

A Stakeholder-Informed Ethical Framework to Guide Implementation of Suicide Risk Prediction Models Derived from Electronic Health Records

Bobbi Jo H. Yarborough & Scott P. Stumbo

To cite this article: Bobbi Jo H. Yarborough & Scott P. Stumbo (2023) A Stakeholder-Informed Ethical Framework to Guide Implementation of Suicide Risk Prediction Models Derived from Electronic Health Records, Archives of Suicide Research, 27:2, 704-717, DOI: [10.1080/13811118.2022.2064255](https://doi.org/10.1080/13811118.2022.2064255)

To link to this article: <https://doi.org/10.1080/13811118.2022.2064255>



Published online: 21 Apr 2022.



Submit your article to this journal [↗](#)



Article views: 347



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 4 View citing articles [↗](#)



A Stakeholder-Informed Ethical Framework to Guide Implementation of Suicide Risk Prediction Models Derived from Electronic Health Records

Bobbi Jo H. Yarborough and Scott P. Stumbo

ABSTRACT

Objective: Develop a stakeholder-informed ethical framework to provide practical guidance to health systems considering implementation of suicide risk prediction models.

Methods: In this multi-method study, patients and family members participating in formative focus groups ($n=4$ focus groups, 23 participants), patient advisors, and a bioethics consultant collectively informed the development of a web-based survey; survey results ($n=1,357$ respondents) and themes from interviews with stakeholders (patients, health system administrators, clinicians, suicide risk model developers, and a bioethicist) were used to draft the ethical framework.

Results: Clinical, ethical, operational, and technical issues reiterated by multiple stakeholder groups and corresponding questions for risk prediction model adopters to consider prior to and during suicide risk model implementation are organized within six ethical principles in the resulting stakeholder-informed framework. Key themes include: patients' rights to informed consent and choice to conceal or reveal risk (autonomy); appropriate application of risk models, data and model limitations and consequences including ambiguous risk predictors in opaque models (explainability); selecting actionable risk thresholds (beneficence, distributive justice); access to risk information and stigma (privacy); unanticipated harms (non-maleficence); and planning for expertise and resources to continuously audit models, monitor harms, and redress grievances (stewardship).

Conclusions: Enthusiasm for risk prediction in the context of suicide is understandable given the escalating suicide rate in the U.S. Attention to ethical and practical concerns in advance of automated suicide risk prediction model implementation may help avoid unnecessary harms that could thwart the promise of this innovation in suicide prevention.

HIGHLIGHTS

- Patients' desire to consent/opt out of suicide risk prediction models.
- Recursive ethical questioning should occur throughout risk model implementation.
- Risk modeling resources are needed to continuously audit models and monitor harms.

KEYWORDS

Ethical framework; implementation; risk models; suicide prevention

INTRODUCTION

Given the escalating suicide rate in the U.S. (Hedegaard, Curtin, & Warner, 2020), there has been a focus on suicide risk detection using novel methods including risk prediction models derived from electronic health records (EHR) data (Barak-Corren et al., 2017; Kessler et al., 2015, 2017, 2020; McCarthy et al., 2015; Simon et al., 2018; Su et al., 2020; Tran et al., 2014; Walsh, Ribeiro, & Franklin, 2017; Walsh, Ribeiro, & Franklin, 2018). Many of these models boast high accuracy in identifying patients at-risk for suicide—some models report classification accuracy (C-statistic) above 80%. Consequently, interest in their use in health care has expanded. Whether this innovation will translate to reduced suicide rates is unknown, but a growing body of literature calls for an anticipatory examination of ethical concerns relevant to risk model implementation (Linthicum, Schafer, & Ribeiro, 2019; McKernan, Clayton, & Walsh, 2018; Tucker, Tackett, Glickman, & Reger, 2019). Ethical considerations of suicide risk prediction may be distinct from other realms of risk prediction, given the potentially serious consequences (e.g., death, inappropriate treatment, stigma) that could result from misclassification or even from well-intended intervention.

Previous scholarship has articulated advantages and concerns about the expanded use of predictive modeling (Cath, 2018; Goldstein, Navar, Pencina, & Ioannidis, 2017; Joyce & Geddes, 2020; Lawrie, Fletcher-Watson, Whalley, & McIntosh, 2019; Nundy, Montgomery, & Wachter, 2019; Obermeyer, Powers, Vogeli, & Mullainathan, 2019; Walsh et al., 2020), including concerns specific to suicide risk prediction (Belsher et al., 2019; Fonseka, Bhat, & Kennedy, 2019; Linthicum et al., 2019; Whiting & Fazel, 2019). Generally, for the purposes here, we are referring to health systems' interest in implementing suicide risk models to identify higher risk patients for appropriate risk assessment, referral, and treatment though risk models could also be used to triage higher risk patients to receive suicide prevention or mental health specialty services faster relative to lower risk patients or for actuarial purposes. One concern about risk models in general is that they can be applied for purposes other than those for which they were developed. Calls have been issued to develop quality measures to ensure risk models are implemented effectively, fairly, and safely in suicide prevention efforts (Eaneff, Obermeyer, & Butte, 2020). There has also been acknowledgement of the need to shift the emphasis of prognostic model research closer to the clinical setting and address questions of practical applicability (Whiting & Fazel, 2019) and to better understand the perspectives of clinicians, individuals with suicide behavior, and familial survivors of suicide (McKernan et al., 2018). The goal of this paper is to present a concise, ethical framework, inclusive of these stakeholder voices not typically represented in implementation planning and absent in previous scholarship.

