun Digumarthi¹, Heather E Strange², Heather B Norman², Derek Ayers³, Raj Patel², Karen E Hegarty¹
- 6
- ⁷ ¹Novant Health Cognitive Computing, Novant Health, Inc, Winston-Salem, North Carolina, USA.
- 8 ²Novant Health Forsyth Medical Center, Novant Health, Inc, Winston-Salem, North Carolina, USA.
- ³Novant Health Charlotte Medical Group, Novant Health, Inc, Charlotte, North Carolina, USA.

10 Abstract

11	Common suicidal ideation screening tools used in healthcare settings rely on the willingness of
12	the patient to express having suicidal thoughts. We present an automatic and data-driven risk model
13	that examines information available in the medical record captured during the normal course of care.
14	This model uses random forests to assess the likelihood of suicidal behavior in patients aged seven or
15	older presenting at any healthcare setting. The Behavioral Health Acuity Risk (BHAR) model achieves an
16	area under the receiver operating curve (AUC) of 0.84 and may be used on its own or as a component of
17	a comprehensive suicidal behavior risk assessment.
18	
19	Introduction

Suicide was among the top four leading causes of death in the United States in 2020 for those between the ages of ten and forty-four years old, and the twelfth leading cause of death among all ages (1). The economic impact is estimated at more than \$4.75 billion per year in the United States (2).

23 Perhaps surprisingly, many decedents of suicide receive some suicide risk assessment or contact 24 with their healthcare provider prior to suicide (3,4). Unfortunately, current standards of care and most 25 existing instruments for suicide risk assessment have been unable to accurately predict risk with high 26 accuracy (5,6,7), and many patients who are assessed for suicidal ideation deny suicidal intent even 27 within one week of completion of suicide (3). For example, the Columbia-Suicide Severity Rating Scale 28 (C-SSRS) (8) was introduced in 2011 and has come to be widely considered as the gold standard for 29 suicide risk assessment. As with other tools relying on self-reported information, its utility is limited by 30 the respondent's willingness to provide truthful answers about intent to commit suicide or self-harm. It 31 has been shown to identify patients at risk at a rate only moderately higher than random, having an area 32 under the receiver operating curve (AUROC) of 0.65 (6), and that the assessment itself is susceptible to 33 navigation errors and misinterpretations of question intent (7).

34 Various computer-aided models have been developed to identify high risk individuals. Tiet et al. 35 developed a decision tree model for individuals with substance use disorder and recent suicidal ideation 36 which evaluates prior suicide attempts, substance use, suicidal ideation, and violent behavior of patients 37 to discriminate between those who will have a future suicide attempt and those who will not (9). Kessler 38 et al. explored regression models for predicting suicide within a population of male non-deployed 39 Regular U.S. Army soldiers having outpatient mental health specialty visits (10). One of the challenges 40 faced by previous researchers is the low prevalence rate of suicide within a general population, which 41 poses difficulty for many statistical and machine learning modeling techniques which suffer under the 42 constraints of a highly imbalanced data set. Typically, the sensitivity and specificity of preceding models 43 has been limited (11,12), which is a barrier to widespread adoption in clinical settings. 44 The objective of the work described here was to develop and validate a predictive model, which 45 we refer to as the Behavioral Health Acuity Risk (BHAR) model, utilizing machine learning to assess 46 likelihood of an individual to complete suicide. The model is designed for any individual aged 7 and older 47 presenting at any healthcare setting for any cause regardless of previous history of suicidal ideation or 48 attempt. The BHAR model generalizes to a wide population and enables behavioral health care team 49 members to direct resources in the most meaningful way to prevent potential harm. It was developed to 50 be used either on its own or in conjunction with a screening instrument such as the aforementioned C-51 SSRS to provide a comprehensive assessment of risk through both self-reported information and 52 automated machine learning driven assessment. We compare the performance of the BHAR model to 53 that of the C-SSRS screening tool and evaluate the effectiveness of the BHAR model in identifying 54 patients at risk over the C-SSRS which is reliant on self-reported information. 55

