

# It is illegal to post this copyrighted PDF on any website.

## Suicide Risk Assessment in Hospitals:

### An Expert System-Based Triage Tool

Isabelle Desjardins, MD<sup>a,\*</sup>; William Cats-Baril, PhD<sup>c</sup>; Sanchit Maruti, MD, MS<sup>a,e</sup>;  
Kalev Freeman, MD, PhD<sup>b</sup>; and Robert Althoff, MD, PhD<sup>a,d</sup>

#### ABSTRACT

**Background:** The November 2010 Joint Commission Sentinel Event Alert on the prevention of suicides in medical/surgical units and the emergency department (ED) mandates screening every patient treated as an outpatient or admitted to the hospital for suicide risk. Our aim was to develop a suicide risk assessment tool to (1) predict the expert psychiatrist's assessment for risk of committing suicide within 72 hours in the hospital, (2) replicate the recommended intervention by the psychiatrist, and (3) demonstrate acceptable levels of participant satisfaction.

**Methods:** The 3 phases of tool development took place between October 2012 and February 2014. An expert panel developed key questions for a tablet-based suicide risk questionnaire. We then performed a randomized cross-sectional study comparing the questionnaire to the interview by a psychiatrist, for model derivation. A neural network model was constructed using 255 ED participants. Evaluation was the agreement between the risk/intervention scores using the questionnaire and the risk/intervention scores given by psychiatrists to the same patients. The model was validated using a new population of 124 participants from the ED and 50 participants from medical/surgical units.

**Results:** The suicide risk assessment tool performed at a remarkably high level. For levels of suicide risk (minimal or low, moderate, or high), areas under the curves were all above 0.938. For levels of intervention (routine, specialized, highly specialized, or secure), areas under the curves were all above 0.914. Participants reported that they liked the tool, and it took less than a minute to use.

**Conclusions:** An expert-based neural network model predicted psychiatrists' assessments of risk of suicide in the hospital within 72 hours. It replicated psychiatrist-recommended interventions to mitigate risk in EDs and medical/surgical units.

*J Clin Psychiatry* 2016;77(7):e874–e882  
[dx.doi.org/10.4088/JCP.15m09881](http://dx.doi.org/10.4088/JCP.15m09881)

© Copyright 2016 Physicians Postgraduate Press, Inc.

<sup>a</sup>Departments of Psychiatry and <sup>b</sup>Surgery, University of Vermont College of Medicine, Burlington

<sup>c</sup>School of Business and <sup>d</sup>Departments of Pediatrics and Psychology, University of Vermont, Burlington

<sup>e</sup>Harvard Medical School, Boston, Massachusetts

\*Corresponding author: Isabelle Desjardins, MD,  
111 Colchester Ave, Burlington, VT 05401  
([isabelle.desjardins@uvmhealth.org](mailto:isabelle.desjardins@uvmhealth.org)).

The Centers for Disease Control and Prevention (CDC) ranks suicide among the top 4 leading causes of death in individuals 10 to 54 years of age.<sup>1</sup> Inpatient suicides consistently rank among the top 5 serious events reported to the Joint Commission with 24.5% occurring in nonpsychiatric units.<sup>2,3</sup> The Joint Commission—formerly known as the Joint Commission on Accreditation of Healthcare Organizations—is a private nongovernmental, independent, not-for-profit agency that sets quality and safety standards and accredits and certifies more than 20,500 health care organizations and programs in the United States. Accreditation and certification by the Joint Commission are recognized as symbols of quality. Their 2004–2014 Sentinel Event database includes 856 reports of inpatient suicides and 69 reports of self-inflicted injuries.<sup>3</sup>

The Joint Commission Sentinel Event Alert issued in November 2010 focused on the prevention of suicides in medical/surgical units and the emergency department (ED). The Event Alert mandates screening every patient treated as an outpatient, in the ED, or admitted to the hospital. It also requires placing at-risk patients in safer environments. Thus, the Event Alert has serious implications for resource utilization.

These Joint Commission requirements increase the burden on hospitals to demonstrate absolute safety around an outcome that is difficult to predict.<sup>4</sup> Suicide has serious and long-lasting consequences for families, staff, and hospitals.<sup>5</sup> It is in the interest of hospitals to fulfill the Joint Commission mandate in a way that is evidence-based and efficient in terms of resource allocation. Current methods of suicide risk screening are not uniformly and systematically implemented.<sup>6</sup> Quantification of the resources required for their implementation remains elusive. Moreover, current screenings are generally designed for research conditions without full consideration for their “real world” applicability.<sup>7</sup>

The bridge between validated measures of suicide-related behaviors and real-time clinical decision-making has not yet been established.<sup>8–11</sup> It is customary for psychiatry professionals to document a number of modifiable, nonmodifiable, and protective risk factors in order to deliver acute and chronic suicide risk estimates. The American Psychiatric Association guidelines intend to provide a guide to clinicians in treating adult patients with suicidal behaviors. However, these guidelines are not meant to serve as a standard of care for such treatment.<sup>12,13</sup> The gold standard to evaluate suicide risk remains the psychiatrist's evaluation.

The goal of this research was to develop a clinically relevant, self-administered, easy-to-use, cost-effective, suicide risk assessment tool that modeled the critical thinking process of board-certified psychiatrists. We hypothesized that the tool would (1) predict the expert psychiatrist's assessment of acute suicide risk in the hospital, (2) replicate the recommended intervention of the expert psychiatrist, and (3) demonstrate acceptable levels of participant satisfaction. We

- Assessment of the acute risk of committing suicide in general hospitals, as mandated by the Joint Commission, requires an effective tool that is not resource intensive.
- This tablet-based suicide risk screening tool replicates a psychiatrist's risk/intervention assessment well, in less time than face-to-face assessment, and with adequate patient satisfaction.
- The tool can be valuable in clinical settings where a shortage of psychiatry staff to assess patient risk exists.

note that the goal of this study was not to predict the actual risk of death by suicide but to replicate common practices.