Early suicide risk model implementation (McCarthy et al., 2015; Reger et al., 2019) and increasing interest in broader deployment has created an urgent need for a practical resource focused on the ethical application of suicide risk models. To date, only the Veteran's Administration has described their consideration of ethical principles prior to suicide risk model implementation (Reger et al., 2019; Tucker, Tackett, Glickman, & Reger, 2019); they did not collect or consider empirical input from important stakeholders, including the patients upon whom the intervention would be conducted. Furthermore, with increasing COVID-19 pandemic-related suicide concerns

(Czeisler et al., 2020), and consequent pressure on health systems to deliver responsive suicide risk screening, an ethical framework could reduce the potential for harms from automated risk identification, keeping the safety and welfare of patients and their trust at the center of health care decision making.

METHOD

Kaiser Permanente Northwest (KPNW) and Washington (KPWA) were the settings. These systems serve approximately 1.3 million members in Oregon, Washington, and Idaho enrolled through individual or employer-sponsored insurance, Medicaid or Medicare, and subsidized low-income programs. Members are representative of each system's service area in age, race/ethnicity, and socioeconomic status. Both settings were engaged in systemic suicide prevention initiatives including implementation of the Zero Suicide Framework (Education Development Center Inc). KPNW was planning future implementation of a suicide risk prediction model developed by the Mental Health Research Network (MHRN) (Simon et al., 2018) and, before proceeding, was awaiting results of a KPWA pilot implementation of the same model. The authors were not involved in implementation planning but received an administrative supplement to use qualitative methods to explore and identify stakeholders' concerns and preferences related to this innovative method of suicide risk prediction and prevention and to develop and disseminate a stakeholder-informed ethical framework to guide the implementation of future risk prediction models. Focus group data, a large patient survey, and stakeholder interviews were collected as part of this funded project and triangulated to provide the foundation for the ethical framework. The KPNW Human Subjects Review Board approved all study activities; all participants provided informed consent.

Focus Groups and Survey

Four two-hour focus groups in December 2019 and January 2020 were part of a formative process to discover and prioritize concerns and topics of importance to patients. Data from these focus groups were analyzed independently to provide qualitative data for this study and were also used to develop a subsequent survey about the use of suicide risk prediction models. Focus group participants were recruited from a random sample, drawn from electronic health record data, of 500 adult KPNW members with past-year suicide ideation (assessed by PHQ-9) or a past-year suicide attempt (assessed by ICD-10 codes); a subset of participants were family members nominated for recruitment by patient focus group participants. Patients received an email inviting them to participate and acknowledged their interest by reading and completing a short REDCap (Harris et al., 2009) survey prior to arriving for the focus group. Participants were given a \$50 gift card in appreciation for their participation.

Focus group findings were used to generate a survey sent to adult KPNW members with and without reported past-year suicide ideation. The survey was emailed to >11,000 Kaiser members half with the same indication of past-year suicide ideation or attempt as for the focus group sample and half randomly selected adult members not meeting those conditions. Focus group questions and survey items, and the recruitment

process were vetted by two patient/consumer advisory boards. Focus group and survey methods and detailed survey results are described elsewhere (Yarborough & Stumbo, 2021), cited throughout this paper, and provide another important source that informs the framework.

Stakeholder Interviews

Stakeholders included KPNW and KPWA administrators with authority to implement suicide prevention initiatives, KPNW clinicians who responded to an email recruitment solicitation (with the behavioral health administrator's imprimatur), and MHRN suicide risk model developers, a bioethicist, and a patient advocate recruited for their work in this field. Interviews were audio recorded and transcribed verbatim; transcripts were loaded into Atlas.ti. (Frieze, 2018). Codes were mostly deductive, based on anticipated ethical issues or focused questions regarding consent, implementation concerns, privacy, etc. Inductive codes, based on novel issues brought up by participants, included domains such as potential misapplication and the need for further study to assess effectiveness. Queries derived from these codes were reviewed by two authors (BJY, SS) for this paper. The authors met to discuss emergent themes and wrote preliminary summaries of themes across the interviews.

Finally, we used all findings—patient focus group results, patient survey results, and stakeholder interviews—to describe from multiple perspectives the landscape of ethical issues for consideration prior to and during implementations of statistical suicide risk prediction models. The addition of patient and clinician perspectives is a novel addition to the literature.

RESULTS

Focus groups included 23 participants, 1,357 individuals responded to the survey, and two administrators, four clinicians, three suicide risk model developers, a bioethicist, and a patient advocate were interviewed. Ethical concerns are detailed below, organized within six domains, along with demonstrative scenarios from the KPWA pilot and illustrative quotes from the stakeholders. The ethical framework (Table 1) poses questions for consideration by risk model adopters prior to and during implementation.

