A. Significance

The burden of suicide in America is great. In 2017 more than 47,000 Americans took their lives.[1] Suicide is the tenth leading cause of death overall in the United States but the second leading cause of death among individuals aged 10-34 and the fourth leading cause of death among those aged 35-54.[12] The burden of suicide falls most heavily on young people and working-age adults. In 2013 the United States spent more than $50 billion on medical and lost productivity costs associated with suicides and suicide attempts, adjustment for underreporting increased the estimate to $93.5 billion.[13]

Gaps in health care leave vulnerable patients at risk. National efforts to curb the rising suicide rate such as the Zero Suicide Initiative[14] recognize that health care visits represent opportunities for suicide prevention. Most individuals who die by suicide make an outpatient health care visit in the year prior to their suicide death and almost half have a visit within a month of their death,[2] but traditional clinical suicide risk prediction is hardly better than chance.[15]

Statistical modeling can improve suicide risk prediction. Using machine learning, the Mental Health Research Network (MHRN) has developed highly accurate suicide risk prediction models that identify patients at risk for a suicide attempt or suicide death in the 90 days following an outpatient health care visit.[10] These models outperform previous suicide risk prediction models and screening instruments routinely used in clinical practice. Several MRHN health systems are planning to implement risk models over next 12-24 months.

However, many stakeholder concerns must be understood and addressed prior to implementation. Included in Table 1 are sample comments, concerns, and questions raised by health system administrators, clinicians, and model developers, collected during several conversations regarding the promise of using the MHRN suicide risk prediction models. Additional implementation and ethical considerations raised in the literature[5, 16, 17] include the need for effective program leadership supporting implementation; the additional burdens on clinicians introduced by these methods—as one article stated, *"history suggests that most forms of clinical decision support add to, rather than replace, the information clinicians need to process;"*[17] the need for clear communication and preparation of clinical staff; the lack of a set of best practices for implementing results from predictive modeling in clinical practice; the difficulty in communicating the complex meaning of statistical risk; clinical ambivalence about communicating suicide risk; the unknown effect of telling someone that they are at high risk of suicide; the lack of transparency in not communicating known risk; the opportunity for suicide risk information to be shared inappropriately; the issue of whether patients should consent to risk analysis or have the option to "opt out" or determine which elements of their data are allowable for risk calculation; the risk of patients feeling violated by health systems' efforts to protect them as these methods *"often involve the collection and combination of information of information in ways patients would never have imagined when freely providing the information for other purposes;"*[16] the possibility that model results may lead to involuntary loss of rights (i.e., forced hospitalization); the selection of appropriate post-identification interventions; and the decision to create a cut point that prioritizes for follow up care those patients whose risk exceeds the threshold, as a result creating a care potential disparity.

Table 1. Early Feedback from Health System Administrators, Clinicians, and Researchers

"Administrators are going to want to know how many people this is going to identify so they can start calculating costs and what they need to manage."
"What should the trigger level be? Top 5% at risk? Broader than that? Should moderate risk trigger any action on the part of the health system or clinician?"
"Would patients who are not insured members, for example those that show up in the emergency department, also trigger the algorithm?"
"Clinicians will want to know what risk factors resulted in identification of high risk, will they be able to see the list of underlying risk factors?"
"When an alert is fired, what is the clinician responsible for doing? And for documenting? And if they document mitigation strategies, does that change the future risk?"
"How will practice alerts fit into clinician workflows?"
"What should a clinician do if an alert suggests the patient is high risk but the patient has just completed a risk assessment that indicates low risk?"
"There is a lot of alert fatigue, how should this alert be prioritized vis a vis others?"
"What about a scenario where a patient comes in for a routine visit, an ankle sprain or something, and the alert fires and the clinician does whatever they are supposed to do to address the suicide risk but they ask the patient to return in a couple of weeks if the ankle isn't feeling better. Will the alert fire again at the second visit? How much will the underlying factors informing the suicide risk percentile have changed? If they have, is the model sensitive enough to pick that up? Will clinicians start to ignore it?"
"What if the risk is based on non-modifiable things or distant historical risks, will the patient keep being identified each visit? Can they be retired out of risk?"
"Can the alert be turned off by the clinician?"
"Should we consider this an acute care tool or a chronic care management tool?"
"Will patients be concerned that they are being monitored? Some patients will think this is a tool to monitor them for psychiatric re-hospitalizations and then they might avoid coming in for care."

