nature medicine

Supplementary information

https://doi.org/10.1038/s41591-025-03603-z

Clinical implementation of AI-based screening for risk for opioid use disorder in hospitalized adults

In the format provided by the authors and unedited

CONSORT-AI checklist of information to include when reporting a randomised trials of AI interventions

Section	Item	CONSORT 2010 Item[a]	CONSORT-AI Item		Addressed on Page No[b]
Title and Abstract					
Title and Abstract	1a	Identification as a randomised trial in the title	CONSORT-AI 1a,b Elaboration	(i) Indicate that the intervention involves artificial intelligence/machine learning in the title and/or abstract and specify the type of model.	1
	1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts)		(ii) State the intended use of the AI intervention within the trial in the title and/or abstract.	1-3
Introduction					
Background and objectives	2a	Scientific background and explanation of rationale	CONSORT-AI 2a (i) Extension	Explain the intended use of the AI intervention in the context of the clinical pathway, including its purpose and its intended users (e.g. healthcare professionals, patients, public).	4-6
	2b	Specific objectives or hypotheses			5-6
Methods					
Trial design	3a	Description of trial design (such as parallel, factorial) including allocation ratio			6, 8, 29
	3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons			n/a
Participants	4a	Eligibility criteria for participants	CONSORT-AI 4a (i) Elaboration	State the inclusion and exclusion criteria at the level of participants.	8, 29, Figure 1
			CONSORT-AI 4a (ii) Extension	State the inclusion and exclusion criteria at the level of the input data.	8, 29-30
	4b	Settings and locations where the data were collected	CONSORT-AI 4b Extension	Describe how the AI intervention was integrated into the trial setting, including any onsite or offsite requirements.	30-32
Interventions	5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered	CONSORT-AI 5 (i) Extension	State which version of the AI algorithm was used.	31
			CONSORT-AI 5 (ii) Extension	Describe how the input data were acquired and selected for the AI intervention.	32-33
			CONSORT-AI 5 (iii) Extension	Describe how poor quality or unavailable input data were assessed and handled.	34
			CONSORT-AI 5 (iv) Extension.	Specify whether there was human-AI interaction in the handling of the input data, and what level of expertise was required of users.	33-34
			CONSORT-AI 5 (v) Extension	Specify the output of the AI intervention	9, Figure 2
			CONSORT-AI 5 (vi) Extension	Explain how the AI intervention's outputs contributed to decision-making or other elements of clinical practice.	10-12, Figure 3
Outcomes	6a	Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed			9-10, 34-35
	6b	Any changes to trial outcomes after the trial commenced, with reasons			na
Sample size	7a	How sample size was determined			38

	7b	When applicable, explanation of any interim analyses and stopping guidelines			n/a
Randomisation					
Sequence generation	8a	Method used to generate the random allocation sequence			n/a
	8b	Type of randomisation; details of any restriction (such as blocking and block size)			n/a
Allocation concealment mechanism	9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned			n/a
Implementation	10	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions			n/a
Blinding	11a	If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how			n/a
	11b	If relevant, description of the similarity of interventions			n/a
Statistical methods	12a	Statistical methods used to compare groups for primary and secondary outcomes			39
	12b	Methods for additional analyses, such as subgroup analyses and adjusted analyses			34-35
Results					
Participant flow (a diagram is strongly recommended)	13a	For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome			8, Figure 1
	13b	For each group, losses and exclusions after randomisation, together with reasons			9
Recruitment	14a	Dates defining the periods of recruitment and follow-up			8
	14b	Why the trial ended or was stopped			8
Baseline data	15	A table showing baseline demographic and clinical characteristics for each group			Table 1
Numbers analysed	16	For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups			Tables 1 and 2
Outcomes and estimation	17a	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)			9-12
	17b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended			9-12
Ancillary analyses	18	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory			11-12

Harms	19	All important harms or unintended effects in each group (for specific guidance see CONSORT for harms)	CONSORT-AI 19 Extension	Describe results of any analysis of performance errors and how errors were identified, where applicable. If no such analysis was planned or done, explain why not.	6-7
Discussion					
Limitations	20	Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses			17-19
Generalisability	21	Generalisability (external validity, applicability) of the trial findings			13-14
Interpretation	22	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence			14-15
Other Information					
Registration	23	Registration number and name of trial registry			3
Protocol	24	Where the full trial protocol can be accessed, if available			Reference 27
Funding	25	Sources of funding and other support (such as supply of drugs), role of funders	CONSORT-AI 25 Extension.	State whether and how the AI intervention and/or its code can be accessed, including any restrictions to access or re-use.	31, 36-37

[a] We strongly recommend reading this statement in conjunction with the CONSORT 2010 Explanation and Elaboration for important clarifications on all the items. [b] Indicates page numbers to be completed by authors during protocol development.

Supplemental Table 1. Associations of the intervention period and patient characteristics with opioid-related Addiction Medicine consultation

Variable	**Beta**	**Std Error**	**Odds Ratio**	**95% CI**	**P-value**
Post-Intervention*	0.086	0.082	1.09	(0.93, 1.28)	0.850
Sex, Female	-0.293	0.041	0.75	(0.69, 0.81)	<.001
Age					
<30 years	0.492	0.091	1.64	(1.37, 1.96)	<.001
30-49 years	0.864	0.067	2.37	(2.08, 2.70)	<.001
50-64 years	0.066	0.074	1.07	(0.92, 1.24)	0.371
Race/Ethnicity					
Hispanic	-0.235	0.254	0.79	(0.48, 1.30)	0.355
Non-Hispanic Black	0.484	0.181	1.62	(1.14, 2.32)	0.008
Non-Hispanic White	0.691	0.164	2.00	(1.45, 2.75)	<.001
Insurance					
Commercial	-0.754	0.078	0.47	(0.40, 0.55)	<.001
Medicaid	1.055	0.069	2.87	(2.51, 3.29)	<.001
Medicare	-0.252	0.093	0.78	(0.65, 0.93)	0.007

Elixhauser Comorbidity Score	-0.094	0.009	0.91	(0.89, 0.93)	<.001

*P-value based on one-tailed non-inferiority test after adjusting for confounding variables; Referent for Age was 65+, Race referent was Mixed/Unknown/Other, Insurance referent was Uninsured/No insurance

Supplemental Table 2. Association of the intervention period and patient characteristics with Receiving Medication for an Opioid Use Disorder (MOUD)

Variable	Beta	Std Error	Odds Ratio	95% CI	P-value
Post-Intervention*	-0.143	0.116	0.87	(0.69, 1.09)	0.109
Sex, Female	-0.212	0.055	0.81	(0.73, 0.90)	<.001
Age					
<30 years	0.510	0.133	1.66	(1.28, 2.16)	<.001
30-49 years	1.054	0.099	2.87	(2.36, 3.49)	<.001
50-64 years	0.131	0.112	1.14	(0.91, 1.42)	0.243
Race/Ethnicity					
Hispanic	-0.135	0.248	0.87	(0.69, 1.09)	0.586
Non-Hispanic Black	0.183	0.163	1.20	(0.87, 1.65)	0.259
Non-Hispanic White	0.519	0.129	1.68	(1.30, 2.17)	<.001
Insurance					
Commercial	-0.818	0.114	0.44	(0.35, 0.55)	<.001
Medicaid	1.229	0.096	3.42	(2.83, 4.13)	<.001
Medicare	-0.269	0.137	0.76	(0.58, 1.00)	0.050
Elixhauser Comorbidity Score	-0.101	0.013	0.90	(0.88, 0.93)	<.001

*P-value based on one-tailed non-inferiority test after adjusting for confounding variables
Referent for Age was 65+, Race referent was Mixed/Unknown/Other, Insurance referent was Uninsured/No insurance

Supplemental Table 3. Association of the intervention period and patient characteristics with 30-Day Readmission among patients receiving an Addiction Medicine consult

Variable	**Beta**	**Std Error**	**Odds Ratio**	**95% CI**	**P-value**
Post-Intervention	-0.640	0.281	0.53	(0.30, 0.91)	0.023
Sex, Female	-0.009	0.261	0.99	(0.59, 1.66)	0.974
Age					
<30 years	-0.181	0.768	0.83	(0.19. 3.77)	0.814
30-49 years	1.063	0.617	2.90	(0.86, 9.73)	0.085
50-64 years	0.607	0.626	1.84	(0.54, 6.27)	0.333
Race/Ethnicity					
Hispanic	-1.072	1.257	0.34	(0.03, 4.04)	0.394
Non-Hispanic Black	-0.588	0.691	0.56	(0.14, 2.16)	0.395
Non-Hispanic White	-0.546	0.602	0.58	(0.18, 1.89)	0.365
Insurance					
Commercial	-0.729	0.490	0.48	(0.18, 1.26)	0.137
Medicaid	-0.934	0.445	0.39	(0.16, 0.94)	0.036
Medicare	-0.435	0.517	0.65	(0.24, 1.79)	0.401
Elixhauser Comorbidity Score	0.034	0.022	1.04	(0.99, 1.08)	0.129

Supplemental Table 4. Association of the intervention period and patient characteristics with Any 30-Day Post-Discharge Hospital or ER Encounter among patients receiving an Addiction Medicine consult

Variable	**Beta**	**Std Error**	**Adjusted Odds Ratio**	**95% CI**	**P-value**
Post-Intervention	-0.404	0.175	0.67	(0.47, 0.94)	0.023
Sex, Female	0.055	0.173	1.06	(0.75, 1.48)	0.751
Age					
<30 years	0.340	0.437	1.41	(0.60. 3.31)	0.438
30-49 years	0.571	0.390	1.77	(0.82, 3.81)	0.146
50-64 years	0.526	0.390	1.69	(0.79, 3.64)	0.181
Race/Ethnicity					
Hispanic	-0.874	0.799	0.42	(0.09, 2.00)	0.276
Non-Hispanic Black	-0.034	0.492	0.97	(0.37, 2.54)	0.945
Non-Hispanic White	–0.232	0.447	0.79	(0.33, 1.91)	0.604
Insurance					
Commercial	0.289	0.396	1.34	(0.61, 2.91)	0.468
Medicaid	0.285	0.371	1.33	(0.64, 2.75)	0.444
Medicare	0.499	0.417	1.65	(0.73, 3.73)	0.234
Elixhauser Comorbidity Score	0.004	0.016	1.00	(0.97, 1.04)	0.789

Referent for Age was 65+, Race referent was Mixed/Unknown/Other, Insurance referent was Uninsured/No insurance

Supplemental Table 5. Association of the intervention period and patient characteristics with of 30-Day Readmission among all patients

Variable	**Beta**	**Std Error**	**Adjusted Odds Ratio**	**95% CI**	**P-value**
Post-Intervention	0.001	0.032	1.00	(0.94, 1.07)	0.985
Sex, Female	-0.008	0..015	0.99	(0.96, 1.02)	0.596
Age					
<30 years	-0.162	0.048	0.85	(0.77. 0.93)	0.001
30-49 years	0.024	0.032	1.02	(0.96, 1.09)	0.452
50-64 years	0.128	0.028	1.14	(1.08, 1.20)	<.001
Race/Ethnicity					
Hispanic	-0.056	0.072	0.95	(0.82, 1.09)	0.441
Non-Hispanic Black	0.276	0.052	1.32	(1.19, 1.46)	<.001
Non-Hispanic White	–0.094	0.036	0.91	(0.85, 0.98)	0.009
Insurance					
Commercial	0.209	0.028	1.23	(1.17, 1.30)	<.001
Medicaid	-0.055	0.043	0.95	(0.87, 1.03)	0.204
Medicare	-0.016	0.032	0.98	(0.93, 1.05)	0.625
Elixhauser Comorbidity Score	0.262	0.011	1.30	(1.27, 1.33)	<.001

Referent for Age was 65+, Race referent was Mixed/Unknown/Other, Insurance referent was Uninsured/No insurance

Supplemental Table 6. Cost Analysis

Fixed Costs			
Personnel	**Median Salary+Benefits**	**Person-Months Devoted to Project**	**Cost**
Analytics Consultant	$133,000-$166,000	3.6	$45,900
Data Scientist	$126,000-$142,000	5.0	$56,100
Machine Learning Engineer	$113,000-$179,000	9.6	$132,300
Personnel Total			$234,300
Storage and Equipment			$107,800
Training			$11,600
Total Fixed Costs			**$355,700**
Incremental Costs			
Personnel	**Median Salary+Benefits**	**Person-Months Devoted to Project**	**Cost**
Machine Learning Engineer	$113,000-$179,000	8.0	$101,400
Personnel Total			$101,400
Storage and Computing			$2,800
Hospital and Addiction Medicine Consulting			$1,900
Total Incremental Costs			**$106,100**

Time spent by providers at each training session accounted for median salaries by specialty type (i.e., surgery versus family medicine).