Autonomy

Informed Consent

Patients are aware that their health care data is analyzed to make decisions about individual- and population-level services, they understand that suicide risk models would access their health information, that there is a level of imprecision inherent in risk estimation, and that risk identification could prompt intervention. In general, patients are supportive of EHR-derived suicide risk models and understand their potential value but prefer to have the option to consent to or opt out of this use of their data (Yarborough & Stumbo, 2021). Patient focus group participants understood that consent is not typically sought prior to predicting other health conditions (e.g., cardiovascular disease), and

TABLE 1. Ethical framework to guide implementation of suicide risk prediction models.

Ethical Principles		Questions Suicide Risk Prediction Model Adopters Need to Consider	
Issues			
1. Autonomy	Informed Consent	How will the health system exercise its obligation to ensure patients have a clear understanding of the types of predictors that will be accessed from their data, or what the inclusion of that data means for benefits and risks associated with a suicide risk identification dataset and the alternatives?	
	Patient Choice to Conceal or Disclose Suicide Risk	How can patients' rights to autonomy (e.g., to exercise choice to participate or be excluded from suicide risk identification, to conceal or disclose suicide behavior) be balanced with the potential lost opportunity to identify at-risk individuals who might opt out?	
2. Explainability	Application of Suicide Risk Models in Populations They Were Not Developed For	How will the health system operationalize informed consent/opt-out in a manner that fosters trust? On what population was the suicide risk prediction model developed and validated?	
		In which populations is implementation appropriate?	
	Limitations of Electronic Health Records Data	How should subgroup representation in the development/ validation datasets inform implementation decisions? Are there subgroups that might not benefit as much as others or could be disproportionately harmed?	
		What kinds of predictors are in the model? What known risk factors are <i>not</i> included because they are not adequately documented in health records?	
		How does the health system ensure that clinicians understand this? How will the health system educate clinicians and patients (when necessary) about what the risk scores mean and do not mean?	
3. Beneficence, Distributive Justice	Ambiguity of Risk Predictors Selecting Actionable Risk Thresholds	How will the health system update the model over time with improved capture of important risk factors (e.g., demographic, social determinants of health)?	
		How frequently does the model need to be run so that recently documented predictors (e.g., yesterday's suicide attempt) are captured?	
4. Privacy	Access to Risk Information & Stigma	How does the health system ensure that clinicians understand this?	
		How should clinicians use the risk score with patients, particularly if risk predictors (and interactions) are not discernable?	
	Risk Models Could Introduce Unanticipated Harms, Lead to Inappropriate Intervention, or Be Used to Deny Services	What is a reasonable threshold where the health system can expect to have sufficient resources to appropriately follow up with at-risk individuals?	
		What resources are available?	
		What resources will be (re-) allocated to support follow up?	
5. Non-maleficence	Risk Models Could Introduce Unanticipated Harms, Lead to Inappropriate Intervention, or Be Used to Deny Services	Is the goal to identify and follow up with the highest risk patients (targeted reach, high intensity intervention) or to prevent the most suicide attempts (broad reach, low intensity intervention)?	
		How, if at all, should risk identification information (i.e., risk score) appear in the electronic health record?	
6. Stewardship	Risk models will drift over time and require evaluation, maintenance, and recalibration	Who should it be visible to within the health system? Outside the system?	
		Should the risk score become part of the patient's permanent record (versus a transient flag that is not stored)?	
	Ongoing oversight	Which staff should be responsible for following up with patients identified high-risk?	
		How would the health system know if patients experienced stigma as a result of risk identification?	
		Have we thoroughly considered and mitigated possible harms?	
	Does the health system have resources allocated (i.e., dedicated staff with modeling expertise) to continuously monitor drift and recalibrate models as needed over time?	How will the health system monitor and respond to unintended consequences?	
		What would cause us to halt automated suicide risk identification?	
	What kind of governing board or ethical oversight committee does the health system need to review implementation plans, monitor appropriate use of the model, surveil harms, and maintain patient trust?	How will the health system measure interventions that would result from automated risk identification so that receipt of the interventions can become predictors in the model?	
		What kind of governing board or ethical oversight committee does the health system need to review implementation plans, monitor appropriate use of the model, surveil harms, and maintain patient trust?	

while some felt suicide risk should not be any different, others felt that because of the stigma associated with suicide behavior and the potential for surprise at being identified as at-risk, requiring permission prior to conducting surveillance was important. One participant noted:

Some people may not be comfortable with their personal data being used in these models. But probably a majority of people would be... I think if you were upfront in the beginning then it's okay. (patient, focus group)

Patients' Choice to Conceal or Disclose Risk

As with any intervention, health systems have an obligation to ensure that patients understand the benefits and risks associated with suicide risk identification. This might include educating patients as to which EHR data is discoverable and which predictors significantly contribute to risk identification. For example, scores from depression and suicide screening instruments significantly influence risk estimates. Absent prediction models, these are the basis for clinical risk assessment. Some patients felt leveraging existing data for automated suicide prediction was no different from, and more efficient than, their provider reviewing records for screening scores and other risk factors. However, patients know, when completing screening questionnaires, that how they respond will likely influence the care they subsequently receive. They maintain autonomy to respond honestly, conceal, or refuse to disclose; sometimes they conceal out of concern about loss of autonomy (Richards et al., 2019). Automated risk identification may limit autonomy when it supersedes patients' deliberate representations of their present risk.