Implementation Science frameworks can help optimize the application of suicide risk prediction models in clinical settings. To maximize the likelihood that suicide risk prediction models are implemented in a thoughtful manner that does not harm patients, supports clinical care management, and is sustainable for

health care delivery systems, it is important to study the broader implementation context in which these models will be used. The Consolidated Framework for Implementation Research (CFIR)[11, 18] offers a pragmatic and systematic structure for assessing factors that may be important to implementation success. The CFIR focuses on five ecological domains: 1) intervention characteristics, 2) outer setting, 3) inner setting, 4) characteristics of individuals involved in implementation, and 5) the implementation process. For example, features of the outer setting such as the escalating suicide rate in the U.S. over the past two decades and the recent national focus on suicide prevention may influence adoption of suicide risk prediction models. Features of the inner setting such as leadership support for participation in the national Zero Suicide initiative, local suicide prevention policies, and established care processes for assessing and addressing suicide risk may positively influence clinicians' receptivity to new suicide risk prediction approaches, whereas overburdened schedules, the expectation of brief medical visits, and poorly planned workflows associated with risk identification may have a negative influence. Lack of knowledge of relative risk and how to communicate risk to patients and lack of understanding best practices for suicide prevention are individual characteristics that may contribute to clinician ambivalence or reluctance to use suicide risk prediction tools. A more nuanced understanding of the implementation context at various levels will result in better preparation to scale up suicide risk prediction models in clinical settings.

3.B. Innovation

Our proposed work is innovative in the following ways:

- Engaging stakeholders to study important implementation factors *prior to* broad scale up of suicide risk prediction models is innovative. Aside from the ongoing prevention study in the VA (REACH VET),[5] we are unaware of any studies of the implementation of suicide risk prediction models in clinical care settings.
- The proposed research will include one site planning outreach independent of visits, one site planning delivery of risk scores at the point of care, and one considering the best implementation approach.
- Importantly, in addition to organizational and clinician feedback, our study adds patients' perspectives which have not been previously described.
- This research will not just influence implementation of existing suicide risk models but will reach upstream and influence the development of future prediction models. Decisions regarding design and modeling methods and implementation processes must be driven by stakeholder requirements.

This research will identify stakeholder priorities and requirements, establish research-practice partnerships to disseminate and implement evidence-based services (NIMH Strategic Objective 4.2.), and evaluate how patient, provider, and organizational factors affect intervention effectiveness (NIMH Strategic Objective 3.3c).

C. Approach

C.1 Overview

We propose a pre-implementation pilot study to explore health system administrators', clinicians', and patients' expectations, concerns, suggestions regarding, and experiences with the use of suicide risk prediction models in clinical settings. At one site considering implementation of the suicide risk models in the next year, interviews will focus on perceived benefits and risks associated with automated risk identification. Two other sites with imminent plans to conduct small implementation pilot studies during the study period will afford opportunities to study the actual experience of different implementation approaches.

C.2 Preliminary Studies

The study team has substantial expertise in all aspects of the proposed work and an extensive history of previous collaborations demonstrated through the research studies described below.

Suicide Risk Modeling. With a database including approximately 20 million visits by 3 million patients, Dr. Simon led MHRN researchers to develop and validate logistic regression models predicting risks of suicide attempt and death in the 90 days following an outpatient visit with a coded mental health diagnosis. Separate models were fit for primary care and mental health specialty visits. The models have accuracy rates ranging from 83% to 86%,[10] exceeding similar published models relying on EHR data[3, 6-9] and outperforming risk stratification based on Patient Health Questionnaire (PHQ-9) item 9 scores. This team is currently extending data collection through 2016, expanding sampling to include emergency department and inpatient visits, and adding additional predictors, including variables related to opioid use.