Supplemental Clinical Trial Protocol

Natural Language Processing for Screening Opioid Misuse

Protocol Number: 2022-0384

Principal Investigators: Majid Afshar, MD, MS

Funding Source: NIH/NIDA R01-DA051464

NCT Number: NCT05745480

Protocol Publication: https://medinform.jmir.org/2023/1/e44977

Table of Contents

1.0 STATEMENT OF COMPLIANCE

I confirm that I have read this protocol. I will comply with the protocol, and applicable regulations, guidelines, laws, and institutional policies.

I agree to ensure that all staff members involved in the conduct of this study are informed about their obligations in meeting the above commitment.

_**Majid Afshar**_____ _____

_____**3/1/2023**_____

Principal investigator

2.0 LIST OF ABBREVIATIONS

AI	Artificial Intelligence
AE	Adverse Event
BPA	Best Practice Alert
EHR	Electronic Health Record
GCP	Good Clinical Practice
HIM	Health Information Management
CDI	Clinical Documentation Integrity
HIPAA	Health Insurance Portability and Accountability Act
CRDS	Clinical Research Data Services at UW Center for Health Informatics Institute
ICTR	Institute for Clinical and Translational Research
IRB	Institutional Review Board
NLP	Natural Language Processing
OHRP	Office for Human Research Protections
PHI	Protected Health Information
PI	Principal Investigator
POC	Point of Contact
REDCap	Research Electronic Data Capture
SMART-AI	Substance Misuse Algorithm for Referral to Treatment using Artificial Intelligence
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SMP	Study Monitoring Plan
UP	Unanticipated Problem

3.0 STUDY SUMMARY

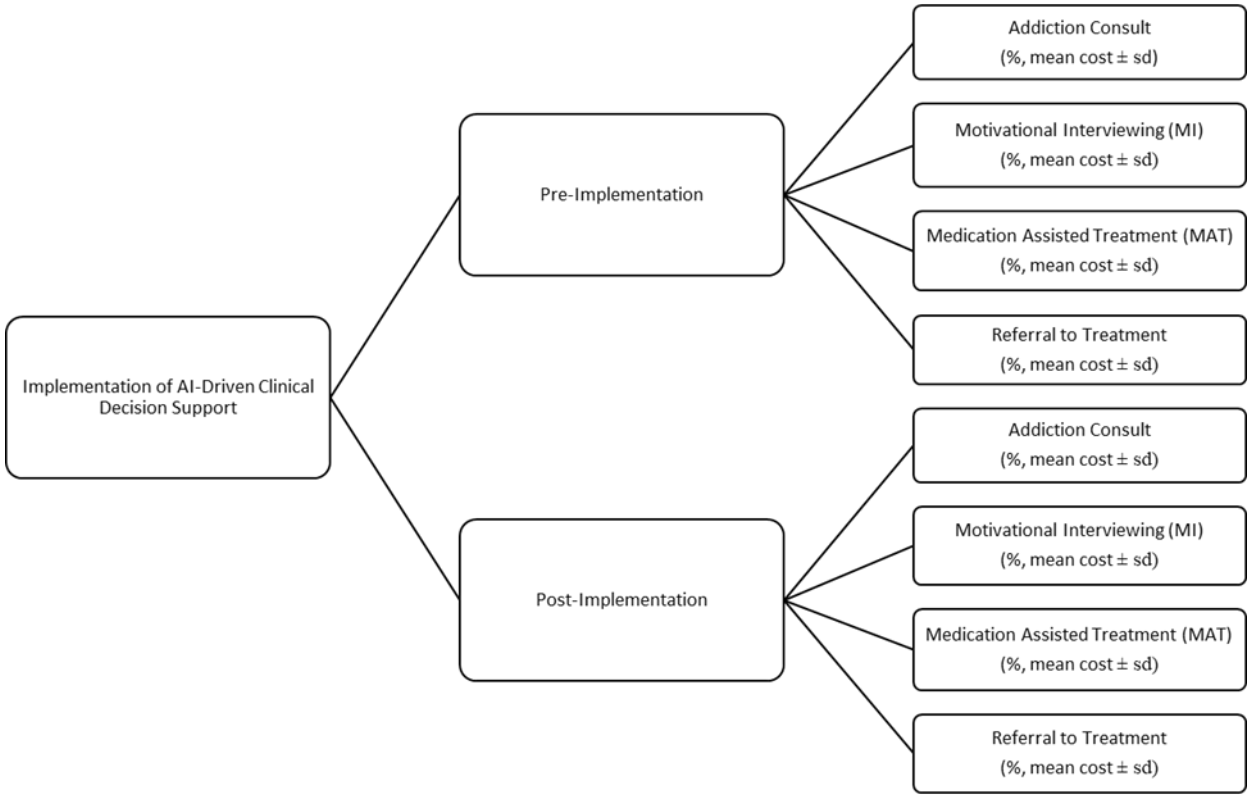
3.1 Synopsis

Full Title	Natural Language Processing for Screening Opioid Misuse
Short Title	AI Screener for Opioid Misuse
Protocol Number	IRB2022-0384
ClinicalTrials.gov Identifier & Summary	NCT05745480 This is a clinical study to implement and evaluate a hospital-wide, operational intervention for a real-time natural language processing (NLP)-driven clinical decision support (CDS) tool, called Substance Misuse Algorithm for Referral to Treatment Using Artificial Intelligence (SMART-AI). The SMART-AI CDS tool will be evaluated via implementation in the UW Health electronic health record (EHR). The CDS tool is meant for screening

	inpatient adults for opioid misuse as part of a best practice alert to nurses and providers for addiction consult service needs.
Number of Site(s)	1 site; University of Wisconsin Hospitals and Clinics
Phase	A pragmatic electronic health record embedded design
Main Inclusion Criteria	• Adult (>18 years) • Admitted to inpatient ward • Hospitalization for a least 24 hours
Main Exclusion Criteria	• Not receiving care in the surgical or medical inpatient wards
Objective(s)	• We aimed to detail a hospital-wide, operational pipeline to implement a real-time NLP-driven CDS tool and describe a protocol for an implementation framework with a user-centered design of the CDS tool
Endpoints	<u>Primary Endpoint</u> • Percentage of inpatients who screened positive (or would have screened positive) based on the NLP CDS tool who received an addiction consult, Percentage of inpatients who screened positive (or would have screened positive) based on the NLP CDS tool who received an addiction consult <u>Secondary Endpoint</u> • 30-day unplanned hospital readmission rate, Criteria for unplanned hospital readmissions were adopted from the Centers for Medicare & Medicaid Services
Study Design	A pre-post quasi-experimental study design over 30 months (24 months of usual care and 6 months for the implementation of automated screening)
Study Intervention	AI Screener composed of the AI model (SMART-AI), EHR-embedded workflow, and Best Practice Alert
Total Number of Subjects	A total sample size of 12,500 patients, with at least 10,000 in the preintervention 2-year period and 2500 in the postintervention 6-month period. Expected 30,000 hospitalizations per year.
Study Population	All hospitalized adults
Statistical Methodology	A pre-post quasi-experimental design to assess the effectiveness of an AI-driven Opioid Use Disorder (OUD) screener embedded within the electronic health record (EHR). Independent sample z-tests to evaluate non-inferiority for the primary outcome, with a one-sided significance level of 0.025. Mixed-effects logistic regression was utilized to estimate the odds ratios, adjusting for patient demographics and comorbidity scores, while a random intercept accounted for repeated hospitalizations by the same patient. Additionally, a cost-effectiveness analysis assessed the incremental costs of the AI screener implementation and its potential to reduce readmissions.
Estimated Subject Duration	30 months

Estimated Enrollment Period & Study Duration	24 months of usual care and 6 months for the implementation of automated screening

3.2 Schematic of Study Design



4.0 KEY ROLES

4.1 Key Personnel and Roles

Co-Principal Investigators	**Majid Afshar, MD, MSCR** **Associate Professor** **UW Health University Hospital** **600 Highland Avenue** **Madison, WI, 53792** **608-263-1280** ***Majid.afshar@wisc.edu***
Co-Investigators	Frank Liao, Brian Patterson, Anne Gravel Sullivan, Felice Resnik 600 Highland Avenue Madison, WI
Biostatistician and Health Economist	Marlon Mundt, PhD Professor UW School of Medicine and Public Health 610 Walnut Street, Ste 701 Madison, WI 53726 (608) 263-5382 marlon.mundt@fammed.wisc.edu

5.0 INTRODUCTION

5.1 Background

The unstructured narrative within electronic health records (EHRs) contains extensive data on patients' conditions, offering a valuable resource for clinical decision support (CDS).[1] Despite the rich detail provided by providers' intake notes, this information is often unorganized and not optimized for bedside decision-making through augmented intelligence.[2] Additionally, the free-text format of clinical notes complicates data analytics, limiting the potential for actionable insights. Natural language processing (NLP) techniques address this challenge by enabling the extraction of meaningful information from clinical notes, which can then be utilized by machine learning algorithms to identify conditions such as opioid misuse.[3] Opioid-related health crises have escalated significantly, with synthetic opioid overdoses rising from 2010 to 2022 and opioid-related emergency department visits increasing nearly 40% between 2018 and 2021.[4] Hospitalized patients with Opioid Use Disorder (OUD) are particularly vulnerable, making hospitalization a critical opportunity for intervention.[5] We developed a convolutional neural network (CNN)-based classifier, trained on EHR notes, that demonstrated over 80% sensitivity and specificity for opioid misuse detection, surpassing rule-based methods.[6] However, there is limited evidence regarding the deployment of clinical NLP models in real-time CDS within operational health systems. The effectiveness of AI interventions depends on the interaction between the system,

users, and implementation context. Despite the promise of NLP-driven CDS, most existing systems are retrospective or simulate clinical settings, lacking integration into real-time clinical workflows.[7] Additionally, frameworks for replicating these CDS implementations are scarce, highlighting a need for standardized processes and best practices in AI-driven CDS planning, development, and deployment within healthcare.