We know this happens; they say they did not have thoughts about self-harm because they didn't want someone to follow-up. There was that autonomy of saying, I'm gonna say no, even though the answer is yes. With risk prediction models, that's not a choice people have. They can't change the answer, or they can't give an answer that they would prefer. The answer comes from a machine... Now, of course, there are people who argue that's an autonomy we don't want people to have. But I think that's at least something you need to explicitly consider. (clinician)

It is possible that if patients are made aware that their suicide screening scores feature prominently in risk estimates it could alter their subsequent behavior. If they become more likely to conceal then screening instruments may be less predictive over time. It is worth carefully considering how patients can exercise the choice to participate or be excluded from risk identification and how informed consent can foster trust.

In the KPWA pilot, patients are not informed that clinicians receive prompts based on a risk prediction model. Clinicians respond to a flag in the EHR just as they would if the patient had screened at-risk using the clinic's standard suicide screening measure, regardless of the screening score at the present visit. One administrator explained that proceeding with a risk assessment in the face of risk denial is common practice:

I do that all the time when I'm seeing patients. Their PHQ-9 is negative, but they were in the hospital two months ago for maybe a suicide attempt or something else. I'm gonna do the [risk assessment]. I didn't use the prediction algorithm for that and I'm not telling them why I did that. I'm just concerned for their safety and that's part of my job. (behavioral health administrator)

An ongoing implementation evaluation at KPWA, including interviews with patients, will help inform decisions about consenting patients and revealing use of the risk models.

Explainability

Suicide risk model developers have an obligation to interpret model performance characteristics so that adopters can make informed implementation decisions.

I have a very strong view that people who develop these models have an absolute obligation to complete transparency. If there's an equation it needs to be public. What's the positive predictive value, what's the sensitivity and so on. That said, many of the people making implementation decisions may not get into things at that technical level. So, I think it's the obligation of model developers to try to accurately and fairly present how accurate is this model and try to put that in terms that most people can understand. (statistical suicide risk model developer)

Adopters need to understand sensitivity/specificity tradeoffs as they relate to clinical implementation. Problems would arise if models overestimated risk and health systems unnecessarily diverted limited resources and potentially caused patients undue distress, just as they would result from underestimation that left vulnerable patients unidentified.

Application of Suicide Risk Models in Populations They Were Not Developed For

A known risk after the deployment of any prediction model is that its use could be extended to populations or situations for which it was not intended or tested. At KPWA, leaders were concerned that patients at high risk for suicide might have avoided care during the early months of the COVID-19 pandemic. Given the urgency of that concern, KPWA used prediction models developed for an outpatient mental health specialty clinic population to inform outreach efforts in the broader population. A model developer shared a concern:

What I'm actually most concerned about is that a large number of resources will be dedicated to this and clinicians will think we are doing a lot when maybe we're not doing as much as we think we're doing, because that's not what the model was built for. The predictors of the whole population may be different. And so, we're not actually catching everyone at risk. (statistical suicide risk model developer)

While not harmful to run the models in the larger population, it may not have been as helpful as anticipated. The model developers articulated their concerns and the leaders proceeded with reasonable expectations. This scenario demonstrates that adopters need to understand the sample parameters the model was trained on to tailor implementation for the conditions in which models will have the greatest impact.

In another example, risk models included suicide attempts and deaths by specific subgroups (e.g., African Americans, Native Americans), but the number of deaths was too small to get a precise prediction for the relationship between the risk factors and suicide death for those subgroups (Coley, Johnson, Simon, Cruz, & Shortreed, 2021). Adopters must be aware of these limitations because adjustments may need to be made to fairly implement risk models and some subgroups may not benefit as much as others.

Limitations of EHR Data

Adopters have an ethical obligation to educate themselves and clinicians interacting with the risk models about their limitations. Models derived from EHR data are only as accurate as their inputs; all the stakeholder groups recognized this shortcoming. Focus group participants expressed concern about risk factors that are not accurately or systematically recorded in the EHR (e.g., gender identity, recent divorce, financial distress). As documentation of demographic characteristics and social determinants of health in the EHR improve, risk models need to be updated. Patient survey respondents agreed social predictors were important to understanding suicide risk, but two-thirds opposed use of externally sourced data (Yarborough & Stumbo, 2021). Model developers underscored the importance of clinicians understanding which predictors are in the models and which are not, including those that might not make it into a prediction calculation due to timing. For example, a recent suicide attempt may not be represented in an estimate if the model was not refreshed between when the event was documented and when the patient was seen.

Patient focus group participants and clinicians were also concerned that EHR data reflects whatever is documented, which may be biased or incomplete. Several stakeholder groups recognized the potential for risk models to amplify socially constructed biases, perpetuate discrimination, or exacerbate disparities in care.

Ambiguity of Risk Predictors

The power of risk prediction models lies in the ability to detect patterns among combinations of correlated predictors interacting in non-linear relationships and to stratify risk based on a probabilistic score. Inscrutable statistical models are useful for prediction but have little value for explanation (Simon et al., 2021). Clinical assessment, on the other hand, can be justified by objective evidence, subjective inference, and can be explained. Clinicians and patients are accustomed to explanation. Adopters must consider whether and how clinicians will be expected to communicate risk scores to patients given that in some models predictors are not discernable. For example, in the KPWA implementation only a risk flag (binary indicator) that the individual may be at risk for suicide appeared in the EHR, clinicians were not notified of the risk factors that increased risk and some of those factors, even if there were identifiable, were immutable (e.g., age, gender, history of suicide attempt).