Suicide Prevention in Health Systems. Dr. Simon and Dr. Yarborough are currently engaged in an evaluation of Zero Suicide (ZS), a systems-level suicide prevention framework and set of evidence-based interventions and strategies designed to close gaps in health care that leave at-risk patients vulnerable. To date, no large-scale evaluation of ZS has been conducted so it is unclear which specific ZS components, bundle of components, or process of care strategies are most effective. As embedded researchers and members of our local ZS implementation teams, we are able to accurately describe suicide care processes, understand implementation context, and measure quality improvements. Further, Dr. Simon serves on Kaiser Permanente's national suicide prevention lead team. In the ZS study, certain sites serve as 'intervention' sites for specific ZS components while others serve as 'controls' allowing a pragmatic interrupted time-series analysis of suicide outcomes. The ZS evaluation takes place across six healthcare systems.

Qualitative and Implementation Research with Health System Leaders, Clinicians, Patients. Our team has used the proposed qualitative methods in several studies assessing quality improvement efforts and while conducting process and implementation evaluations with administrators, clinicians, and patient stakeholders. For example, Dr. Yarborough's qualitative team interviewed 30 health system leaders and clinicians across four sites about clinician beliefs, behaviors, and perceived health system/organizational barriers and facilitators of preventive care for patients with serious mental illnesses;[19] 163 patients were also interviewed.[20] In another study of first episode psychosis care, Dr. Yarborough's team conducted and analyzed 15 interviews with health system leaders and clinicians in two of the sites involved in this proposal, and 22 interviews with patients and family members.[21] In both of these studies 100% of planned health system leader and clinician interviews were conducted using the same procedures proposed here. Dr. Rossom also has experience using qualitative or mixed-methods to study implementation factors relevant to collaborative care for depression.[22, 23] Our consultant, Rinad Beidas, a national leader in implementation science has conducted similar research (regarding implementation of firearm safety counseling) in two MHRN health systems.

C.3 Settings

The three health systems in this application (Kaiser Permanente Northwest [KPNW], Kaiser Permanente Washington [KPWA], and Health Partners [HP]) are part of the Health Care Systems Research Network (HCSRN) and the NIMH-funded Mental Health Research Network (MHRN). These sites provide comprehensive medical and psychiatric specialty care to defined populations of healthcare system members who are enrolled through employer-sponsored or individual insurance plans, or capitated Medicaid or Medicare programs. Members are representative of their respective regional populations in terms of age, sex, and race/ethnicity. Each of these health systems maintains an electronic health record (EHR) database that captures members' demographic, enrollment, and clinical/diagnostic information, prescription dispensing data, internal service utilization, external health care claims data, and mortality data (via linkage to state death data and/or the National Death Index). Within each health system, EHR data are organized into a research virtual data warehouse (VDW) with data specifications and formats that are shared across the health care systems. These are the data accessed by the MHRN suicide risk prediction algorithms.

The participating sites were selected on the basis of their expressed interest in rapid implementation of the MHRN suicide risk prediction models. KPNW is awaiting deployment of a national Epic tool that will leverage the MHRN models. They plan to deploy this tool as part of their Zero Suicide initiative, first in behavioral health clinics before broader deployment in other departments (e.g., primary care, emergency medicine). KPNW is interested in the outcomes of this research as they consider different approaches to using these models and determine whether they can build a business case for a population health approach. KPWA has decided to conduct a single-clinic (behavioral health) implementation pilot of a visit-based alert in the medical record that will notify clinicians of patients with a suicide risk score in the top 2%. At HP a team of behavioral health care managers will use suicide risk models to identify high risk patients for outreach. Variability in implementation plans across these sites allows for a broad range of clinician and patient experiences and deep understanding of the factors that influence different implementation approaches. This is not a study of pre-implementation research for an intervention we seek to promote in health systems; rather, health systems are leading the way and we seek to maximize learning from their experiences. No funding is requested for actual implementation of the risk models but merely for evaluation of the different implementation approaches and processes.

