2022

An Examination Of Clinical Decision Support For Discharge Planning: Systematic Review, Simulation, And Natural Language Processing To Elucidate Referral Decision Making

Erin Elizabeth Kennedy
University of Pennsylvania

An Examination Of Clinical Decision Support For Discharge Planning: Systematic Review, Simulation, And Natural Language Processing To Elucidate Referral Decision Making

Abstract

Statement of the Problem: As healthcare data becomes increasingly prolific and older adult patient needs become more complex, there is opportunity for evidence-based technology such as clinical decision support systems (CDSS) to improve decision making at the point of care. Although CDSS for discharge planning is available, few published tools have been translated to new settings. Existing studies have not explored discordance between recommended and actual discharge disposition. Understanding the reasons why patients do not receive optimal post-acute care referrals is critical to improving the discharge planning process for older adults and their families. Methods: Three-paper dissertation examining CDSS. Paper 1 is a systematic review of studies with prediction models for post-acute care (PAC) destination. Paper 2 is a retrospective simulation of a discharge planning CDSS on electronic health record (EHR) data from two hospitals to examine differences in patient characteristics and 30-day readmission rates based on a CDSS recommendation among patients discharged home to self-care. Paper 3 is a natural language processing (NLP) study including retrospective analysis of narrative clinical notes to identify barriers to PAC among hospitalized older adults and create an NLP system to identify sentences containing negative patient preferences. Results: Most prediction models in the literature were developed for specific surgical populations using retrospective structured EHR data. Most models demonstrated high risk of bias and few published follow-up studies. In the simulation study, surgical patients identified by the CDSS as needing PAC but discharged home to self-care experienced adjusted 51.8% higher odds of 30-day readmission compared to those not identified. In the NLP study, the top three barriers were patient has a caregiver, negative preferences, and case management clinical reasoning. Most patients experienced multiple barriers. The negative preferences NLP system achieved an F1-Score of 0.916 using a deep learning model after internal validation. Conclusions: Future prediction modeling studies should follow TRIPOD guidelines to ensure rigorous reporting. Findings from the simulation and NLP studies suggest transportability of the CDSS to large urban academic health systems, especially among surgical patients. Incorporating natural language processing variables into CDSS tools may aid the identification of barriers to PAC.

Degree Type

Dissertation

Degree Name

Doctor of Philosophy (PhD)

Graduate Group

Nursing

First Advisor

Kathryn H. Bowles

Keywords

Clinical Decision Support, Geriatrics, Natural Language Processing, Nursing, Transitions in Care

Subject Categories

Databases and Information Systems | Medicine and Health Sciences | Nursing

AN EXAMINATION OF CLINICAL DECISION SUPPORT FOR DISCHARGE PLANNING: SYSTEMATIC REVIEW, SIMULATION, AND NATURAL LANGUAGE PROCESSING TO ELUCIDATE REFERRAL DECISION MAKING

Erin Elizabeth Kennedy

A DISSERTATION

in

Nursing

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2022

Supervisor of Dissertation

Kathryn Bowles, PhD, RN, FAAN, FACMI

Professor and van Ameringen Chair in Nursing Excellence

Graduate Group Chairperson

J. Margo Brooks Carthon, PhD, RN, FAAN

Tyson Family Endowed Term Chair for Gerontological Research, Associate Professor of Nursing

Dissertation Committee

Danielle Mowery, PhD, Assistant Professor and Chief Research Information Officer

Sarah J. Ratcliffe, PhD, Associate Chair for Research and Director of the Division of Biostatistics, University of Virginia

Subhash Aryal, PhD, Research Associate Professor

DEDICATION

To my childhood friend, Andee, for inspiring me to become a nurse when we were kids.

To my parents, Dan and Lisa, my brother Ryan, and my sister Lauren, who have always supported my career and sent me words of encouragement when I've needed it most.

To my grandmothers, Margaret and Carole, for being my original nursing role models.

To my extended family and friends who have supported me on my Penn Nursing journey.

ACKNOWLEDGMENT

I would like to acknowledge my dissertation chair, Dr. Kathy Bowles for supporting my growth as a researcher and believing in me when I first approached her about my interest in informatics as undergraduate at Penn Nursing. She selflessly dedicated so much of her time to mentoring me and provided me with many opportunities to collaborate with health systems, other universities, and informatics experts that I would not have had otherwise. Although a natural language processing dissertation was unconventional, Dr. Bowles enthusiastically supported my goals and helped me learn how to gracefully navigate this new territory. I am so thankful for Danielle Mowery for welcoming me into her natural language processing lab and teaching me everything I needed to know about NLP and python. Even on the most challenging days, her positive attitude motivated me to keep coding. I would also like to thank Drs. Sarah Ratcliffe and Subhash Aryal for helping me navigate statistical analysis and interpretation, as well as their expertise in analyzing electronic health record data. I am grateful for my readers of my Qualifying Exam, Candidacy Exam, and Dissertation, Dr. Pam Cacchione, Dr. George Demiris, and Dr. Karen Lasater for their excellent feedback over the years.

I am eternally grateful to Dr. Catherine McDonald, who helped me discover my passion for nursing research as an undergraduate and encouraged me to apply to the Hillman program. She has been an amazing mentor over the last 6 years and always supported my journey to informatics. I would like to thank Dr. Nancy Hodgson, Dr. Peggy Compton, and the Rita and Alex Hillman Foundation for fostering my interest in nursing innovation and supporting my journey from BSN to PhD. I am grateful to Dr. Mary Naylor for welcoming me into the NewCourtland Center for Transitions in Care, where I have been exposed to so many amazing leaders in care transitions and inspired to think about the impact of our work on a much larger policy scale. I would like to thank Dr. Max Topaz for introducing me to nursing NLP research and providing me with home health research experiences. I am thankful for librarians Richard James and Sherry Morgan for all of their help on my literature searches throughout college and graduate school.

I would also like to thank the Rita and Alex Hillman Foundation and the National Institutes of Health for funding this research. Research reported in this dissertation was supported by the National Institute Of Nursing Research of the National Institutes of Health under Award Numbers F31NR019919 and T32NR009356. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

I will close by acknowledging my friends and family for their unwavering support throughout my journey at Penn Nursing. Together, all of these people have profoundly influenced the course of my career and I feel so lucky to be surrounded by you all.

ABSTRACT

AN EXAMINATION OF CLINICAL DECISION SUPPORT FOR DISCHARGE PLANNING: SYSTEMATIC REVIEW, SIMULATION, AND NATURAL LANGUAGE PROCESSING TO ELUCIDATE REFERRAL DECISION MAKING

Erin Elizabeth Kennedy

Kathryn H. Bowles

Statement of the Problem: As healthcare data becomes increasingly prolific and older adult patient needs become more complex, there is opportunity for evidence-based technology such as clinical decision support systems (CDSS) to improve decision making at the point of care. Although CDSS for discharge planning is available, few published tools have been translated to new settings. Existing studies have not explored discordance between recommended and actual discharge disposition. Understanding the reasons why patients do not receive optimal post-acute care referrals is critical to improving the discharge planning process for older adults and their families.

Methods: Three-paper dissertation examining CDSS. Paper 1 is a systematic review of studies with prediction models for post-acute care (PAC) destination. Paper 2 is a retrospective simulation of a discharge planning CDSS on electronic health record (EHR) data from two hospitals to examine differences in patient characteristics and 30-day readmission rates based on a CDSS recommendation among patients discharged home to self-care. Paper 3 is a natural language processing (NLP) study including retrospective analysis of narrative clinical notes to identify barriers to PAC among hospitalized older adults and create an NLP system to identify sentences containing negative patient preferences.

Results: Most prediction models in the literature were developed for specific surgical populations using retrospective structured EHR data. Most models demonstrated high risk of bias and few published follow-up studies. In the simulation study, surgical patients identified by the CDSS as needing PAC but discharged home to self-care experienced adjusted 51.8% higher odds of 30-

day readmission compared to those not identified. In the NLP study, the top three barriers were patient has a caregiver, negative preferences, and case management clinical reasoning. Most patients experienced multiple barriers. The negative preferences NLP system achieved an F1-Score of 0.916 using a deep learning model after internal validation.