Beneficence, Distributive Justice

Selecting Actionable Risk Thresholds

A very important consideration for suicide risk model adopters is determining where to set the threshold for intervention. A cohort study that explored feasibility of suicide risk model implementation demonstrated how the work burden to clinicians depends largely on the population size, risk threshold selected, number of unique alerts (i.e., patients not already identified at-risk), and protocol for responding to alerts (Kline-Simon et al., 2020). Health systems have limited resources and statistical risk models produce continuous scores; adopters can select an absolute risk threshold (e.g., absolute risk $\geq 10\%$) or a percentile risk (e.g., ≥ 95 th percentile). The latter allows estimation of patients/visits expected to require follow up. In the KPWA pilot, patients exceeding the 95th percentile

for statistical risk are flagged, corresponding roughly to the amount that would be determined high risk by their existing clinical screening instrument, therefore not significantly increasing workload.

It is crucial that adopters and clinicians recognize that individuals below the threshold should not be considered “not at-risk” and those above the threshold may not be at imminent risk. For example, in the KPWA pilot, individuals above the 95th percentile had only an approximately 6% risk of suicide attempt in the 90 days following a visit (Simon et al., 2018). Adopters need to consider the volume of at-risk individuals their system is equipped to manage, determine appropriate interventions for various risk levels (an outstanding empirical question), and estimate the burden of resources required for various interventions. One approach is to intervene with the highest risk patients, another strategy involves delivering less intensive intervention(s) to a larger proportion at lower risk to maximize prevented suicide attempts. The goals of the suicide prevention program and the needed resources to meet them should influence these decisions. Stakeholders were keenly aware of how limited resources affect access to care and were concerned about adequate follow up:

What’s begging the question for me, is what are you gonna do with that information? ... I feel real cynical. We’ve identified you as a risk. Then what? (patient, focus group)

Privacy

Access to Risk Information, Stigma

Adopters should consider who will have access to risk information, how it will be displayed and/or stored in the EHR, and who will be expected to respond. Patients prefer only their trusted clinicians have access to suicide risk model results (Yarborough & Stumbo, 2021). Patients and clinicians voiced concerns about stigma within the health system:

I won’t talk to my primary doctor about it. I will not. I don’t have a counselor. But I don’t feel safe with my primary doctor to talk about my mental status. (patient, focus group)

I do believe that within primary care there can be different degrees of treatment based on the mental health diagnosis. Not everyone and not every time, but I do believe that sometimes that stigma of having a mental health diagnosis carries over into their care they may receive by nurses or by MAs [medical assistants] or whether or not people take their complaints or illnesses seriously. (clinician)

There were also concerns from a range of stakeholders about the risk score becoming part of the EHR and/or being shared outside the health system:

But that whole fiction of what happens in this room, I will keep things confidential ... You can’t provide that anymore in a learning healthcare system. You have to negotiate either keeping things out of the record, cause once you put it in the record everybody knows. But then also if you keep it out of the record then their potential benefits from a learning system, you’re then depriving a patient of those. (bioethicist)

The big fear is could it be used to deny people health insurance because they’re high [risk]. We would like to believe that most interventions would be protective or helpful or supportive, although they may not be perceived that way. You could think about discriminatory things outside the healthcare system, like denying people car insurance or life insurance or not being able to buy firearms. (clinician)

And from a focus group discussion between patient participants:

Participant 1: It makes me very uneasy. I mean, that might be the intent now, but I don't know down the road how that information might be used.

Participant 2: Could my insurance rates go up?

Participant 3: So, like employers, or law agencies, there's no sharing of that information?

Non-maleficence

Risk Models Could Introduce Unanticipated Harms, Lead to Inappropriate Intervention, or Be Used to Deny Services

In addition to potential harms introduced by their private information being shared, patients were very concerned that suicide risk models, particularly if used by clinicians without adequate training (generally referring to clinicians outside of specialty mental health), could prompt coercive or inappropriate treatment.

I would anticipate what would happen as a result of being identified as high risk for suicide that I would be... coerced. I will use that word. Coerced into a mind treatment. Yeah, because of legal concerns on [health system's] part, or any other healthcare provider. And possibly the moral bias of the person who contacts me. (patient, focus group)

And if they're misunderstood and misinterpreted, are we suddenly gonna start prescribing a million members to take anti-depressants? And now we have all these people on medications that they don't necessarily need to be on. But because of models that they were at high risk for suicide we need to push these drugs on them. (patient, focus group)

Model developers also expressed the importance of interventions being appropriately mapped to the risk and a concern that if clinicians did not understand relative risk or the low absolute risk of patients flagged as high-risk then patients could be elevated to inappropriate or unnecessarily higher levels of care. Some stakeholders expressed concern for the opposite—that risk scores could be used to ration health care or deny services:

My biggest fear in all of this is that we're [committed] to this predictive tool and then at the end they're like, well, the predictive tool says that your risk is actually very low so we're not gonna proceed [with treatment]. (patient, focus group)

It is imperative that adopters take time to thoroughly consider and mitigate any potential harms and consider systems that will monitor for and respond to unintended negative consequences.