56 Materials and methods

57 Data collection

58	All research was conducted according to a protocol approved by the Novant Health Institutional
59	Review Board and in accordance with the ethical standards of the Helsinki Declaration of 1975. Data was
60	sourced from existing historical electronic health records housed in the health system's Epic Clarity
61	database (13) and included any patient having been seen by a Novant Health provider between January
62	1, 2016 and August 1, 2020. Novant Health is an integrated health system located in the southeastern
63	United States consisting of 15 medical centers and a medical group of >1,800 physicians at >800
64	ambulatory clinics (14). In addition to these health record data, mortality records from the state of
65	North Carolina Department of Health and Human Services (15) during the same period were used to
66	identify suicide completions which the health system may not otherwise have documented in the
67	medical record. The data set included all encounters for all patients aged 7 years or older. In total, data
68	from 2,884,214 patients were used for development of the model, with the model trained and tested at
69	the encounter level. On average there were 15 encounters per patient.

70

71 Feature and label construction

72 A set of 81 variables were investigated for predictive value in identifying patients who would 73 later have completed suicide with a prediction horizon of 1 week from a given encounter. The predictors 74 included patient demographic and socio-economic characteristics such as age at encounter, gender, 75 race, ethnicity, and marital status. Behavioral characteristics and social determinants of health included 76 access to firearms, feelings of hopelessness or helplessness, use of tobacco, use of alcohol, use of illegal 77 drugs, volume of alcohol consumed per week, receiving routine care through a primary care physician, 78 and prior history of self-injury. Total scores extracted from behavioral assessments such as the Patient 79 Health Questionnaire-9 (PHQ-9) (16) and Generalized Anxiety Disorder-7 (GAD-7) (17) were also 80 included. Predictors related to patient utilization history included prior outpatient visit to a behavioral 81 health specialty clinic in the last 3 days, 30 days and 365 days, as well as having an inpatient admission