C.4 Aim 1: Identify administrator and clinician expectations for or experiences with suicide risk prediction tools, any current or future implementation plans, and clinician preferences for risk tool development.

There are very few published descriptions of the experiences of health system stakeholders implementing suicide risk prediction models in clinical settings and these are limited to patients receiving care in the VA.[4, 5, 16] Conducting interviews with administrators and clinicians in non-Veteran health care settings at different stages of pre-implementation and pilot implementation will reveal examples of how health systems' plan to deploy suicide risk prediction models, clinicians' preferences for risk prediction tools, and the extent to which guidance about the appropriate uses, strengths, and limitations of predictive modeling is needed. Preliminary discussions with administrator and clinician stakeholders suggest a need for appropriate expectation-setting.

C.4.1 Population, Recruitment, Sample Size

We propose to interview six administrators (≈ two per site) with interest in suicide prevention and/or authority to implement risk models. Interviewees will be identified by the site PIs, all of whom are embedded clinician-researchers with ability to identify relevant interview candidates. Interviewees may also nominate other key informants (snowball sampling[24]). Additionally, we propose to identify approximately 30 clinicians (10 per site) to interview about their concerns and hopes for how suicide risk models would be deployed (KPNW) or about their experiences using a visit-based alert triggered by risk models (KPWA) or a patient registry for outreach to high risk patients (HP). At KPNW we will randomly select clinicians from behavioral health, primary care, and emergency medicine as these departments are potential implementation settings. At KPWA, we propose to interview clinicians who are expected to interact with the visit-based alert during the implementation pilot study; clinicians may have the option to opt in or out of participating in the implementation pilot. In that case, we will attempt to recruit those who do and do not participate, and we will seek to understand motivation and reservations about participation. Another important selection criterion at KPWA is based on patient risk. The proposal to identify the top 2% at highest risk means that more than half of patients flagged would likely have already been identified through existing screening processes (i.e., PHQ-9 item 9). It will be important to identify and focus some interviews on clinicians who interacted with patients identified by the risk flag who were <u>not</u> identified by other means—that is, patients who would have been missed if it weren't for the risk flag. These interviews will help us understand how results from the risk models are used in unexpected scenarios and how that affects the patient visit, risk conversation, and decisions about follow up. At HP, we will seek to sample the behavioral health care managers with the most experience doing these specific outreach calls. At both KPWA and HP, we will pace the interviews so that some are conducted at the beginning of each site's pilot and others toward the end of the interview window. By doing so, later experiences can be compared to earlier impressions to examine how attitudes and concerns change over time after experience having risk conversations with patients. We may interview willing clinicians at two different time points. We will first email or staff message clinicians with study information and follow up by email and phone to recruit those who do not respond.

C.4.2 Interviews, Interview Guide

Based on our prior experience interviewing busy health system administrators and clinicians, we plan for a 20-minute phone interview, with flexibility to follow up by email when needed or requested. We will send a study information sheet in advance so those who are able to view it prior to the interview understand the context and goals and can think ahead about the content they would like to prioritize in the short time available.

Guided by CFIR domains, we will ask administrators to describe their familiarity with risk prediction models, the value of suicide risk prediction models specifically (CFIR domains: characteristics of individuals, inner setting), their vision and leadership support for their health system's use of these models and how they will be implemented (CFIR domain: process). For example, in which departments/populations the models would be applied and why; whether they have considered a threshold for identification (e.g., top 2% at risk) and how they determined that threshold; what follow up of high risk patients would entail and how it would affect existing workflows; what, if any, preparation, communication, and training they plan for clinicians; and what concerns they have about burdening clinicians with additional risk assessment and safety planning work (CFIR domains: process, inner setting).