Stewardship

Risk Models Will Drift Over Time and Require Evaluation, Maintenance, and Recalibration

Model developers and the bioethicist were especially concerned with making sure adopters understood that models drift over time. Adopters need to plan for expertise and resources to conduct ongoing assessment of whether the model continues to perform as intended, identifies risk appropriately, and ultimately reduces suicide outcomes while

not producing negative unintended consequences. Documentation of intervention receipt must also be considered:

Once the risk prediction model goes live, if an intervention is being done you can no longer assess the quality of your model based on the predictive performance because your hope would be that the intervention is working, and the individuals' risk is decreasing as a result. You would have worse performance if you just compared your predictive probability to observed outcomes after implementation. Prospective monitoring of a model and how you change it over time to respond to changes in how the data are selected as well as the responses to your intervention ... that's a huge issue. (statistical suicide risk model developer)

Ongoing Oversight

Finally, in the interest of anticipatory ethics, avoiding harms, and safeguarding patient trust, adopters should consider appointing a governing board or oversight committee, ideally with representation by each of the perspectives that have contributed to this framework. Adequate stewardship involves preliminary review and endorsement of implementation plans, continuous auditing, monitoring for harms, and authority to redress grievances.

DISCUSSION

EHR-derived risk prediction is a relatively new innovation in suicide prevention. As Morley et al. (Morley, Floridi, Kinsey, & Elhalal, 2020) note in their review of artificial intelligence ethics tools, the gap between principles and practice is large yet not impossible if the right questions are asked. Informed by stakeholder feedback, this ethical framework serves as a practical resource to help adopters discipline themselves to consider how to ethically implement suicide risk identification models. During the writing of this manuscript, an important paper was published acknowledging the potential for well-intended suicide risk prediction models to inadvertently perpetuate health disparities (Coley et al., 2021). The group of researchers who developed the MHRN suicide risk models undertook an examination to determine whether two risk models that estimated suicide death following an outpatient visit performed as accurately across race and ethnicities as they did across the whole population. The study demonstrated that implementation of either suicide risk model would disproportionately benefit certain subgroups compared to others. The authors concluded that health system stakeholders "must carefully consider disparities in benefits and harms posed when deciding whether and how to implement a prediction model." (Coley et al., 2021) These findings underscore the importance of an ethical framework that creates an intentional pause in the implementation process to take time to consider critical issues such as this. The framework questions should be considered by adopters before and throughout implementation in a recursive manner. Given that contextual and technical knowledge should inform implementation and given this study was conducted in only two health care systems considering or implementing one specific model, the generalizability may be limited. The framework is not meant to be exhaustive but rather to prompt deliberate and thoughtful consideration of consequential ethical issues that may be relevant in any specific context.

Enthusiasm for automated risk prediction in the context of suicide is understandable given the high personal and social costs of suicide, but the costs of proceeding with implementation without careful ethical consideration are also high, particularly if mistakes outweigh the benefits of success, reduce adoption of risk models for suicide prevention, or result in the severance of public trust.

ACKNOWLEDGEMENTS

The authors wish to acknowledge Ms. Leah Harris and Drs. Danton Char and Gregory Simon for their collaboration and assistance in interpreting focus group themes, developing and refining survey items, and interpreting survey results.

DISCLOSURE STATEMENT

No potential competing interest was declared by the authors.

FUNDING

This study was supported by the National Institute on Drug Abuse [DA047724].

AUTHOR NOTES

Dr. Bobbi Jo H. Yarborough, PsyD and Scott P. Stumbo, MA, Kaiser Permanente Northwest Center for Health Research, Portland, OR, USA.

Correspondence concerning this article should be addressed to Dr. Bobbi Jo H. Yarborough, PsyD, Kaiser Permanente Northwest Center for Health Research, 3800 N Interstate Avenue, Portland, OR 97227 USA. Email: bobbijo.h.yarborough@kpchr.org

DATA AVAILABILITY STATEMENT

Due to the nature of this research, participants of this study did not agree for their data to be shared publicly, so supporting data is not available.

REFERENCES

- Barak-Corren, Y., Castro, V. M., Javitt, S., Hoffnagle, A. G., Dai, Y., Perlis, R. H., ... Reis, B. Y. (2017). Predicting suicidal behavior from longitudinal electronic health records. *The American Journal of Psychiatry*, 174(2), 154–162. doi:[10.1176/appi.ajp.2016.16010077](https://doi.org/10.1176/appi.ajp.2016.16010077)
- Belsher, B. E., Smolenski, D. J., Pruitt, L. D., Bush, N. E., Beech, E. H., Workman, D. E., ... Skopp, N. A. (2019). Prediction models for suicide attempts and deaths: A systematic review and simulation. *JAMA Psychiatry*, 76(6), 642–651. doi:[10.1001/jamapsychiatry.2019.0174](https://doi.org/10.1001/jamapsychiatry.2019.0174)
- Cath, C. (2018). Governing artificial intelligence: Ethical, legal and technical opportunities and challenges. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376(2133), 20180080. doi:[10.1098/rsta.2018.0080](https://doi.org/10.1098/rsta.2018.0080)
- Coley, R. Y., Johnson, E., Simon, G. E., Cruz, M., & Shortreed, S. M. (2021). Racial/Ethnic disparities in the performance of prediction models for death by suicide after mental health visits. *JAMA Psychiatry*, 78(7), 726–734. doi:[10.1001/jamapsychiatry.2021.0493](https://doi.org/10.1001/jamapsychiatry.2021.0493)
- Czeisler, M. É., Lane, R. I., Petrosky, E., Wiley, J. F., Christensen, A., Njai, R., ... Rajaratnam, S. M. W. (2020). Mental health, substance use, and suicidal ideation during the COVID-19

