

Original Paper

Developing a Suicide Risk Prediction Algorithm Using Electronic Health Record Data in Mental Health Care: Real-World Case Study

Linda Hummel^{1,2}, MSc; Karin C A G Lorenz-Artz^{1,2}, PhD; Joyce J P A Bierbooms¹, PhD; Inge M B Bongers¹, PhD

¹Tranzo Scientific Center for Care and Wellbeing, Tilburg School of Social and Behavioral Sciences, Tilburg University, Tilburg, The Netherlands

²GGzE Mental Health Care Institute, Eindhoven, The Netherlands

Corresponding Author:

Linda Hummel, MSc
Tranzo Scientific Center for Care and Wellbeing
Tilburg School of Social and Behavioral Sciences, Tilburg University
Prof. Cobbenhagenlaan 125, Reitse Poort, Room RP 204
Tilburg 5037 DB
The Netherlands
Phone: 31683662495
Email: l.hummel@tilburguniversity.edu

Abstract

Background: Artificial intelligence (AI) offers potential solutions to address the challenges faced by a strained mental health care system, such as increasing demand for care, staff shortages, and pressured accessibility. While developing AI-based tools for clinical practice is technically feasible and has the potential to produce real-world impact, only a few are actually implemented into clinical practice. Implementation starts at the algorithm development phase, as this phase bridges theoretical innovation and practical application. The design and the way the AI tool is developed may either facilitate or hinder later implementation and use.

Objective: This study aims to examine the development process of a suicide risk prediction algorithm using real-world electronic health record (EHR) data through a qualitative case study approach for clinical use in mental health care. It explores which challenges the development team encountered in creating the algorithm and how they addressed these challenges. This study identifies key considerations for the integration of technical and clinical perspectives in algorithms, facilitating the evolution of mental health organizations toward data-driven practice. The studied algorithm remains exploratory and has not yet been implemented in clinical practice.

Methods: An exploratory, multimethod qualitative case study was conducted, using a hybrid approach with both inductive and deductive analysis. Data were collected through desk research, reflective team meetings, and iterative feedback sessions with the development team. Thematic analysis was used to identify development challenges and the team's responses. Based on these findings, key considerations for future algorithm development were derived.

Results: Key challenges included defining, operationalizing, and measuring suicide incidents within EHRs due to issues such as missing data, underreporting, and differences between data sources. Predicting factors were identified by consulting clinical experts; however, psychosocial variables had to be constructed as they could not directly be extracted from EHR data. Risk of bias occurred when traditional suicide prevention questionnaires, unequally distributed across patients, were used as input. Analyzing unstructured data by natural language processing was challenging due to data noise, but ultimately enabled successful sentiment analysis, which provided dynamic, clinically relevant information for the algorithm. A complex model enhanced predictive accuracy but posed challenges regarding understandability, which was highly valued by clinicians.

Conclusions: To advance mental health care as a data-driven field, several critical considerations must be addressed: ensuring robust data governance and quality, fostering cultural shifts in data documentation practices, establishing mechanisms for continuous monitoring of AI tool usage, mitigating risks of bias, balancing predictive performance with explainability, and maintaining a clinician "in-the-loop" approach. Future research should prioritize sociotechnical aspects related to the development, implementation, and daily use of AI in mental health care practice.

Keywords: implementation science; artificial intelligence; prediction algorithms; electronic health records; mental health services; suicide prevention

Introduction

Advances in artificial intelligence (AI; see [Textbox 1](#) for a glossary of key terms) may provide promising opportunities for many actual and future challenges that mental health care faces [1], such as increasing demand for care and staff

shortages that affect the quality of care and accessibility of mental health care. AI may enhance multiple domains, such as triage and screening, diagnostics, clinical decision-making, care delivery, chronic care management, self-care, and prevention of (psychiatric) symptoms [2].

Textbox 1. Glossary.

Artificial intelligence refers to a computerized system (hardware or software) that is able to perform tasks or reasoning processes that we usually associate with intelligence in a human being [3,4].

Machine learning is part of artificial intelligence. It is able to build algorithms that can detect patterns in large training datasets to explain variation in data or predict outcomes (predictive algorithms) [3,5-7].

Predictive analytics aims to build models which allow for individual (ie, single subject) predictions, thereby moving from the description of patients (hindsight) and the investigation of statistical group differences or associations (insight) toward models capable of predicting current or future characteristics for individual patients (foresight), thus allowing for a direct assessment of a model's clinical use [8].

Natural language processing refers to ML techniques capable of parsing unstructured texts, such as free written texts or spoken language [9]. Natural language processing enables the use of unstructured data that is otherwise largely inaccessible, such as the dynamic fluctuations of symptoms (eg, by analyzing clinical notes) and can identify social and behavioral factors, which are not commonly registered in EHR [10].

Predictive algorithms are a promising subfield of AI, in which machine learning (ML) models generate predictions on an individual patient level to support decision-making in the diagnosis, prognosis, and treatment of patients [5,8,11]. Predictive algorithms have been shown to be more successful than traditional statistical models, as they provide a new way to gain knowledge from a data set. For example, in the field of predicting suicide risk at the individual level, predictive algorithms can yield more accurate predictions [12-15] than traditional statistical models, which predict only slightly better than chance and whose predictive ability has not improved over time, even after 50 years of research [16]. Not all ML models for suicide risk are adequate, although, as performance can vary depending on various factors, such as data quality, nature of the features used as input, preprocessing steps, and the used techniques [12]. In other mental health care domains, ML models are also able to perform better than traditional models, such as the prediction of psychosis, treatment response [17], or opioid use disorder [18].

Algorithms designed for clinical care often use patient data collected during regular clinical care in the organization's electronic health record (EHR). Combining the expanding possibilities of AI with EHR data fuels the transformation into a data-driven mental health care, as the EHR contains rich, real-world, and real-time data. Reusing EHR data to develop algorithms that support clinical practice on an everyday basis opens many possibilities. However, it may be challenging, as these data are usually noisy, heterogeneous, dynamic, and contain both structured and unstructured data (eg, text) and require careful preprocessing [19]. Developing [20] and implementing AI-based algorithms using real-world EHR data has been proven to be feasible and can have a

significant clinical impact [21-23]. However, the studies of algorithms that are actually implemented and used in mental health clinical practice are still scarce [24]. Also, there is a gap between the creation of scientifically sound algorithms and their meaningful application in real-world settings, a challenge often referred to as the "AI chasm" [25].

For mental health care to benefit from AI opportunities and the wealth of EHR data, developing algorithms based on real-world EHR patient data intended for clinical use is a crucial first step toward successful implementation and use. The development phase bridges theoretical innovation and practical application. Therefore, the development phase is not only a preparatory step but an integral part of implementation, as the design and development of the AI tool may either facilitate or hinder later implementation and use (eg, through alignment with organizational workflows and engendering trust) [26].

To enable the successful development, implementation, and actual use of AI and ML in mental health clinical practice, implementation science provides a valuable framework [3], as it studies the contextual factors that enable the process of embedding and integrating new practices into actual care [27,28]. Also, it fosters greater attention for the sociotechnical aspects of technology, focusing not only on the development of technological improvements but also the social changes required for their successful implementation and use [17,29-33]. Multiple barriers and facilitators have been identified in the literature, highlighting aspects of AI technology itself (eg, accuracy, transparency, and generalizability); organizational factors (eg, integration with workflows); governance, ethical, and legal issues; perception of staff and patients; or the implementation process

(eg, engagement, user-centered design) [34,35]. However, these barriers and facilitators have, for a large part, been identified not primarily on implementation data but rather based on hypothetical implementation, white papers, and thought articles [34,35]. Therefore, more empirical studies are required to fill this knowledge gap [6,17].

To better understand how this development phase unfolds within the specific context of mental health care, we conducted a case study. This study explores the challenges encountered by a development team in creating a suicide risk prediction algorithm using real-world EHR data for clinical application in close collaboration with clinicians, as well as the team's responses to these issues. Ultimately, we aim to identify key considerations relevant to bridging technical and clinical needs in algorithm development, offering insights for mental health professionals and organizations seeking to move toward data-driven care.

Methods

Research Design

This study is a detailed case study aimed to explore the challenges involved in the development of a suicide risk prediction algorithm using real-world EHR data in mental health care. Using an exploratory, multimethod approach with both inductive and deductive analysis suited to this emerging field, as the study seeks to identify key development considerations applicable to broader contexts [36].

Case Description

Setting and Background

Suicide prevention in the Netherlands is organized through a national policy agenda focused on public awareness, early detection, and professional care. The national helpline 113 Suicide Prevention is available 24/7 by phone or chat. This organization also provides public campaigns and training for professionals and citizens. The National Suicide Prevention Agenda, developed by the Ministry of Health, Welfare and Sport together with 113 and other partners, offers a framework for collaboration between health care, education, municipalities, and the media. Locally, Municipal Health Services, youth health care, schools, and social services work on prevention and early risk identification. In mental health care, suicide prevention is part of standard practice, including staff training and procedures following a suicide incident. The Health and Youth Care Inspectorate monitors suicide prevention and reporting, while knowledge institutes, such as Trimbos, support professionals through guidelines and research.