82	into behavioral health units in the last 3 days, 30 days and 365 days. Presence or absence of diagnoses
83	such as anxiety disorders, post-traumatic stress disorder (PTSD), and substance use disorders were
84	included as binary variables. For diagnosis related predictors, two features were created for each
85	diagnosis; first, presence or absence of a history of a diagnosis in the patient's lifetime, and second, if
86	the patient was diagnosed for the first time in the last 6 months. In addition to these discrete diagnosis
87	related predictors, the calculated Charlson Comorbidity Index (CCI) score was also included as a
88	predictor. Predictors related to prescribed medications include, for example, antipsychotics,
89	antidepressants, benzodiazepines, and hypnotics. Medication features utilized a six-month lookback of
90	medications reconciled or prescribed during both inpatient and outpatient clinical settings and were
91	indicated as binary variables.
92	
93	Model training and evaluation
94	A random forest model (18) was selected to model the likelihood of suicide due to suitability to
94 95	A random forest model (18) was selected to model the likelihood of suicide due to suitability to this problem and alignment to technical deployment constraints within the EHR platform. A random
94 95 96	A random forest model (18) was selected to model the likelihood of suicide due to suitability to this problem and alignment to technical deployment constraints within the EHR platform. A random forest consists of many independently built decision trees. Each tree is built from a randomly selected
94 95 96 97	A random forest model (18) was selected to model the likelihood of suicide due to suitability to this problem and alignment to technical deployment constraints within the EHR platform. A random forest consists of many independently built decision trees. Each tree is built from a randomly selected subset of features and by bootstrap aggregation of training data. This helps to prevent overfitting
94 95 96 97 98	A random forest model (18) was selected to model the likelihood of suicide due to suitability to this problem and alignment to technical deployment constraints within the EHR platform. A random forest consists of many independently built decision trees. Each tree is built from a randomly selected subset of features and by bootstrap aggregation of training data. This helps to prevent overfitting without compromising on accuracy. Moreover, random forest models can be natively hosted in the EPIC
94 95 96 97 98 99	A random forest model (18) was selected to model the likelihood of suicide due to suitability to this problem and alignment to technical deployment constraints within the EHR platform. A random forest consists of many independently built decision trees. Each tree is built from a randomly selected subset of features and by bootstrap aggregation of training data. This helps to prevent overfitting without compromising on accuracy. Moreover, random forest models can be natively hosted in the EPIC EHR platform (13), enabling the risk score to calculate in near real time and be immediately visible to
94 95 96 97 98 99 100	A random forest model (18) was selected to model the likelihood of suicide due to suitability to this problem and alignment to technical deployment constraints within the EHR platform. A random forest consists of many independently built decision trees. Each tree is built from a randomly selected subset of features and by bootstrap aggregation of training data. This helps to prevent overfitting without compromising on accuracy. Moreover, random forest models can be natively hosted in the EPIC EHR platform (13), enabling the risk score to calculate in near real time and be immediately visible to clinical staff. As mentioned previously, due to the low prevalence of suicide within the general
94 95 96 97 98 99 100 101	A random forest model (18) was selected to model the likelihood of suicide due to suitability to this problem and alignment to technical deployment constraints within the EHR platform. A random forest consists of many independently built decision trees. Each tree is built from a randomly selected subset of features and by bootstrap aggregation of training data. This helps to prevent overfitting without compromising on accuracy. Moreover, random forest models can be natively hosted in the EPIC EHR platform (13), enabling the risk score to calculate in near real time and be immediately visible to clinical staff. As mentioned previously, due to the low prevalence of suicide within the general population this dataset is highly imbalanced. Balancing can be achieved by various techniques, and we
94 95 96 97 98 99 100 101	A random forest model (18) was selected to model the likelihood of suicide due to suitability to this problem and alignment to technical deployment constraints within the EHR platform. A random forest consists of many independently built decision trees. Each tree is built from a randomly selected subset of features and by bootstrap aggregation of training data. This helps to prevent overfitting without compromising on accuracy. Moreover, random forest models can be natively hosted in the EPIC EHR platform (13), enabling the risk score to calculate in near real time and be immediately visible to clinical staff. As mentioned previously, due to the low prevalence of suicide within the general population this dataset is highly imbalanced. Balancing can be achieved by various techniques, and we
94 95 96 97 98 99 100 101 102 103	A random forest model (18) was selected to model the likelihood of suicide due to suitability to this problem and alignment to technical deployment constraints within the EHR platform. A random forest consists of many independently built decision trees. Each tree is built from a randomly selected subset of features and by bootstrap aggregation of training data. This helps to prevent overfitting without compromising on accuracy. Moreover, random forest models can be natively hosted in the EPIC EHR platform (13), enabling the risk score to calculate in near real time and be immediately visible to clinical staff. As mentioned previously, due to the low prevalence of suicide within the general population this dataset is highly imbalanced. Balancing can be achieved by various techniques, and we chose to employ the down sampling method. Down sampling the majority class to achieve a 1:16 ratio
94 95 96 97 98 99 100 101 102 103 104	A random forest model (18) was selected to model the likelihood of suicide due to suitability to this problem and alignment to technical deployment constraints within the EHR platform. A random forest consists of many independently built decision trees. Each tree is built from a randomly selected subset of features and by bootstrap aggregation of training data. This helps to prevent overfitting without compromising on accuracy. Moreover, random forest models can be natively hosted in the EPIC EHR platform (13), enabling the risk score to calculate in near real time and be immediately visible to clinical staff. As mentioned previously, due to the low prevalence of suicide within the general population this dataset is highly imbalanced. Balancing can be achieved by various techniques, and we chose to employ the down sampling method. Down sampling the majority class to achieve a 1:16 ratio was found to be optimal through a hyperparameter search. That is, all the positive cases were retained, and then negative cases were randomly sampled into the data set until the desired ratio was met. The

106	Hyperparameter tuning was performed using a grid search method with 3-fold cross validation having
107	max tree depth between 15 and 30, max bins between 16 and 32, number of trees between 20 and 200
108	and feature subset strategy between 10 and 50. For binary classification tasks, model discrimination
109	ability is frequently evaluated using techniques such as the Area Under the Receiver Operating Curve
110	(AUROC) or the Area Under the Precision-Recall Curve (AUPRC). As even the down sampled data set was
111	imbalanced in terms of outcome label, and because a prioritization is placed on the identification of
112	positive cases, we chose to optimize the model using the AUPRC. Packages from pyspark machine
113	learning were used to create a pipeline for training and testing. Matplotlib was used for creating
114	visualizations. All extraction, analyses, transformations, and training were performed using spark
115	architecture on 8 Azure-hosted compute-optimized standard F4s with 4-core CPUs and 8 GB RAM. For
116	model explainability and interpretability, we examine feature importance which captures the effect that
117	each variable imparts to calculating risk. Feature importance is calculated using Gini impurity.
118	