5.2 SMART-AI Study Intervention

5.3.1 Overview of SMART-AI

The SMART-AI protocol outlines a cloud-based service designed to ingest, process, and store clinical notes from a widely used electronic health record (EHR) system, transforming them into standardized and interoperable messages. Utilizing an elastic cloud infrastructure, SMART-AI leverages open-source tools, including a natural language processing (NLP) engine, to analyze EHR notes and feed relevant information into a deep learning model for screening opioid misuse. This integrated NLP and deep learning pipeline enables clinical notes to be processed and provides decision support at the bedside within minutes of note entry, enhancing the efficiency and timeliness of care.

5.3.2 Mechanism of Action and Safety

The SMART-AI system is designed to function within a secure, elastic cloud environment that supports the ingestion, processing, and storage of clinical notes. This cloud-based infrastructure utilizes an open-source NLP engine to extract structured information from free-text clinical notes, which is then processed by a deep learning model trained to screen for opioid misuse. The resultant data is fed back into the EHR to support clinical decision-making. Ensuring safety, the pipeline operates under stringent data security and compliance standards, and continuous monitoring evaluates its performance and effectiveness, with adjustments made as needed to maintain high standards of accuracy and reliability.

5.3.3 Availability to Participants

The intervention phase of the SMART-AI study includes all hospitalized patients flagged for opioid misuse by the NLP-driven clinical decision support tool. Participants have access to the AI-based screening system as part of their routine clinical workflow, allowing providers to document and assess patient needs in real-time. This setup enables a controlled evaluation of the tool's effectiveness in identifying high-risk patients and delivering timely interventions, ultimately aiming to improve patient outcomes and streamline clinical processes.

5.3 Rationale

The rise in opioid-related hospital visits and overdose deaths highlights the urgent need for effective interventions for patients with Opioid Use Disorder (OUD). Hospitalization presents a critical opportunity for intervention, as many hospitalized patients with OUD are at elevated risk for adverse outcomes, including overdose and readmission.[8–10] However, identifying these patients within the fast-paced hospital setting is challenging[5], given that essential clinical information is often buried within unstructured EHR notes. Natural language processing (NLP) techniques and machine learning algorithms offer a scalable approach to screen patients for OUD by leveraging data embedded in EHR narratives. While various NLP-based tools have been developed, few are fully integrated within clinical workflows to provide real-time decision support. This study aims to implement and evaluate an AI-driven clinical decision support system, embedded within the EHR, to facilitate the timely identification of patients at risk for opioid misuse, enabling healthcare providers to initiate appropriate interventions. By offering a standardized and automated approach, the AI screener seeks to address gaps in current OUD care, enhancing clinical outcomes and contributing to a scalable solution for health systems managing the opioid crisis.

5.4.1 Study Objectives

The primary objective of this study is to evaluate the effectiveness of an AI-driven opioid misuse screener embedded in the EHR for improving clinical outcomes among hospitalized patients at risk for an OUD. Specifically, the study aims to determine if the AI screener can match or surpass the effectiveness of traditional provider-driven consult requests for Addiction Medicine services. The primary outcome will assess the proportion of hospitalized adults who receive an Addiction Medicine consultation after the implementation of the AI screener. Secondary objectives include evaluating the impact of the AI screener on 30-day hospital readmission rates and conducting a cost-effectiveness analysis. This study will explore whether the AI-enabled tool can enhance the scalability and accessibility of OUD care by automating patient identification and consultation recommendations within the clinical workflow.

6.0 STUDY OBJECTIVES AND ENDPOINTS

Objectives	Outcomes and Hypotheses
Primary	
The percentage of inpatients who screened positive (or would have screened positive) based on the AI screener who received an addiction consult.	*H1*: that the proportion with the primary outcome was non-inferior in the post-intervention period compared to pre-intervention
Secondary	
To evaluate and 30-day hospital readmission (all-cause and all hospitalizations)	*H1*: Patients in the post-period will have lower rates of readmission

7.0 STUDY DESIGN

7.1 General Design – IT Integration and Interoperability

Component 1: Transferring Clinical Notes From the EHR to Cloud Computing

Health Level 7 (HL7) refers to the standards for transferring health care data between data sources. Cloverleaf (Infor Cloverleaf Integration Suite) was the UW's vendor solution that served as an application programming interface (API) gateway for accessing the clinical narratives in the EHR using HL7 version 2. To initialize the data feed, a UW Health interface analyst created a new entry in the Cloverleaf vendor software detailing the desired record information, which included clinical note text and identifiers. The analyst then "activated" the data feed, which began a continuous Transmission Control Protocol message–generating process. The Transmission Control Protocol messages were communicated using the HL7 application protocol to the Azure Virtual Machine (version 2022; Microsoft Corp) host at a port designated by the data science engineering team. This port was reserved by a NET program (the "TCP listener"), which wrote the message to the cloud file system and replied to the Cloverleaf server with acknowledgment messages, conforming to the HL7 version 2 specification. Ultimately, the API extracted clinical notes from Epic and transferred them to a Microsoft Azure cloud computing environment that was under a business associate agreement with the UW. On-premise relays with the File Transfer Protocol were used to transfer the clinical notes to a specified location in the Azure cloud environment.

Component 2: Cloud Analytic Computing Platform

In the Microsoft© Azure framework (Microsoft, 2022), UW Health invoked the Databricks© (Databricks Inc., San Francisco) analytic resource and services for scalable computing, data storage and querying. The open-source tools from the NLP engine and our trained, publicly available machine-learning model were hosted in Databricks. The Machine Learning model lifecycle management (MLFlow) tool in Databricks© supported the data flow for the deep learning model. MLFlow created and scored models when clinical notes were received and subsequently reported the results upon request. The final infrastructure was a scaleable and failure-resistant environment for analytic computation.

Component 3: Natural Language Processing Pipeline

The clinical Text Analysis and Knowledge Extraction System (cTAKES, Apache Software Foundation)[11] was built on multiple open-source Apache projects and incorporated technologies with the Unstructured Information Management Architecture (UIMA) framework and the Apache OpenNLP natural language processing toolkit. This configuration contained several engines for sentence detection, tokenization, part-of-speech tagging, concept detection, and normalization to extract information from the clinical narrative in the EHR. We did not use the negation module because it was not used in the current use case but this can be turned on for other use cases. cTAKES is one of the most ubiquitous NLP engines used in the clinical domain. cTAKES provided named entities from the free text that were mapped from the National Library of Medicine's Unified Medical Language System (UMLS), which were groups of words with relevant clinical context (e.g., Drugs, Diseases, Symptoms, Anatomical Sites, and Procedures). Each named entity mapped to a concept unique identifier (CUI) using the UMLS SNOMED and RxNORM ontologies. For instance, 'heroin

misuse' from the text was assigned C0600241 as its CUI and was a separate CUI from 'history of heroin misuse', which was C3266350.

As clinical notes were entered into the EHR for an individual patient, Cloverleaf© relayed the notes via HL7 from Epic© EHR and used the Azure FTP server running on a virtual machine to place them in a known location within the Azure cloud environment. In 15-minute intervals, DataBricks triggered a custom python script to extract the text and fed it into the cTAKES pipeline to map and extract the CUIs. The CUIs were stored in the Azure Data Lake with appended data including patient ID, encounter ID, and note timestamp and were ready to be fed into any machine learning model. The code being executed for the pipeline consisted of several services that operated independently and communicated through data stores. These services were 'always on', but each had a trigger condition that initiated code execution. We provided the pseudocode for these services in the Appendix.

Component 4: Text Feed from NLP Pipeline into Deep Learning Model

We previously published a substance misuse screening algorithm using CUIs fed into a CNN called the Substance Misuse Algorithm for Referral to Treatment using Artificial Intelligence (SMART-AI).[6] SMART-AI was trained on the first 24 hours of all clinical notes entered into the EHR starting with the patient's arrival time. This approach provided enough time for robust training but still time for the Addiction consult service to intervene before hospital discharge. For ease of implementation, the model was not trained on any specific note type and followed a timestamp approach for all notes filed within 24 hours from arrival at the hospital. SMART-AI was a supervised model with target labels that were derived from manual screening data on over 50,000 patients who self-reported on the validated Drug Abuse Screening Test[12] with follow-up questions about opioid use. SMART-AI is publicly available to run the trained model and more details about the model architecture and development may be found in the original development and validation publication.

Temporal validation of the classifier (trained on data between 2017 and 2019 and tested on data from 2020) at an outside hospital demonstrated an area under the precision-recall curve of 0.87 (95% CI 0.84–0.91) for screening opioid misuse. Similar results were shown in external validation at a second, independent health system. Multiple cutoff points were examined for optimal threshold selection for the BPA, including the point on the AUROC curve that minimized the difference between sensitivity and specificity. During validation on the full cohort of hospitalized patients, the optimal cutoff point for screening opioid misuse was 0.05. At that cutoff, the sensitivity was 0.87 (95% Confidence Interval [CI]: 0.84–0.90), specificity 0·99 (0·99–0·99), negative predictive value 0.99 (95% CI: 0.99–0.99), and positive predictive value 0.76 (95% CI: 0.72–0.88). The number needed to alert to detect a true positive screen was 1.4 and would create 26 alerts per 1000 hospitalized patients. This was deemed as an acceptable workload for consultation requests in live production for the UW Addiction Medicine clinicians. Additional silent testing was performed at UW Health to examine sensitivity and specificity with 95% CI in the practice setting.

All notes from the first 24 hours of arrival at the UW hospital were combined into a single document per patient encounter and converted into sequences of vector representations (e.g., embeddings). The CUI embeddings defined the input layer to the SMART-AI model at the encounter level. The model provided prediction probabilities for opioid misuse and stored them in a DataBricks© table with the predefined cutpoint for screen positives.

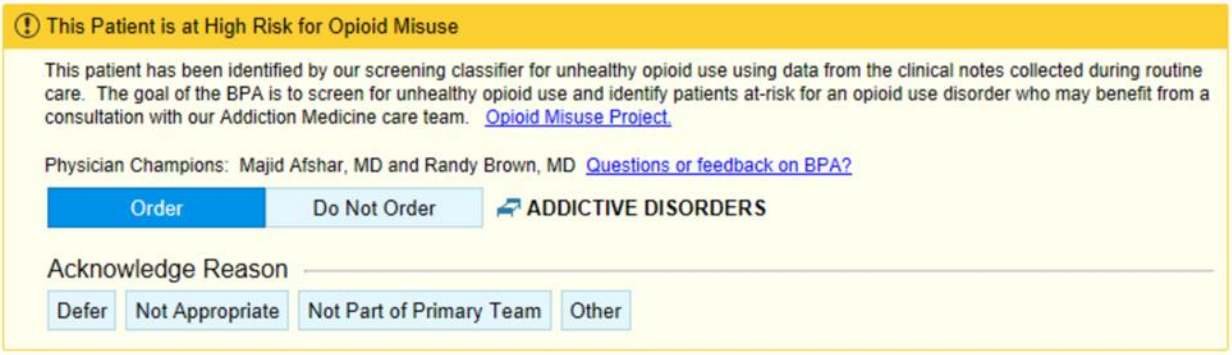
Component 5: Real-Time Delivery of Prediction Results

Nebula Cloud Platform was Epic©'s Software as a Service (SaaS) platform for integrating new technology and specifically supported clinical prediction modeling. Nebula capabilities

included the deployment of machine learning models, including a library of Epic-curated models for healthcare and custom algorithms. Our solution leveraged the latter to facilitate triggers from Epic© to call out to the Databricks© environment and provided the predictions for the BPAs.