## METHODS

### Study Design and Setting

We designed a randomized cross-sectional study to evaluate the performance of a tablet-based suicide risk assessment tool in the ED setting. We hypothesized that the tool would predict the assessment of experienced psychiatrists in evaluating a patient's risk of committing suicide in the next 72 hours in the hospital. The study was conducted at our university-affiliated academic medical center. It consisted of 3 phases and was approved by the University Human Research Subjects Institutional Review Board.

### Phase 1: Model Development

Phase 1, conducted October 9–10, 2012, consisted of literature review and model development using nominal group technique in collaboration with an expert panel of recognized suicidologists during a 2-day workshop.<sup>14,15</sup> Nominal group technique is a facilitated, structured group process to build consensus. It typically consists of 3 phases: problem identification, solution generation, and prioritizing of next steps.<sup>16</sup> Nominal group technique can be used effectively in groups with up to 12 members. The method gives equal weight to the opinions of all members of the group and arrives at consensus through a series of converging, anonymous, "round-robin" voting sessions. Nominal group technique is designed to allow the full and equal participation of members. It is also relatively efficient as the high structure of the process allows the exploration of complex issues in a short period of time. Moreover, it generates the production of a large number of ideas and a sense of closure that are often not achieved in less-structured group methods.

Experts were chosen based on their seminal contributions to the literature and policy making on suicide risk assessment. The experts were Jan Fawcett, MD; David A. Jobes, PhD, ABPP; Peter D. Mills, PhD, MS; Morton M. Silverman, MD; and Douglas G. Jacobs, MD (who was a contributor but was unable to attend the meeting). The experts discussed current evidence for suicide screening,<sup>17</sup> reviewed best practices in suicide risk mitigation, and reviewed simulated cases.

Using the nominal group technique for the group consensus process, they selected relevant variables and assigned ranges and weights. A preliminary questionnaire was created and translated into a computerized version. Separate models of risk and intervention were created. Acute suicide risk (referred to hereafter as "Risk") was defined as the patient's risk of committing suicide within 72 hours in the hospital and was classified as minimal, low, moderate, or high. Interventions (referred to hereafter as "Intervention") were classified as routine, specialized, highly specialized, or secured (Table 1).

### Phase 2: Discovery Sample

The receiver operating characteristic (ROC) size module, Stata 12.1, StataCorp, was used to estimate power with varying assumptions.<sup>18</sup> We aimed to achieve 90% positive predictive validity and 90% negative predictive validity under assumptions of  $\alpha = .05$ , a base rate of suicidal ideation of 5% or less,<sup>1</sup> a null hypothesis of positive predictive validity and negative predictive validity = 0.5, and  $1 - \beta$  greater than 0.8 would be achieved with a sample size of 200.

A total of 801 participants were screened by research associates in the ED during a 6-week period. Individuals 18 years or older and able to provide informed consent were approached for participation, regardless of their chief complaint, which was assessed by the ED triage nurse (Figure 1). The screening of participants took place sequentially between Monday and Friday from 2 PM–10 PM. Exclusion criteria included individuals who were unable to consent for participation in the study, intoxicated or unconscious, in severe pain or agitation, escorted by law-enforcement officers, or held in the hospital on an involuntary basis. Two hundred fifty-five participants were included in the analysis.

Research associates obtained informed consent from eligible participants. Participants were randomly assigned to the order in which they completed the suicide risk questionnaire and the face-to-face evaluation by 1 of 4 psychiatrists who were faculty of the University of Vermont College of Medicine's Department of Psychiatry, were board certified in general psychiatry, had at least 5 years of experience in post-residency training, and worked as either inpatient or consultation-liaison psychiatrists. Three of 4 psychiatrists were independent and blinded to all aspects of the study. One was involved in the risk assessment process and the model development and was aware of the final model that was developed. Participants completed satisfaction surveys about their experience with the electronic tool and with the psychiatrist's interview (Table 1).

Psychiatrists were randomly assigned to data collection shifts. The questions and conduct of the interviews were not scripted with a template, as the study team was trying to capture the gestalt of the psychiatrist's assessment in the course of a regular clinical encounter. Psychiatrists independently entered their risk rating and recommendation for intervention on an electronic interface and recorded a summary of their critical thinking for each evaluation. The