119 **Results and Discussion**

- 120 The final model incorporates 69 features. Table 1 details these features and their corresponding
- 121 importance in calculating the risk score.
- 122 Table 1. Model features, description, and feature importance.
- 123

Feature	Description	Importance
Medicare/Medicaid	If the patient has active Medicare or Medicaid	0.022150
	coverage at the time of the prediction encounter	
Has Primary Care Provider	If the patient has a Primary Care Provider	0.027619
	indicated in Epic at the time of the prediction	
	encounter.	
Age	The patient's age at the time of the prediction	0.132686
	encounter.	
Gender - Male	If the patient's gender is Male at the time of the	0.036439
	prediction encounter.	

Single	If the patient relationship status is indicated as single at	0.018067
	the time of the prediction encounter.	
Married or significant other	If the patient relationship status is indicated as married	0.020761
	or has a significant other at the time of the prediction	
	encounter.	
Divorced, separated, or	If the patient relationship status is indicated as	0.015483
widowed	divorced, separated, or widowed at the time of the	
	prediction encounter.	
Has Coronary Artery Disease	The patient is active on the Coronary Artery Disease	0.008645
(CAD)	(CAD) registry at the time of the prediction encounter.	
Has Congestive heart failure	The patient is active on the Congestive heart failure	0.003730
(CHF)	(CHF) registry at the time of the prediction encounter.	
Has Chronic kidney disease	The patient is active on the Chronic kidney disease	0.005192
(CKD)	(CKD) registry at the time of the prediction encounter.	
Has Chronic liver disease	The patient is active on the Chronic liver disease (CLD)	0.004033
(CLD)	registry at the time of the prediction encounter.	
Has Chronic obstructive	The patient is active on the Chronic obstructive	0.013216
pulmonary disease (COPD)	pulmonary disease (COPD) registry at the time of the	
	prediction encounter.	
Has diabetes	The patient is active on the diabetes registry at the	0.013223
	time of the prediction encounter.	
Has Human	The patient is active on the Human Immunodeficiency	0.000007
Immunodeficiency Virus	Virus (HIV) registry at the time of the prediction	
(HIV)	encounter.	
Has obesity	The patient is active on the obesity registry at the time	0.024823
	of the prediction encounter.	
Chronic opioids	The patient is active on the chronic opioids registry at	0.003502
	the time of the prediction encounter.	
Asian	The patient primary race is indicated as Asian	0.000376
Black or African American	The patient primary race is indicated as Black or African	0.010433
	American	
White or Caucasian	The patient primary race is indicated as White or	0.017762
	Caucasian	
Hispanic or Latino	The patient primary race is indicated as Hispanic or	0.001588
	Latino	
Alcohol use	If the social history in the last 3 years from time of	0.025474
	prediction encounter indicates alcohol use.	
Alcohol ounce (oz.) per	Maximum consumption of reported ounces per week	0.024881
week	of alcohol in the last 3 years from time of prediction	
	encounter.	
Tobacco use	If the social history in the last 3 years from time of	0.026711
	prediction encounter indicates tobacco use.	
Illicit drug use	If the social history indicates use of illicit drugs before	0.007396
	the time of prediction encounter.	
Access to firearms	If the patient indicated that they have access to	0.005630
	firearms on the Behavioral Health Access form in the	
	last 1 year from time of prediction encounter.	