Conclusions: Future prediction modeling studies should follow TRIPOD guidelines to ensure rigorous reporting. Findings from the simulation and NLP studies suggest transportability of the CDSS to large urban academic health systems, especially among surgical patients. Incorporating natural language processing variables into CDSS tools may aid the identification of barriers to PAC.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

CHAPTER 1: INTRODUCTION

Background

Each year, clinicians refer 41.7% of Medicare's 13 million total hospital discharges to post-acute care (PAC) services[1,2] including long term acute care hospitals, inpatient rehabilitation, skilled nursing facilities, and home health care.[3] Coordinated discharge planning is an effective readmission risk reduction strategy that improves patient satisfaction and health outcomes.[4] Ideally, this coordination includes a complete assessment of the patients' needs planned collaboratively with, and in the context of the patient's caregivers, in a culturally competent way to facilitate informed decision making.[5]

Despite the high proportion of PAC referrals, discharge processes vary significantly at the patient, provider, and hospital level. According to a survey of over 1,000 hospitals enrolled in the National Hospital-to-Home quality improvement initiative to reduce unnecessary cardiovascular-related hospital readmissions, only 34% of participating hospitals estimate readmission risk for their patients in a standardized way.[6]

This problem is exacerbated by the complicated nature of discharge planning, which involves complex decision making, coordination across inpatient and outpatient settings as well as communication between patients, multiple disciplines of healthcare providers, and insurance companies.[7] A recent human factors study identified 14 roles involved in discharge planning and most patients had 1 to 6 people in their discharge planning network. The number of people in the discharge planning network among readmitted versus non-readmitted patients was not statistically significant, meaning that the size of the team is not associated with reductions in negative outcomes.[8] One explanation could be that clinicians increasingly face time constraints and communication issues on teams. Comprehensive training about PAC is not prevalent in medical education, which increases the need for consults to other specialties like social work and physical therapy.[9,10] Some providers do not value PAC,[11] and racial and gender disparities in PAC referrals are well-documented in cardiology and orthopedics.[12,13] Clinicians from different disciplines and/or different levels of experience might assess patients differently or have different perspectives about risk tolerance, contributing to subjective decision making.[4]

Patient and caregiver perspectives also influence PAC referrals. Although nearly 75% of older adults will need formal care at some point, only 40% of Americans expect to need it.[14,15] Family or caregiver preferences may impact the decision to pursue formal care.[16,17] Even when clinicians do recommend appropriate PAC, patients refuse up to 28% of the time, and these patients were readmitted at twice the rate of those who received PAC in one study.[18]

At the system level, hospital characteristics, geography, and insurance coverage are known barriers to PAC referrals.[19] Rural areas may have capacity constraints or limited PAC availability.[19,20] Insurance barriers include type of coverage, benefit limits for PAC, authorization requirements, narrow provider networks, ambiguity in medical necessity definitions, and lack of insurance.[19]

Together, these discharge planning problems contribute to unplanned hospital readmissions, unnecessary treatments,[21] increased costs,[22] and decreased patient satisfaction.[19] Although the focus tends to be on emergency care and hospitalization, in 2019 the president of the American Hospital Association encouraged policymakers to place more consideration on the PAC phase in which a patient begins the recovery process and may need help to improve function and transition back to their life prior to hospitalization.[23] Patients who receive coordinated discharge planning with evidence-based PAC referrals have better outcomes including reductions in hospital readmissions[24] and these patients experience fewer errors.[25] A recent systematic review of discharge communication practices found that well-designed technology solutions in discharge planning improve patient satisfaction and outcomes.[26]

In the years following the HITECH Act which promoted the implementation of electronic health records (EHRs) and health information exchanges, many researchers saw an opportunity to leverage this new electronic data source to develop solutions to improve discharge planning.[27] A 2009 literature review of opportunities for informatics in discharge planning identified outcomes related to information exchange, satisfaction, and communication.[28] Most tools utilize predictive analytics or data mining to build prediction models to reduce readmissions, determine discharge disposition, or minimize healthcare costs.[29] One of the most successful and widely implemented applications of predictive analytics has been clinical decision support systems (CDSS).

CDSS take prediction models and makes them into tools that enhance decision-making.[30] CDSS equip clinicians, patients, and other stakeholders with relevant and/or person-specific knowledge at appropriate times to improve decision making to ultimately enhance health and healthcare.[31] CDSS are frequently integrated into EHR workflows to improve decision making at the point of care. Common examples of CDSS include "computerized alerts and reminders to care providers and patients; clinical guidelines; condition-specific order sets; focused patient data reports and summaries; documentation templates; diagnostic support, and contextually relevant reference information (para. 1)."[30] Most CDSS tools utilize structured data, which includes EHR fields like age, laboratory values, and sociodemographic data. This is because structured data are easier to extract and process than unstructured data like clinical notes or images.

An early systematic review of computerized CDSS tools found that they improved clinical practice in 68% of overall clinical trials and up to 90% among trials with CDSS integrated into existing workflows with an actionable recommendation at the point of care.[32] Recently, CDSS have emerged for discharge planning in specific diseases and settings including chronic obstructive pulmonary disease,[33] pediatric gastroenteritis,[34] chest pain in the emergency department,[35] and transition to homecare.[36] As a result, newer systematic reviews of CDSS in specific clinical domains have emerged,[37-39] but there is little synthesis in CDSS research for PAC prediction models as a whole. Paper 1 of this study addresses this gap.

Conceptual Underpinnings of the Study

There have been over 50 models predicting PAC in discharge planning published in the literature, and over 30 have been internally and/or externally validated.[40] Although several of these models have been incorporated into CDSS, few follow-up studies have been published. This study will build upon a CDSS called the Discharge Referral Expert System for Care Transitions (DIRECT), which aims to help discharge planning teams determine older adults' PAC needs.[41,42] DIRECT was selected for three reasons: it is rooted in theory, is an expert-developed system, and was rigorously tested in real world clinical practice in a quasi-experimental study.

Before explaining how DIRECT was developed, it is critical to understand the importance of theory. The Cochrane Prognosis Methods Group recommends incorporating predictor variables grounded in theory, literature, and/or clinical expertise into prediction models.[43] Although some prediction modeling studies incorporate clinical experts in the variable selection process, very few cite theory. One of the major strengths of DIRECT is that to our knowledge, it is the only CDSS for PAC that incorporates all of Cochrane's recommended components, including being theory driven. It incorporates Orem's Self-Care Deficit Theory,[44] which is well-positioned for a CDSS study because it is an action theory with clear nurse and patient roles, as well as measurable concepts and relationships that may impact a patient's PAC needs.[45]

Orem's Self-Care Deficit Theory aims to answer why people need nursing, and posits that nursing is required when a patient is unable to perform continuous self-care. According to Orem, there are two categories of people: those who need nursing care (patients), and those who provide it (nurses).[45] Orem [44] viewed self-care as "the practice of activities that individuals initiate and perform on their own behalf in maintaining life, health, and well-being" (pg. 43). The self-care deficit can arise from health states that cause internal or external conditions and therefore activity limitations. The 10 basic conditioning factors that may impact self-care include age, gender, developmental state, health state, sociocultural orientation, health care system factors, family system factors, patterns of living, environmental factors, and socioeconomic factors.[46] When this occurs, the role of the nurse is to meet the dependent patient's health needs and/or help develop their ability to perform self-care.[45] During a hospitalization, a patient in a negative health state combined with other conditioning factors may limit their ability to perform self-care after discharge. Getting patients to PAC provides that needed nursing care. For example, in home health care a registered nurse may teach the patient how to manage their new medications, and a physical therapist may help them strengthen and overcome functional limitations. DIRECT incorporates predictor variables based on Orem's conditioning factors.

Next, it is important to understand why DIRECT is an expert driven CDSS.[41] Prediction modeling studies can include a variety of data sources such as randomized trials or cohorts, and it is common for the original data source to have a different purpose than the modeling study.[47] EHR data is the norm, and although it is compelling for its high volume of patients, it has several quality issues including heterogeneity, missing values, and lack of emphasis on expert knowledge.[48] When the models are trained based on how all clinicians make decisions about PAC needs in the EHR, the decisions of less experienced or biased clinicians are weighted equal to those of experts in the field.[49] When those models are used to drive CDSS in practice, they are recommending common decisions rather than the best-practice decisions, which could have negative implications for healthcare quality and patient outcomes.