**It is illegal to post this copyrighted PDF on any website.**

**Table 1. Questionnaires and Risk Ratings**

Suicide Risk Questionnaire				
Questionnaire Item	Question Wording		Response Choices	
Pain	How much pain are you in?		5-point scale: 1 (No pain)–5 (Worst pain ever)	
Coping	How are you coping?		Not well/ So, so/Well	
Upset	How upset are you?		A lot/A little/Not at all	
Wish to live	How would you rate your wish to live?		None (0)/Weak (1)/Moderate to strong (2)	
Wish to die	How would you rate your wish to die?		None (0)/Weak (1)/Moderate to strong (2)	
	Have you been considering suicide in the last month?		Yes/No	
	Have you had a recent loss? (eg, job loss/change, relationship change, divorce, change of custody of children, death of a loved one, death of a pet, loss of home, forced to move, bankruptcy, dismissal from school, legal troubles, etc)		Yes/No	
	Have you recently abused substances? (eg, alcohol, sedatives or tranquilizers, stimulants, pain killers, marijuana or hashish, cocaine, club drugs, hallucinogens, opioids, inhalants or solvents, other drugs, etc)		Yes/No	
	Have you ever attempted to commit suicide?		Yes/No	
	Have you ever been diagnosed with depression or panic, anxiety, or bipolar disorders?		Yes/No	
	Have you ever had a previous inpatient psychiatric hospitalization?		Yes/No	
Suicide Risk Rating				
High	Moderate		Low	Minimal
Even considering mitigating protective factors, the acute risk of suicide—in the following 72 hours—is high (suicide is possible and likely).	The combination of risk factors and mitigating protective factors yield a risk of suicide in the following 72 hours that is only moderate (suicide possible, but unlikely).		Although there may be suicide risk factors present, the combination of risk factors and mitigating protective factors yields a relatively low risk of suicide in the following 72 hours (suicide very unlikely).	The risk of suicide in the following 72 hours is minimal.
Intervention Rating				
	Routine	Specialized	Highly Specialized	Secured
Psychiatric assessment	No	Psychiatry or crisis consultation	Psychiatry or crisis consultation	Psychiatry or crisis consultation
Environment of care modification	No	No	Environment of care free of harmful hazards	Environment of care free of harmful hazards
Level of observation	Routine	Routine	Routine	Constant observation
After care referral	Routine	Yes	Yes	Yes
Education re: suicide prevention resources at discharge	Yes	Yes	Yes	Yes
Patient Satisfaction Questionnaire for Self-Administered Questionnaire and Psychiatrist's Interview				
Question	Choices			
Overall, how difficult or easy did you find this task?	Easy/Neither easy nor difficult/Difficult			
In terms of length, this task was	Too short/About right/Too long			
Did you answer the questions truthfully?	All of the time/Some of the time/None of the time			
Do you think asking you these questions helped to improve your care?	Yes/ Somewhat/ No			
What do you think about answering these questions again in the future?	I would be happy to answer these questions again/ I do not mind one way or the other/ I do not ever want to answer these questions again			

time required to complete the suicide risk questionnaire and the interview was collected.

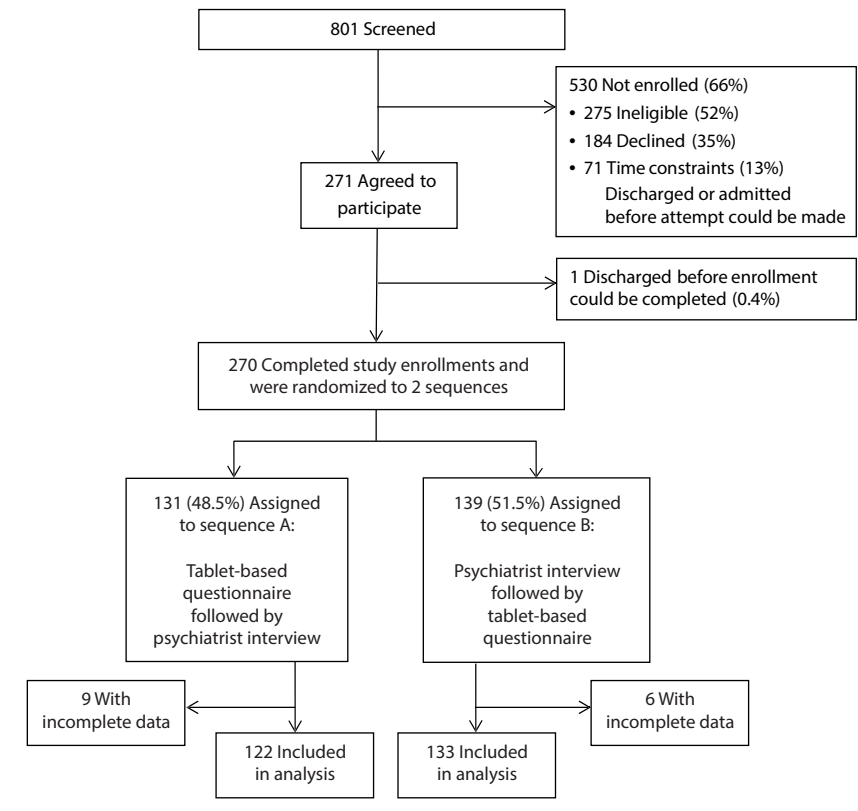
Data and screening records were completed using REDCap (Research Electronic Data Capture, Vanderbilt University), secure web-based software designed to support digital data collection for research purposes. Phase 2 was conducted from May 1, 2013, to June 24, 2013.

**Statistical analysis.** All analyses were conducted using IBM SPSS version 21. Specific measures to test the hypotheses included calculating the agreement between the risk scores using the instrument and the risk scores given by psychiatrists to the same patients. Instrument characteristics were first examined, including descriptive statistics and discriminative power using discriminant analysis. To account

for potential nonlinear interactions among variables, neural network models were used to separately predict the level of risk assigned by the psychiatrist and the appropriate level of intervention.