In the case of SMART-AI, we designed a BPA (**Figure 1**) to trigger once a clinician opened the patient chart in the EHR. Epic© called its Nebula component to see if a BPA should be generated. Nebula made an HTTP call to DataBricks© to request the score. The RESTful HTTP API provided the SMART-AI model score that was serviced using MLFlow. Parameters included UMLS dictionaries, model results, patient identifiers, and other necessary attributes for individual-level predictions. The score was returned to Nebula which was used to trigger a BPA if it met the cutoff score for opioid misuse. For screen positives, the alert provided the recommendation to the clinician for consultation with UW's Addiction Medicine consult service. The following were internal targets to meet the real-time needs of the end-user at the bedside: (1) a throughput of 1000 notes per minute (<60ms each); (2) three nines (99.9%) availability – equivalent of fewer than nine hours downtime annually; and (3) an established error rate threshold.

Figure 1. BPA



In an iterative design with feedback from end users, a final BPA was implemented for bedside care. The BPA triggers upon opening a chart for a patient that meets the cutpoint predicted probability for opioid misuse from the NLP and deep learning model (SMART-AI).

Component 6: Cybersecurity

Two principles of security were applied: (1) defense-in-depth and (2) zero-trust. Zero trust architecture was outlined in the National Institute of Standards and Technology publication SP 800-207. To secure access between Azure Databricks© MLFlow and Epic©'s Nebula we employed an authentication token and IP range restriction (Databricks© admin utility). The authentication token was issued via Databricks© standard authentication. As a security best practice, we employed the Databricks service principal and its Databricks© access token to give automated tools and systems access to Databricks© resources.

The full pipleine is visualized in **Figure 2** and pseudocode for IT implementation is in **Appendix B**.

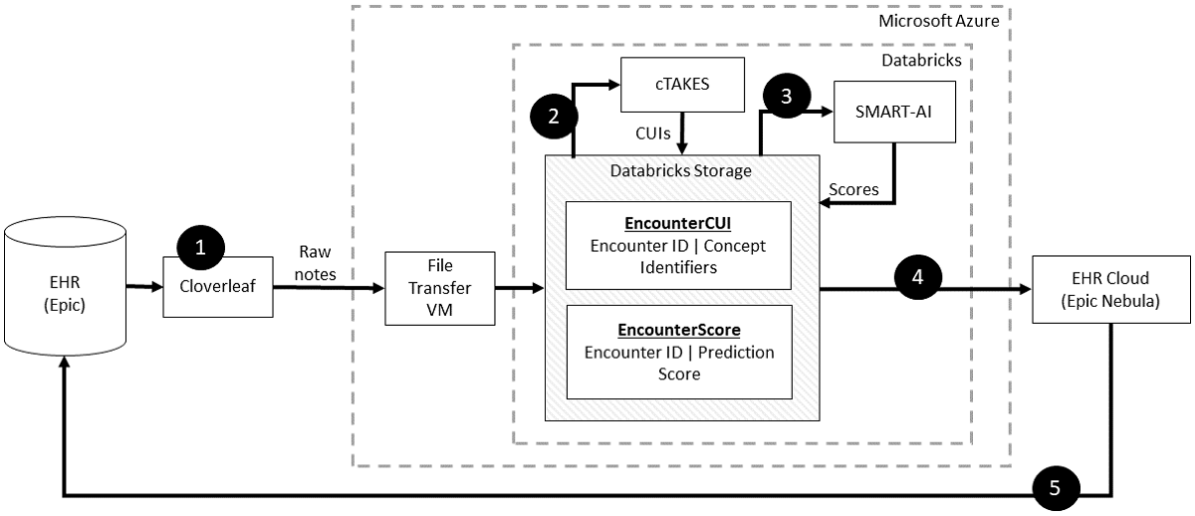


Figure 2. Step-by-step implementation of clinical natural language processing (NLP) pipeline. Step 1 ran a scheduled program to ingest notes from the EHR for each patient organized the notes, and relayed them via an HL7 data feed (Cloverleaf) into the cloud computing environment and data lake (Microsoft Azure and Databricks) onto a VM (Step 2). The NLP engine (cTAKES) processed the text stored on the VM and mapped them to medical concepts from the National Library of Medicine's metathesaurus (CUIs). The machine learning model received the CUIs as inputs and stored the results in DataBricks. At regular intervals, a custom Python script in Databricks performed the text extraction and linguistic feature engineering via cTAKES and stored CUIs with the appended data of patient identifiers. The CUIs served as the input to the machine learning model (SMART-AI) at the encounter level. The output of prediction probabilities and classification was stored in a Databricks table (Step 3). An API call from the EHR cloud is made to determine whether the cutpoint threshold from the machine learning model was met to trigger a BPA. In Step 4, the EHR cloud made an HTTP call to Databricks to request the score. The score was returned to the EHR cloud and subsequently delivered as a BPA when the provider opened the patient's chart in our on-premise instance of the EHR (Step 5). API: application programming interface; BPA: best practice alert; cTAKES: Clinical Text Analysis and Knowledge Extraction System; CUI: concept unique identifier; EHR electronic health record; HL7: Health Level 7; SMART-AI: Substance Misuse Algorithm for Referral to Treatment Using Artificial Intelligence; VM: virtual machine.

7.2 Study Description

To our knowledge, this is the first protocol for a bedside implementation of an AI screener. We anticipate that our protocol will serve as a guide for other health systems to utilize open-source tools within interoperable data standards and ontologies. This study provides an implementation framework and a cost-effectiveness analysis of an AI screener developed for the automated screening of hospitalized adults for opioid misuse. Our aim is to outline a hospital-wide protocol and computing architecture for implementing a real-time AI screener.

7.3 Study Groups/Arms

Preintervention Period: Usual Care With Ad Hoc Addiction Consultations:
The UW Hospital launched an inpatient Addiction Medicine consult service in 1991 to address the high prevalence of substance use disorders among hospitalized adults. A screening, brief intervention, and referral to treatment program was implemented for alcohol misuse, incorporating BPAs, intervention flow sheets, and consult order sets within EHR workflows for inpatient nurses and social workers. This program utilized the Alcohol Use Disorders Identification Test–Concise (AUDIT-C)[13] for identifying patients at risk for alcohol use disorder and facilitating order sets for withdrawal treatment. However, for opioid misuse, no formal screening process existed, and consultations were typically ad hoc, initiated at the discretion of the primary provider based on perceived risk.

Postintervention Period: Computing Architecture and Real-time Implementation:

The technical infrastructure enabling real-time AI screener implementation was built on advanced, industry-leading technologies. **Figure 2** illustrates the AI screener architecture, which involves exporting clinical notes from the EHR, processing them through an NLP pipeline to extract text features, inputting these features into the opioid screening deep learning model, and returning the resulting risk scores to the EHR in real-time as a BPA. This real-time AI screener consists of six key components, each described in detail in the Section 7.1.

The study will be conducted across multiple UW Health clinics in Madison, Wisconsin, covering diverse hospital wards and specialties. Each patient encounter will be evaluated independently across both study periods, enabling a comprehensive assessment of the intervention's impact within the UW Health system.

7.4 Data Collection

Data on study participants will be gathered through multiple established clinical operation methods:

- **Electronic Health Record**: Baseline and follow-up data will be extracted from the EHR, including clinician documentation practices, patient and provider demographics, billing codes, Addiction Medicine consult notes, medications prescribed, procedures and orders performed, follow-up encounters, Elixhauser comorbidity scores, provider documentation metrics, clinical data, payor data, vital signs, and problem lists. This data will be supplied by the Clinical Research Data Service (CRDS). Additional patient data will be collected via REDCap surveys, which directly access EHR data with provider review.
- **Surveys**: REDCap surveys will pull data from a daily Epic Workbench Report on screening results. The survey structure and data elements are detailed in the Data Dictionary, located in the Appendix C.

7.5 Duration of Intervention and Subject Participation

The intervention will be implemented over approximately 30 months (24 months of usual care and 6 months with implementation of automated screening).

7.6 Primary and Secondary Outcomes

The SMART-AI study included all hospitalized patients flagged by the AI screener as at-risk for opioid misuse. The primary outcome was the proportion of hospitalized patients receiving an Addiction Medicine consult following the AI screener’s recommendation, comparing the post-intervention period with the pre-intervention usual care period. The control sample was derived from hospitalized patients in the two years prior to the AI screener's implementation, who received ad hoc addiction consults based on provider discretion.

For the primary outcome, we assessed the percentage of inpatients flagged by the AI screener who received an Addiction Medicine consult, with any of the following interventions: (1) opioid use intervention or motivational interviewing (MI); (2) medication for opioid use disorder (MOUD); and/or (3) referral to substance use disorder treatment. This outcome will be reported as a percentage of total admissions in the pre- and post-intervention periods, reflecting the engagement of substance use screening and treatment services for at-risk hospitalized patients. Secondary outcomes include the 30-day unplanned

hospital readmission rate, with readmission criteria defined according to the Centers for Medicare & Medicaid Services (CMS) standards.[14]

7.7 Implementation and Education

The Consolidated Framework for Implementation Research (CFIR) guided the development of pre-implementation assessments and will continue to be utilized throughout the rapid Plan-Do-Study-Act (PDSA) cycles following the AI screener deployment.[15,16] Key stakeholder interviews were conducted to better understand the context, as well as to identify barriers and facilitators for implementing the Best Practice Alert (BPA) tool. Selected strategies from the Expert Recommendations for Implementing Change (ERIC) were employed to address identified barriers.[17] During pilot implementation, the implementation team held regular meetings to reflect on and evaluate challenges in adopting the BPA. A process evaluation was conducted with providers and addiction specialists to identify persistent barriers to BPA utilization and action. Clinical performance data will be collected and reviewed during PDSA cycles to provide guidance for clinicians and administrators in monitoring, evaluating, and modifying provider behavior.

Using the CFIR-ERIC matching tool, we will customize implementation strategies to enhance provider engagement with the AI screener. Additionally, interviews will be conducted with providers in non-pilot units to explore their perspectives and anticipated barriers to using the BPA. Following a three-month pilot period, efforts will focus on optimizing provider training, improving educational materials, and establishing quality monitoring processes in preparation for a hospital-wide rollout.

Clinical Operations Review Committee

This protocol will be presented to the UW Health Clinical AI and Predictive Analytics (CAIPA) committee for approval. Educational materials will be reviewed and refined based on feedback from key stakeholders. The primary goal is to ensure the effective integration of the AI screener into clinical workflows, enabling timely identification and intervention for hospitalized patients at risk for opioid misuse. The committee's review will focus on aligning the implementation with clinical operations to optimize patient outcomes and support sustainable improvements in OUD care.