- pandemic - United States, June 24–30, 2020. *MMWR. Morbidity and Mortality Weekly Report*, 69(32), 1049–1057. doi:10.15585/mmwr.mm6932a1
- Eaneff, S., Obermeyer, Z., & Butte, A. J. (2020). The case for algorithmic stewardship for artificial intelligence and machine learning technologies. *JAMA*, 324(14), 1397–1398. doi:10.1001/jama.2020.9371
- Education Development Center Inc. *Zero suicide in health and behavioral health care*. Retrieved from <https://zerosuicide.sprc.org/>
- Fonseka, T. M., Bhat, V., & Kennedy, S. H. (2019). The utility of artificial intelligence in suicide risk prediction and the management of suicidal behaviors. *The Australian and New Zealand Journal of Psychiatry*, 53(10), 954–964. doi:10.1177/0004867419864428
- Friese, S. (2018). *User's manual for ATLAS.ti 8.0*. Berlin: ATLAS.ti Scientific Software Development GmbH. Retrieved from http://www.atlasti.com/uploads/media/atlasti_v6_manual.pdf
- Goldstein, B. A., Navar, A. M., Pencina, M. J., & Ioannidis, J. P. (2017). Opportunities and challenges in developing risk prediction models with electronic health records data: A systematic review. *Journal of the American Medical Informatics Association : JAMIA*, 24(1), 198–208. doi:10.1093/jamia/ocw042
- Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009). Research electronic data capture (REDCap)-a metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics*, 42(2), 377–381. doi:10.1016/j.jbi.2008.08.010
- Hedegaard, H., Curtin, S. C., & Warner, M. (2020). Increase in suicide mortality in the United States, 1999–2018. *NCHS Data Brief*, (362), 1–8.
- Joyce, D. W., & Geddes, J. (2020). When deploying predictive algorithms, are summary performance measures sufficient? *JAMA Psychiatry*, 77(5), 447–448. doi:10.1001/jamapsychiatry.2019.4484
- Kessler, R. C., Bauer, M. S., Bishop, T. M., Demler, O. V., Dobscha, S. K., Gildea, S. M., ... Bossarte, R. M. (2020). Using administrative data to predict suicide after psychiatric hospitalization in the Veterans Health Administration System. *Frontiers in Psychiatry*, 11, 390. doi:10.3389/fpsy.2020.00390
- Kessler, R. C., Stein, M. B., Petukhova, M. V., Bliese, P., Bossarte, R. M., Bromet, E. J., ... Ursano, R. J., Army STARRS Collaborators (2017). Predicting suicides after outpatient mental health visits in the Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS). *Molecular Psychiatry*, 22(4), 544–551. doi:10.1038/mp.2016.110
- Kessler, R. C., Warner, C. H., Ivany, C., Petukhova, M. V., Rose, S., Bromet, E. J., ... Ursano, R. J., Army STARRS Collaborators (2015). Predicting suicides after psychiatric hospitalization in US Army soldiers: The Army Study To Assess Risk and Resilience in Servicemembers (Army STARRS). *JAMA Psychiatry*, 72(1), 49–57. doi:10.1001/jamapsychiatry.2014.1754
- Kline-Simon, A. H., Sterling, S., Young-Wolff, K., Simon, G., Lu, Y., Does, M., & Liu, V. (2020). Estimates of workload associated with suicide risk alerts after implementation of risk-prediction model. *JAMA Network Open*, 3(10), e2021189. doi:10.1001/jamanetworkopen.2020.21189
- Lawrie, S. M., Fletcher-Watson, S., Whalley, H. C., & McIntosh, A. M. (2019). Predicting major mental illness: Ethical and practical considerations. *BJPsych Open*, 5(2), e30. doi:10.1192/bjo.2019.11
- Linthicum, K. P., Schafer, K. M., & Ribeiro, J. D. (2019). Machine learning in suicide science: Applications and ethics. *Behavioral Sciences & the Law*, 37(3), 214–222. doi:10.1002/bsl.2392
- McCarthy, J. F., Bossarte, R. M., Katz, I. R., Thompson, C., Kemp, J., Hannemann, C. M., ... Schoenbaum, M. (2015). Predictive modeling and concentration of the risk of suicide: Implications for preventive interventions in the US Department of Veterans Affairs. *American Journal of Public Health*, 105(9), 1935–1942. doi:10.2105/ajph.2015.302737
- McKernan, L. C., Clayton, E. W., & Walsh, C. G. (2018). Protecting life while preserving liberty: Ethical recommendations for suicide prevention with artificial intelligence. *Frontiers in Psychiatry*, 9, 650–650. doi:10.3389/fpsy.2018.00650
- Morley, J., Floridi, L., Kinsey, L., & Elhalal, A. (2020). From what to how: An initial review of publicly available AI ethics tools, methods and research to translate principles into practices. *Science and Engineering Ethics*, 26(4), 2141–2168. doi:10.1007/s11948-019-00165-5