DIRECT evolved from an earlier CDSS developed by the same team.[50,51] Both studies are described because the learnings from the prior CDSS informed the development of DIRECT. The Bowles et al.[51,52] studies mitigated the data quality issue by deriving the prediction models through interdisciplinary experts judging case studies derived from EHR data. The EHR data was drawn from nursing documentation of patient data during their holistic assessments over the course of a hospitalization. Bowles et al.[51,52] recruited physicians, nurses, social workers, and physical therapists with extensive discharge planning experience to evaluate case studies and reach consensus through Delphi rounds to create the data source for model development.

The model was trained with predictor variables from the EHR data and outcomes from the case study process. The original model included 6 predictor variables (availability of help, walking function, subjective help rating, length of stay, age, and number of comorbidities) to determine whether or not the patient needed PAC services.[51]

Typical performance measures in prediction modeling studies include discrimination, sensitivity, and specificity. Discrimination is the ability of a model to distinguish individuals who go to PAC from those who do not. Studies frequently plot sensitivity-specificity pairs in a receiver operating characteristic (ROC) curve for multiple probabilities and express discrimination as the area under the curve (AUC).[53] It is commonly assessed with the concordance index (c-index), which represents the area under the receiver-operating characteristic curve in logistic regression models.[43] Traditionally, AUC <0.70 is poor, 0.70-0.79 is fair, and ≥0.80 is good.[54] Sensitivity is the ability of the predictive model to correctly classify an individual will go to PAC, and specificity is the ability of a predictive model to correctly classify an individual who will not go to PAC.[55,56] The original Bowles et al.[51] model demonstrated good performance, with AUC 0.863, sensitivity 0.876, specificity 0.652, and overall predictive value 0.832 (0.801 in the cross-validated sample).[51]

Bowles et al.[52] expanded upon prior work by developing DIRECT. This tool was built on a larger sample and not only identifies which patients need PAC, but also recommends the level of care as facility-level or home-health care.[52] The algorithm within DIRECT was developed using the same approach as the prior study, based on consensus of expert multidisciplinary clinicians (doctors, nurses, physical therapists, and social workers) about the discharge disposition of 1498 case studies from 6 hospitals of adults aged 55 years or older, hospitalized for at least 48 hours, and discharged alive. The findings created a 2-step algorithm calculated from the values of structured predictor variables in the EHR that profile the patient characteristics associated with need for post-acute care (PAC). The first step of the algorithm that recommends whether a patient needs PAC (yes/no) is calculated from 17 predictor variables including activities of daily living, fall risk, equipment use at home, and other clinical variables. If the first step recommends care, the second step of the algorithm recommends the level of care as home health care or facility level care. The second step is calculated from 13 predictor variables including caregiver information, Braden pressure ulcer risk, functional status, activities of daily living and other clinical variables. DIRECT provides discharge planning teams with advice to indicate which patients need post-acute care (PAC) services, and the recommended level of care as home health care or facility level care.[52]

The other unique strength of these studies is that Bowles et al.[51,52] tested both models in real world clinical settings in NIH-funded studies. Few prediction modeling studies receive any funding, and even fewer are rigorously tested in clinical settings, which contributes to quality issues given the lack of peer review and reduces implementation potential.[48] Testing prediction models in experimental studies is crucial for dissemination and translation to other settings. The earlier CDSS was successfully commercialized and launched nationally[50,57] after demonstrating a 33% relative reduction in 30-day readmission rates in a quasi-experimental pre-post study in 3 hospitals. DIRECT was tested in a quasi-experimental pre-post study to evaluate the effects of DIRECT on PAC referrals and the patient outcome of acute care utilization in a suburban and community hospital in one health system.[42] The study demonstrated statistically significant reductions in readmissions at 7-, 14-, and 30- days with DIRECT compared to without DIRECT. Patient outcomes are optimized when the discharge disposition matches the algorithm’s recommendation, achieving a 22% relative reduction in readmission rates.[42]

Statement of the Problem

Although DIRECT demonstrated statistically significant reductions in 30-day readmissions in the quasi-experimental study in the two community hospitals, it has not been tested in other settings. Furthermore, it identified 25.6% more patients for PAC than actual discharge disposition, and reasons for this discordance in decision making and its consequences for patients were not explored.

One of the biggest shortcomings of current CDSS tools is that important clinical information lies in the unstructured (narrative) text of clinical notes, inaccessible to existing algorithms.[58] In addition to other implementation barriers, this could be one reason why a recent meta-analysis of 122 CDSS controlled clinical trials found that CDSS only increased the proportion of patients receiving recommended care by 5.8%.[59] Until recently, manual chart review was necessary to extract additional information from the clinical notes, which is costly, time-consuming, error-prone, and limits the number of notes that can be processed. Developing natural language processing (NLP) algorithms to read and classify clinical notes can automate this process, enabling thousands of notes to be processed systematically by a computer much more efficiently than a human. NLP “provides a means of ‘unlocking’ this important data source, converting unstructured text to structured, actionable data for use in applications for clinical decision support, quality assurance, and public health surveillance.”[60] A recent call to action from CDSS experts recommended the use of NLP to improve CDSS.[61]

Members of the interdisciplinary discharge planning team document relevant information about patients’ barriers to PAC over the course of a hospitalization in the unstructured clinical notes. Extracting this information using NLP could illuminate why patients do not receive appropriate CDSS-recommended PAC and inform future algorithm refinement or clinical interventions to get patients the care they need. Discharge planning notes are good candidates for NLP analysis and several NLP systems to analyze discharge planning notes have recently emerged, including those that identify social risk in adults,[62] psychiatric readmission risk,[63] adverse drug reactions,[64] care coordination,[65] and patient experience.[66]

Although it is likely that health systems refine CDSS at a local level to better serve their patients, these efforts are rarely disseminated. It is crucial to rigorously study CDSS beyond the organization where it is developed in order to better understand its weaknesses, identify areas for refinement, and publish the results so others can determine the potential applicability in different settings.[61] A recent study found that only 0.3% of published CDSS tools are replicated in the literature. Replication has the potential to improve both efficiency and effectiveness of CDSS as well as minimize harms related to technology in healthcare delivery.[67] Furthermore, widespread implementation of CDSS aligns with the Office of the National Coordinator's Health Information Technology's interoperability goals.[68] Papers 2 and 3 of the study address these gaps.

Purpose of the Study

This dissertation study expands the application of the DIRECT CDSS to a new setting and facilitate a deeper exploration of decision making. We conducted a systematic review of the literature to understand the state of prediction models driving CDSS in this domain in terms of clinical populations, predictors, development methods, quality, and performance. Then, we advanced DIRECT CDSS by applying the tool in a new population in a large urban academic health system with different leadership, resources, and a more diverse patient population than the suburban community hospitals in the original study. We computed the algorithm on a retrospective dataset to compare 30-day readmission rates among patients discharged without PAC who were identified by DIRECT as needing PAC to those where DIRECT and clinicians agreed on no referral for PAC. Further, among patients discharged home without services when DIRECT recommends care, we analyzed case management, social work, and discharge summary narrative notes to develop a reference standard of barriers to PAC among older adults. This method presented the opportunity to uncover novel barriers to PAC that are not well-understood in the literature. An automatic NLP classifier was developed to identify sentences from clinical notes containing the highest-value barrier, negative preferences. Negative preferences were defined as statements from the patient or family indicating that they prefer not to have PAC or are unsure. Findings support future algorithm refinement and may inform interventions to target patients at risk of not receiving appropriate PAC.

Aims and Hypotheses

Aim 1: Conduct a systematic review of studies reporting development and validation of models predicting PAC after adult inpatient hospitalization, summarize areas of model development and variables in the final models, evaluate model performance, and assess risk of bias and applicability using the Prediction Model Study Risk of Bias Assessment Tool (PROBAST).

Aim 2: Among patients discharged home without PAC, compare patient characteristics and 30-day readmission rates between those identified by DIRECT as needing PAC and those not identified as needing a PAC referral

Hypothesis: Among patients discharged home without services, those identified by the algorithm as needing PAC will be older, with more limitations in activities of daily living, more comorbidities, and more hospitalizations in the 6 months prior to hospitalization compared to patients not flagged for PAC. Those flagged by the algorithm as needing PAC will also experience higher rates of 30-day readmissions compared to patients not flagged for PAC.

Aim 3: Conduct an annotation study to identify common barriers to post-acute care, then develop and evaluate an NLP system to encode sentences containing negative preferences among hospitalized older adults.

Hypothesis: Patient refusal and insurance will be the most frequent barriers to post-acute care.