Neural network modeling is a technique capable of modeling complex functions that are nonlinear and interactive. The prediction analysis was conducted using a multilayer perceptron network in the neural network analysis routine in SPSS. Neural networks are used in situations in which a relationship between the independent variables and predicted variables exists, is strongly suspected, but appears very complex, and is not obvious. The input/output relationship is “learned” through “training.” Training data, including inputs and corresponding outputs,

Figure 1. Phase 2: Participant Enrollment and Design



are assembled to create the weight attribution. The neural network then learns to infer the relationship between the two. This initial training is performed only on a subset of the data so that the predictive capacity of the model can be tested on a new set. Given that training may result in the overfitting of the model to the training data, once the initial algorithm is created, it is then “tuned” on a small sample of participants in order to adjust weights and thresholds to minimize the generalizability of the model obtained during training. When the neural network has learned to model the “unknown” function that relates the input variables to the output variables, it is considered properly trained. At this point, it can then be used to make predictions for which the output is not known. The final step is one of validation. The algorithm is tested with an additional set of data that was not used for training.

Questionnaire responses were entered as variables along with demographic information and chief complaint. Linear age and quadratic age were entered as covariates. Analysis was run separately using psychiatrist risk assessment as the outcome variable and psychiatrist intervention as the outcome variable. For each of these outcomes, a training set of 70% of the data was used to train the neural network that was then tuned with 10% of the sample and finally tested on a holdout set of 20% of the sample. Hidden layer units were allowed to be determined automatically in the analysis program, and batch training was used. Classification outputs for risk and intervention for each of the sets of data were compared.

Outputs of the ROC curve for each level of risk and intervention were created taking into account the entire sample of Phase 2. Differences in participant satisfaction were analyzed using the Bowker test of internal symmetry.<sup>16</sup> We report on the ROC curve from each of these data sets (training, tuning, and holdout) whereby, for each level of Risk or Intervention, we examined the sensitivity and specificity of the model to predict the expert response. Within SPSS, a classification table is created for each categorical dependent variable that gives the number of cases correctly classified for that category. A ROC curve is then created for each categorical dependent variable. Because both the Risk and the Intervention variables have more than 2 categories, each category was treated as a positive state versus the aggregate of all other categories (the equivalent of dummy coding). It is from these curves (1 for each level of Risk or Intervention) that the areas under the curves were constructed.

### Phase 3: Replication and Extension Samples

To determine if the model fit in Phase 2 was replicable and could be extended to non-ED patients, we studied a new set of 124 participants from the ED and 50 participants from medical/surgical units using the inclusion/exclusion criteria and methods of data collection outlined in Phase 2 (Figures 2 and 3).

Of the 4 psychiatrists who completed the interviews in Phase 3 (Phase 3A conducted December 4, 2013–January 8, 2014, and Phase 3B conducted February 3, 2014–February 27, 2014), 3 had also conducted them in Phase 2. The fourth

Figure 2. Phase 3A: Participant Enrollment and Design

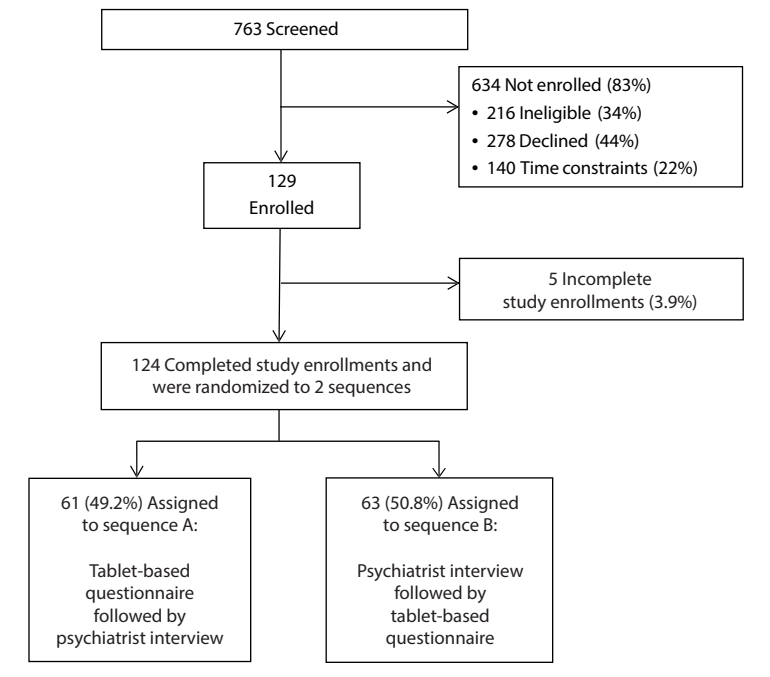
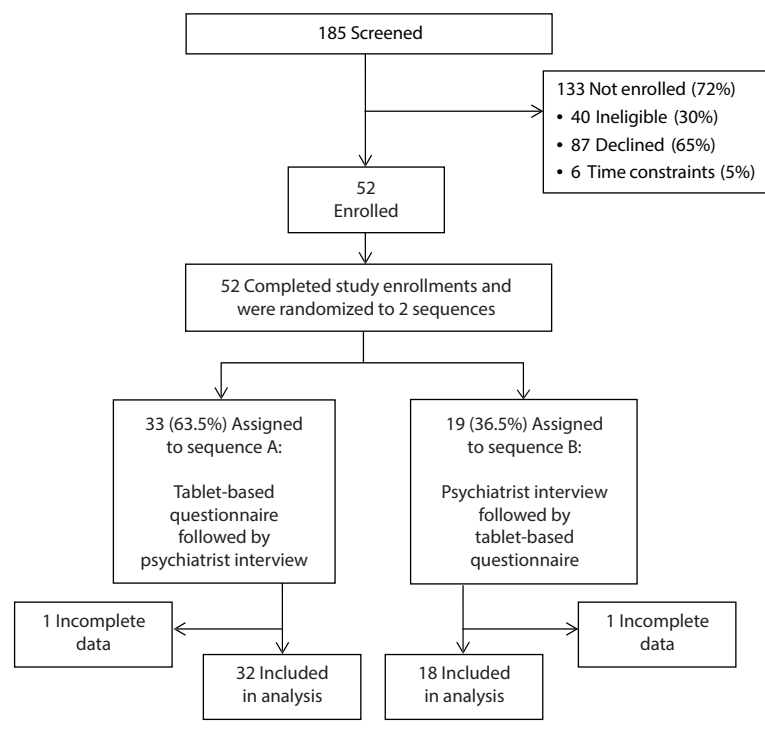


Figure 3. Phase 3B: Participant Enrollment and Design



psychiatrist met all the qualifications mentioned under Phase 2 but was less than 5 years out of residency training.