GGz Eindhoven and the Kempen (Geestelijke Gezondheidszorg Eindhoven en de Kempen [GGzE]) is a mental health institution in southern Netherlands, providing specialized care to around 17,000 patients annually. At GGzE, a protocol is in place in accordance with the national multidisciplinary guideline [37]. All health care professionals are required to complete suicide prevention training courses. Suicide risk assessments are conducted at the start of

each treatment trajectory and at other critical moments, for example, during a crisis; at the beginning, end, and after hospitalization; or prior to the extension of patient privileges during admission. A standardized suicide assessment tool, the Suicide Prevention Questionnaire (SPQ), is embedded within the EHR and should be administered at these key points, at least once or twice per year.

To enhance suicide prevention, GGzE explored the feasibility of developing a predictive algorithm using EHR data to support clinicians in identifying suicide risk in individual patients. Key objectives included establishing a reliable measure of suicide incidents, identifying predictive factors, and creating an algorithm with an accuracy target of 70%-90% [38]. The development team, consisting of 2 data scientists with support from data science, governance, law, and clinical suicide experts (multiple psychologists and psychiatrists), operated under a structured project organization with a steering committee and project group. Both the steering committee and the project group comprised health care professionals as well as staff members with legal, organizational, and other forms of specialized expertise.

The project was embedded in the organization-wide Suicide Prevention Program, which provided consultation and training, developed organizational guidelines and protocols, organized evaluation of suicide incidents and subsequent quality enhancement of clinical practice, ensured evidence-based treatment options regarding suicidality, and developed a reflective learning culture regarding suicidality within the organization. The organization was part of the Suicide Prevention Action Network (Supernet GGZ), in which mental health institutions work together to make the quality of care for suicidal patients transparent, to learn what can be improved, and to improve care where necessary. Within this network, data on suicide incidents were structurally collected and shared, including core indicators on suicide risk.

The Suicide Prediction Algorithm: Description and Performance

The development team designed the algorithm for hospitalized patients with severe mental illness, using the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework [39], a standard method in data mining and ML, to predict suicide risk within 1 month. Data for the algorithm was taken from the EHR and included data from 5044 patients (December 31, 2015, to August 31, 2021) included demographics (eg, age and sex), administrative details (eg, financing type), treatment history (eg, hospitalization count), psychiatric factors (eg, primary diagnosis), current treatment details (eg, duration and intensity), and social factors (eg, social contacts). Consultations with clinicians, interviews, and a literature search resulted in a list of possible predictors of suicide incidents that were derived from the EHR.

The team applied natural language processing (NLP) to clinicians' daily reports, identifying unreported suicide incidents and capturing patient sentiment as an additional predictor, using topic modeling (BERTopic) [40]. Ensemble learning (extreme gradient boosting) [41] strengthened

the model. The development team deemed the algorithm successful based on Hicks et al criteria [38]. The team acknowledged that the performance indicator outcomes were exploratory, as the algorithm would be retrained prior to implementation in a pilot phase and subsequently retrained annually in accordance with quality standards.

The predictive algorithm for suicide incidents will be gradually tested in a small-scale pilot project conducted in close collaboration with health care professionals. The algorithm will be implemented solely as an add-on to existing suicide prevention policies. Health care professionals will remain fully responsible for clinical assessment and decision-making. The pilot project will be evaluated at multiple stages throughout its implementation.

Sampling and Participants

The study's participants included 2 core data science professionals who led the development of the suicide risk prediction algorithm, supported by internal and external experts. The team worked closely with clinicians (multiple psychologists, psychiatrists, and the Medical Director's office) and experts in data science, law, and data governance, as well as with external data science collaborators from other mental health care institutions. The participants were selected by expert sampling, in which participants are included who have demonstrable experience and expertise regarding the study's aim and objective [42,43]. These participants offered crucial insights into the challenges and considerations of algorithm development. The team worked closely with clinicians (eg, psychologists, psychiatrists, and the medical director) and experts in data science, law, and data governance, as well as with external data science collaborators from other mental health care institutions.

Data Collection and Data Analysis

Data for our study on the development process were collected through desk research, reflective meetings, and iterative feedback sessions with the development team. The study proceeded in three phases: (1) desk research and meetings with the development team to draw up a detailed process description, (2) identification of challenges and responses by thematic analysis, and (3) formulation of key considerations.

The detailed process description formed the basis for thematic analysis, which was conducted following the Braun and Clarke approach [44]. An exploratory, multimethod design combining inductive and deductive analysis was used to identify challenges during the algorithm development process. Deductive analysis was guided by the innovation domain of the Consolidated Framework for Implementation Research (CFIR) [28,45] and concepts from AI-related literature.

Data on process steps, challenges, and responses were analyzed and organized in a chart by 1 researcher (LH). Data was discussed by all 4 researchers to derive the main challenges and describe the team's responses. Communicative validation was achieved through discussions with the development team. Triangulation across data sources strengthened the credibility and transferability of findings.

Data collection and analysis were iterative and integrated throughout all phases.

Ethical Considerations

Ethical approval was obtained from the Ethical Review Board of Tilburg University (reference TSB_RP1038). Participants received no financial compensation and provided written informed consent prior to participation. All procedures adhered to principles of privacy and confidentiality: no identifying information was collected or published, and any nonessential identifying details were omitted. No images or supplementary materials contain identifiable participant features.

Results

Overview

The encountered challenges and the team's responses are described below, related to the construction of the suicide incidents (dependent variable) and possible predictors (independent variables), and the building of the model. At the start, the concept of the suicide incidents and the predictors had to be defined, operationalized, and measured. Several challenges were encountered during this phase: data was hard to find, data sources yielded different results, and underreporting was suspected. Regarding predicting factors, psychosocial predictors were not available and had to be estimated. To calculate the algorithm, the team used complex models that improved predictive performance but reduced understandability for clinicians.

Regular consultations between the project team and clinical experts during the development process were important to gain an accurate and profound understanding of the phenomena of suicide incidents and their predictive factors.

The Results section will focus on the development team's challenges and their responses to them. In the Discussion section, the results will be contextualized within the literature to formulate key considerations for algorithm development in mental health care.

Defining, Operationalizing, and Measuring Suicide Incidents

Different Data Sources Provide Different Outcomes

The data science team collaborated with health care professionals to define suicide incidents as the dependent variable. Within the organization, a suicide incident was defined as either a fatal suicide or a nonfatal attempt with a certain intention to die. This also covered self-harming or risky behavior that involved a risk of death or failure to avoid such risk. Severe suicide incidents were those that were fatal or caused, or could cause, severe or permanent physical injury [source: GGzE Protocol Reporting suicides and suicide attempts; consulted 2025-09-19]. Clinicians had to

distinguish these from self-harm driven by other motives, such as reducing emotional pain.

Organizational protocols prescribed the registration of every suicide incident in a digital registration system. The Medical Director's office kept its own registration. The 2 data sources within the organization yielded different outputs on the total number of suicide incidents:

In the Client Details source there is a variable called Suicide Attempt. In addition, a variable called Suicide Incident has been created based on the incident data. [...] there are more numbers [...] for Suicide Incident than for Suicide Attempt." [Report of the development team and external partners on characteristics of the suicide incident risk algorithm, p. 10].

The medical director's office clarified to the data science team that they reviewed the registered suicide incidents in the reporting system and recorded only those suicide incidents classified as severe. The team chose to include all suicide incidents, providing more data to train the algorithm and enabling the prediction of both severe and less severe cases. Clinical experts reflected on the overall numbers and considered them to be low, suggesting underregistration, particularly of less severe cases, possibly due to noncompliance with protocols or sensitivities around suicide. Moreover, ambiguity about the intention behind the behavior in less severe cases may have played a role. For example, behavior without serious injury or risk of injury that primarily served a communicative function.

Using Unstructured Data to Improve Detection of Suicide Incidents Is Challenging

To detect underreported suicide incidents, the development team added text mining using NLP of clinicians' unstructured daily reports. However, the method produced significant data noise, identifying many reports related to suicidality but failing to pinpoint actual suicide incidents. Text analysis was hindered by clinicians' writing habits, such as negation terms ("not feeling suicidal"), the use of synonyms (eg, "jumping off a bridge"), and the use of standard templates (eg, "has the patient expressed suicidality today?").