- Nundy, S., Montgomery, T., & Wachter, R. M. (2019). Promoting trust between patients and physicians in the era of artificial intelligence. *JAMA*, 322(6), 497. doi:[10.1001/jama.2018.20563](https://doi.org/10.1001/jama.2018.20563)
- Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science (New York, N.Y.)*, 366(6464), 447–453. doi:[10.1126/science.aax2342](https://doi.org/10.1126/science.aax2342)
- Reger, G. M., McClure, M. L., Ruskin, D., Carter, S. P., & Reger, M. A. (2019). Integrating predictive modeling into mental health care: An example in suicide prevention. *Psychiatric Services (Washington, D.C.)*, 70(1), 71–74. doi:[10.1176/appi.ps.201800242](https://doi.org/10.1176/appi.ps.201800242)
- Richards, J. E., Whiteside, U., Ludman, E. J., Pabiniak, C., Kirlin, B., Hidalgo, R., & Simon, G. (2019). Understanding why patients may not report suicidal ideation at a health care visit prior to a suicide attempt: A qualitative study. *Psychiatric Services (Washington, D.C.)*, 70(1), 40–45. doi:[10.1176/appi.ps.201800342](https://doi.org/10.1176/appi.ps.201800342)
- Simon, G. E., Johnson, E., Lawrence, J. M., Rossom, R. C., Ahmedani, B., Lynch, F. L., ... Shortreed, S. M. (2018). Predicting suicide attempts and suicide deaths following outpatient visits using electronic health records. *The American Journal of Psychiatry*, 175(10), 951–960. doi:[10.1176/appi.ajp.2018.17101167](https://doi.org/10.1176/appi.ajp.2018.17101167)
- Simon, G. E., Matarazzo, B. B., Walsh, C. G., Smoller, J. W., Boudreaux, E. D., Yarborough, B. J. H., ... Schoenbaum, M. (2021). Reconciling statistical and clinicians' predictions of suicide risk. *Psychiatric Services (Washington, D.C.)*, 72(5), 555–562. doi:[10.1176/appi.ps.202000214](https://doi.org/10.1176/appi.ps.202000214)
- Su, C., Aseltine, R., Doshi, R., Chen, K., Rogers, S. C., & Wang, F. (2020). Machine learning for suicide risk prediction in children and adolescents with electronic health records. *Translational Psychiatry*, 10(1), 413. doi:[10.1038/s41398-020-01100-0](https://doi.org/10.1038/s41398-020-01100-0)
- Tran, T., Luo, W., Phung, D., Harvey, R., Berk, M., Kennedy, R. L., & Venkatesh, S. (2014). Risk stratification using data from electronic medical records better predicts suicide risks than clinician assessments. *BMC Psychiatry*, 14, 76. doi:[10.1186/1471-244x-14-76](https://doi.org/10.1186/1471-244x-14-76)
- Tucker, R. P., Tackett, M. J., Glickman, D., & Reger, M. A. (2019). Ethical and practical considerations in the use of a predictive model to trigger suicide prevention interventions in healthcare settings. *Suicide & Life-Threatening Behavior*, 49(2), 382–392. doi:[10.1111/sltb.12431](https://doi.org/10.1111/sltb.12431)
- Walsh, C. G., Chaudhry, B., Dua, P., Goodman, K. W., Kaplan, B., Kavuluru, R., ... Subbian, V. (2020). Stigma, biomarkers, and algorithmic bias: Recommendations for precision behavioral health with artificial intelligence. *JAMIA Open*, 3(1), 9–15. doi:[10.1093/jamiaopen/ooz054](https://doi.org/10.1093/jamiaopen/ooz054)
- Walsh, C. G., Ribeiro, J. D., & Franklin, J. C. (2017). Predicting risk of suicide attempts over time through machine learning. *Clinical Psychological Science*, 5(3), 457–469. doi:[10.1177/2167702617691560](https://doi.org/10.1177/2167702617691560)
- Walsh, C. G., Ribeiro, J. D., & Franklin, J. C. (2018). Predicting suicide attempts in adolescents with longitudinal clinical data and machine learning. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 59(12), 1261–1270. doi:[10.1111/jcpp.12916](https://doi.org/10.1111/jcpp.12916)
- Whiting, D., & Fazel, S. (2019). How accurate are suicide risk prediction models? Asking the right questions for clinical practice. *Evidence-Based Mental Health*, 22(3), 125–128. doi:[10.1136/ebmental-2019-300102](https://doi.org/10.1136/ebmental-2019-300102)
- Yarborough, B. J. H., & Stumbo, S. P. (2021). Patient perspectives on acceptability of, and implementation preferences for, use of electronic health records and machine learning to identify suicide risk. *General Hospital Psychiatry*, 70, 31–37. doi:[10.1016/j.genhosppsych.2021.02.008](https://doi.org/10.1016/j.genhosppsych.2021.02.008)