**Statistical analysis.** The neural network models, built off Phase 2 data, were fitted to the new population of ED participants (direct replication) and to medical/surgical unit participants (external validity). By using the weights determined separately for levels of risk and intervention, prediction accuracy of the models for the Phase 3 samples was determined.

## Phase 2

Basic descriptive statistics for the variables entering the analysis and demographics are presented in Table 2. Individuals presenting with a psychiatric chief complaint had lower wish to live (median 2.00 vs 3.00, Mann-Whitney  $U=920.5$ ,  $n_1=235$ ,  $n_2=20$ ,  $P<.001$ ) and higher wish to die (median 2.50 vs 1.00, Mann-Whitney  $U=1,130.5$ ,  $n_1=235$ ,  $n_2=20$ ,  $P<.001$ ) than those without a psychiatric chief complaint, respectively. The result range for “wish to live” and “wish to die” is 0–2 on the questionnaire (Table 1). However, the data were recoded to be on a scale of 1–3. Measure of time taken for interview and questionnaire completion was available in 244 cases (95.7%). Mean time for the interview from beginning to end was 7:57 minutes (SD=6:46 minutes). Mean time for questionnaire completion was significantly lower at 0:56 minutes (SD=0:41 minutes;  $t_{243}=16.34$ ,  $P<.001$ ). Segmenting the sample by chief complaint shows that, for those patients with a psychiatric chief complaint, both the expert evaluation (15:42 minutes, SD=7:39 minutes) and the questionnaire assessment (1:14 minutes, SD=0:39 minutes) took longer. The difference in expert evaluation time was statistically significant ( $t_{251}=-5.64$ ,  $P<.001$ ) while the difference for the questionnaire was not ( $t_{244}=-1.78$ ,  $P=.076$ ). While several measures were not normally distributed, parametric and nonparametric tests showed the same significance; only the parametric tests are presented here.

On the basis of a 3-point level of satisfaction, 91% of participants found the electronic questionnaire easy to complete, whereas 95% found the interview easy to complete (Kendall tau-b=0.30,  $P=.024$ ). Twelve percent of participants found the length of the electronic questionnaire too short, whereas only 6% found the interview too short (Kendall tau b=0.28,  $P=.015$ ). More participants thought that the interview was likely to help improve their care (53%) compared to the questionnaire (44% of participants) (Kendall tau b=0.62,  $P<.001$ ).

In terms of the ability of the tool to accurately predict psychiatrist opinion, initial discriminant analysis demonstrated that the group centroids for the minimal and low risk groups were nearly identical. Consequently, these 2 groups were collapsed into a “low or minimal” risk group (Risk). Similar analyses for the intervention groups did not demonstrate clear overlap among categories in the group

**Table 2. Demographics**

Variable	Phase 2: ED—Discovery (n=255)	Phase 3A: ED—Replication (n=124)	Phase 3B: Med/ Surg Extension (n=50)
Sex, mean	.48	.48	.54
proportion male			
Psychiatric, mean	.08	.07	.00
proportion <sup>a</sup>			
Age, mean (SD), y	42.58 (17.09)	46.60 (18.69)	63.96 (13.25)
Coping, mean (SD)	2.37 (0.691)	2.43 (0.689)	2.40 (0.700)
Upset, mean (SD)	2.28 (0.702)	2.28 (0.705)	2.30 (0.735)
Wish to live, mean (SD)	2.89 (0.371)	2.93 (0.317)	2.96 (0.198)
Wish to die, mean (SD)	1.19 (0.500)	1.16 (0.449)	1.12 (0.385)
Expert time, mean (SD), min	7:57 (6:46)	5:52 (4:25)	6:24 (4:00)
iPad time, mean (SD), min	0:56 (0:41)	1:15 (0:58)	1:56 (2:23)

<sup>a</sup>Proportion with a psychiatric chief complaint.

Abbreviations: ED=emergency department, med/surg=medical/surgical.

centroids, thus intervention (Intervention) was entered as a 4-category output variable in the neural network analysis. We estimated the clinical implications of collapsing the (Risk) categories to be reasonable. In reality, the clinical attention afforded to an acute risk estimate of “very unlikely” versus “minimal” is not different in the context of a busy clinical setting.

For Risk, the neural network optimization converged on 1 hidden layer with 6 nodes. The classification accuracy for the training set was 94% and 94% for the testing set. For the holdout set, which was not involved in creating the model, classification accuracy was 94%. ROC curves were created and areas under the curves were constructed. Areas under the curves represent the percent agreement between the risk assessment resulting from the interview and the risk assessment resulting from the questionnaire. Areas under the curves for the Risk levels were 0.97 for the minimal or low group, 0.94 for the moderate group, and 0.98 for the high group (Table 3).