Manual labeling of cases was tried, but appeared to be very time-consuming and unfeasible as a permanent solution, especially as the algorithm had to be retrained every 6 months due to quality requirements:

We still need to be 100% sure that there is/isn't a suicide incident in the daily report data. Possible solution: Read the data yourself and assess whether there is/isn't a suicide incident (label data). Disadvantages: [...] Takes a lot of time (for a small dataset it took 3 days)

The development team responded to these challenges by making efforts to enhance data quality. The organization was in the process of transitioning to another provider of an EHR system. The team recommended including a check

mark for suicide incidents [source: Presentation Suicide Predictive Model_2nd iteration, p. 36] within this new system and advocated this solution in the new EHR project team, underlining the importance of these interventions to transform into a data-driven health care organization. Also, a new executive board took the lead of the organization and, consequently, a new multiyear strategy was developed, in which "working data-driven to enhance decision support" was one of the main ambitions [source: GGzE, multiyear strategy 2024-2028; consulted 2025-09-19]. This prompted management to take a more active role in ensuring data quality, such as completing suicide-related information. Also, the organization implemented an alternative instrument for the SPQ. This way, the organization aimed to improve data quality. The algorithm would be retrained when the new EHR system was implemented.

However, the team still had to deal with the fact that possible underregistration of suicide incidents, especially less severe cases, was an issue. From a data science perspective, they could interpret this as missing data, which threatened data quality. From a clinical perspective, the possibility that mild incidents were not always reported was a reflection of the ambiguity of the interpretation of a suicide incident: did the patient actually aim to end their life, or should the behavior be interpreted more as self-harming behavior with the intent to decrease pain or to communicate distress?

Defining, Operationalizing, and Measuring Potential Predicting Factors

The development team compiled a list of potential predicting factors by performing a literature review, reviewing factors from well-known suicide risk assessment methods, and interviewing clinicians to leverage their expertise and to involve and engage clinicians in the development process. The list of predictors then had to be defined, operationalized, and measured. Reflective meetings with the development team revealed several challenges they faced, which are described below.

The EHR Lacks Psychosocial Factors: Construction of Predictors is Necessary

Some predicting factors, like previous suicide attempts or substance dependence, could be directly derived from the EHR. Other predicting factors were psychosocial and could not be directly obtained from the EHR, thus requiring the construction of an estimator.

For example, information on the number of contacts of a patient, as registered in the EHR, along with a selection of characterization of these relationships (eg, spouse, sibling, child, and neighbor) was used to construct the factor "social network." Also, the factor "quality of the therapeutic relationship" was operationalized as the number of primary [1] practitioners during hospitalization. Although it was recognized that these operationalizations did not necessarily reflect the actual factors, in the absence of a better alternative, it was decided to operationalize and measure these concepts in this way.

Data Availability Differs Between Subgroups: Risk of Bias

Clinicians considered the SPQ, available in the EHR, the most promising source of predictors. Although protocols required annual administration for all patients, the data science team found a recent SPQ in only 21.95% of cases. Clinicians confirmed that SPQ variables were highly explanatory but acknowledged their limited availability. They suspected that SPQs were often completed only when suicide risk was already suspected, creating bias and overrepresenting high-risk patients. To include all patients, the data science team had to derive estimators from other EHR data. Additional disparities emerged: SPQs were missing more often for youth (0-17 y) than for older patients, despite comparable incident rates. These findings were shared with the Child & Youth Department to improve data quality. Ethnicity was also explored as a predictor but was excluded due to frequent missing values.

Overall, the results exposed a gap between organizational protocols (eg, annual SPQ administration, registration of ethnicity) and actual practice. They also revealed different perspectives: clinicians valued SPQ data but were less familiar with data science principles, such as representativeness and external validity, while data scientists emphasized these risks. Close collaboration was deemed essential to identify predictors and minimize bias.

To address these challenges, the development team worked with the Suicide Prevention Program and a national suicide prevention organization to design a digital toolkit as a replacement for the SPQ. This decision support tool aimed to make risk assessment more user-friendly, link assessments to targeted treatment advice based on the national guideline, and reduce clinician workload through robotic process automation that transferred data directly to the EHR.

Sentiment Analysis to Construct a Time- and Context-Sensitive Predictor

As the literature review and interviews with clinicians emphasized the role of time- and context-related factors in suicide risk (eg, relationship breakup and intoxication), the development team explored ways to integrate these into the model. Since collecting daily data directly from patients was impractical due to severe psychiatric conditions and low cooperation, daily reports by clinicians about patients were identified as the best alternative. These allowed real-time data to strengthen predictive accuracy.

Although detecting suicide incidents in daily reports was challenging, sentiment analysis proved feasible. Patients with suicide incidents showed lower sentiment in the daily notes by clinicians, especially in the most recent entries, compared to those without incidents (internal document: PowerPoint Suicide Predictive Model_2nd iteration, p. 29).

When the development team presented the results to the medical director's office, new applications were discussed. The algorithm was initially designed as an add-on to existing protocols, such as administering an SPQ annually. Given the

time-intensive nature of the SPQ, 1 proposal was to let the algorithm preselect high-risk patients for SPQ administration. This would save time by limiting assessments to flagged patients.

Clinical experts, including the medical director, rejected this option, noting that the algorithm was still in development and should not replace established tools. They did, however, acknowledge that SPQs were not consistently administered in practice. As an alternative, they proposed using sentiment analysis as an alert system: if emotional decline was detected and no recent SPQ was available, clinicians would be prompted to administer one. This approach was accepted, as it integrated the algorithm while reinforcing compliance with existing suicide prevention protocols.

Building the Algorithm: High-Performance Model Appears Hard to Understand by Clinicians

After identifying dependent and independent variables, the development team built the algorithm using ensemble learning, which combines weaker models into a stronger one. This approach delivered high predictive power, meeting the project's feasibility goal, but the model was too complex to interpret directly.

The data science team shared the results with clinicians and the steering group, including the model's performance metrics (eg, accuracy, precision, and recall), acknowledging that the results were exploratory as the model would later be retrained. Clinicians appreciated the potential value of the algorithm but emphasized that understanding why suicide risk was elevated was essential for fostering trust, ensuring accountability, and supporting treatment planning. They highlighted that a well-performing algorithm alone would not be sufficient to generate meaningful benefits in clinical practice. The data science team, therefore, identified explainability as a core requirement. They aimed to develop case-level reports that not only flagged risk but also explained the reasons behind it: "Not only: alarm! But also: alarm, for this reason" [Reflective meeting, 2023-11-07].

To address this challenge, a team member engaged in additional training in explainable AI, noting the scarcity of knowledge within mental health care. Additionally, the data science team planned to organize several collaborative sessions with clinicians to co-create an explainability dashboard that contextualizes the algorithm's outputs. Also, collaborative sessions were planned to facilitate the integration of the algorithm into existing workflows and jointly reflect on the tasks, roles, and responsibilities of each stakeholder in its use.

Discussion

Principal Findings

This study provided a detailed case study of the development of a suicide risk prediction algorithm using real-world EHR data. Its unique contribution lies in demonstrating how

algorithm development is carried out within the clinical practice of mental health care, where technical innovation must be aligned with clinical needs, professional values, and organizational structures. By describing the challenges of a development team and their responses to them, this study offers an in-depth description of the development of a predictive algorithm, in which the development process is seen as a first step in the implementation process.

This paper highlights the sociotechnical and organizational aspects of algorithm development, showing how collaboration, joint efforts to enhance data quality, and clinician engagement are essential prerequisites for algorithm development.

Reflection on the Principal Results

Addressing Data Quality Challenges in the EHR

High data quality is essential for AI in mental health care, yet EHR data are often low in quality, incomplete, “soft,” and indirect [46,47]. Also, handling unstructured EHR data requires unique approaches [48,49].

Regarding data quality, EHR data quality in mental health is typically problematic. In the current case, suicide incident data were hard to construct and possibly underreported, although this could also be a reflection of real-world clinical assessment issues. Low data quality is noted in the literature, where suicide incidents are frequently underdocumented [50] due to clinicians’ workload [7,46] and departmental workflow inconsistencies [51]. Addressing this, data quality issues span all 5 dimensions identified by Weiskopf and Weng [52]: completeness, correctness, concordance, plausibility, and currency. Organizations are faced with challenges on data quality according to each of these dimensions.

Mental health data often rely on “soft” psychological constructs (eg, social support), which are harder to quantify than “hard” somatic measures (eg, heart rate) [47]. This raises risks for construct validity [53]. Moreover, much of the data reflects professionals’ perceptions rather than patients’ own, which may diverge significantly [54-57]. Future approaches using direct patient data (eg, real-time sampling and devices) could help improve accuracy [13].

Analysis of unstructured data, like staff notes, showed high noise, with frequent mentions of suicidality but few actual incidents, a challenge also noted in other mental health domains [13,58]. Variations in writing style and negations further complicate analysis [59-61]. Still, NLP offers potential to capture real-time risk fluctuations and enhance predictive accuracy [62-65].