Hopelessness/Helplessness	0.013006		
	the Behavioral Health Access form in the last 1 year		
	from time of prediction encounter.		
Attempt or self-injury	If the patient has indicated that they have a	0.006491	
composite	current/recent/history of suicide attempt or self-injury		
	on the Behavioral Health Access form in the last 1 year		
	from time of prediction encounter.		
Patient Health	The total score for the most recent Patient Health	0.038088	
Questionnaire-9 (PHQ9)	Questionnaire-9 (PHQ9) assessment which is not more		
total score	than 1 year from the time of prediction encounter.		
Generalized Anxiety	The total score for the most recent Generalized Anxiety	0.010135	
Disorder (GAD) Assessment	Disorder (GAD) assessment which is not more than 1		
total score	vear from the time of prediction encounter.		
On antidepressants	If the nation has a medication order for a drug in the		
	ACO ERX General Depression Medications grouper 6		
	months prior to the prediction encounter		
On antinsychotics	If the nation thas a medication order for a drug in the	0.018418	
on anapsychotics	Antinsychotics pharmacy class 6 months prior to the		
	prediction encounter		
Behavioral health discharge	If the nation three discharged from a behavioral health	0.001677	
in the past 72 hours	unit at any Novant Health facility within 72 hours prior	0.001077	
	to the prediction encounter		
Robavioral boalth discharge	If the patient was discharged from a behavioral health	0.00/579	
in the past 20 days	unit at any Novant Health facility within 20 days prior	0.004375	
In the past 50 days	to the prediction opcounter		
Robaviaral boalth discharge	If the nationt was discharged from a helpovioral health	0.020410	
in the past year	In the patient was discharged from a behavioral field in	0.020410	
in the past year	to the prediction opcounter		
Major Doprossivo Disordor	If the nationt has had an encounter diagnosis for Major	0.019/89	
	Depressive Disorder (MDD) at any time prior to the	0.019489	
	prediction oncounter		
Dineler diserder	If the notion has had an ansauntan diagnosis for	0.000427	
Bipolar disorder	If the patient has had an encounter diagnosis for	0.009437	
	Bipolar disorder at any time prior to the prediction		
Cabizanterania	The notions had an encounter diagnosis for	0.001308	
Schizophrenia	The patient has had an encounter diagnosis for	0.001398	
	Schizophrenia at any time prior to the prediction		
	encounter.	0.000741	
Schizoaffective	If the patient has had an encounter diagnosis for	0.000741	
	Schizoaffective at any time prior to the prediction		
· · ·	encounter.	0.022222	
Insomnia	If the patient has had an encounter diagnosis for	0.022332	
	Insomnia at any time prior to the prediction encounter.		
Personality	If the patient has had an encounter diagnosis for	0.001843	
	Personality disorder at any time prior to the prediction		
	encounter.		
Anxiety	If the patient has had an encounter diagnosis for	0.029484	
	Anxiety at any time prior to the prediction encounter.		

Post-traumatic stress	If the patient has had an encounter diagnosis for Post-	0.002371
disorder (PTSD)	traumatic stress disorder (PTSD) at any time prior to	
	the prediction encounter.	
Major depressive disorder	If the patient has had an encounter diagnosis for Major	0.012966
diagnosed in last 6 months	depressive disorder for the first time within 6 months	
	prior to the prediction encounter.	
Bipolar disorder diagnosed	If the patient has had an encounter diagnosis for	0.000796
in last 6 months	last 6 months Binolar disorder for the first time within 6 months prior	
	to the prediction encounter.	
Schizophrenia diagnosed in	If the patient has had an encounter diagnosis for	0.001182
last 6 months	Schizophrenia for the first time within 6 months prior	
	to the prediction encounter.	
Schizoaffective diagnosed in	If the patient has had an encounter diagnosis for	0.000071
last 6 months	Schizoaffective for the first time within 6 months prior	
	to the prediction encounter.	
Insomnia diagnosed in last 6	If the patient has had an encounter diagnosis for	0.010209
months	Insomnia for the first time within 6 months prior to the	
	prediction encounter.	
Personality diagnosed in last	If the patient has had an encounter diagnosis for	0.003656
6 months	Personality disorder for the first time within 6 months	
	prior to the prediction encounter.	
Anxiety diagnosed in last 6	If the patient has had an encounter diagnosis for	0.017339
months	Anxiety for the first time within 6 months prior to the	
	prediction encounter.	
Post-traumatic stress	If the patient has had an encounter diagnosis for Post-	0.000406
disorder (PTSD) diagnosed	traumatic stress disorder (PTSD) for the first time	
in last 6 months	within 6 months prior to the prediction encounter.	
Drug or Alcohol abuse	If the patient has had an encounter diagnosis for Drug	0.024698
	or Alcohol abuse at any time prior to the prediction	
	encounter.	
Drug or Alcohol abuse	If the patient has had an encounter diagnosis for Drug	0.005854
diagnosed in last 6 months	or Alcohol abuse for the first time within 6 months	
	prior to the prediction encounter.	
Prior suicide attempt or self-	If the patient has had an encounter diagnosis for	0.072841
harm in last 6 months	Suicide and Intentional Self-Inflicted Injury within 6	
	months prior to the prediction encounter.	
Prior suicide attempt or self-	If the patient has had an encounter diagnosis for	0.014476
harm more than 6 months	Suicide and Intentional Self-Inflicted Injury more than 6	
ago	months ago from the prediction encounter.	
On Benzodiazepines	If the patient has a medication order for a drug in the	0.023327
	Benzodiazepines grouper 6 months prior to the	
	prediction encounter.	
On Hypnotics	If the patient has a medication order for a drug in the	0.018933
	Hypnotics pharmacy class 6 months prior to the	
	prediction encounter.	
1	p. ca.otion encounteri	1