Assumptions, Limitations, and Design Controls

This study operated under a set of assumptions drawn from Orem's Self-Care Deficit Theory and the informatics literature. These assumptions include:

1. Nursing is needed when a patient is unable to perform self-care.[44]
2. Providing patients appropriate PAC after hospitalization provides needed nursing care.
3. Appropriate PAC prevents negative outcomes like 30-day readmissions.
4. CDSS aids clinician decision making.
5. 30-day readmission is an appropriate outcome to evaluate discharge planning interventions.
6. The EHR contains structured data entered by clinicians
7. Interdisciplinary discharge planning team members document barriers to PAC in their unstructured notes.

This study had limitations that we attempted to mitigate. Data from two hospitals in one health system may not be representative of the population. This concern was addressed by including a broad sample from a full year of data without limiting the study to a specific clinical population. Study data was limited to the instruments currently used in the health system's EHR, and EHR data has been associated with variations in quality. We interviewed health system stakeholders and used literature review[69] of EHR data reliability and validity studies which outlined several strategies to improve the quality of this data as a guide for our extraction strategy. The health system uses Elsevier's Clinical Practice Model EHR modules, which combine clinical practice guidelines and standardized assessments into daily nursing documentation.[70] Since important clinical information is recorded in unstructured notes, we analyzed clinical notes with NLP to add context to the structured information and improve accuracy of the data. Another limitation of EHR data is the inability to capture the full scope of readmission rates or mortality if patients are readmitted elsewhere and mortality outside of the health system. However, the study was conducted in a health system with six hospitals in the region with an integrated EHR system to capture readmissions data beyond the two hospitals in the study in five of the six hospitals.

This question-driven approach to secondary analysis had several strengths to improve internal and external validity. Our partnership with the health system strengthened internal validity by having access to detailed descriptions of the study population, codebooks, survey instruments embedded in the EHR, and documentation among different professions in the EHR.[71] Inclusion criteria and variables for analysis were determined a priori based on the original DIRECT study in order to externally simulate the tool and broadly encompassed most of the inpatient population at the two hospitals.[71] Finally, secondary analysis of existing EHR data provided access to a larger dataset than a typical experimental study, which enhances generalizability.[72]

Compared to traditional experimental studies, using existing EHR data is low-cost, provides access to a much larger dataset, and has time-saving benefits. The study utilized a team science approach between nursing, biomedical informatics, and biostatistics to gain a deeper understanding of CDSS algorithms, manage and analyze complex EHR data, as well as leverage NLP and advanced statistics methodology.

Definitions of Key Terms

There are several key terms that are important to understand the dissertation study, and they are defined as follows.

Annotation schema is a representation developed by the research team to identify, define, and provide examples of classes and sub-classes associated with a concept of interest in a natural language processing study. The schema is refined over the course of the annotation study.

Annotation study is the first stage of a natural language processing study, which includes the development and refinement of feature sets that encode relevant sentences according to annotation classes, representing structural descriptions and properties of text.[73]

Area under the Receiver Operating Curve (AUROC) is a common way of evaluating overall performance of a prediction model or algorithm across all possible classification thresholds. The x-axis is false positive rate (100-specificity) and the y-axis is the true positive rate (sensitivity). The highest possible AUC is 1.0 (perfect predictions) and the lowest possible score is 0 (model is 100% incorrect).[43]

Automatic Machine Learning Classifiers (AutoML) are applications that enable a user to input a dataset and test several machine learning models within one tool. The user can compare and select models based on different performance measures. This approach operates much more quickly than manually building machine learning models (decision trees, logistic regression, etc.).

Automation study is the second stage in a natural language processing study which involves the process of training and testing different algorithms based on the annotation study to learn a prediction model that correctly classifies sentences from text according to the annotation schema.[73]

Clinical decision support systems (CDSS) provide clinicians, staff, patients, or other stakeholders with person-specific information, intelligently filtered at the point of care, to enhance healthcare.[31] CDSS are developed based on prediction models.

Deep learning is a subfield of machine learning including algorithms inspired by the brain called neural networks. Deep learning can include both supervised approaches which rely on labeled data, and unsupervised approaches which rely on unlabeled datasets.

Discharge Referral Expert System for Care Transitions (DIRECT) is a 2-step discharge planning CDSS algorithm developed by Bowles et al. that identifies 1) whether or not a patient needs post-acute care, and 2) the level of care in which the patient should be referred to as home health care or facility care.[41]

Discharge planning is the individualized, interdisciplinary process of transitioning a patient from one level of care to another with the goal of ensuring continuity of care. The process ideally includes identification, goal setting, planning, implementation, coordination, and evaluation.[74]

Electronic health records (EHRs) are real-time, patient-centered records securely shared among relevant people and settings that integrate medical history, diagnoses, medications, care plans, allergies, imaging, test results, and unstructured notes.[75]

Extensible Human Oracle Suite of Tools (eHOST) is a java-based computer application that enables researchers to annotate text and compare annotations among annotators in a natural language processing study, creating a reference standard. To use, researchers highlight sentences in unstructured texts and tag them as classes.[76] eHOST supports comparison of annotations between annotators and calculation of inter-annotator agreement metrics.

F1-score is the harmonic mean of precision and recall, which is used to select a machine learning classifier with optimal performance. The highest possible F1- score is 1.0 (perfect precision and recall) and the lowest possible score is 0. This is often visualized using a precision-recall curve (PR-curve), which are preferred over AUROC curves when datasets are imbalanced.[77,78]

Feature engineering is the process of creating features for a machine learning model using raw text data.[79] Relevant types of features include:

> *Lexical features:* number of words shared by statement pairs (for example, grouping words into different lengths like one word, two words, sentences).
>
> *Semantic features:* higher-level text features about the meaning of a word or phrase including what something is or the role they play (for example, clinical role, sentiment, positive, negative).

Inter-annotator agreement is a measure of how well two or more annotators make the same annotation decision for a given class or subclass in the annotation schema.

Natural language processing explores how computers can be used to understand and manipulate natural language text or speech in order to perform desired tasks such as machine translation, text processing, information classification, and artificial intelligence.[80]

Natural language processing pipeline is the process of breaking up a large problem into small pieces and using machine learning to solve each smaller piece separately, then chaining several machine learning models that feed into each other to do complex tasks.[81]

Post-acute care (PAC) includes rehabilitation or palliative services that patients receive after an acute care hospitalization. Treatment may include skilled nursing care in a facility or at home.[3]

Precision is the fraction of examples classified as positive that are true positives. This is calculated as the number of true positives divided by the number of all positives, including false positives.[82]

Recall is the fraction of true positives that were classified as positive. It is calculated as the number of true positives divided by all samples that should have been classified as positive, including false negatives.[82]

Structured vs. unstructured data: Structured data is highly organized and formatted so it is easily searchable in relational databases (for example, vital signs, demographic data). Unstructured data has no predefined format or organization, making it more difficult to collect, process, and analyze (for example, clinical notes and images).[83]

Supervised learning requires labeled data, meaning that in natural language processing, annotators previously tagged unstructured data with correct labels. These methods help one use previous data to predict outcomes in new data.

Unsupervised learning utilizes unlabeled data to discover information by performing more complex processing. The downside is that it is more unpredictable, but it saves time compared to preparing a labeled dataset in supervised methods. These methods are often used to supplement supervised learning methods.[84]

30-day readmissions are unplanned readmissions that happen within 30 days of discharge from the index admission, or patients who are readmitted to the same hospital, or another applicable acute care hospital for any reason.[85]

Summary

The literature suggests that discharge planning and decision making are highly variable and when patients do not receive adequate PAC, they are more likely to experience poor outcomes, such as readmission. A body of research on CDSS for discharge referral decision making demonstrates the positive effect of these tools on patient outcomes.[42,50,52,57,86] This dissertation fulfilled the opportunity to apply the DIRECT algorithm in a new setting, examine the impact on patient outcomes, and to explore the reasons why up to 25.6% of patients do not receive recommend care.

The study is innovative for three reasons: 1) It is one of the first nursing studies to use NLP to understand discharge planning decision making. 2) It advances the science of discharge planning CDSS by examining the impact and reasons for decisional discordance, which holds the potential for future intervention development to target at-risk patients. 3) The study also leverages nurse-generated EHR data to run the algorithm and support decision making.