Clinician interviews will assess how well clinicians understand the risk information presented to them (CFIR domain: characteristics of individuals). For example, setting a threshold at the top 2% corresponds to a 6% average risk of suicide attempt in the following 90 days. Even though these patients are at the highest risk, the actual risk is relatively low and should not be overstated. If a clinician concluded "My patient is in the top 2% of risk, maybe they should be in hospital" that could be a potentially harmful inference. We will also assess broad familiarity with how models derived from machine learning produce risk estimates (CFIR domain: characteristics of individuals). This is important because concerns have been raised in the literature that the "black box" nature of these models that *"generate insights via unobservable methods"* could impede uptake of

these tools into practice.[17] It will be important to assess whether clinicians expect to see predictors that put their patients at risk and whether not having that information is problematic (CFIR domain: intervention characteristics). We will also assess clinicians' expectations for using the risk tools over time. For example, some patients may remain at high risk due to non-modifiable characteristics (e.g., middle-aged male, history of suicide attempts) and it will be important to understand if and when clinicians begin to ignore risk indicators for such patients. It will be equally important to understand health systems' expectations for how these patients are managed over time. We are interested in how suicide risk will be communicated to patients, particularly when patients are not self-identified at risk, are not seeking suicide prevention care at the visit, report conflicting information (e.g., patient is identified high risk but reports low risk behaviors or denies suicide ideation), or when they have difficulty understanding the complexity of statistical risk. Interviews will include a discussion of clinician preferences for implementation of risk models and associated processes and tools (CFIR domain: intervention characteristics), and whether or not they feel they have adequate post-identification interventions and resources at mitigate risk. We will explore concerns expressed in our preliminary stakeholder discussions (see Table 1) such as whether these tools threaten to overwhelm clinicians with more risk assessment and safety planning work, how alert fatigue may influence clinicians' receptivity to these tools and what would cause them to respond or ignore an alert, whether these tools are perceived as undermining clinician autonomy, whether clinicians are concerned about a false sense of security from low risk scores, and whether they feel ambivalent about risk discussions or use of machine learning to identify risk without patients' knowledge/consent. Interviews at KPWA and HP will elicit clinicians' actual experiences using the visit-based alert or patient registry and will inquire about their own experiences and their perceptions of patients' experiences of risk conversations. Interviews with administrators and clinicians will conclude with opportunity for interviewees to express concerns about liabilities that might accompany use of risk prediction tools.

C.5 Aim 2: Interview at-risk patients regarding preferences for or experiences with health system-initiated suicide risk discussions based on suicide risk prediction algorithms.

Currently there are no descriptions in the peer-reviewed literature of how patients have experienced discussions with a clinician about their suicide risk following identification by a suicide risk prediction model. Thus, the goal of Aim 2 is to interview patients to understand the potential benefits and risks of being identified as having, and having conversations about, suicide risk.

C.5.1 Population, Recruitment, Sample Size

Across the three sites we will seek to interview 60 patients (20 per site). The goal will be to identify a representative sample of higher risk patients, including those for whom traditional risk factors (e.g., previous suicide attempts, high PHQ-9 item 9 response) are not apparent and who might be surprised to learn they are identified as high risk. At KPNW where a suicide risk model intervention is aspirational, patients who could be expected to be identified by the suicide risk prediction models will be interviewed about the hypothetical use of these models in clinical care. At KPWA and HP, patients who were exposed to each system's intervention will be eligible for interviews. We will attempt to balance the samples, across the sites, on age, sex, and race/ethnicity in an effort to get as representative a sample as possible. We will attempt to recruit and interview participants within a two-week window following a visit where they were identified as high risk by the suicide risk models. Potentially eligible participants will be sent a recruitment letter/email providing details about the study including a number to call to indicate interest in participation. We will phone individuals who do not respond, making up to 8 attempts at different times of the day on different days of the week over a 2-3 week period. We have successfully used this recruitment strategy in several studies and expect to send about 180 letters to achieve our recruitment goal. Patients will receive a $50 gift card in gratitude for their participation.