Outpatient Behavioral	If the patient had an outpatient behavioral health	0.002036
health encounter in past 72	encounter from unit at any Novant Health facility with	
hours	certain specialties within 72 hours prior to the	
	prediction encounter.	
Outpatient Behavioral	If the patient had an outpatient behavioral health	0.002226
health encounter in past 30	encounter from unit at any Novant Health facility with	
days	certain specialties within 30 days prior to the	
	prediction encounter.	
Outpatient Behavioral	If the patient had an outpatient behavioral health	0.005396
health encounter in past	encounter from unit at any Novant Health facility with	
year	certain specialties within a year prior to the prediction	
	encounter.	
Attention deficit disorder	If the patient has had an encounter diagnosis for	0.008516
(ADD)	Attention deficit disorder (ADD) at any time prior to the	
	prediction encounter.	
Autism spectrum disorder	If the patient has had an encounter diagnosis for	0.000802
(ASD)	Autism spectrum disorder (ASD) at any time prior to	
	the prediction encounter.	
Conduct disorder	If the patient has had an encounter diagnosis for	0.000367
	Conduct disorder at any time prior to the prediction	
	encounter.	
Dementia	If the patient has had an encounter diagnosis for	0.005813
	dementia at any time prior to the prediction	
	encounter.	
Eating disorder	If the patient has had an encounter diagnosis for Eating	0.004055
	disorder at any time prior to the prediction encounter.	
Traumatic Brain Injury (TBI)	If the patient has had an encounter diagnosis for	0.004714
	Traumatic Brain Injury (TBI) at any time prior to the	
	prediction encounter.	
Attention deficit disorder	If the patient has had an encounter diagnosis for	0.000451
(ADD) diagnosed in last 6	Attention deficit disorder (ADD) for the first time	
months	within 6 months prior to the prediction encounter.	
Dementia diagnosed in last	If the patient has had an encounter diagnosis for	0.002033
6 months	Dementia for the first time within 6 months prior to	
	the prediction encounter.	
Traumatic Brain Injury (TBI)	If the patient has had an encounter diagnosis for	0.005943
diagnosed in last 6 months	Traumatic Brain Injury (TBI) for the first time within 6	
	months prior to the prediction encounter.	
Charlson Comorbidity Index	The patient's Charlson Comorbidity Index (CCI) score is	0.055663
(CCI) Score	calculated prior to the prediction encounter.	

124

125

The Receiver Operating Characteristic curve is shown in Fig 1. As previously mentioned, as we

are more interested in evaluating the model in terms of predicting true positives rather than true

127 negatives, we focused mainly on the area under precision recall curve (AUPRC) which avoids the



128 influence of true negatives. The Precision-Recall curve is shown in **Fig 2**.

- 130 Fig 1. Receiver Operating Characteristic curve of the random forest model.
- 131





133 Fig 2. Precision-Recall curve of the random forest model.

134



136 which a patient can be considered as at high risk. As the F1 score is the harmonic mean of precision and

recall, higher F1 scores indicate higher precision and recall, which in turn implies lower false positive and
false negative rates. The F1 score is calculated at every threshold with intervals of 0.01 and plotted as
shown in **Fig 3**. The threshold at which the F1 score peaks, at 44% risk, was selected as the ideal
threshold for classifying a patient as high risk or not, though generally a plateau in F1 score occurs
between approximately 30% and 60% risk. The F1 score and other performance metrics at thresholds of

142 30%, 44%, and 60% are listed in **Table 2**.