For Intervention, classification accuracy numbers were similar. The optimization converged on a 2 hidden layer network, with 7 nodes in the first layer and 6 in the second. The classification accuracy for the training set was 95% and 91% for the testing set. For the holdout set, classification accuracy was 88% (Table 3). If the psychiatrists assessed Intervention as “routine,” there were no cases for which the model predicted anything except “routine.” There was only 1 classification error by the model for which the psychiatrist’s recommendation was “secured.” In this single case, the model classified the intervention as “highly specialized.” Areas under the curves for the 4 levels were 0.95 for “routine,” 0.92 for “specialized,” 0.91 for “highly specialized,” and 0.99 for “secure,” based on all sets.

### Phase 3: Replication (3A) and Extension (3B) Samples

Basic descriptive statistics for the variables entering the models and demographics are presented in Table 2. The medical/surgical sample (3B) was, on average, 17.4 years older than the ED sample (3A),  $t_{172} = -5.98$ ,  $P < .001$ .

Average scores on wish to live and wish to die were similar to the discovery sample. Measure of time for both interview and questionnaire completion was available for 171 subjects (98.2%), and measure of time for questionnaire completion was available for all subjects (100%). Mean time for the interview from beginning to end was 5:52 minutes (SD=4:25 minutes) in the ED and 6:24 minutes (SD=4:00 minutes) in medical/surgical units. Mean time for questionnaire completion was significantly lower at 1:15 minutes (SD=0:58 minutes) in the ED as compared to 1:56 minutes (SD=2:23 minutes) in medical/surgical units ( $t_{172} = -2.68$ ,  $P = .008$ ).

Placing variables from Phase 3 into the neural network models designed in Phase 2 and providing no further training of the network demonstrated high levels of continued accuracy. Overall, on the new Phase 3 samples, the model predicted the psychiatrists’ assessment of Risk 91% of the time (Table 3) and predicted the psychiatrists’ assessment of Intervention 89% of the time (Table 3). In a single case, the model predicted “routine” intervention when the psychiatrist recommended “secured” intervention. While this error may seem significant, this participant was not labeled “high” Risk by either the model or the psychiatrist. On further investigation, this participant was not at acute risk for suicide but was at risk for falls and confusion secondary to a history of stroke, a documented history of cognitive slowing, and a remote history of depression, attention-deficit/hyperactivity disorder, and narcotic abuse, resulting in the psychiatrist’s recommendation for a “secured” intervention.

For the medical/surgical sample (3B), randomization for the 2 sequence groups was 1:1. However, due to random variation in accessing the participants in the midst of their medical care and coordinating it with expert availability, more participants ended up receiving the tablet-based assignment first.

## DISCUSSION

Our data indicate that a neural network model can accurately predict a psychiatrist’s assessment of risk of suicide in the hospital within 72 hours and that it can accurately replicate psychiatrist-recommended interventions to mitigate the risk of suicide in ED and medical/surgical units. We have shown that this self-administered, electronically based assessment tool is clinically accurate and convenient. To our knowledge, this is the first study of its kind to use neural network models to replicate the current gold standard—evaluation of acute suicide risk and level of intervention made by a psychiatrist.

The strength of this research study is the use of both statistical replication within the data analysis (using a holdout set) and physical replication (using new data collection) in testing the predictive models. In the neural networks, both the Intervention and the Risk models predicted the holdout sets with high levels of accuracy. This high level of accuracy was also seen in the replication and extension sets,

**Table 3. Levels of Agreement**

Level of Agreement on Risk Level (Risk)		Model Predicted			Total
Sample	Psychiatrist Assessment	Low or Minimal	Moderate	High	
Discovery (Phase 2)					
Training	Low or Minimal	136	1	0	137
	Moderate	3	3	3	9
	High	0	1	3	4
	Total	139	5	6	150
Testing	Low or Minimal	30	0	0	30
	Moderate	1	1	0	2
	High	0	1	0	1
	Total	31	2	0	33
Holdout	Low or Minimal	64	0	0	64
	Moderate	2	2	2	6
	High	0	0	2	2
	Total	66	2	4	72
Replication (Phase 3)					
Training	Low or Minimal	110	2	0	112
	Moderate	7	1	1	9
	High	0	1	2	3
	Total	117	4	3	124
Extension (Phase 3)					
Training	Low or Minimal	46	1	0	47
	Moderate	3	0	0	3
	High	0	0	0	0
	Total	49	1	0	50

Level of Agreement on Intervention Modality (Intervention)		Model Predicted				Total
Sample	Psychiatrist Assessment	Routine	Specialized	Highly Specialized	Secure	
Discovery (Phase 2)						
Training	Routine	133	1	0	0	134
	Specialized	2	2	1	0	5
	Highly Specialized	1	1	2	2	6
	Secure	0	0	0	5	5
	Total	136	4	3	7	150
Testing	Routine	28	0	0	0	28
	Specialized	2	1	0	0	3
	Highly Specialized	1	0	0	0	1
	Secure	0	0	0	1	1
	Total	31	1	0	1	33
Holdout	Routine	62	0	0	0	62
	Specialized	3	0	2	2	7
	Highly Specialized	1	0	0	0	1
	Secure	0	0	1	1	2
	Total	66	0	3	3	72
Replication (Phase 3)						
Training	Routine	105	2	2	0	109
	Specialized	7	0	2	0	9
	Highly Specialized	1	1	0	1	3
	Secure	1	0	1	1	3
	Total	114	3	5	2	124
Extension (Phase 3)						
Training	Routine	44	0	1	0	45
	Specialized	4	0	1	0	5
	Highly Specialized	0	0	0	0	0
	Secure	0	0	0	0	0
	Total	48	0	2	0	50

suggesting high external validity and potential to export the results to new samples and alternative settings. Further, the design of the study balanced experimental design (randomizing initial assessment modality and assigned psychiatrist) in an actual clinical setting.