C.5.2 Interviews, Interview Guide

We plan for a 60-minute phone or in-person interview with patients. Semi-structured interview guides will be developed by the team at KPNW. The interview guide will be organized using a funnel approach[25] and will explore patients' perceived value of risk prediction models generally, and suicide risk prediction models specifically. Questions will include, for example, "In health care, risk prediction includes using patients' data from the electronic medical record to predict which patients might need additional care or prevention. What do you think about computers predicting a person's risk of suicide?" "How would you feel about your health information being used to calculate a suicide risk score?" "If your doctor told you that you were at high risk for suicide, what other information, if any, would you like to know from him/her?" "How might that affect the clinical visit or your relationship with your doctor?" "Would you rather your doctor alert you to this risk or would you be comfortable receiving a call from another health professional tasked with following up with high risk patients?"

"Do you think this kind of outreach is caring and compassionate or creepy and intrusive?" We will explore concerns about risk information being released (e.g., to employer, insurance) or used to justify coercive or restrictive treatments. The interview guides employed at KPWA and HP will ask similar questions but will be anchored to the patient's recent experience with their clinician and the actual conversation which took place regarding their suicide-related risks. We will ask participants about any conversations regarding suicidal ideation or risks with clinician(s), how those conversations happened, any outcomes, and if they have any suggestions for how to improve those discussions. Interviews will be conducted by masters' level trained interviewers with significant experience working with individuals with serious mental illnesses and screening for suicide risk. All individuals identified with suicidal ideation at the time of the interview will be screened using the Columbia Suicide Severity Rating Scale (C-SSRS). Interviewers will follow suicide risk assessment and care facilitation protocols already implemented at all three sites in the MHRN Suicide Prevention Outreach Trial.

C.6 Analyses

Interviews will be recorded using HIPAA-compliant encrypted recorders and transcribed verbatim. Transcripts will be reviewed for accuracy by the qualitative analytic team (led by Dr. Yarborough) who will read each transcript independently then meet to develop initial descriptive codes.[26-28] The analytic team will begin by using open coding[27] to capture important concepts that emerge spontaneously from participant reports or that were not anticipated by the questionnaire domains. Thereafter the team will meet regularly to discuss new emerging codes, and to discuss important themes for development of the coding scheme. Once all team members agree on the full set of descriptive codes, and no new codes emerge, a final coding dictionary will include definitions and example text. Primary coding of the transcripts, using Atlast.ti.,[29] a qualitative software program, will be conducted; all transcripts will be double-coded by an independent coder in order to assess reliability. We will aim for ≥ 85% agreement between the two coders; if agreement is below this level we will review and clarify code definitions. Once all interviews are coded, we will use the query feature in Atlas.ti to produce reports of text associated with codes. We will use thematic analysis[30, 31] to identify congruent and discordant patterns in administrator, clinician, and patient transcripts.

C.7 Anticipated Outcomes. Aim 1 interviews will help us better understand how health systems plan to utilize suicide risk prediction models; how well administrators and clinicians understand appropriate uses, strengths, and limitations of predictive modeling of suicide risk; and any unanticipated or unintended consequences that occurred at the pilot implementation sites. Results from Aim 1 interviews will also summarize clinicians' preferences for how risk prediction tools (outreach vs. point-of-care) should be developed for use and we will return this information to the health system administrators to inform their decision making. Aim 2 interview results will summarize patients' preferences for being notified of their suicide risk, including how they would/not like to be notified, by whom, and situations where risk notification would be unwelcomed. As with Aim 1 clinician preferences, these results will be returned to the health system administrators to inform their implementation plans.

C.8 Timeline.	Year 1				Year 2				Year 3	
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2
IRB agreements, project start up, finalize interview guides										
Administrator, clinician, patient interviews										
Interview coding										
Qualitative analysis										
Manuscript preparation, dissemination activities, project close-out										