143

144 Fig 3. F1 score by risk threshold.

145

146 **Table 2. Performance metrics for select thresholds.**

	Threshold		
	30%	44%	60%
Sensitivity	0.65	0.53	0.41
Specificity	0.96	0.99	0.99
Precision	0.52	0.74	0.85
Negative Predictive Value	0.98	0.97	0.96
Accuracy	0.95	0.96	0.96
F1 Score	0.58	0.62	0.56

- 148 It is worth noting that, as with other probabilistic models, it is not strictly necessary to choose a
- single threshold for clinical intervention. The BHAR score could be used to rank patients by decreasing

150 risk to prioritize limited resources such as behavioral health specialists to the highest risk patients first 151 and proceed down the list in reverse rank order. Similarly, a low-medium-high categorization could be 152 used to address the highest risk patients with the most resource intensive interventions and medium 153 risk patients, comprising a larger proportion of patients, with less resource intensive interventions. 154 At all emergency departments included in this study, the C-SSRS is employed as a suicidal 155 ideation screening tool. The C-SSRS assigns a risk level (No Risk, Low Risk, Moderate Risk, or High Risk) 156 based on patient responses to questions about suicidal thoughts. When examining the one year period 157 over 2021, a total of 386,130 patients were seen at a Novant Health emergency department and 158 completed the C-SSRS screening. Of those, six individuals completed suicide within a week of the ED 159 visit. All six individuals were assessed as No Risk based on their responses to the C-SSRS. The BHAR 160 model would have assigned risk scores of 72%, 51%, 26%, 11%, 2%, and 0% to these six individuals. 161 However, the C-SSRS is a very valuable tool for identifying patients ready to talk about suicidal ideation. 162 In fact, out of the patients screened using the C-SSRS who were rated as High Risk (6,684 or 1.7%), none 163 were identified as having completed suicide within a week of discharge, supporting that identifying 164 patients ready to speak to someone and connecting them to those resources is a successful strategy for 165 suicide prevention. However, an objective data-based risk score such as the BHAR score can supplement 166 self-reported screeners by identifying patients at risk based on data in their medical record. Some 167 patients may respond better to a more personal and gentle inquiry into whether they would like to 168 speak to someone, whereas the C-SSRS, delivered as a standard set of questions asked to every patient, 169 may be perceived as procedural and impersonal.

170

171 Limitations and future work

172 One limitation to this work is the way in which suicide completions are labeled in the medical 173 record and in mortality records. Oftentimes coding for accidental overdose or accidental self-inflicted

174 injury is used when intent is not clear, and these cases would have been omitted from the training set

used to train the model to identify patients at risk.

- 176 There are a few opportunities for model improvement which would be valuable enhancements
- to this work. One is the exploration of sexual orientation and gender identity (SOGI) data as model
- 178 features. Members of the LGBTQ+ community are known to be disproportionately affected by
- depression and suicide (19), so inclusion of SOGI information could be reasonably expected to have
- value in a suicide risk model. SOGI aspects are not well documented in the Novant Health EHR at the
- 181 time of this study, though efforts are underway to increase this documentation for the purposes of
- 182 improving patient care and patient experience.
- 183 Natural language processing could also increase model performance by examining text
- 184 communication between patients and providers, such as through MyChart, and documentation notes
- 185 made by care providers during the normal course of care. One approach could be to finetune the

186 publicly available Clinical BERT model (20) for this task.

187

188 Acknowledgments

- 189 The authors would like to acknowledge colleagues in Novant Health Digital Products and
- 190 Services who enabled this research through support of the Novant Health data warehouse and data
- 191 platform. The authors extend gratitude to Sean Ganon who assisted in obtaining mortality records for
- 192 the state of North Carolina and for literature review relevant to this work.

193

194 **References**

195

 Centers for Disease Control and Prevention. WISQARS Leading Causes of Death Reports, 1981 -2020. [Online]. 2020 [cited 2022 May 17]. Available from: <u>https://wisqars.cdc.gov/fatal-leading</u>.