This study is a 3-paper dissertation. Chapter 2—Paper 1 is the Systematic Review of Prediction Models for Postacute Care Destination Decision-Making, published in the Journal of the American Medical Informatics Association.[87] Chapter 3—Paper 2 is the Comparison of Clinical Decision Support Recommendation for Discharge Disposition to Usual Decision Making: Evaluation of 30-Day Readmissions. Chapter 4—Paper 3 is Identifying Barriers to Post-Acute Care Referral and Characterizing Negative Patient Preferences Among Hospitalized Older Adults Using Natural Language Processing. This paper is currently under review in the student paper competition for the American Medical Informatics Association Symposium.

CHAPTER 2: PAPER 1

SYSTEMATIC REVIEW OF PREDICTION MODELS FOR POSTACUTE CARE

DESTINATION DECISION-MAKING

*A published version of Chapter 2 appears in: Kennedy EE, Bowles KH, Aryal S. Systematic review of prediction models for postacute care destination decision-making. *Journal of the American Medical Informatics Association.* 2021.

Abstract

Objective: This article reports a systematic review of studies containing development and validation of models predicting post-acute care destination after adult inpatient hospitalization, summarizes clinical populations and variables, evaluates model performance, assesses risk of bias and applicability, and makes recommendations to reduce bias in future models.

Materials and Methods: A systematic literature review was conducted following PRISMA guidelines and the Cochrane Prognosis Methods Group criteria. Online databases were searched in June 2020 to identify all published studies in this area. Data was extracted based on the CHARMS checklist, and studies were evaluated based on predictor variables, validation, performance in validation, risk of bias, and applicability using the Prediction Model Risk of Bias Assessment (PROBAST) tool.

Results: The final sample contained 28 articles with 35 models for evaluation. Models focused on surgical (22), medical (5), or both (8) populations. Eighteen models were internally validated, 10 were externally validated, and 7 models underwent both types. Model performance varied within and across populations. Most models used retrospective data, the median number of predictors was 8.5, and most models demonstrated risk of bias.

Discussion and Conclusion: Prediction modeling studies for post-acute care destinations are becoming more prolific in the literature, but model development and validation strategies are inconsistent, and performance is variable. Most models are developed using regression, but machine learning methods are increasing in frequency. Future studies should ensure rigorous variable selection and follow TRIPOD guidelines. Only 14% of the models have been tested or implemented beyond original studies, so translation into practice requires further investigation.

Background and Significance

Each year, approximately 29 million American adults are discharged from acute care hospitalizations.[88] Among 13 million Medicare beneficiaries discharged annually, 41.7% receive referrals to post-acute care (PAC).[1,2] PAC includes long-term acute care hospitals, inpatient rehabilitation facilities, skilled nursing facilities, and home health care.[3] PAC referral is a central part of discharge planning, which ideally includes an assessment of patient needs and shared decision making in a culturally competent manner.[5] This decision making is complex and requires communication between multiple clinical disciplines, insurance companies, patients, and families, as well as coordination across inpatient and outpatient settings.[7] The logistics of discharge planning are challenging, and a human factors study found that patients have 1-6 other people involved in the process.[89] Although referrals occur daily and directly impact the patient's health and outcomes, there is significant variation in discharge destination decision making at the patient, provider, and system level.[90]

Providers may be biased in their PAC referrals. A mixed-methods study revealed that some cardiologists do not value sending patients to cardiac rehabilitation, while others consider it the standard of care.[11] Patient preferences and expectations also impact PAC. Even though 75% of older Americans will use formal services at some point, only 40% expect to.[3,90] External factors like family members, insurance coverage, and geographic location may influence decision-making. Without standardized discharge planning and PAC referrals, patients face serious risks including unplanned readmissions, increased costs, unnecessary treatment, and decreased satisfaction.[7,19] One study found that patients who refuse PAC have twice-higher odds of 30-day readmissions.[18] Prediction modeling is a time- and cost-efficient strategy to reduce bias in decision making with standardized clinical decision support approaches and has become increasingly common[91] with new technologies.[61]

Systematic reviews that evaluate models predicting 30-day readmissions[92] and discharge disposition in specific populations including stroke[93] are becoming more common as healthcare incorporates more clinical decision support technology. Only one systematic review of models predicting supportive care after hospitalization has been published, but it is limited to medical patients and excluded rehabilitation and long-term acute care destinations.[94] To our knowledge, no systematic reviews of models predicting PAC across patient populations and all discharge destinations exist. In this review, discharge disposition, supportive care after hospitalization, and PAC destination represent the same concept of where a patient transitions to after a hospitalization and will be referred to as PAC destination.

Objective

The goal was to conduct a systematic review of studies reporting development and validation of models predicting PAC destinations, summarize areas of model development and variables in the final models, evaluate model performance, trends over time, and assess risk of bias and applicability using the Prediction Model Risk of Bias Assessment Tool (PROBAST).[43] This information could be used for those implementing models in real-world healthcare settings at the point of care through technology like clinical decision support systems and/or future model development or refinement efforts.

Materials and Methods

All study procedures were conducted according to the PRISMA guidelines[95] (**Appendix A**) and Cochrane Prognosis Methods Group systematic review tools, which include guidance for search strategies, data extraction, risk of bias, and reporting.[96]

Search Strategy and Study Selection

The goal was to identify studies that developed and evaluated models to predict PAC destination for hospitalized adults and described the predictors, development and validation statistics. We excluded studies in non-hospitalization or pediatric settings, studies with models predicting outcomes other than PAC destination, studies that aimed to identify predictors without validation, or studies that only performed external validation with or without model updating because they require different evaluation criteria.[97] A research librarian was consulted to develop a search strategy in PubMed, CINAHL, and Embase for English language studies published before June 5, 2020. Complete strategies are in **Appendix B**. The main search term combination included adult, inpatient, prediction models/algorithms/clinical decision support, referral, discharge/post-acute care. Additional papers were added by cross-checking reference lists.

References were imported into DistillerSR for article screening.[98] One investigator (EEK) conducted study selection in a two-step process, which was verified by a second investigator (KHB). Articles were screened for eligibility by reading titles and abstracts. Prediction models were defined as prognostic models that predict PAC destination among hospitalized adults using regression or non-regression techniques.[43] Data was extracted according to the Checklist for Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies (CHARMS).[97]

Risk of Bias Assessment

PROBAST guided the risk of bias and applicability assessment. Bias is defined as the “systematic error in a study that leads to distorted or flawed results and hampers the study’s internal validity” and applicability compares the population, predictors, or outcomes in the study question to the review question.[43] The PROBAST contains four domains (participants, predictors, outcomes, and analysis) with 20 “signaling” questions that guide evaluation.

The PROBAST tool is very conservative. Risk is considered high for the entire domain if ≥1 signaling question(s) poses a risk, and overall models are at risk of bias if one or more domains demonstrates high risk of bias. One investigator (EEK) completed the PROBAST assessment, and a second investigator (KHB) verified findings. A third, biostatistician investigator (SA) evaluated statistical concerns.