7.8 End of Study Definition

The end of the study will occur after 6 months in the post-period.

8.0 SUBJECT SELECTION

8.1 Inclusion & Exclusion Criteria

The eligibility of clinicians participating will be determined by the inclusion criteria listed below. All participants will be a part of the UW Health system.

8.1.1 Inclusion Criteria

1. Healthcare Encounters in the UW Health Inpatient Clinic Setting
2. Individuals at least 18 years of age
3. Admitted to inpatient ward

4. At least 24 hours in hospital

8.1.2 Exclusion Criteria

1. Not receiving care in the surgical or medical inpatient wards

8.2 Vulnerable Populations

This study does not specifically include or exclude adult patients who may belong to vulnerable populations, such as pregnant women or incarcerated individuals, within the data analysis derived from electronic medical and billing records. These individuals are included as part of routine care, and no data will be intentionally recorded to identify members of these groups. Identifying participants as belonging to a vulnerable group (e.g., nursing home residents, incarcerated individuals) could increase their risk of confidentiality or privacy breaches regarding their vulnerable status. Health Information Management (HIM) and Clinical Documentation Improvement (CDI) staff will ensure that all clinical care documentation complies with the standards and requirements of the Centers for Medicare and Medicaid Services (CMS), maintaining patient privacy and confidentiality throughout the study.

8.3 Subject Identification

The AI tool will be implemented at the University of Wisconsin (UW) Hospital across the surgical and medical hospital inpatient wards. The EHR system at UW Health is Epic© (Epic Systems Corporation, Verona, WI, USA). All hospitalizations will be identified and tracked through the EHR.

8.4 Subject Recruitment

All adult inpatients at UW Health facilities, regardless of demographics or specific health conditions, will be automatically screened for opioid misuse risk using the AI screener embedded in the EHR. The study will include patients admitted to all hospital wards over the designated study period. This approach enables comprehensive, real-time screening within routine clinical workflows and does not require direct recruitment or consent from patients, as the intervention aligns with standard care processes.

As the AI screener generates BPAs for potential OUD, healthcare providers will review these alerts and decide whether to initiate an Addiction Medicine consultation based on the AI recommendations.
For broader study implementation, clinical staff, including residents, hospitalists, and primary care providers, will be informed of the AI screener's capabilities and the study objectives through a series of informational sessions and training materials. Additionally, updates on the study progress and functionality of the AI screener will be provided via regular communications from UW Health's IS operations team. The goal is to ensure provider engagement and familiarize staff with the screener's integration within the clinical workflow to optimize the identification and care of patients with potential OUD.

8.5 Remuneration and Retention Strategies

8.5.1 Monitoring and Support

The clinical operations team will monitor the AI screener's performance weekly through a real-time data dashboard embedded within the EHR. This dashboard will track key metrics, including the rate of AI screener alerts and provider response rates, and will include a control chart to detect any outlier activity. These metrics will serve as process measures to maintain fidelity to the study protocol and to identify any potential issues, such as false-positive or false-negative alerts that may affect clinical decision-making. Any adverse events, such as incorrect risk identification or unintended consequences of alert fatigue, will be promptly addressed by the study team in collaboration with the UW Health IS operations.

Provider support will be facilitated by the UW Health IT team, who will oversee technical aspects of the AI screener, as well as by the implementation team, via PDSA cycles and provide educational resources to enhance tool adoption and proper use. Any technical or administrative issues encountered by providers will be handled by the UW Health IT team to ensure consistent and effective support throughout the study.

8.5.2 Compensation and Incentives

- Participants in this study will not receive direct financial compensation, as their involvement is part of routine clinical care and documentation practices. However, several non-financial incentives are provided to support smooth participation and engagement:
- Enhanced Clinical Workflow: Providers benefit from the AI screener's integration within the EHR, which assists in identifying patients at risk for opioid misuse, potentially improving patient care and workflow efficiency.
- Technical and Administrative Support: Continuous support from the UW Health IT team and the AI screener technical team will be available to address any issues, ensuring seamless operation within clinical workflows.
- Professional Development: Participation offers providers the chance to be early adopters of advanced AI technology in healthcare, providing valuable experience in integrating AI into clinical practice, which may enhance their professional skills and knowledge.

8.5.3 Retention Strategies

- Retention strategies will focus on maintaining engagement and satisfaction among participants throughout the study period:

- Regular Feedback Opportunities: Providers will have the chance to share feedback on the AI screener's effectiveness and usability, allowing the study team to address any issues promptly and make continuous improvements.

- Ongoing Education: Training sessions and educational resources will be provided to ensure providers are comfortable with the AI screener and can effectively integrate it into their daily workflows, maximizing the tool’s clinical value.

- Performance Updates: Providers will receive updates on the AI screener’s impact on patient care, reinforcing the importance of their participation and the tool’s benefits to clinical operations.

9.0 STUDY AGENT AND PROCEDURAL INTERVENTION

9.1 Study Agent and Control Description

The study agent, SMART-AI, is implemented as a quality improvement initiative within UW Health to enhance clinical operations focused on screening and treating patients with substance use disorders. This AI-driven screening model was procured as part of UW Health’s commitment to identifying and addressing opioid misuse among hospitalized patients. Integrated directly into the EHR system (Epic, Inc.), the AI screener is designed to function seamlessly within UW Health’s existing infrastructure, supporting real-time clinical workflows for efficient and timely patient risk identification..

9.2 Study Procedural Intervention(s) Description

The AI screener processes clinical notes using NLP to identify and categorize opioid misuse-related information. Within the EHR, each note is analyzed to recognize medical terms and concepts, mapped to the Unified Medical Language System (UMLS). The UMLS Metathesaurus organizes named entities such as diseases, symptoms, medications, and procedures into Concept Unique Identifiers (CUIs). For example, the term "substance misuse" is mapped to CUI C4540992, distinguishing it from other related entities such as "family history of substance misuse" (C4540855). These CUIs are fed into machine learning models to build a predictive classifier for opioid misuse.

The classifier uses an embedding layer model with 300 dimensions. For model training, a multi-label classifier that can detect multiple substance misuse types simultaneously with binary labels (yes/no) was trained in a convolutional neural network (CNN). The model was trained on clinical notes from the first 24 hours of hospitalization to predict the likelihood of opioid misuse. The training dataset consisted of the first 26 months of hospitalizations from a large health system, with a 90/10 split for training and model selection. The primary evaluation metric was the macro-averaged area under the precision-recall curve (AUPRC), chosen to prioritize high positive predictive value and minimize false negatives.

9.3 Administration of Procedural Intervention

The AI Screener will be fully implemented after PDSA cycles are completed.

9.4 Procedures for Training of Clinicians on Procedural Intervention

Clinicians may attend a training session to go over the tool. In addition, educational flyers will be available for clinicians that contains information about the AI screener.

9.5 Method for Assigning to Treatment Groups

This was a quasi-experimental design with matched seosonal periods before and after AI screener implementation.

10.0 STUDY VISITS AND PROCEDURES

10.1 Study Calendar

The procedures performed at each study visit are listed in the table below.

Procedure	Pre-Period	Post-Period	Study Termination	Early Withdrawal
Visit Window	**2 years**	**6 months**	**12 ± 1 weeks**	
Demographics and Baseline Characteristics				

Procedure	Pre-Period	Post-Period	Study Termination	Early Withdrawal
Chart Review/Epic Workbench Audits		X	X	
Documentation Metrics (Secondary Outcomes)	X	X	X	
Training Session & EHR Integration Check		X		
Randomization	Na	Na	Na	na
Administer Study Intervention		X	X	
EHR Data Review	X	X	X	
Technical Support	X	X	X	
Optional Interviews		X	X	
Adverse Event Review and Evaluation		X	X	X

10.2 Screening and Enrollment

All adults hospitalized receiving the AI screening tool.

10.2.1 Informed Consent

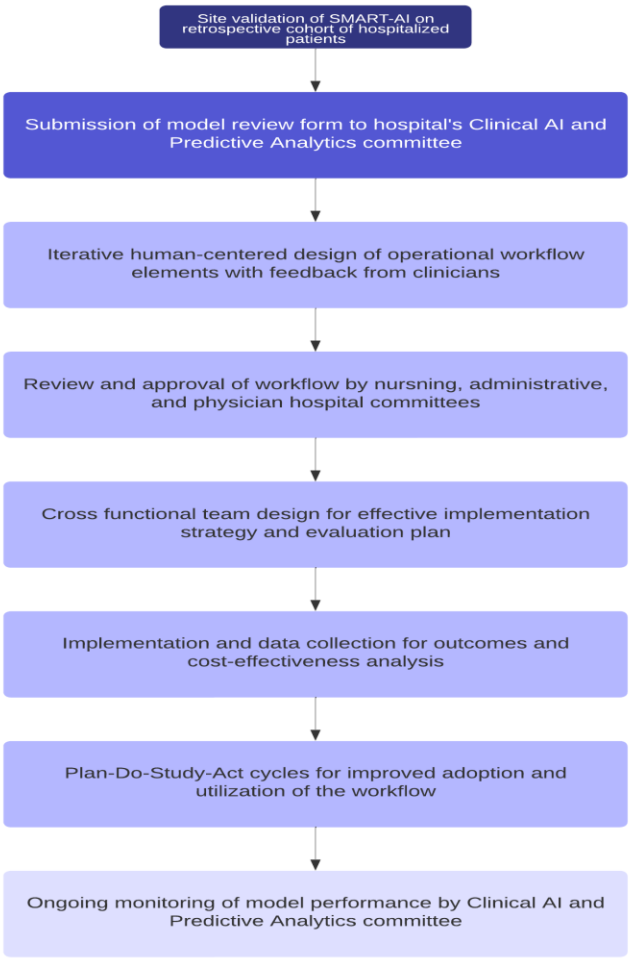
This study qualifies for waiver of consent and is exempt from IRB as it is a quality improvement initiative for the health system.

- Minimal Risk to Patients: The study involves only the use of de-identified medical record data, with robust safeguards in place to protect patient confidentiality. These activities present minimal risk to patients, as the data is used solely for quality improvement purposes.
- No Adverse Impact on Patients' Rights and Welfare: The study does not adversely affect the rights and welfare of patients, as data collection is restricted to authorized individuals with valid access to medical records. The use of patient data is for operational improvements and is not expected to impact the care provided to the patients involved.
- Necessity of Identifiable Information: The research cannot be conducted without identifiable information, as account numbers are required to link operational software data with the enterprise data warehouse. This linkage is necessary for accurate data aggregation and for conducting quality checks to ensure the integrity of the study data.

10.3 On-Study/Follow-up Visits

After patients have been screened, the on-study and follow-up visits will occur using all the EHR data during the study period.

Flow Diagram for the Process to Bedside Implementation and Evaluation



10.3.1 Data Collection

The following data will be collected during the on-study and follow-up visits:

- **Survey Data**: Collected daily, these surveys will capture chart audits on the AI Screener's peformance. See Appendix C.
- **EHR Data**: Continuous monitoring of AI tool usage through the EHR system will provide quantitative data on compliance and utilization rates. (See Data Dictionary in Appendix A)
- **Qualitative Feedback**: Insights from provider interviews and feedback forms will help contextualize the quantitative data and provide a comprehensive understanding of the BPA's impact.
- **Technical Support Records:** Documentation of any technical issues and their resolutions will be reviewed to ensure ongoing compliance and functionality of the AI tool.