While low data quality in mental health is a barrier for AI, algorithm development may also support quality improvement. In the current case, feedback on data inconsistencies spurred organizational motivation to enhance quality and adapt the EHR system.

Mitigating Risk of Bias to Ensure Performance for Subgroups

When algorithms are trained on data of low quality, risk of bias (RoB) occurs, in which the algorithm may not perform as well for each subgroup, and human biases are repeated and even amplified. When predictors are constructed from EHR data as data are not readily available, selection bias may occur. Algorithms require balanced representation across subgroups, though, to perform effectively. In this case, using SPQ items as predictors risked overrepresenting high-risk patients and weakening performance for others. While EHR data enables inclusion of all hospitalized patients, including minority groups often underrepresented in research [66], bias remains a major concern. As social factors are often key predictors in mental health care [15,67], social data is often missing due to true absence or nonrandom issues, such as patient refusal, confidentiality concerns, or language barriers, and excluding such cases may worsen bias [15,67]. Selection bias also threatens performance when samples do not reflect the broader population [10,68,69]. Bias is therefore considered an additional dimension of data quality, alongside the 5 identified by Weiskopf and Weng [52,70].

By building algorithms from reused EHR data, human biases in the data may be reinforced. The development team acknowledged this risk but considered the potential benefits sufficient to proceed, provided mitigating measures were applied. Since algorithms are trained on subjective EHR data, cognitive biases, such as stereotypes or in-group favoritism [71-73], may become embedded and systematically applied [66,68,69]. In practice, if an algorithm’s outcome aligns with a clinician’s initial diagnosis, confirmation bias may lead the clinician to ignore contradictory information. If the outcomes differ, precautionary bias might cause them to trust the algorithm over their own judgment [73]. Automation bias could also emerge, where clinicians rely more on the tool than on other information, potentially diminishing their skills [7]. The question remains what mitigating measures are adequate to prevent systematically disadvantaging subgroups. At the same time, algorithms may also reduce bias by integrating fragmented information across professionals and systematically processing both subjective and objective data [74]. Determining under which conditions algorithms amplify or mitigate bias remains an open question for future research.

Actionability, Trust, Accountability, and Communication in Complex Models

Predictive power and understandability have to be balanced in algorithm development. In our study, data scientists prioritized predictive performance, applying a complex model to enhance accuracy, while clinicians emphasized the need for understandability. This reflects the broader trade-off between complex “black box” models, which yield higher predictive power but limited explainability, and simpler “glass box” models, which are more interpretable but often less accurate [75-77]. Initially, the team focused on feasibility, whether an algorithm with sufficient predictive performance could be built from organizational data, rather than

on clinical applicability. Clinician feedback prompted the inclusion of explainability as a key objective, as it enhances the actionability of the outcomes, that is, whether clinicians can use the information to identify actions to change the outcome predicted by the algorithm [78]. In suicide prevention, detecting elevated risk is insufficient; clinicians must also understand which factors contributed most, which are modifiable, and which interventions are likely to reduce risk [79,80]. Explainability is crucial for building trust [81-83], and providing an appropriate level of transparency (not too little and not too much) is necessary in designing for trust [84]. Clinicians should be well informed about the actual capabilities of the algorithm, as responsible use of AI systems also depends on the level of correspondence between a user's trust and the technology's capabilities, known as calibration [83]. Explainability may also enhance patient communication and support shared decision-making [76,79,85]. Also, mental health institutions are accountable for care quality, effectiveness, and safety. In our study, clinicians feared that using complex, nonexplainable models would undermine accountability, as evidence-based protocols are easier to justify than black-box algorithms. Given the high-stakes nature of suicide

care, caution is warranted when applying complex models [80].

Whether performance or explainability should be prioritized in mental health care remains unresolved, and future research should examine context-specific trade-offs as well as post hoc strategies to enhance explainability [77,86]. Future research should explore what information should be communicated to clinicians and how it should be presented. Moreover, fostering mutual trust and collaboration between clinicians and data scientists throughout the algorithm development process is essential. Technical performance alone does not guarantee clinical usefulness. Instead, cocreation of both the algorithm and its user interface, combined with joint efforts to integrate it into everyday practice, is crucial for enabling clinicians to trust and meaningfully use the algorithm's outcomes.

Below, we present [Table 1](#) summarizing the challenges encountered during the development phase, the corresponding responses by the team and organization, and the broader considerations derived from these experiences.

Table 1. Challenges, responses, and considerations during algorithm development in mental health care.

The development phase	Challenges	Responses	Considerations
Defining, operationalizing, and measuring suicide incidents	<ul style="list-style-type: none"> Inconsistent detection of suicide incidents across data sources. 	<ul style="list-style-type: none"> Enhanced suicide incident data by integrating clinician insights. Broadened incident inclusion. Tested NLP^a approaches. Improved EHR^b registration. 	<ul style="list-style-type: none"> Ensure sufficient data availability for the predicted outcome. Consult clinicians to improve data interpretation. Consider clinical explanations for data quality issues. Provide user-friendly EHR options for registering suicide incidents.
Defining, operationalizing, and measuring potential predicting factors	<ul style="list-style-type: none"> Limited and uneven availability of psychosocial and contextual predictors constrains robust and unbiased model development. 	<ul style="list-style-type: none"> Developed high-availability estimators. Addressed data gaps with clinicians. Implemented a guideline-aligned assessment tool to simultaneously enhance data registration and support clinical decision-making. Developed a sentiment analysis for contextual risk signals. 	<ul style="list-style-type: none"> Ensure predictor validity by prioritizing high-availability EHR data. Address subgroup gaps through data governance. Supplement psychosocial information via clinicians or patients. Develop user-friendly assessment tools that integrate required data registration with clinical decision-making support.
Building the algorithm	<ul style="list-style-type: none"> Complex models improve prediction but are difficult for clinicians to interpret. 	<ul style="list-style-type: none"> Established explainability as a core requirement. Planned cocreation of an explainability dashboard with clinicians before the pilot. 	<ul style="list-style-type: none"> Balance predictive performance and explainability through close collaboration between data scientists and clinicians.
Toward an implementation pilot	<ul style="list-style-type: none"> Effective use requires full integration into clinical workflows and sufficient clinician experience. Algorithm use may introduce cognitive and decision biases. 	<ul style="list-style-type: none"> Planned joint data scientist–clinician sessions to integrate the algorithm into workflows and clarify roles. Positioned the algorithm as an add-on to suicide prevention care, with clinicians' judgment remaining central. 	<ul style="list-style-type: none"> Organize data scientist–clinician sessions to support workflow integration and clarify roles. Clearly communicate the algorithm's role and its support function for clinical decision-making. Calibrate clinicians' expectations to the algorithm's actual capabilities and limitations.

^aNLP: natural language processing.

^bEHR: electronic health record.

Limitations

This study provided a detailed case study of the development of a suicide risk prediction algorithm using real-world EHR data, highlighting the tension between technical requirements and clinical needs. Using a qualitative, real-world approach, the study captured contextualized knowledge from the development team, who collaborated closely with clinicians. While generalizability is limited by the single-case design, triangulation and detailed process descriptions support transferability. A key limitation is that clinicians' perspectives were obtained indirectly via the development team rather than through direct engagement, although this was partly mitigated by desk research and targeted questions to include broader stakeholder views.

Conclusions and Future Research

Conclusions

Based on the findings of our case study, we highlight the following 5 main recommendations for mental health organizations that aim to evolve into data-driven practice.

1. Data governance and quality: high-performing algorithms require robust data governance and timely, deliberate data preparation to ensure data quality.
2. Cultural change in data entry: a cultural shift is needed to regard and manage data as a strategic asset for algorithm development.
3. Bias awareness and monitoring: organizations must remain sensitive to potential bias in both data and use, requiring careful data handling and regular evaluation of real-world application.
4. Balancing performance and transparency: predictive performance using complex models must be balanced with clinicians' needs for transparency and explainability to ensure actionability, trust, communication, and accountability.
5. Clinician-in-the-loop: it is essential to define the knowledge, skills, and workflows that enable clinicians to use algorithms responsibly while maintaining final clinical judgment.