Assessment of Model Performance

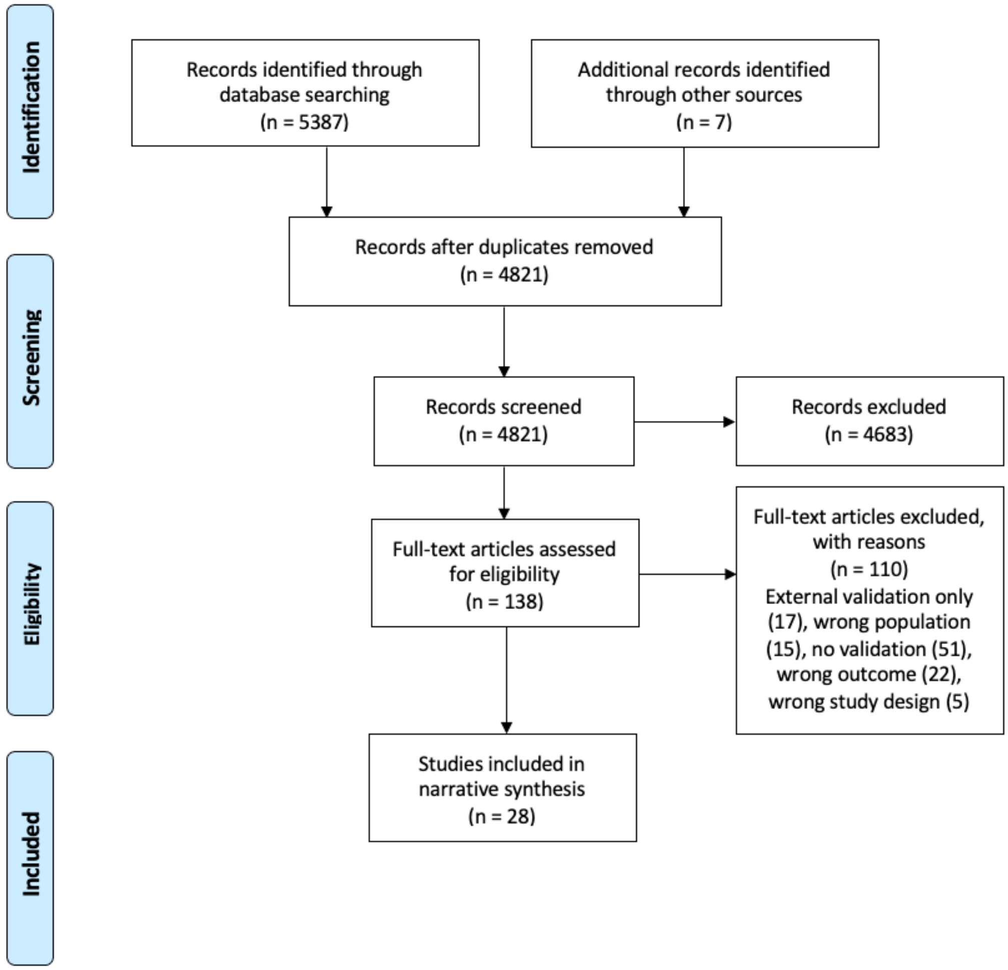
In accordance with PROBAST recommendations, discrimination and calibration of internally and/or externally validated models determined model performance.[43] Discrimination is a model’s ability to distinguish between individuals who go to PAC from those who do not represented as a plot of sensitivity-specificity pairs in a receiver operating characteristic (ROC) curve and is expressed as the concordance index (c-index).[43,53] In this review, an AUC <0.70 is poor, 0.70-0.79 is fair, and ≥0.80 is good.[54] Calibration is the agreement between model predictions and observed outcomes commonly reported as the Hosmer-Lemeshow goodness of fit test, or with additional detail using calibration plots or tables.[54]

Results

Study Selection

Figure 1 illustrates article screening and selection. 5,387 records were identified through the PubMed, CINAHL, and Embase databases. Seven articles were identified from reference lists.

Figure 1: Article Screening from Literature Search to Final Sample



566 duplicates were removed, then 4,821 records were screened for title and abstract. 137 articles were screened for full text, and 17 articles were excluded for performing only external validation, 15 for wrong population (ICU only, pediatric), 51 for not performing validation (model development only), 22 for wrong outcome (hospital readmissions, composite negative outcome, long-term outcomes), and 5 for design (editorial, presentation). The final sample contained 28 articles with 35 models (6 articles developed 2 models;[99-104] 1 article developed a two-step model).[105]

Study and Model Characteristics

Table 1 reports the model characteristics. Fourteen models were derived from retrospective electronic health record (EHR) data,[51,100,102,104-110] 14 from registries or data warehouses,[99,101,111-120] and 4 used a combination.[103,121,122] Three models collected data prospectively.[123-125] Six models focused on older adults.[51,105,121,122,125] Four studies were international.[115,123-125] Populations included 5 medical (2 general,[123,125] 3 stroke[99,118]), 22 surgical (5 cardiac,[103,108,119,121] 3 gastrointestinal,[101,112] 2 gynecologic,[111,117] 8 total joint arthroplasty (TJA),[102,104,107,109,122,124] 1 transplant,[110] 3 spine[113,114,116]), and 8 models included both medical and surgical populations (1 cardiac,[120] 2 falls,[100] 4 general,[51,105,106] 1 isolated lower extremity fracture[115]). Thirty-four models used binary outcomes (including combined outcomes), and only 1 study created multinomial prediction models for each outcome.[119] Twenty-seven models created combined outcomes to predict non-home discharge, with great variability in outcome definitions. For example, one study's combined binary outcome was facility (including skilled nursing facility and inpatient rehabilitation facility) or home (including home health)[123] while another study's binary outcome was inpatient rehabilitation facility or home.[124] Among models that used individual outcomes, 2 models predicted home discharge,[118,124] 3 predicted inpatient rehabilitation facility[110,115,116] and 1 predicted skilled nursing facility.[121]

Table 1: Study Characteristics

Author	Model	Prediction Model Type+	Purpose	Data Source & Study Population (P)	Inclusion (I) and Exclusion (E) Criteria	N*	Study Design #	Outcome~
AlHilli, M. M. et al (2013)[111]		LR	Identify independent risk factors for non-home discharge and propose a risk-scoring tool to identify patients at increased risk of non-home discharge	American College of Surgeons (ACS) National Surgical Quality Improvement Program (NSQIP) database **P:** Ovarian cancer surgical patients	**I:** Surgical staging and/or primary cytoreduction for EOC, primary peritoneal carcinoma, or fallopian tube cancer **E:** Neoadjuvant chemotherapy, recurrent disease, nonepithelial malignancy, previous surgical diagnosis of their cancer	587	RO	NHD (SNF, IRF, D, hospice) vs. home
Bailey, E. A. et al (2017)[112]		LR	Examine the association between discharge status, hospital duration of stay, and cost for colorectal operation patients without complications and uses risk factors to predict the need for post-acute are	New York Statewide Planning and Research Cooperative System and California HCUPS Databases **P:** Colorectal cancer surgery patients	**I:** Discharged home or post-acute care after operative resection for colorectal cancer, 18 years of age **E:** Postoperative complication, LOS greater than 75th percentile, not discharged home or to PAC	23,942	RO	PAC (SNF, IRF) vs. home
Ballester, N. et al (2018)[106]		LR	Develop an approach that could serve as an early warning decision aid to care providers for predicting, within 24 hours of admission, the discharge disposition of hospitalized veterans based on the available clinical and health utilization factors at index and previous hospitalizations	VA Boston Healthcare System (VA-BHS) corporate data warehouse (US) **P:** General inpatients	**I:** Adult, inpatient medical service including general, cardiac intensive care unit, medical ICU, medical step down, telemetry, and hospice for acute care **E:** Patients with missing discharge disposition	4,760 (D: 3,351; V: 1,409)	RO	NHD (VA nursing home care units, non-VA NH) vs. home

Barsoum, W. K. et al (2010)[107]		LR	Develop an easily administered tool to preoperatively predict patient discharge disposition after total joint arthroplasty	EHR data from 1 hospital, external validation from 3 hospitals in 1 health system **P:** Total hip/knee arthroplasty patients	**I:** Primary TKA, revision TKA, bilateral TKA, primary THA, revision THA **E:** Mortality (although none died)	667 (D: 517; V: 150)	RO	NHD vs. home
Bowles, K. H. et al (2009)[51]		LR	Elicit expert knowledge about factors important to referral decision making and identify characteristics of hospitalized patients who need a PAC referral	Retrospective and prospective EHR data from 6 northeast hospitals (urban, suburban, rural) from other research studies and 8 case study experts **P:** Hospitalized medical and surgical patients	**I:** 65 years or older, English speaking, cognitively intact, expected to be discharged home **E:** Missing data, not readable, cases too similar, cases used to train abstractors	355	MM	PAC (HHC, ORF, NH, SNF, IRF) vs. home
Bowles, K. H. et al (2017)[105]	Step 1	PR	Build and validate a clinical decision support (CDS) algorithm for discharge decisions regarding referral for post-acute care and to what site of care	EHR data from 6 hospitals in New England, Mid-Atlantic, Midwest US **P:** Hospitalized medical, surgical, critical care patients	**I:** Age 55 or older; medical, surgical, or critical care units **E:** Observation stays, admissions to skilled rehabilitation, obstetrics, and pediatrics	1,496 (D: 1,251; V: 245)	MM	Step 1: Need for PAC
	Step 2							If Step 1 is yes: Step 2: Need for facility PAC vs. HHC
Chang, D. C. et al (2007)[121]		LR	Identify which preoperative risk factors were associated with admission to SNF and develop a predictive index from these data to help clinicians counsel older patients considering CABG	California Hospital discharge database and EHR data from Johns Hopkins (validation) **P:** CABG surgery patients	**I:** 65 years or older, ICD-9 procedure code 36.1 for CABG cases **E:** Valve repairs, patients admitted from SNF/residential care facility/other hospital setting	D: 26,040; V: not specified	RO	SNF (including IC) vs. not SNF