- 2. Centers for Disease Prevention and Control. WISQARS Cost of Injury. [Online]. 2020 [cited 2022 5 17]. Available from: <u>https://wisqars.cdc.gov/cost/</u>.
- 3. Smith EG, Kim HM, Ganoczy D, Stano C, Pfeiffer PN, Valenstein M. Suicide Risk Assessment Received Prior to Suicide Death by Veterals Health Administration Patients With a History of Depression. J Clin Psychiatry. 2013 March; 74(3): 226-232.
- 4. Juurlink DN, Herrmann N, Szalai JP, Kopp A, Redelmeier DA. Medical Illness and the Risk of Suicide in the Elderly. JAMA Internal Medicine. 2004 June; 164(11): 1179-1184.
- 5. Runeson B, Odeberg J, Pettersson A, Edbom T, Adamsson IJ, Waern M. Instruments for the assessment of suicide risk: A systematic review evaluating the certainty of the evidence. PLoS One. 2017 July; 12(7): e0180292.
- Lindh AU, Waern M, Breckman K, Renberg ES, Dahlin M, Runeson B. Short term risk of non-fatal and fatal suicidal behaviours: the predictive validity of the Columbia-Suicide Severity Rating Scale in a Swedish adult psychiatric population with a recent episode of self-harm. BMC Psychiatry. 2018 October; 18: 319.
- 7. Giddens JM, Sheehan KH, Sheehan DV. The Columbia-Suicide Severity Rating Scale (C–SSRS): Has the "Gold Standard" Become a Liability? Innov Clin Neurosci. 2014 Sep-Oct; 11(9-10): 66-80.
- 8. Posner K, Brown GK, Stanley B, Brent DA, Yershova KV, Oquendo MA, et al. The Columbia-Suicide Severity Rating Scale: Initial Validity and Internal Consistency Findings From Three Multisite Studies With Adolescents and Adults. Am J Psychiatry. 2011 December; 168(12): 1266-1277.
- 9. Tiet QQ, Ilgen MA, Byrnes HF, Moos RH. Suicide Attempts Among Substance Use Disorder Patients: An Initial Step Toward a Decision Tree for Suicide. Alcoholism: Clinical and Experimental Research. 2006 June; 30(6): 998-1005.
- 10. Kessler RC, Stein MB, Petukhova MV, Bliese P, Bossarte RM, Bromet EJ, et al. Predicting suicides after outpatient mental health visits in the Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS). Mol Psychiatry. 2017 April; 22(4): 544-551.
- 11. Bernert RA, Hilberg AM, Melia R, Kim JP, Shah NH, Abnousi F. Artificial Intelligence and Suicide Prevention: A Systematic Review of Machine Learning Investigations. Int J Environ Res Public Health. 2020; 17: 5929.
- 12. Kessler RC, Bossarte RM, Luedtke A, Zaslavsky AM, Zubizarreta JR. Suicide prediction models: a critical review of recent research with recommendations for the way forward. Mol Psychiatry. 2020 January; 25(1): 168-179.
- 13. Epic Systems Corporation. Epic. 2021.
- 14. Novant Health. Company information. [Online]. 2022 [cited 2022 5 17]. Available from: https://www.novanthealth.org/home/about-us/company-information.aspx.

- 15. NCDHHS State Center for Health Statistics. N.C. Vital Records database. North Carolina Department of Health and Human Services, State Center for Health Statistics; 2020.
- 16. Health Resources & Services Administration. Patient Health Questionaire (PHQ) Screeners. [Online]. [cited 2022 7 5]. Available from: <u>https://www.hrsa.gov/behavioral-health/patient-health-</u> <u>questionnaire-phq-screeners</u>.
- Health Resources & Services Administration. GAD-7 (General Anxiety Disorder-7). [Online]. [cited 2022 7 5]. Available from: <u>https://www.hrsa.gov/behavioral-health/gad-7-general-anxiety-disorder-7</u>.
- 18. Breiman L. Random Forests. Machine Learning. 2001; 45: 5-32.
- Haas AP, Eliason M, Mays VM, Mathy RM, Cochran SD, D'Augelli AR. Suicide and Suicide Risk in Lesbian, Gay, Bisexual, and Transgender Populations: Review and Recommendations. Journal of Homosexuality. 2011 January; 58(1): 10 - 51.
- 20. Alsentzer E, Murphy J, Boag W, Weng WH, Jindi D, Naumann T, et al. Publicly Available Clinical BERT Embeddings. Association for Computational Linguistics. 2019 June; Proceedings of the 2nd Clinical Natural Language Processing Workshop: 72 - 78.