For mental health organizations aspiring to evolve toward data-driven care, early preparation is essential. First, building predictive algorithms requires sufficient and well-structured data, which demands timely and deliberate data collection. Reliance on existing organizational protocols alone is insufficient; explicit attention must be given to how relevant variables are defined, recorded, and embedded in the EHR system. A cultural transformation is also critical. Second, data should not be viewed merely as an administrative or accountability tool but as a strategic resource for predictive modeling. However, this is not always aligned with clinical practice and may require a change in working processes for clinicians. This transition requires awareness and commitment from clinicians, whose engagement from the earliest stages of algorithm development is crucial. Collaboration between data scientists and clinicians must begin in an early stage of the development process: data

scientists contribute technical and data-analytical expertise, while clinicians bring contextual understanding, interpret data inconsistencies, identify meaningful predictors, and help foster a data-driven culture. Organizations must also recognize potential biases within their data and proactively assess how information from subgroups is collected and represented. Third, ongoing attention to bias is crucial, as it may stem not only from data or model design but also from real-world use. Regular evaluation of both performance and application is essential to ensure quality and responsibility in practice. Decisions regarding model design should balance predictive accuracy with interpretability. Fourth, explainable AI techniques may help translate complex model outputs into clinically meaningful insights, allowing clinicians to understand and act upon algorithmic predictions. Clinicians require not only accurate models but also transparency to understand the reasons for the outcome to ensure actionability, trust, and clinician-patient communication. Institutions should further consider how accountability for algorithm use is shared between individual practitioners and the organization. Finally, advancing toward data-driven mental health care requires investment in clinicians' education and professional development regarding AI. Defining the knowledge, skills, and workflows that enable responsible algorithm use, while maintaining clinical judgment as the ultimate decision point, will be key to realizing the added value of predictive technologies in practice.

Future Research

Future research should further explore how robust data governance structures can be established to ensure data quality and readiness for algorithm development. This includes examining how AI techniques may contribute to improving data quality through automation, standardization, and enhanced data validation. Moreover, studies should investigate how cultural change can be fostered to promote viewing data as a strategic resource rather than an administrative by-product, and how such shifts can be maintained throughout the organization.

Attention to bias should extend beyond data collection and model design to the actual use of algorithms. Future studies are needed to determine under which conditions algorithms amplify or mitigate bias, and how systematic monitoring of real-world application can safeguard fairness and responsible use.

The ongoing tension between predictive performance and explainability also warrants investigation. Research should identify context-specific trade-offs between complex, high-performing models and simpler, more transparent approaches and explore how explainability dashboards can enhance actionability, trust, accountability, and clinician-patient communication.

Finally, future research should examine what knowledge, skills, and workflows clinicians need to engage responsibly with predictive algorithms while maintaining final clinical judgment. Integrating sociotechnical considerations

throughout algorithm development will be key to ensuring that AI tools are both effective and ethically grounded in clinical practice.

While much focus has been placed on the technical aspects of ML in mental health care, sociotechnical factors related to

successful algorithm development, implementation, and use have been largely overlooked [6,13,15,17,24,25,30,33,87]. More research is needed to explore these issues in real-world settings to ensure responsible and sustainable innovation.

Acknowledgments

All authors contributed to study design and manuscript revision. LH collected and analyzed the data and drafted the initial manuscript, after which all authors contributed to multiple revisions. We thank the GGzE development team for their cooperation.

Data Availability

The datasets generated or analyzed during this study are not publicly available due to the inability to fully anonymize the raw data but are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

References

1. WRR [Netherlands Scientific Council for Government Policy]. Sustainable healthcare, a matter of choice. People, resources and public support. The Hague: Netherlands Scientific Council for Government Policy; 2021. URL: <https://english.wrr.nl/publications/reports/2022/05/03/sustainable-healthcare-a-matter-of-choice.-people-resources-and-public-support> [Accessed 2025-12-05]
2. Transforming healthcare with AI: the impact on the workforce and organisations. EIT Health and McKinsey & Company; 2020. URL: <https://www.mckinsey.com/industries/healthcare/our-insights/transforming-healthcare-with-ai/#/> [Accessed 2025-12-05]
3. Nilsen P, Svedberg P, Nygren J, Frideros M, Johansson J, Schueller S. Accelerating the impact of artificial intelligence in mental healthcare through implementation science. *Implement Res Pract*. 2022;3:26334895221112033. [doi: [10.1177/26334895221112033](https://doi.org/10.1177/26334895221112033)] [Medline: [37091110](#)]
4. AI HLEG. A definition of AI: main capabilities and disciplines. European Union; 2019. URL: https://ec.europa.eu/futurium/en/system/files/ged/ai_hleg_definition_of_ai_18_december_1.pdf [Accessed 2025-12-18]
5. Graham S, Depp C, Lee EE, et al. Artificial intelligence for mental health and mental illnesses: an overview. *Curr Psychiatry Rep*. Nov 7, 2019;21(11):31701320. [doi: [10.1007/s11920-019-1094-0](https://doi.org/10.1007/s11920-019-1094-0)] [Medline: [31701320](#)]
6. Koutsouleris N, Hauser TU, Skvortsova V, De Choudhury M. From promise to practice: towards the realisation of AI-informed mental health care. *Lancet Digit Health*. Nov 2022;4(11):e829-e840. [doi: [10.1016/S2589-7500\(22\)00153-4](https://doi.org/10.1016/S2589-7500(22)00153-4)] [Medline: [36229346](#)]
7. Monteith S, Glenn T, Geddes J, Whybrow PC, Achtyes E, Bauer M. Expectations for artificial intelligence (AI) in psychiatry. *Curr Psychiatry Rep*. Nov 2022;24(11):709-721. [doi: [10.1007/s11920-022-01378-5](https://doi.org/10.1007/s11920-022-01378-5)] [Medline: [36214931](#)]
8. Hahn T, Nierenberg AA, Whitfield-Gabrieli S. Predictive analytics in mental health: applications, guidelines, challenges and perspectives. *Mol Psychiatry*. Jan 2017;22(1):37-43. [doi: [10.1038/mp.2016.201](https://doi.org/10.1038/mp.2016.201)] [Medline: [27843153](#)]
9. Khurana D, Koli A, Khatter K, Singh S. Natural language processing: state of the art, current trends and challenges. *Multimed Tools Appl*. 2023;82(3):3713-3744. [doi: [10.1007/s11042-022-13428-4](https://doi.org/10.1007/s11042-022-13428-4)] [Medline: [35855771](#)]
10. Goodday SM, Kormilitzin A, Vaci N, et al. Maximizing the use of social and behavioural information from secondary care mental health electronic health records. *J Biomed Inform*. Jul 2020;107:103429. [doi: [10.1016/j.jbi.2020.103429](https://doi.org/10.1016/j.jbi.2020.103429)] [Medline: [32387393](#)]
11. Dwyer DB, Falkai P, Koutsouleris N. Machine learning approaches for clinical psychology and psychiatry. *Annu Rev Clin Psychol*. May 7, 2018;14(91–118):91-118. [doi: [10.1146/annurev-clinpsy-032816-045037](https://doi.org/10.1146/annurev-clinpsy-032816-045037)] [Medline: [29401044](#)]
12. Ehtemam H, Sadeghi Esfahlani S, Sanaei A, et al. Role of machine learning algorithms in suicide risk prediction: a systematic review-meta analysis of clinical studies. *BMC Med Inform Decis Mak*. May 27, 2024;24(1):138. [doi: [10.1186/s12911-024-02524-0](https://doi.org/10.1186/s12911-024-02524-0)] [Medline: [38802823](#)]
13. Kirtley OJ, van Mens K, Hoogendoorn M, Kapur N, de Beurs D. Translating promise into practice: a review of machine learning in suicide research and prevention. *Lancet Psychiatry*. Mar 2022;9(3):243-252. [doi: [10.1016/S2215-0366\(21\)00254-6](https://doi.org/10.1016/S2215-0366(21)00254-6)] [Medline: [35183281](#)]
14. Kusuma K, Larsen M, Quiroz JC, et al. The performance of machine learning models in predicting suicidal ideation, attempts, and deaths: a meta-analysis and systematic review. *J Psychiatr Res*. Nov 2022;155:579-588. [doi: [10.1016/j.jpsychires.2022.09.050](https://doi.org/10.1016/j.jpsychires.2022.09.050)] [Medline: [36206602](#)]