Cho, J. S. et al (2017)[99]	Extended Model	LR	Evaluate the association of selected patient characteristics with hospital discharge disposition status and predict such status at the time of an acute stroke admission	Hospital Discharge Data System maintained by Tennessee Department of Health **P:** Hospitalized stroke patients	**I:** Principal diagnosis of stroke **E:** Missing data for any field, deceased, discharged to hospice, discontinued care/court	127,581 (D: 101,223; V: 26,358)	RO	Facility (SNF, IC, IRF, hospital) vs. home (including HHC)
	Simplified Model							
James, M. K et al (2018)[100]	Early	LR	Examine clinical and non-clinical factors that may predict discharge disposition after hospitalization for a fall	EHR data from an urban level one trauma center **P:** Hospitalized falls patients	**I:** Adult, falls patients including trauma activations or surgical consult **E:** Under 18 years of age, not admitted to hospital, if fall was secondary to a medical condition, left AMA, died, or were transferred to another hospital	1,121	RO	Facility (IRF/SNF) vs. home (including HHC)
	Late							

Karhade, A. V. et al (2018)[113]	ML (ANN)	Use machine learning algorithms to develop an open-access web application for preoperative prediction of nonroutine discharges in surgery for elective inpatient lumbar degenerative disc disorders	National Surgical Quality Improvement Program (NSQIP) database **P:** Elective lumbar degenerative disk disorder surgery patients	**I:** Inpatient operation, elective surgery, current procedural terminology code for decompression or decompression and fusion at lumbar levels, primary post-op diagnosis of ICD for lumbar disc displacement or lumbar disc degeneration, general anesthesia, ASA classification I-IV 7, operation 2011-2016 **E:** preoperative wound infection; preoperative SIRS, sepsis or septic shock; emergency surgery; admission from any setting other than home; ventilator dependent preoperatively	26,364 (D: 21,091; V: 5,273)	RO	Non-routine discharge (NHD) vs. routine discharge
Karnuta, J. M. et al (2020)[114]	ML	Develop a Naïve Bayes machine-learning model to predict inpatient payments, LOS, discharge disposition following dorsal and lumbar fusion for non-scoliosis indications	New York State Department of Health's Statewide Planning and Research Cooperative System (SPARCS) administrative database **P:** Surgical dorsal and lumbar fusion patients	**I:** Dorsal and lumbar fusion for reasons other than curvature of the back, Medicare beneficiaries **E:** Patients with a CCS diagnosis code that contained only one patient, curvature of the back, osteoporosis, acquired foot deformities, upper limb fracture, lower limb fracture indication	38,070	RO	NHD (SNF, IRF, other) vs. Home (including HHC)

Kimmel L.A. et al (2011)[115]	LR	Develop a prognostic model for discharge to inpatient rehabilitation	Victorian Orthopedic Trauma Outcomes Registry (VOTOR) **P:** Isolated lower limb fracture patients	**I:** Admitted to VOTOR participating hospital between 3/07 and 11/08, aged 18 or older, admitted for management of isolated lower limb fracture **E:** Neck of femur fracture, additional injuries other than minor lacerations/abrasions, brief loss of consciousness without neurological sequelae, died during hospital stay	1,429 (D: 690; V: 739)	RO	IRF vs. home
Louis Simonet, M. et al (2008)[123]	LR	Develop and validate a score predicting discharge to a PAC facility and to determine its best assessment time	Prospective patients in one hospital **P:** Hospitalized medical patients	**I:** Medical, discharge to home or PAC facility (SNF or IRF) **E:** Comatose or terminally ill on admission, died in the hospital after enrollment, transferred to other acute care settings, or discharged to a nursing home where they lived prior	460 (D: 299; V: 161)	PO	Facility (SNF, IRF) vs. home (including HHC)

McGirt, Matthew J. et al (2017)[116]		LR	Develop a grading scale that effectively stratifies risk of costly events (LOS, unplanned hospital readmission, need for inpatient rehabilitation) after elective surgery for degenerative lumbar pathologies	Quality and Outcomes Database (QOD) registry from 74 hospitals in 26 US states **P:** Elective lumbar spine surgery patients	**I:** 1-3 level lumbar surgery for stenosis, spondylolisthesis, symptomatic mechanical disc collapse, revision surgery including same-level disc herniation and adjacent segment disease **E:** Spinal infection, tumor, fracture, traumatic dislocation, deformity, pseud-arthrosis, recurrent multilevel stenosis, neurological paralysis due to preexisting spinal disease, age <18, incarceration, deformity/herniation	D: 6,921; V: not specified	RO	Facility vs. home
Nassour, I. et al (2017)[101]	Pre-Op	LR	Determine the rate of non-home discharge (NHD) following pancreaticoduodenectomy (PD) in a national cohort of patients and develop preoperative and postoperative predictive models for NHD	American College of Surgeons (ACS) National Surgical Quality Improvement Program (NSQIP) database **P:** Pancreaticoduodenectomy surgical patients	**I:** Whipple-type procedure or without pancreatojejunostomy and a pylorus-sparing, Whipple-type procedure with and without pancreatojejunostomy **E:** All other diagnoses, including lesions for intraductal papillary mucinous neoplasm, mucinous cystic neoplasm, or serous cystadenoma; other procedures not usually performed during Whipple-type procedures, ASA class 5, ventilator dependence, SIRS,	11,510 (D: 6,856; V: 4,654)	RO	NHD (SNF, NH, acute care, IRF) vs. home (including facilities that were home prior to hospitalization)
	Post-Op							

				sepsis, septic shock, pneumonia, open wound, acute renal failure, coma, receipt of blood transfusion, dialysis, disseminated cancer, died before discharge or discharged to unknown location			
Oldmeadow, L. B et al (2003)[124]	LR	Develop and validate an easily administered and accurate method of predicting, at or before admission, a patient's risk of needing extended inpatient rehabilitation services after elective hip or knee arthroplasty	Prospective patients from one hospital **P:** Hip/knee arthroplasty patients	**I:** Hip or knee arthroplasty **E:** Discharged to country hospital, transferred to private hospital, admitted secondary to complications, missing data, dead	650 (D: 520; V: 130)	PO	Home vs. IRF
Pattakos, G. et al (2012)[108]	LR	Identify preoperative factors associated with non-home discharge and develop a validated prediction tool for advance planning of non-home discharge	EHR data from Cleveland Clinic **P:** Cardiac surgery patients	**I:** Cardiac surgery, discharged alive **E:** None	5,313 (D: 4,031; V: 1,282)	RO	NHD (IRF, LTCH, NH, hospital, transitional care) vs. home (including HHC and hotel)

Penn, C. A. et al (2017)[117]		LR	Develop a preoperative risk scoring model predicting non-home discharge after surgery for gynecologic malignancy	Michigan Surgical Quality Collaborative, external validation using National Surgical Quality Improvement Program (NSQIP) database **P:** Hysterectomy for gynecologic malignancy patients	**I:** ≥18 years old, hysterectomy for malignant indications (explicit indication as uterine, cervical or ovarian malignancy; any ICD 9 code for gynecologic malignancy; or presence of gynecologic cancer diagnosis on a pathology report) between Jan. 1, 2013 and May 15, 2015 **E:** Died prior to discharge, left AMA	6,382 (D: 2,134; V: 4,248)	RO	NHD (SNF, IRF, LTCH, hospice, other) vs. home (including HHC and home hospice)
Ramkumar, P. N. et al (2019)[122]		ANN	Develop and test an artificial neural network (ANN) that learns and predicts length of stay, inpatient charges, and discharge disposition for THA. Secondary: create a patient-specific payment model (PSPM) accounting for patient complexity	National Inpatient Sample (NIS) database, external validation using the Orthopedic Minimal Data Set Episodes of Care (OrthoMiDas) database; **P:** Total hip arthroplasty patients	**I:** Primary diagnosis of OA who underwent THA and were subsequently discharged, Medicare, age ≥65 years **E:** Patients without Medicare, missing more than one predictor variable, cost/charge/LOS greater than 99th percentile or less than 1st percentile	81,106 (D: 78,335; V: 2,771)	RO	NHD (all other disposition including IRF or SNF) vs. home (including HHC)
Rondon, A. J. et al (2018)[102]	Pre-Op	LR	Identify factors for discharge to PAC facilities with an institutional protocol for discharging TKA patients home	EHR records from a single institution **P:** Elective unilateral primary total knee arthroplasty	**I:** Elective unilateral primary TKA, discharge disposition (home, home health, IRF, SNF), surgeons where pre-operative expectation is to go home in over 95% of patients **E:** Simultaneous bilateral TKA, revision	2,281	RO	PAC (IRF/SNF) vs. home (including HHC)
	Hospital							