10.4 Long-Term Follow-up | Re-contacting Subjects

No long-term follow-up or re-contacting providers is planned in this pragmatic design.

11.0 DATA HANDLING AND RECORD KEEPING

11.1 Data Collection

11.1.1 Data Collection Forms

The EHR data collected on patients ensure data collected are consistent and compliant with the protocol and IRB application.

Data collection is the responsibility of the clinical operations team in maintaining the EHR and its certification requirements for clinical operations. The PI and study team is responsible for ensuring the accuracy, completeness, legibility, and timeliness of the recorded and reported data.

EHR data are maintained in the subject files and retained as described in Section 11.3: Records Retention.

- Source records that will be used to collect data about participants:
  UW Health medical or billing records via ICTR's Clinical Research Data Service (CRDS)
  Data from UW Health Enterprise Data Warehouse (EDW)

11.1.2 Data Management Software System(s)

All Clinical data are collected in the UW Health Epic EHR. These data are provided back to the operations team via secure FTP in a secure and HIPAA-compliant data directory at UW Health and are considered metadata for UW Health's original data source. No new provider or patient data are provided beyond UW Health's own EHR and provider cohort.

11.1.2.1 *REDCap*

The School of Medicine and Public Health (SMPH) Research Electronic Data Capture (REDCap) system is used to manage the survey data for this study.

REDCap is a largely self-service, secure, web-based application for building and managing data collection forms. REDCap provides data management functionality by allowing the development of instruments and surveys to support data capture for research studies. All the surveys, including the PFI and secondary endpoints will be collected in REDCap.

11.2 Confidentiality and Privacy

Subject confidentiality and privacy are strictly held in trust by the participating investigators, UW Health staff, and ICTR staff. The study protocol, documentation, data, and all other information generated will be held in strict confidence

All study staff engaged in the conduct of this project have completed training on the protection of human subjects and the Health Insurance Portability and Accountability (HIPAA) Privacy Rule. In addition, all key personnel (i.e., Principal Investigator, individuals involved in identifying/recruiting subjects, obtaining informed consent, or interacting and intervening with subjects) have undergone Good Clinical Practice (GCP) training.

Information about study subjects will be kept confidential and managed according to HIPAA requirements. Study data will be maintained per federal, state, and institutional data policies.

The investigator(s) will ensure that the identities of providers are protected by using coded subject information. The log of subject identifying information that links subjects to their study-specific identification number will be maintained by the principal investigator. Electronic study records/files will be stored on UW Health Servers and secure SMPH devices or the cloud Platform X, and accessed via networked computers that are password-protected with access provided only to authorized study personnel. Identifiers on provider data will be destroyed at study closure or at the time of publication (whichever happens first). The patient data will be a limited dataset with timestamps but all other identifiers will be deidentified with pseudoIDs by the CRDS team.

Authorized representatives of the following groups may need to review this research as part of their responsibilities to protect research subjects: authorized representatives of UW Health, representatives of the IRB, and regulatory agencies or the operational oversight team. The clinical study site will permit access to such records.

Study staff may use e-mail to communicate with enrolled providers. The information contained in emails will be limited to study visit time and date information, and general questions pertaining to the study. All emails to subjects will be sent from secure institution-based accounts; personal, home or Gmail email accounts will not be used.

11.3 Records Retention

The investigator and ICTR team will retain study essential documents for a minimum period of 10 years following completion of the study per UW-Madison institutional policy.

11.4 Retention for Future Research: Data, Image, Audio- or Video-Recording

11.4.1 Purpose of Storage

As part of the UW Learning Health System, the pragmatic study data will be stored with the Clinical Research Data Service (CRDS) at UW School of Medicine and Public Health (SMPH) for secondary research purposes under an IRB approval to use the data.

11.4.2 Data Being Stored

As part of the UW Learning Health System, the data queries, data dictionary, and data procedure will be mirrored over from the UW Health GitHub instance to SMPH GitLab instance for secondary research use. All patient data except timestamps will be deidentified (limited dataset) and managed and stored on SMPH servers and devices and provided through Clinical Research Data Service (CRDS) team in ICTR's Center for Health Informatics Institute and the UW SMPH Honest Broker. Provider data will be stored with identifiers for the follow-up focus groups and interviews.

11.4.3 Location of Storage

All SQL code for data queries to extract the data element from the enterprise data warehouse (not containing any PHI) and the analytic code (in R or Python) to perform the statistics and data cleaning to ensure reproducibility will be maintained at

https://git.doit.wisc.edu/smph-public/dom/uw-icu-data-science-lab-public/smart-ai

11.4.4 Duration of Storage

Duration of storage will be at least 10 years.

11.4.5 Access to Data and Security Measures

Access to data, andrecordings and security measures will require IRB approval and request through CRDS.

11.4.6 Procedures to Release Data

Procedures to release data will only be provided after IRB approval and a data request form is submitted and approved by CRDS at https://ictr.wisc.edu/consults/biomedical-informatics/

A protocol deviation is any noncompliance with the clinical trial protocol or investigational plan requirements. The noncompliance may be either on the part of the subject, the investigator, or the study site staff. As a result of deviations, corrective actions are to be developed by the site and implemented promptly.

It is the responsibility of the Principal Investigator/site investigator/study staff to use continuous vigilance to identify and report deviations. The Principal Investigator is responsible for assessing whether the deviation constitutes noncompliance as defined by the reviewing IRB and if so, reporting it within the required time frame(s). The site investigator is responsible for knowing and adhering to the reviewing IRB requirements.

11.5 Publication and Data Sharing Policies

This study will be conducted in accordance with the following publication and data sharing policies and regulations. Final peer-reviewed journal manuscripts will be archived to PubMed Central (*https://www.ncbi.nlm.nih.gov/pmc/*) upon acceptance for publication. This study will comply with the NIH Data Sharing Policy and Policy on the Dissemination of NIH-Funded Clinical Trial Information and the Clinical Trials Registration and Results Information Submission rule. As such, this trial will be registered at ClinicalTrials.gov, and results information from this trial will be submitted to ClinicalTrials.gov. In addition, every attempt will be made to publish results in peer-reviewed journals

12.0 STUDY ANALYSIS

12.1 Statistical Hypotheses

Hypothesis testing for the intervention effects will be conducted using independent tests to compare the proportion of patients receiving motivational interviewing, medication for opioid use disorder (MOUD), and/or referrals for substance use disorder treatment. The primary hypothesis is framed as a non-inferiority test. The null hypothesis posits that the proportion of patients achieving the primary outcome (receiving these interventions) is lower (inferior) in the post-intervention period compared to the pre-intervention period. Specifically, the null hypothesis (H0) states that the difference in proportions (p1 for the pre-intervention period and p2 for the post-intervention period) exceeds a non-inferiority margin, M: H0: p1 - p2 > M.

The alternative hypothesis (H1) is that the proportion in the post-intervention period is non-inferior to the pre-intervention period, indicating that automated screening may be as effective as manual processes at identifying patients needing intervention while potentially offering cost-saving benefits: H1: p1 - p2 < M. This hypothesis will be tested using a one-tailed z-test for non-inferiority.

The study employs a non-inferiority design to demonstrate that the AI screener is as effective as manual screening methods in achieving key clinical outcomes while being less resource-intensive. In this context, achieving statistically superior performance may not be necessary for the AI screener to be adopted for hospital operations. Instead, if the AI screener is demonstrated to be non-inferior in terms of clinical outcomes and is cost-effective, it may be a desirable solution due to the operational benefits of efficiency, scalability, and reduced costs associated with IT infrastructure and vendor support.

12.2 Sample Size Justification

In hospital-wide screening, we expected a prevalence of 3% of adult inpatients with opioid misuse based on prior findings of hospital-wide analyses. A total sample size of 12500 patients with 10000 in the pre-intervention 2-year period and 2500 in the post-intervention 6-month period had 85% power to detect a difference of +0.75% in the post-intervention period (3.75%) compared to the pre-intervention period (3.0%) with a non-inferiority difference of -0.5% using a one-sided z-test with significance level = 0.025.

12.3 Subject Population(s) for Analysis

The Intention-to-Treat (ITT) population for this study will include all hospitalized patients screened by the AI screener, regardless of whether an Addiction Medicine consult was ultimately initiated. This population is essential for evaluating the effectiveness of the AI screener within real-world clinical conditions, as it reflects the population initially targeted by the automated screening process. Including all patients who were flagged by the AI screener preserves the applicability of the findings across various clinical settings and provides insight into the tool’s impact on patient care and operational efficiency within routine hospital workflows. Additionally, this approach allows for an accurate assessment of the screener’s scalability and cost-effectiveness as a screening tool embedded in the EHR..

12.4 Statistical Methods

The analytic cohort included all adult hospitalizations during the study period, with the primary outcome being the proportion of patients who received an inpatient Addiction Medicine consult, as documented by consult notes in the EHR. A one-tailed z-test was used to assess non-inferiority, testing whether the AI screener's impact on consult rates was comparable to traditional ad hoc consultations. The non-inferiority margin was set to ensure that the intervention's effectiveness in recommending Addiction Medicine consults would not fall below a clinically acceptable threshold compared to usual care.

For secondary outcomes, we examined the 30-day unplanned hospital readmission rate among patients who received an Addiction Medicine consult. Criteria from the Centers for Medicare and Medicaid Services (CMS) were applied to identify unplanned readmissions. To capture a comprehensive view of follow-up encounters, sensitivity analyses were conducted to include observation stays and emergency department visits post-discharge, addressing potential rehospitalizations that could occur outside traditional inpatient settings.

We employed a mixed-effects logistic regression model to estimate the odds of receiving an Addiction Medicine consult in the post-intervention period, adjusting for relevant confounders such as age, sex, race/ethnicity, insurance status, and comorbidity score. This model included a random intercept to account for repeated hospitalizations within individual patients, thereby adjusting for the within-subject correlation. The Elixhauser comorbidity index was used to quantify patient health status based on 31 comorbid conditions, providing an adjusted risk profile that enhances the reliability of outcome estimates. This approach allowed us to evaluate the AI screener's effectiveness while controlling for variations in patient acuity and baseline health characteristics.

12.5 Cost Effectiveness Analysis

Cost-effectiveness analysis was reported in terms of the incremental cost-effectiveness ratio (ICER) per additional patient who receives substance use treatment. For this study, the ICER was calculated as the difference between pre-implementation and post-implementation intervention costs divided by the difference between pre-implementation and post-implementation intervention effectiveness as measured by rates of patient engagement with substance use treatment services (i.e., primary outcome) and 30-day hospital readmission (i.e., secondary outcome).

The usual care control group and AI screener intervention group were characterized by the pathway probabilities of substance use treatment receipt and meeting the primary outcome. The pathway probabilities of patients' engagement with inpatient substance use consult, brief intervention/motivational interviewing (MI), medication assisted treatment (MAT), and referral to substance use treatment for both study groups would result in eight treatment combinations.