15. Nordin N, Zainol Z, Mohd Noor MH, Chan LF. Suicidal behaviour prediction models using machine learning techniques: a systematic review. *Artif Intell Med*. Oct 2022;132:102395. [doi: [10.1016/j.artmed.2022.102395](https://doi.org/10.1016/j.artmed.2022.102395)] [Medline: [36207078](#)]
16. Franklin JC, Ribeiro JD, Fox KR, et al. Risk factors for suicidal thoughts and behaviors: a meta-analysis of 50 years of research. *Psychol Bull*. Feb 2017;143(2):187-232. [doi: [10.1037/bul0000084](https://doi.org/10.1037/bul0000084)] [Medline: [27841450](#)]
17. Salazar de Pablo G, Studerus E, Vaquerizo-Serrano J, et al. Implementing precision psychiatry: a systematic review of individualized prediction models for clinical practice. *Schizophr Bull*. Mar 16, 2021;47(2):284-297. [doi: [10.1093/schbul/sbaa120](https://doi.org/10.1093/schbul/sbaa120)] [Medline: [32914178](#)]
18. Fouladvand S, Talbert J, Dwoskin LP, et al. A comparative effectiveness study on opioid use disorder prediction using artificial intelligence and existing risk models. *IEEE J Biomed Health Inform*. Jul 2023;27(7):3589-3598. [doi: [10.1109/JBHI2023.3265920](https://doi.org/10.1109/JBHI2023.3265920)] [Medline: [37037255](#)]
19. Liu F, Panagiotakos D. Real-world data: a brief review of the methods, applications, challenges and opportunities. *BMC Med Res Methodol*. Nov 5, 2022;22(1):287. [doi: [10.1186/s12874-022-01768-6](https://doi.org/10.1186/s12874-022-01768-6)] [Medline: [36335315](#)]
20. Kessler RC, Hwang I, Hoffmire CA, et al. Developing a practical suicide risk prediction model for targeting high-risk patients in the Veterans Health Administration. *Int J Methods Psych Res*. Sep 2017;26(3). [doi: [10.1002/mpr.1575](https://doi.org/10.1002/mpr.1575)]
21. Delgadillo J, Ali S, Fleck K, et al. Stratified care vs stepped care for depression: a cluster randomized clinical trial. *JAMA Psychiatry*. Feb 1, 2022;79(2):101-108. [doi: [10.1001/jamapsychiatry.2021.3539](https://doi.org/10.1001/jamapsychiatry.2021.3539)] [Medline: [34878526](#)]
22. Garriga R, Mas J, Abraha S, et al. Machine learning model to predict mental health crises from electronic health records. *Nat Med*. Jun 2022;28(6):1240-1248. [doi: [10.1038/s41591-022-01811-5](https://doi.org/10.1038/s41591-022-01811-5)] [Medline: [35577964](#)]
23. McCarthy JF, Cooper SA, Dent KR, et al. Evaluation of the recovery engagement and coordination for health-veterans enhanced treatment suicide risk modeling clinical program in the Veterans Health Administration. *JAMA Netw Open*. Oct 1, 2021;4(10):e2129900. [doi: [10.1001/jamanetworkopen.2021.29900](https://doi.org/10.1001/jamanetworkopen.2021.29900)] [Medline: [34661661](#)]
24. DECIDE-AI Steering Group. DECIDE-AI: new reporting guidelines to bridge the development-to-implementation gap in clinical artificial intelligence. *Nat Med*. Feb 2021;27(2):186-187. [doi: [10.1038/s41591-021-01229-5](https://doi.org/10.1038/s41591-021-01229-5)] [Medline: [33526932](#)]
25. Keane PA, Topol EJ. With an eye to AI and autonomous diagnosis. *NPJ Digit Med*. 2018;1(40):40. [doi: [10.1038/s41746-018-0048-y](https://doi.org/10.1038/s41746-018-0048-y)] [Medline: [31304321](#)]
26. Golden G, Popescu C, Israel S, et al. Applying artificial intelligence to clinical decision support in mental health: what have we learned? *Health Policy Technol*. Jun 2024;13(2):100844. [doi: [10.1016/j.hpt.2024.100844](https://doi.org/10.1016/j.hpt.2024.100844)]
27. Proctor EK, Landsverk J, Aarons G, Chambers D, Glisson C, Mittman B. Implementation research in mental health services: an emerging science with conceptual, methodological, and training challenges. *Adm Policy Ment Health*. Jan 2009;36(1):24-34. [doi: [10.1007/s10488-008-0197-4](https://doi.org/10.1007/s10488-008-0197-4)] [Medline: [19104929](#)]
28. Damschroder LJ, Reardon CM, Widerquist MAO, Lowery J. The updated consolidated framework for implementation research based on user feedback. *Implementation Sci*. Oct 29, 2022;17(1). [doi: [10.1186/s13012-022-01245-0](https://doi.org/10.1186/s13012-022-01245-0)] [Medline: [36309746](#)]
29. Wieringa M. What to account for when accounting for algorithms. Presented at: FAT* '20: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency; Jan 27-30, 2020:1-18; Barcelona Spain. [doi: [10.1145/3351095.3372833](https://doi.org/10.1145/3351095.3372833)]
30. Wiens J, Saria S, Sendak M, et al. Do no harm: a roadmap for responsible machine learning for health care. *Nat Med*. Sep 2019;25(9):1337-1340. [doi: [10.1038/s41591-019-0548-6](https://doi.org/10.1038/s41591-019-0548-6)] [Medline: [31427808](#)]
31. Group T, Vasey B, Clifton DA. DECIDE-AI: new reporting guidelines to bridge the development-to-implementation gap in clinical artificial intelligence. *Nat Med*. Feb 2021;27(2):186-187. [doi: [10.1038/s41591-021-01229-5](https://doi.org/10.1038/s41591-021-01229-5)] [Medline: [33526932](#)]
32. Sujan M, Furniss D, Grundy K, et al. Human factors challenges for the safe use of artificial intelligence in patient care. *BMJ Health Care Inform*. Nov 2019;26(1):31780459. [doi: [10.1136/bmjhci-2019-100081](https://doi.org/10.1136/bmjhci-2019-100081)] [Medline: [31780459](#)]
33. O'Dell B, Stevens K, Tomlinson A, Singh I, Cipriani A. Building trust in artificial intelligence and new technologies in mental health. *Evid Based Ment Health*. May 2022;25(2):45-46. [doi: [10.1136/ebmental-2022-300489](https://doi.org/10.1136/ebmental-2022-300489)] [Medline: [35444002](#)]
34. Baldwin H, Loebel-Davidsohn L, Oliver D, et al. Real-world implementation of precision psychiatry: a systematic review of barriers and facilitators. *Brain Sci*. Jul 16, 2022;12(7):35884740. [doi: [10.3390/brainsci12070934](https://doi.org/10.3390/brainsci12070934)] [Medline: [35884740](#)]
35. Hassan M, Kushniruk A, Borycki E. Barriers to and facilitators of artificial intelligence adoption in health care: scoping review. *JMIR Hum Factors*. Aug 29, 2024;11:e48633. [doi: [10.2196/48633](https://doi.org/10.2196/48633)] [Medline: [39207831](#)]
36. Fereday J, Muir-Cochrane E. Demonstrating rigor using thematic analysis: a hybrid approach of inductive and deductive coding and theme development. *Int J Qual Methods*. Mar 2006;5(1):80-92. [doi: [10.1177/160940690600500107](https://doi.org/10.1177/160940690600500107)]

37. Kerkhof AM, Verwey B, Hummelen JW, et al. Multidisciplinaire Richtlijn Diagnostiek En Behandeling van Suïcidaal Gedrag [Multidisciplinary Guideline on the Assessment and Treatment of Suicidal Behaviour] [Book in Dutch]. Trimbos Institute; 2012. URL: https://psychiatrieweb.nl/pw.spoed/files/docs/suicidaal%20gedrag/richtlijn/120801_MDR_DBSC.pdf [Accessed 2025-12-05]

38. Hicks SA, Strümke I, Thambawita V, et al. On evaluation metrics for medical applications of artificial intelligence. *Sci Rep.* Apr 8, 2022;12(1):5979. [doi: [10.1038/s41598-022-09954-8](https://doi.org/10.1038/s41598-022-09954-8)] [Medline: [35395867](https://pubmed.ncbi.nlm.nih.gov/35395867/)]

39. Wirth R, Hipp J. CRISP-DM: towards a standard process model for data mining. Presented at: Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining; Apr 18-20, 2000; Kyoto, Japan. 2000. URL: <https://www.semanticscholar.org/paper/CRISP-DM%3A-Towards-a-Standard-Process-Model-for-Data-Wirth-Hipp/48b9293cf4297f855867ca278f7069abc6a9c24#extracted> [Accessed 2025-12-05]

40. Grootendorst M. BERTopic: neural topic modeling with a class-based TF-IDF procedure. *arXiv.* Preprint posted online on Mar 11, 2022. [doi: [10.48550/arXiv.2203.05794](https://doi.org/10.48550/arXiv.2203.05794)]

41. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. Presented at: KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; Aug 13-17, 2016:785-794; San Francisco, California, USA. [doi: [10.1145/2939672.2939785](https://doi.org/10.1145/2939672.2939785)]

42. Etikan I, Musa SA, Alkassim RS. Comparison of convenience sampling and purposive sampling. *AJTAS.* 2016;5(1):1. [doi: [10.11648/j.ajtas.20160501.11](https://doi.org/10.11648/j.ajtas.20160501.11)]

43. Singh K. Quantitative Social Research Methods. Sage Publications; 2007. [doi: [10.4135/9789351507741](https://doi.org/10.4135/9789351507741)]

44. Braun V, Clarke V. Using thematic analysis in psychology. *Qual Res Psychol.* Jan 2006;3(2):77-101. [doi: [10.1191/1478088706qp063oa](https://doi.org/10.1191/1478088706qp063oa)]

45. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. *Implementation Sci.* Dec 2009;4(1):50. [doi: [10.1186/1748-5908-4-50](https://doi.org/10.1186/1748-5908-4-50)] [Medline: [19664226](https://pubmed.ncbi.nlm.nih.gov/19664226/)]

46. Hersh WR, Weiner MG, Embi PJ, et al. Caveats for the use of operational electronic health record data in comparative effectiveness research. *Med Care.* Aug 2013;51(8 Suppl 3):S30-7. [doi: [10.1097/MLR.0b013e31829b1dbd](https://doi.org/10.1097/MLR.0b013e31829b1dbd)] [Medline: [23774517](https://pubmed.ncbi.nlm.nih.gov/23774517/)]

47. Wickramarathne TL, Premaratne K, Murthi MN, Scheutz M, Kubler S, Pravia M. Belief theoretic methods for soft and hard data fusion. Presented at: ICASSP 2011 - 2011 IEEE International Conference on Acoustics, Speech and Signal Processing; May 22-27, 2011; Prague, Czech Republic. [doi: [10.1109/ICASSP.2011.5946964](https://doi.org/10.1109/ICASSP.2011.5946964)]

48. Sedlakova J, Daniore P, Horn Wintsch A, et al. Challenges and best practices for digital unstructured data enrichment in health research: a systematic narrative review. *PLOS Digit Health.* Oct 2023;2(10):e0000347. [doi: [10.1371/journal.pdig.0000347](https://doi.org/10.1371/journal.pdig.0000347)] [Medline: [37819910](https://pubmed.ncbi.nlm.nih.gov/37819910/)]

49. Tayefi M, Ngo P, Chomutare T, et al. Challenges and opportunities beyond structured data in analysis of electronic health records. *WIREs Computational Stats.* Nov 2021;13(6). [doi: [10.1002/wics.1549](https://doi.org/10.1002/wics.1549)]

50. Anderson HD, Pace WD, Brandt E, et al. Monitoring suicidal patients in primary care using electronic health records. *J Am Board Fam Med.* 2015;28(1):65-71. [doi: [10.3122/jabfm.2015.01.140181](https://doi.org/10.3122/jabfm.2015.01.140181)] [Medline: [25567824](https://pubmed.ncbi.nlm.nih.gov/25567824/)]

51. Parsons A, McCullough C, Wang J, Shih S. Validity of electronic health record-derived quality measurement for performance monitoring. *J Am Med Inform Assoc.* 2012;19(4):604-609. [doi: [10.1136/amiajnl-2011-000557](https://doi.org/10.1136/amiajnl-2011-000557)] [Medline: [22249967](https://pubmed.ncbi.nlm.nih.gov/22249967/)]

52. Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. *J Am Med Inform Assoc.* Jan 1, 2013;20(1):144-151. [doi: [10.1136/amiajnl-2011-000681](https://doi.org/10.1136/amiajnl-2011-000681)] [Medline: [22733976](https://pubmed.ncbi.nlm.nih.gov/22733976/)]

53. De Choudhury M, Kiciman E. Integrating artificial and human intelligence in complex, sensitive problem domains: experiences from mental health. *AI Mag.* Sep 2018;39(3):69-80. [doi: [10.1609/aimag.v39i3.2815](https://doi.org/10.1609/aimag.v39i3.2815)]

54. Baune BT, Christensen MC. Differences in perceptions of major depressive disorder symptoms and treatment priorities between patients and health care providers across the acute, post-acute, and remission phases of depression. *Front Psychiatry.* 2019;10(335):335. [doi: [10.3389/fpsyg.2019.00335](https://doi.org/10.3389/fpsyg.2019.00335)] [Medline: [31178765](https://pubmed.ncbi.nlm.nih.gov/31178765/)]

55. Ilkiw-Lavalle O, Grenyer BFS. Differences between patient and staff perceptions of aggression in mental health units. *Psychiatr Serv.* Mar 2003;54(3):389-393. [doi: [10.1176/appi.ps.54.3.389](https://doi.org/10.1176/appi.ps.54.3.389)] [Medline: [12610249](https://pubmed.ncbi.nlm.nih.gov/12610249/)]

56. Rossberg JI, Friis S. Patients' and staff's perceptions of the psychiatric ward environment. *Psychiatr Serv.* Jul 2004;55(7):798-803. [doi: [10.1176/appi.ps.55.7.798](https://doi.org/10.1176/appi.ps.55.7.798)] [Medline: [15232020](https://pubmed.ncbi.nlm.nih.gov/15232020/)]

57. Van Der Merwe M, Muir-Cochrane E, Jones J, Tziggili M, Bowers L. Improving seclusion practice: implications of a review of staff and patient views. *J Psychiatr Ment Health Nurs.* Apr 2013;20(3):203-215. [doi: [10.1111/j.1365-2850.2012.01903.x](https://doi.org/10.1111/j.1365-2850.2012.01903.x)] [Medline: [22805615](https://pubmed.ncbi.nlm.nih.gov/22805615/)]

58. Hazewinkel MC, de Winter RFP, van Est RW, et al. Text analysis of electronic medical records to predict seclusion in psychiatric wards: proof of concept. *Front Psychiatry*. 2019;10(188):188. [doi: [10.3389/fpsyg.2019.00188](https://doi.org/10.3389/fpsyg.2019.00188)] [Medline: [31031650](https://pubmed.ncbi.nlm.nih.gov/31031650/)]

59. Velupillai S, Hadlaczky G, Baca-Garcia E, et al. Risk assessment tools and data-driven approaches for predicting and preventing suicidal behavior. *Front Psychiatry*. 2019;10(36):36. [doi: [10.3389/fpsyg.2019.00036](https://doi.org/10.3389/fpsyg.2019.00036)] [Medline: [30814958](https://pubmed.ncbi.nlm.nih.gov/30814958/)]

60. Zhang D, Yin C, Zeng J, Yuan X, Zhang P. Combining structured and unstructured data for predictive models: a deep learning approach. *BMC Med Inform Decis Mak*. Oct 29, 2020;20(1):280. [doi: [10.1186/s12911-020-01297-6](https://doi.org/10.1186/s12911-020-01297-6)] [Medline: [33121479](https://pubmed.ncbi.nlm.nih.gov/33121479/)]

61. Boggs JM, Kafka JM. A critical review of text mining applications for suicide research. *Curr Epidemiol Rep*. 2022;9(3):126-134. [doi: [10.1007/s40471-022-00293-w](https://doi.org/10.1007/s40471-022-00293-w)] [Medline: [35911089](https://pubmed.ncbi.nlm.nih.gov/35911089/)]

62. Boudreux ED, Rundensteiner E, Liu F, et al. Applying machine learning approaches to suicide prediction using healthcare data: overview and future directions. *Front Psychiatry*. 2021;12:707916. [doi: [10.3389/fpsyg.2021.707916](https://doi.org/10.3389/fpsyg.2021.707916)] [Medline: [34413800](https://pubmed.ncbi.nlm.nih.gov/34413800/)]

63. Fernandes AC, Dutta R, Velupillai S, Sanyal J, Stewart R, Chandran D. Identifying suicide ideation and suicidal attempts in a psychiatric clinical research database using natural language processing. *Sci Rep*. May 9, 2018;8(1):7426. [doi: [10.1038/s41598-018-25773-2](https://doi.org/10.1038/s41598-018-25773-2)] [Medline: [29743531](https://pubmed.ncbi.nlm.nih.gov/29743531/)]

64. Levis M, Levy J, Dent KR, et al. Leveraging natural language processing to improve electronic health record suicide risk prediction for veterans health administration users. *J Clin Psychiatry*. Jun 19, 2023;84(4):22m14568. [doi: [10.4088/JCP.22m14568](https://doi.org/10.4088/JCP.22m14568)] [Medline: [37341477](https://pubmed.ncbi.nlm.nih.gov/37341477/)]

65. Cusick M, Adekkattu P, Campion TR Jr, et al. Using weak supervision and deep learning to classify clinical notes for identification of current suicidal ideation. *J Psychiatr Res*. Apr 2021;136(95–102):95-102. [doi: [10.1016/j.jpsychires.2021.01.052](https://doi.org/10.1016/j.jpsychires.2021.01.052)] [Medline: [33581461](https://pubmed.ncbi.nlm.nih.gov/33581461/)]

66. Geneviève LD, Martani A, Shaw D, Elger BS, Wangmo T. Structural racism in precision medicine: leaving no one behind. *BMC Med Ethics*. Feb 19, 2020;21(1):17. [doi: [10.1186/s12910-020-0457-8](https://doi.org/10.1186/s12910-020-0457-8)] [Medline: [32075640](https://pubmed.ncbi.nlm.nih.gov/32075640/)]