This suicide risk assessment tool is of potential use in settings with high volume and rapid turnover where timely psychiatric expertise is unavailable.

This tool has the potential to serve as a clinical decision support system, contributing to increase quality and reliability in hospitals, allowing them to efficiently meet safety and regulatory requirements and to optimize the use of limited resources by eliminating the need for excessive screening of low risk groups. Instead, hospitals may focus these resources on settings where risk is high.<sup>19</sup>

More participants perceived that the interview would improve their care than would the questionnaire. In most situations, however, the risk and intervention from both the interview and the tool were the same, and participant satisfaction for both was high. Given that the time needed for the electronic questionnaire is a fraction of the time for face-to-face assessment, future research will determine whether this small trade-off in perceived satisfaction is worth the cost of the human resources and increased wait times for patients. Patient education around the usefulness of electronic expert systems may be helpful in changing perception.

**Limitations**

Psychiatrists were instructed to conduct a face-to-face interview aiming at estimating the subject’s risk to commit suicide in the hospital in the next 72 hours. This interview is an extremely specific task, narrow in scope, which is usually embedded in a lengthier and broader interview when performing a psychiatric evaluation and suicide risk assessment.

In this study, “measure of time” is defined as the time between the examiner’s entrance and exit from the patient’s room. It does not include the time spent by the examiner to review the subject’s personal health information in the electronic medical record prior to and after the interview. It also does not include the time spent by the examiner to think through the data before determining the risk and intervention recommendations. Despite the significant time difference spent with the subjects with a “psychiatric chief complaint,” the “measure of time” results may lead to the perception by the reader of a suboptimal risk assessment, given that the time is admittedly faster than a regular comprehensive evaluation. The chosen definition of “measure of time” was used to avoid an overestimation of the length of the face-to-face interview, given that the aim of a psychiatric interview is not solely a suicide

**It is illegal to post this copyrighted PDF on any website.**

risk assessment. Instructing the psychiatrists to perform a full psychiatric evaluation would have, on the other end, biased the time comparisons between interview and tool by bringing less specificity to the task studied.

Although 4 of the 5 psychiatrists doing the face-to-face evaluations were independent and blinded to all aspects of the study, 1 was involved in the risk assessment process and in the model development and was aware of the final model developed. Although the role of this 1 psychiatrist may have potentially biased the results toward greater agreement between the model and the expert psychiatrists' risk assessments, there was in fact no difference in results across all 5 psychiatrists.

All psychiatrists doing the face-to-face evaluations are academic psychiatrists with extensive clinical experience in the assessment of high risk, complex psychiatric patients at a tertiary health care facility. As a group, the examiners have more exposure to the full range of suicide risk spectrum than a "typical" psychiatrist. Moreover, they practice psychiatry in the state of Vermont, which is a very liberal state when it comes to putting more weight on the respect for civil liberties and the right to self-determination over the treatment of psychiatric illness against objection. This represents potential biases away from high-risk determinations. To address the small sample size of patients predicted as moderate to high risk, future work should include replication of the results at another institution, in a different state, or in patient

populations with a psychiatric chief complaint in the ED, or who are admitted to inpatient psychiatry units.

In a single case, the model predicted "routine" intervention when the psychiatrist recommended "secured" intervention. While this error may seem significant, this participant was not labeled "high" risk by either the model or the psychiatrist. Although this participant was not at acute risk for suicide per se, this intervention misclassification represents a limitation.

The tool presented here is meant to "merely" model the psychiatrist's risk assessments and intervention recommendations for a very specific patient population and clinical setting. The tool does not assist in the prediction of suicide. It is a triage and risk mitigation tool. Since inpatient suicides are rare, it is difficult to use inpatient suicide or inpatient suicide attempts as a dependent variable in evaluating the power of the implementation of this tool without conducting a large scale, longitudinal, multicenter study. These large data collection enterprises require a standardized, easy-to-deliver assessment method that can be broadly disseminated. We think that the tool we have developed will aid to further research into suicide prevention.

In addition, the cost-benefit ratio of universal screening and patient perception of quality of care with these models will need to be established, along with the satisfaction of health care providers with this type of tool.

Finally, the participation rate and the exclusion criteria may also limit generalization of the findings.

**Submitted:** February 10, 2015; accepted August 3, 2015.

**Online first:** June 7, 2016.

**Author contributions:** Dr Desjardins assumes accountability for the integrity of the work as a whole. She was involved in the conception and design of the work as well as in the acquisition, analysis, and interpretation of the data. She was also involved with drafting and revising the manuscript.

**Potential conflicts of interest:** Drs Desjardins, Cats-Baril, Maruti, and Althoff are partners in WISER Systems, LLC, which, with the University of Vermont, has ownership rights to the Systematic Electronic Risk Assessment for Suicide. Dr Althoff receives grant or research support from the National Institute of Mental Health, the National Institute of General Medical Sciences (NIGMS), and the Klingenstein Third Generation Foundation; receives honoraria from CME presentations for Oakstone Medical Publishing; and is employed, in part, by the nonprofit Research Center for Children, Youth, and Families that has developed and publishes the Achenbach System of Empirically-Based Assessment. Dr Freeman is funded by the NIGMS (K08GM098795), the Luoxis Corporation, and the Totman Trust.

**Funding/support:** This research was supported by grants from the Fletcher Allen Foundation and from the University of Vermont Medical Group (UVMMG).