The differential costs pre- and post- AI screener intervention were determined as the difference in the weighted sum of the individual pathway costs, using the pathway probabilities as weights for the intervention and control groups. Effectiveness was determined as the difference in rates of hospitalized patients engaging with substance use disorder treatment pre- and post-implementation of AI screener for the intervention and control groups. The ICER was calculated as follows:

$$ICER = \frac{(Intervention\ Group\ Costs) - (Control\ Group\ Costs)}{(Intervention\ Group\ Treatment\ Rates) - (Control\ Group\ Treatment\ Rates)}$$

Sensitivity analyses will introduce uncertainty in substance use treatment receipt rates and costs for the intervention and control groups. Monte Carlo-based simulation estimation used the rates of substance use treatment service uptake observed in the intervention and control groups as a reference to simulate a cohort of post-implementation hospitalized patients and a cohort of usual care hospitalized patients. The ICER per additional individual who received an inpatient substance use consult, brief intervention, MI, MAT, or referral to substance use treatment was calculated by drawing a random sample with replacement from the observed distributions for health-care costs (μCOSTi) and substance use treatment services (μTRTi) for intervention and control groups. This process was repeated (N=1000) to produce bootstrap estimates of the 95% CI for the ICER per additional individual who received an inpatient substance use consult, brief intervention, MI, MAT, or referral to substance use treatment. These probabilistic sensitivity analyses estimated the elasticity of the differential cost per patient relative to differential substance use treatment service rates for intervention and control groups.

12.6 Handling of Missing Data

12.6.1 Documentation of Reasons for Missing Data

All instances of missing data will be thoroughly documented, with detailed notes on the reasons for the missing data.

12.6.2 Sensitivity Analysis

To evaluate the impact of missing data on study findings, sensitivity analyses will be performed. These analyses will estimate the effects of missing data under various assumptions regarding missingness, such as:

- **Missing Completely at Random (MCAR):** If the missingness has no relation to any data, observed or missing.
- **Missing at Random (MAR):** Assuming that the missingness is related to observed data but not to the missing data itself.
- **Missing Not at Random (MNAR):** Assuming that the missingness is related to the unobserved data.

Sensitivity analyses will involve comparing results obtained from multiple imputed data sets with those obtained from complete case analysis and other imputation methods. This comparison will help in assessing the robustness of the study findings to different assumptions about the missing data.

13.0 RISK/BENEFIT ASSESSMENT

13.1 Known Potential Benefits to the Subjects

13.1.1 Indirect Benefits to Patients

In addition to direct benefits, the study may yield several indirect advantages for patients and the healthcare environment as a whole:

- *Timely Identification and Intervention*: The AI screener enhances the hospital's ability to identify patients at risk for opioid misuse promptly, leading to more

timely Addiction Medicine consults. This early intervention may improve patient outcomes by reducing delays in care.

- *Improved Continuity of Care*: By integrating substance use risk identification into routine workflows, the AI screener supports a more seamless transition from hospitalization to follow-up care, potentially reducing rehospitalizations and supporting sustained recovery.
- *Increased Provider Efficiency*: By automating the screening process, the AI screener reduces the time providers spend manually identifying at-risk patients, allowing them to focus more on patient care. This can lead to better overall clinical workflow efficiency and improved resource allocation within the hospital setting.
- *Contribution to Patient Safety*: The systematic approach to screening may help reduce instances of missed diagnoses for opioid use disorder, which can contribute to increased patient safety and better quality of care.

13.1.2 Benefits to Science and Society

Beyond the immediate clinical benefits, this study holds potential for significant scientific and societal contributions:

- *Advancement of AI in Healthcare*: This study provides valuable data on the real-world application of AI in identifying substance misuse, informing best practices for future AI implementations in clinical environments.
- *Evidence-Based Insights*: The findings will contribute to the growing evidence base regarding the effectiveness of AI tools in automating substance misuse screening. This can guide future policy and clinical guidelines on integrating AI into healthcare.
- *Enhanced Healthcare Efficiency*: If successful, the study demonstrates the potential of AI to optimize hospital workflows and reduce resource strain, providing a model for other healthcare systems to enhance care delivery.
- *Support for Public Health Initiatives*: By addressing opioid misuse in a systematic manner, this study aligns with broader public health efforts to mitigate the opioid crisis, potentially contributing to reduced hospitalizations, decreased overdose rates, and improved long-term patient outcomes.

13.2 Known Potential Risks

Participating in this study poses minimal risks compared to standard care practices. However, certain potential risks associated with the AI screener should be acknowledged and mitigated. These risks fall into physical, psychological, sociological, economic, and legal categories.

13.2.1 Immediate and Long-Range Risks

- Physical Risks: There are no physical risks to patients or providers, as the AI screener operates within the EHR system without involving any direct or invasive interaction.
- Psychological Risks: Providers may experience stress or discomfort as they adapt to the AI screener in their daily workflow, particularly if they are unfamiliar with AI-driven tools. This risk will be mitigated through comprehensive training sessions and ongoing support from the UW Health IT team.

- Sociological Risks: Minimal sociological risks are expected. While the AI screener automates certain tasks, it does not interfere with patient-provider interactions. In fact, by reducing the time spent on manual screening, it may allow providers more time for direct patient care.
- Economic Risks: There are no economic risks to participants, as the AI screener is implemented within the UW Health system and does not incur additional costs for providers or patients.
- Legal Risks: The main legal risk involves the potential for data breaches, which could result in unauthorized access to sensitive patient information. This risk will be managed with robust data security measures, as outlined below.

13.2.2 Confidentiality and Privacy Risks

- The primary confidentiality risk is an inadvertent breach of patient information through unauthorized access. To minimize this risk, the following safeguards will be employed:
- Secure Data Storage: All data related to the study will be stored securely on password-protected servers maintained by UW Health. Access will be restricted to authorized research and operations team members only.
- Data Access Control: Access to raw data and patient records will be tightly controlled. Only specific, authorized individuals will have access to sensitive information, with all data access monitored and logged for compliance.

13.3 Alternatives Considered

Several alternative approaches were considered for improving substance use disorder screening, such as using manual screening protocols or other AI tools. However, these alternatives were not selected due to the health system's focus on implementing a scalable, real-time solution that integrates directly into the EHR. The AI screener was chosen based on its ability to provide timely, automated recommendations for Addiction Medicine consults, aligned with UW Health's quality improvement objectives and strategic priorities for addressing the opioid crisis.

13.4 Risk/Benefit Analysis

This study presents minimal risks to participants, mainly concerning data confidentiality and potential psychological stress as providers adapt to the AI screener within their workflow. These risks are mitigated through robust strategies, including secure data storage, restricted access, ongoing monitoring, and comprehensive training and support. The minimal risks involved are outweighed by the substantial potential benefits to providers, patients, and the broader healthcare system. Providers may experience enhanced efficiency and greater clinical support, while patients benefit from timely identification and intervention for opioid misuse. Additionally, the study contributes valuable insights into the integration of AI tools within clinical workflows, potentially setting a precedent for other healthcare systems. With strong protective measures in place, the study's benefits in improving healthcare delivery and operational efficiency justify the minimal risks involved.

14.0 DATA AND SAFETY MONITORING

This minimal-risk intervention study evaluating the impact of an AI-based screener poses no greater risks than those ordinarily encountered in daily life or during the performance of usual care. The AI Screener is a SaaS and is not regulated by the FDA for clinical use.

No adverse or serious adverse events are anticipated because of participating in this minimal-risk study. Annual debriefing to review the following plan with all study team members will be performed to ensure that expectations are clear and to provide refresher training on data safety monitoring processes as needed. In addition, the clinical operations team is continually monitoring with compliance of safety of the tool.

14.1 Reporting and Documentation

At predetermined intervals the study team will prepare reports as part of clinical operations for review by compliance with operational leaders. These reports will aggregate any adverse events or present them by treatment groups as requested. Serious adverse events will be reported in an expedited manner as they occur.

14.2 Study Monitoring

Monitoring for this study will be performed by the clinical operations governance team with the Analytic Workgroup, including members of the research team reviewing daily charts all the screens performed during the duration of the study. The monitoring occurs daily through Epic Workbench reports by research coordinators that are trained to examine for inaccuracies and issues with the software performance in documentation. In addition, monitoring by the IT staff is performed continually on data access, security, and compliance. The activities are all performed on-site by the operations team.

The AI model's predictions were assessed using Integrated Gradients (IG) to understand feature contributions from the medical concepts.[18] IGs operate by comparing the model's prediction on a given input with a baseline input that represents the absence of features (neutral state). This method calculates the gradients of the model's output relative to the input, tracing a path from the baseline to the actual input. By integrating these gradients, attribution scores were derived to highlight the contribution of each text feature to the final prediction. For this analysis, padded tokens were used as the baseline input for medical concepts. The analytics team examined the IGs post-hoc to monitor the model during deployment.

Study Stopping Rules
This study may be temporarily suspended or prematurely terminated if there is sufficient reasonable cause. Written notification, documenting the reason for study suspension or termination, will be provided by the operational leaders of the project. If the study is prematurely terminated or suspended, the PI will promptly inform the IRB and will provide the reason(s) for the termination or suspension.
Circumstances that may warrant termination or suspension include, but are not limited to:

- Determination of unexpected, significant, or unacceptable risk to subjects
- Demonstration of efficacy that would warrant stopping
- Determination of futility

The study may resume once concerns about safety and data quality are addressed and satisfy the applicable federal and institutional regulatory authorities.

15.0 STUDY FEASIBILITY

15.1 Economic Burden to Subjects

No anticipated costs that subjects may be responsible for.

15.2 Facilities and Locations

Providers from all UW Health Hospitals and Clinics are eligible to be subjects where study procedures will be performed. Recruitment procedures will be performed within UW Health.

15.3 Feasibility of Recruiting the Required Number of Subjects

With the large health system and infrastructure to implement the tool and administer the surveys, feasibility has been demonstrated.

15.4 Principal Investigator Considerations

15.4.1 Time Devoted to Conducting the Research

PI Afshar is the Sponsor for the clinical operation deployment of the Software and the Clinical Informaticist with dedicated effort to implement and study these emerging technologies. PI Afshar is also the Director of the Learning Health System in the SMPH Institute for Clinical and Translational Research and has effort dedicated to support the supervision and execution of this clinical trial.

16.0 REFERENCES

1. Osheroff, J., Osheroff, J. A., Levick, D. & Teich, J. M. *Improving Outcomes with Clinical Decision Support: An Implementer's Guide, Second Edition*. (HIMSS, 2012).
2. Gao, Y. *et al.* A scoping review of publicly available language tasks in clinical natural language processing. *J. Am. Med. Inform. Assoc. JAMIA* **29**, 1797–1806 (2022).
3. Afshar, M. *et al.* External validation of an opioid misuse machine learning classifier in hospitalized adult patients. *Addict. Sci. Clin. Pract.* **16**, 19 (2021).
4. Langabeer, J. R. *et al.* Prevalence and charges of opioid-related visits to U.S. emergency departments. *Drug Alcohol Depend.* **221**, 108568 (2021).
5. Calcaterra, S. L. *et al.* Management of opioid use disorder and associated conditions among hospitalized adults: A Consensus Statement from the Society of Hospital Medicine. *J. Hosp. Med.* **17**, 744–756 (2022).
6. Afshar, M. *et al.* Development and multimodal validation of a substance misuse algorithm for referral to treatment using artificial intelligence (SMART-AI): a retrospective deep learning study. *Lancet Digit. Health* **4**, e426–e435 (2022).
7. Ahmed, M. I. *et al.* A Systematic Review of the Barriers to the Implementation of Artificial Intelligence in Healthcare. *Cureus* **15**, e46454 (2023).
8. King, C., Cook, R., Korthuis, P. T., Morris, C. D. & Englander, H. Causes of Death in the 12 Months After Hospital Discharge Among Patients With Opioid Use Disorder. *J. Addict. Med.* **16**, 466–469 (2022).
9. P, M. *et al.* Discharge locations after hospitalizations involving opioid use disorder among medicare beneficiaries. *Addict. Sci. Clin. Pract.* **17**, (2022).