67. Cook LA, Sachs J, Weiskopf NG. The quality of social determinants data in the electronic health record: a systematic review. *J Am Med Inform Assoc*. Dec 28, 2021;29(1):187-196. [doi: [10.1093/jamia/ocab199](https://doi.org/10.1093/jamia/ocab199)] [Medline: [34664641](https://pubmed.ncbi.nlm.nih.gov/34664641/)]

68. Timmons AC, Duong JB, Simo Fiallo N, et al. A call to action on assessing and mitigating bias in artificial intelligence applications for mental health. *Perspect Psychol Sci*. Sep 2023;18(5):1062-1096. [doi: [10.1177/17456916221134490](https://doi.org/10.1177/17456916221134490)] [Medline: [36490369](https://pubmed.ncbi.nlm.nih.gov/36490369/)]

69. Suresh H, Guttag J. A framework for understanding sources of harm throughout the machine learning life cycle. Presented at: EAAMO '21: Proceedings of the 1st ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization; Oct 5-9, 2021:1-9; New York, United States. [doi: [10.1145/3465416.3483305](https://doi.org/10.1145/3465416.3483305)]

70. Lewis AE, Weiskopf N, Abrams ZB, et al. Electronic health record data quality assessment and tools: a systematic review. *J Am Med Inform Assoc*. Sep 25, 2023;30(10):1730-1740. [doi: [10.1093/jamia/ocad120](https://doi.org/10.1093/jamia/ocad120)] [Medline: [37390812](https://pubmed.ncbi.nlm.nih.gov/37390812/)]

71. Akata Z, Balliet D, de Rijke M, et al. A Research Agenda for Hybrid Intelligence: Augmenting Human Intellect With Collaborative, Adaptive, Responsible, and Explainable Artificial Intelligence. *Computer* (Long Beach Calif). 2020;53(8):18-28. [doi: [10.1109/MC.2020.2996587](https://doi.org/10.1109/MC.2020.2996587)]

72. Mouchabac S, Conejero I, Lakhlii C, et al. Improving clinical decision-making in psychiatry: implementation of digital phenotyping could mitigate the influence of patient's and practitioner's individual cognitive biases. *Dialogues Clin Neurosci*. 2021;23(1):52-61. [doi: [10.1080/19585969.2022.2042165](https://doi.org/10.1080/19585969.2022.2042165)] [Medline: [35860175](https://pubmed.ncbi.nlm.nih.gov/35860175/)]

73. Choudhury A, Chaudhry Z. Large language models and user trust: consequence of self-referential learning loop and the deskilling of health care professionals. *J Med Internet Res*. Apr 25, 2024;26:e56764. [doi: [10.2196/56764](https://doi.org/10.2196/56764)] [Medline: [38662419](https://pubmed.ncbi.nlm.nih.gov/38662419/)]

74. Vitacca M, Giardini A, Gazzi L, Vitacca M. Hidden biases in clinical decision-making: potential solutions, challenges, and perspectives. *Monaldi Arch Chest Dis*. Sep 7, 2022;93(2):36069639. [doi: [10.4081/monaldi.2022.2339](https://doi.org/10.4081/monaldi.2022.2339)] [Medline: [36069639](https://pubmed.ncbi.nlm.nih.gov/36069639/)]

75. Hulsen T. Explainable artificial intelligence (XAI): concepts and challenges in healthcare. *AI*. 2023;4(3):652-666. [doi: [10.3390/ai4030034](https://doi.org/10.3390/ai4030034)]

76. Amann J, Blasimme A, Vayena E, Frey D, Madai VI, Precise4Q consortium. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. *BMC Med Inform Decis Mak*. Nov 30, 2020;20(1):310. [doi: [10.1186/s12911-020-01332-6](https://doi.org/10.1186/s12911-020-01332-6)] [Medline: [33256715](https://pubmed.ncbi.nlm.nih.gov/33256715/)]

77. Bodria F, Giannotti F, Guidotti R, Naretto F, Pedreschi D, Rinzivillo S. Benchmarking and survey of explanation methods for black box models. *Data Min Knowl Disc*. Sep 2023;37(5):1719-1778. [doi: [10.1007/s10618-023-00933-9](https://doi.org/10.1007/s10618-023-00933-9)]

78. Singh R, Miller T, Sonenberg L, Velloso E, Vetere F, Howe P, et al. An actionability assessment tool for explainable AI. *arXiv*. Preprint posted online on Jun 19, 2024. [doi: [10.48550/arXiv.2407.09516](https://doi.org/10.48550/arXiv.2407.09516)]

79. Menger VJ, Spruit MR, Scheepers FE. Kennisontwikkeling in de klinische psychiatrie: leren van elektronische patiëntendossiers [Article in Dutch]. *Tijdschr Psychiatr.* 2021;63(4):294-300. URL: <https://www.tijdschriftvoorspsychiatrie.nl/media/1/63-2021-4-artikel-menger.pdf> [Accessed 2025-12-05]
80. Vale D, El-Sharif A, Ali M. Explainable artificial intelligence (XAI) post-hoc explainability methods: risks and limitations in non-discrimination law. *AI Ethics.* Nov 2022;2(4):815-826. [doi: [10.1007/s43681-022-00142-y](https://doi.org/10.1007/s43681-022-00142-y)]
81. Reddy S, Shaikh S. The long road ahead: navigating obstacles and building bridges for clinical integration of artificial intelligence technologies. *J Med Artif Intell.* 2025;8:7-7. [doi: [10.21037/jmai-24-148](https://doi.org/10.21037/jmai-24-148)]
82. Kaplan AD, Kessler TT, Brill JC, Hancock PA. Trust in artificial intelligence: meta-analytic findings. *Hum Factors.* Mar 2023;65(2):337-359. [doi: [10.1177/00187208211013988](https://doi.org/10.1177/00187208211013988)] [Medline: [34048287](https://pubmed.ncbi.nlm.nih.gov/34048287/)]
83. Glikson E, Woolley AW. Human trust in artificial intelligence: review of empirical research. *Ann R Coll Physicians Surg Can.* Jul 2020;14(2):627-660. [doi: [10.5465/annals.2018.0057](https://doi.org/10.5465/annals.2018.0057)]
84. Kizilcec RF. How much information?: Effects of transparency on trust in an algorithmic interface. Presented at: CHI '16: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems; May 7-12, 2016:2390-2395; San Jose, California, USA. [doi: [10.1145/2858036.2858402](https://doi.org/10.1145/2858036.2858402)]
85. Luk JW, Pruitt LD, Smolenski DJ, Tucker J, Workman DE, Belsher BE. From everyday life predictions to suicide prevention: clinical and ethical considerations in suicide predictive analytic tools. *J Clin Psychol.* Feb 2022;78(2):137-148. [doi: [10.1002/jclp.23202](https://doi.org/10.1002/jclp.23202)] [Medline: [34195998](https://pubmed.ncbi.nlm.nih.gov/34195998/)]
86. van der Veer SN, Riste L, Cheraghi-Sohi S, et al. Trading off accuracy and explainability in AI decision-making: findings from 2 citizens' juries. *J Am Med Inform Assoc.* Sep 18, 2021;28(10):2128-2138. [doi: [10.1093/jamia/ocab127](https://doi.org/10.1093/jamia/ocab127)] [Medline: [34333646](https://pubmed.ncbi.nlm.nih.gov/34333646/)]
87. Barredo Arrieta A, Díaz-Rodríguez N, Del Ser J, et al. Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion.* Jun 2020;58:82-115. [doi: [10.1016/j.inffus.2019.12.012](https://doi.org/10.1016/j.inffus.2019.12.012)]

Abbreviations

AI: artificial intelligence

CFIR: Consolidated Framework for Implementation Research

CRISP-DM: Cross-Industry Standard Process for Data Mining

EHR: electronic health record

GGzE: Geestelijke Gezondheidszorg Eindhoven en de Kempen

ML: machine learning

NLP: natural language processing

RoB: risk of bias

SPQ: Suicide Prevention Questionnaire

Edited by Caroline Perrin; peer-reviewed by Hsueh-Han Yeh, Ross Jacobucci; submitted 20.Mar.2025; final revised version received 18.Nov.2025; accepted 18.Nov.2025; published 14.Jan.2026

Please cite as:

Hummel L, Lorenz-Artz KCAG, Bierbooms JJPA, Bongers IMB

Developing a Suicide Risk Prediction Algorithm Using Electronic Health Record Data in Mental Health Care: Real-World Case Study

JMIR Med Inform 2026;14:e74240

URL: <https://medinform.jmir.org/2026/1/e74240>

doi: [10.2196/74240](https://doi.org/10.2196/74240)

© Linda Hummel, Karin C A G Lorenz-Artz, Joyce J P A Bierbooms, Inge M B Bongers. Originally published in JMIR Medical Informatics (<https://medinform.jmir.org>), 14.Jan.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on <https://medinform.jmir.org/>, as well as this copyright and license information must be included.