**Role of the sponsor:** The supporting agencies had no role in the conduct and publication of the study.

**Disclaimer:** The content is solely the responsibility of the authors and does not necessarily represent the official views of the Fletcher Allen Foundation or UVMMG.

**Acknowledgments:** The authors thank the University of Vermont Emergency Medicine Research Associate Program and the following University of Vermont-affiliated staff and faculty: Ms Diantha Howard; Ms Abigail Wager; Ms Chelsea Manning; Donna Rizzo, PhD; Judy Lewis, MD; Isabel Norian, MD; Anne Rich, MD; Tobey Horn, MD; and Conor Carpenter, MD, for their efforts on this project. The following were among the contributors to the expert panel, and the authors acknowledge them for their contributions: Jan Fawcett, MD, University of New Mexico School of Medicine; Douglas G. Jacobs, MD, Screening for Mental Health Inc; Peter D. Mills, PhD, MS, The Geisel School of Medicine at Dartmouth; Morton M. Silverman, MD, The University of Colorado at Denver; and David A. Jobes, PhD, ABPP, The Catholic University of America; members of the expert panel were compensated for their time. None of these individuals have any conflicts of interest to report.

## REFERENCES

1. Crosby AE, Han B, Ortega LAG, et al; Centers for Disease Control and Prevention (CDC). Suicidal thoughts and behaviors among adults aged  $\geq 18$  years—United States, 2008–2009. *MMWR Surveill Summ*. 2011;60(13):1–22.
2. The Joint Commission. A follow-up report on preventing suicide: focus on medical/surgical units and the emergency department. Sentinel Event Alert. Issue 46; November 17, 2010.
3. Summary data of sentinel events reviewed by The Joint Commission. The Joint Commission's Web site. [http://www.jointcommission.org/assets/1/18/2004\\_to\\_2014\\_4Q\\_SE\\_Stats\\_-\\_Summary.pdf](http://www.jointcommission.org/assets/1/18/2004_to_2014_4Q_SE_Stats_-_Summary.pdf). Updated January 14, 2015. Accessed February 25, 2016.
4. Busch KA, Fawcett J, Jacobs DG. Clinical correlates of inpatient suicide. *J Clin Psychiatry*. 2003;64(1):14–19.
5. Ballard ED, Pao M, Horowitz L, et al. Aftermath of suicide in the hospital: institutional response. *Psychosomatics*. 2008;49(6):461–469.
6. Horowitz LM, Ballard ED, Pao M. Suicide screening in schools, primary care and emergency departments. *Curr Opin Pediatr*. 2009;21(5):620–627.
7. Boudreaux ED, Horowitz LM. Suicide risk screening and assessment: designing instruments with dissemination in mind. *Am J Prev Med*. 2014;47(suppl 2):S163–S169.
8. Bongiovi-Garcia ME, Merville J, Almeida MG, et al. Comparison of clinical and research assessments of diagnosis, suicide attempt history and suicidal ideation in major depression. *J Affect Disord*. 2009;115(1–2):183–188.
9. Ronquillo L, Minassian A, Vilke GM, et al. Literature-based recommendations for suicide assessment in the emergency department: a review. *J Emerg Med*. 2012;43(5):836–842.
10. Randall JR, Colman I, Rowe BH. A systematic review of psychometric assessment of self-harm risk in the emergency department. *J Affect Disord*. 2011;134(1–3):348–355.
11. Horowitz LM, Bridge JA, Teach SJ, et al. Ask Suicide-Screening Questions (ASQ): a brief instrument for the pediatric emergency department. *Arch Pediatr Adolesc Med*. 2012;166(12):1170–1176.
12. Jacobs DG, Baldessarini RJ, Conwell Y, et al; Work Group on Suicidal Behaviors. *Practice Guideline for the Assessment and Treatment of Patients with Suicidal Behaviors*. 2nd ed. Arlington, VA: American Psychiatric Association; 2004.
13. Jacobs D, Brewer M. APA Practice Guideline provides recommendations for assessing and treating patients with suicidal behaviors. *Psych*



**It is illegal to post this copyrighted PDF on any website.**

- Annals*. 2004;34(5):373–380.
14. Bridwell KH, Cats-Baril W, Harrast J, et al. The validity of the SRS-22 instrument in an adult spinal deformity population compared with the Oswestry and SF-12: a study of response distribution, concurrent validity, internal consistency, and reliability. *Spine (Phila Pa 1976)*. 2005;30(4):455–461.
  15. Sanders JO, Polly DW Jr, Cats-Baril W, et al; AIS Section of the Spinal Deformity Study Group. Analysis of patient and parent assessment of deformity in idiopathic scoliosis using the Walter Reed Visual Assessment Scale. *Spine (Phila Pa 1976)*. 2003;28(18):2158–2163.
  16. Delbecq AL, VandeVen AH. A group process model for problem identification and program planning. *J Appl Behav Sci*. 1971;7(4):466–492.
  17. Brown GK. A review of suicide assessment measures for intervention research in adults and older adults. Technical report submitted to NIMH under Contract No. 263- MH914950. 2002. <http://www.sprc.org/sites/sprc.org/files/library/BrownReviewAssessmentMeasuresAdultsOlderAdults.pdf>. Accessed March 7, 2016.
  18. Bowker AH. A test for symmetry in contingency tables. *J Am Stat Assoc*. 1948;43(244):572–574.
  19. Olfson M, Marcus SC, Bridge JA. Focusing suicide prevention on periods of high risk. *JAMA*. 2014;311(11):1107–1108.

**It is illegal to post this copyrighted PDF on any website.**