10. Fairley, M. *et al.* Cost-effectiveness of Treatments for Opioid Use Disorder. *JAMA Psychiatry* **78**, 767–777 (2021).
11. Savova, G. K. *et al.* Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. *J. Am. Med. Inform. Assoc. JAMIA* **17**, 507–513 (2010).
12. Yudko, E., Lozhkina, O. & Fouts, A. A comprehensive review of the psychometric properties of the Drug Abuse Screening Test. *J. Subst. Abuse Treat.* **32**, 189–198 (2007).
13. Bradley, K. A. *et al.* AUDIT-C as a brief screen for alcohol misuse in primary care. *Alcohol. Clin. Exp. Res.* **31**, 1208–1217 (2007).
14. Zuckerman, R. B., Sheingold, S. H., Orav, E. J., Ruhter, J. & Epstein, A. M. Readmissions, Observation, and the Hospital Readmissions Reduction Program. *N. Engl. J. Med.* **374**, 1543–1551 (2016).
15. Plan-Do-Study-Act (PDSA) Directions and Examples. https://www.ahrq.gov/health-literacy/improve/precautions/tool2b.html.
16. Damschroder, L. J. *et al.* Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. *Implement. Sci. IS* **4**, 50 (2009).
17. Powell, B. J. *et al.* A refined compilation of implementation strategies: results from the Expert Recommendations for Implementing Change (ERIC) project. *Implement. Sci. IS* **10**, 21 (2015).
18. Sundararajan, M., Taly, A. & Yan, Q. Axiomatic Attribution for Deep Networks. in *Proceedings of the 34th International Conference on Machine Learning* 3319–3328 (PMLR, 2017).

17.0 APPENDICES

17.1 Appendix A. EHR Data Dictionary

Groups	EPIC Location	*Variable Name*	Description
1	**Demographics**		
	Demographics	*age*	Age
	Demographics	*sex, race, etc.*	Other dems
	Surgical	*loc_cat (OR category)*	Prior operation/ intervention
	ADT	*loc_cat*	Location of patient
	Demographics	*discharge disposition*	Survived, expired, hospice, home, etc.
	Demographics	*Insurance status*	Private/Medicare/Medicaid/Uninsured/Other
	Epic Death Status		
2	**ICD10 Codes & DRGs**		All ICD 10 codes and MS DRGs
3	**Vital Signs- from nursing flowsheets**		
	Vital signs	*rr*	Respiratory rate

	Vital signs	*temp*	Temperature
	Vital signs	*pulse*	Heart rate
	Vital signs	*sbp*	SBP
	Vital signs	*dbp*	DBP
	Vital signs	*avpu*	Combination of 4 NA mental status variables
	Vital signs	*rass*	Richmond agitation sedation scale
	Vital signs	*gcs_score*	Glasgow coma scale
	Vital signs	*gcs_eye*	Glasgow coma scale - eye opening
	Vital signs	*gcs_motor*	Glasgow coma scale - best motor
	Vital signs	*gcs_verbal*	Glasgow coma scale - adult verbal response
	Vital signs	*Height*	
	Vital signs	*Weight*	
	Vital signs	*BMI*	
	Vital signs	*pain_scale*	Pain scale
4	**Laboratory Values**		
	Urine Drug Screen	*Urine toxicology*	Urinary toxicology result
	Laboratory results	*blood alcohol concentration*	Alcohol
	Laboratory results	*potassium*	Potassium
	Laboratory results	*chloride*	Chloride
	Laboratory results	*co2*	Bicarbonate
	Laboratory results	*bun*	BUN
	Laboratory results	*creatinine*	Creatinine
	Laboratory results	*glucose*	Glucose
	Laboratory results	*wbc*	WBC count
	Laboratory results	*hb*	Hemoglobin
	Laboratory results	*platelet*	Platelets
	Laboratory results	*protein*	Total protein
	Laboratory results	*albumin*	Albumin
	Laboratory results	*bili_total*	Total bilirubin
	Laboratory results	*sgot*	AST
	Laboratory results	*alt*	ALT
	Laboratory results	*alk_phos*	Alkaline phosphatase
	Laboratory results	*inr*	INR
	Laboratory results	*ptt*	PTT
	Laboratory results	*hematocrit*	Hematocrit
	Laboratory results	*lipase*	Lipase
	Laboratory results	*ammonia*	Ammonia
5	**Medications**		

	Medications	*all opioids*	Opioids
	Medications	*all benzos*	Benzos
	Medications	*all amphetamines*	Ampheta
	Medications	*All Naloxone*	Naloxone
	Medications	*drip_dexmedetomidine*	Dexmedetomidine drip
	Medications	*drip_fentanyl*	Fentanyl drip
	Medications	*drip_midazolam*	Midazolam drip
	Medications	*drip_morphine*	Morphine drip
	Medications	*drip_propofol*	Propofol drip
6	**Discharge Medications**		
7	**Diagnostics- from procedures/ orders in EPIC**		
	Orders	*admission*	Admission orders
	Orders	*Treatment limitations*	Treatment limitations
	Orders	*AODA*	Alcohol and Other Drug Abuse
	Orders	*DNR/I*	DNR DNI; do not resuscitate
8	**Nursing Assessments- from nursing flowsheet**		
	Nursing Assessments	*braden_scale*	Braden scale score
	Nursing Assessments	*braden_activity*	Degree of physical activity
	Nursing Assessments	*braden_friction*	Friction and shear
	Nursing Assessments	*braden_mobility*	Ability to change and control body position
	Nursing Assessments	*braden_moisture*	Degree to which skin is exposed to moisture
	Nursing Assessments	*braden_sensory*	Ability to respond meaningfully to pressure related discomfort
	Nursing Assessments	*braden_nutrition*	Usual food intake pattern
	Nursing Assessments	*morse_score*	Total Morse score
	Nursing Assessments	*morse_dateoffall*	Date of last fall
	Nursing Assessments	*morse_ambaid*	Ambulatory aid
	Nursing Assessments	*morse_iv*	IV/Heparin lock
	Nursing Assessments	*morse_gaittransfer*	Gait/transferring
	Nursing Assessments	*morse_fallhistory*	History of falling; immediate or within 3 months
	Nursing Assessments	*morse_mentalstat*	Mental status
	Nursing Assessments	*morse_secdiag*	Secondary diagnosis
	Nursing Assessments	CIWA_TOTAL	CIWA
	Nursing Assessments	CIWA_NAUSEA	NAUSEA AND VOMITING
	Nursing Assessments	CIWA_TACTILE	TACTILE DISTURBANCES
	Nursing Assessments	CIWA_AUDITORY	AUDITORY DISTURBANCES
	Nursing Assessments	CIWA_VISUAL	VISUAL DISTURBANCES

	Nursing Assessments	CIWA_TREMOR	TREMOR
	Nursing Assessments	CIWA_PAROXYSMAL	PAROXYSMAL SWEATS
	Nursing Assessments	CIWA_ANXIETY	ANXIETY
	Nursing Assessments	CIWA_HEADACHE	HEADACHE, FULLNESS
	Nursing Assessments	CIWA_AGITATION	AGITATION
	Nursing Assessments	CIWA_ORIENTATION	ORIENTATION AND CLOUDING
	Nursing Assessments	ALCOHOL	DO YOU DRINK ALCOHOL|TYPICAL WEEK ALCOHOL
	Nursing Assessments	BAC	BLOOD ALCOHOL CONTENT
	Nursing Assessments	AUDIT-C_1	HOW OFTEN DO YOU HAVE A DRINK CONTAINING ALCOHOL
	Nursing Assessments	AUDIT-C_2	HOW MANY DRINKS TYPICAL DAY
	Nursing Assessments	AUDIT-C_3	HOW OFTEN SIX OR MORE DRINKS
	Nursing Assessments	MARIJUANA	MARIJUANA|HASHISH
	Nursing Assessments	DRUG	COCAINE|OPIATE|METHAMP|HEROIN|STREET DRUG|MARIJUANA|HASHISH
	Nursing Assessments	AUDIT_TOTAL	AUDIT-C_1 + AUDIT-C_2 + AUDIT-C_3

9	**Financial Data**

17.2 Appendix B. Pseudocode for Independent Services to ingest, score, and report the NLP model

The code does not include the processing of the notes nor the trained model. Those are open-source and the online resources are reported separately at *https://git.doit.wisc.edu/smph-public/dom/uw-icu-data-science-lab-public/smart-ai*

HL7Listener

The purpose of the HL7 Listener is to respond to data pushed to it by Cloverleaf, parse the message and store relevant parts both to an encounter database table and to a file containing only the note text
When a new clinical note is authored by a clinician, Cloverleaf sends it to the HL7Listern service using TCP/IP:

```
WHEN BYTESTREAM received over TCP/IP:
        # BYTESTREAM is in HL7v2.3 pipe-delimited format; parse out the parts we need
        contactSerialNumber, noteID, noteText, inpatientStatus, patientDOB,
        admissionServiceLocation, inpatientStatus, patientAge, admissionServiceLocation
                = parse(BYTESTREAM)
        patientAge = CURRENT_DATE() - patientDOB
        IF timeSinceAdmission < 24 hours AND patientAge >= 18
                AND admissionServiceLocation = "University Hospital":
                INSERT contactSerialNumber, noteID, noteText INTO TABLE cnlp.EncounterText
                WRITE noteText TO FILE contactSerialNumber IN DIRECTORY CTAKES_INPUT
```

Scoring

The purpose of the Scoring module is to run CTAKES on files containing clinical notes to create CUIs, and then run SMART-AI on those CUIs to generate encounter scores. The results are stored in a database table

```
EVERY 5 MINUTES:
        CALL_USING_SHELL CTAKES
                with INPUT DIRECTORY CTAKES_INPUT and OUTPUT DIRECTORY CTAKES_OUTPUT
        newcuis = pandas.DataFrame()
        FOR EACH FILE file in CTAKES_OUTPUT:
                cui_string = PARSE_AND_EXTRACT_CUIS(file)
                data = {'csn': int(file.name), 'concepts': cui_string }
        newcuis.append(data)
        p = SMART_AI.predict_batch(newcuis)         # Score cuis with SMART-AI
        data = {'csn': p.csn,  p.opioid }
        INSERT data INTO TABLE cnlp.EncounterScores
```

MLFlowRESTAPIService

This service is a RESTful service that reports scores for a given encounter:

```
WHEN contactSerialNumber received over HTTP:
```

```
SELECT csn, scores FROM cnlp.EncounterScores WHERE csn = contactSerialNumber
RETURN as httpPost(csn, scores)
```