	Course				TKA, and traumatic indication for TKA, patients of attending surgeons who routinely discharge patients to IRF or SNF or allow patients a choice in their desired discharge destination			
Sharma, B. S. et al (2018)[109]		LR	Create a predictive model for discharge to PAC facilities in patients undergoing unilateral total hip replacement	EHR data from University of California San Diego (UCSD) healthcare system **P:** Unilateral primary hip arthroplasty	**I:** Elective unilateral primary hip arthroplasty **E:** Missing data, non-elective surgeries, peripheral nerve blocks for postoperative analgesia	960	RO	PAC (SNF) vs. home
Stineman, M. G. (2014)[118]		LR	Develop an index for establishing probability of being discharged home after hospitalization for acute stroke using information about previous living circumstances, comorbidities, hospital course, and physical grades and cognitive stages of independence achieved	110 VA facilities databases **P:** Hospitalized stroke patients	**I:** Primary diagnosis of stroke **E:** Hospitalized ≥365 days, evidence of previous stroke within a year, missing discharge data	6,515 (D: 3,909; V: 2,606)	RO	Home vs. non-home (death, non-VA hospital, extended care facility, another location)
Stuebe, J. et al (2018)[103]	Pre-Op	LR	Identify the primary predictors of NHD after cardiac operations to generate a robust preoperative and postoperative prediction tool for those at greatest risk	Research Patient Data Registry (one institution) and Society of Thoracic Surgeons Adult Cardiac Surgery Database **P:** Cardiac surgery patients External validation: retrospective cohort from Brigham and Women's Hospital	**I:** Cardiac operation **E:** Died in hospital, missing data for discharge/predictors, emergency or salvage operation, heart transplantation, VAD placement, transcatheter aortic valve replacement, or other operation	6,660 (D: 4,800; V: 1,860)	RO	NHD (IRF, SNF, NH, hospital) vs. home
	Post-Op							

Sultana, I. et al (2019)[119]		LR	Identify factors associated with PAC referral decisions at acute care discharge	EHR data from Cerner Health-Facts Data Warehouse **P:** CABG or valve replacement surgery patients	**I:** CABG or valve replacement surgery, discharged alive, 20 years or older, admitted through ED or transferred from other clinical facility **E:** Expired, left AMA, discharged for outpatient service, LOS >75 days, age <20 years, missing data for predictor variables	14,224	RO	6 categories: home, LTCH, SNF, IRF, HHC, Other
Tan, T. L. et al (2019)[104]	Pre-op	LR	Create two predictive models based on preoperative and postoperative risk factors to identify which patients require PAC facilities	EHR data from a single institution **P:** Elective unilateral total hip arthroplasty patients	**I:** Elective unilateral primary THA with documented discharge to home, IRF, SNF **E:** Simultaneous bilateral THA, revision THA, arthroplasty for fracture, patients of surgeons who did not routinely send patients to IRF/SNF	D: 2,372; V: not specified	RO	Facility (SNF/IRF) vs. home
	Post-op							
Tapper, E. B. et al (2015)[110]		LR	Evaluate the predictive role of frailty in an observational cohort study of inpatients with decompensated cirrhosis	EHR data from Beth Israel Deaconess Medical Center **P:** Liver transplant patients	**I:** Admitted or discharged from liver unit from January 1, 2010 to September 1, 2013 **E:** Missing data	734 (D: 489; V: 245)	RO	IRF vs. home
Wasfy, J. H. et al (2018)[120]		LR	Develop predictive models that estimate patient's likelihood of prolonged hospitalization and need for post-acute services from data available at the beginning of the index hospitalization	Acute Coronary Treatment and Intervention Outcomes Network (ACTION) registry **P:** Patients hospitalized with ST-segment-elevation MI (STEMI) and non-ST-segment-elevation MI (NSTEMI)	**I:** Discharged alive between July 1, 2008 and March 31, 2017 **E:** Discharged to hospice, transferred to different acute care facility, discharged AMA, patients in the limited ACTION data collection	906,324 (D: 633,737; V: 272,587)	RO	Facility (IRF, LTCH, transitional care unit) vs. home

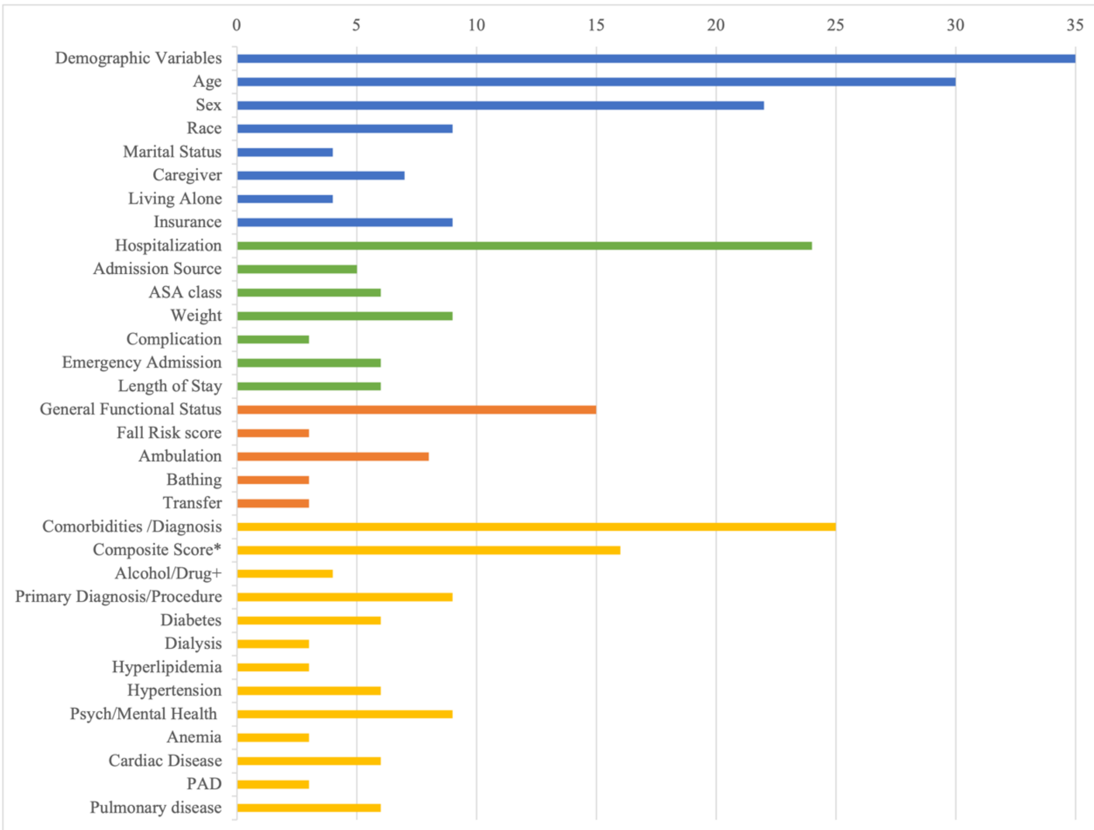
Zureik, M. et al (1997)[125]	LR	Develop a simple index able to identify at an early stage those elderly patients at high risk of requiring discharge to a residential or nursing home after admission to hospital for acute care	Prospective data from 2 hospitals in Paris **P:** Medical patients admitted from home	**I:** Age 75 or older, admitted from home through ED, medical care **E:** Directly admitted to ICU, died during hospitalization, missing data	354 (D: 210; V: 144)	PO	Residential NH (SNF, LTCH, Intermediate care) vs. home (including HHC)

Legend

+Type of prediction model: LR = logistic regression | PR = penalized regression | ANN = artificial neural network | ML = machine learning
***N:** D = development | V = validation
 # Study Design: RO = retrospective observational | PO = prospective observational | MM = mixed-methods
~ Outcome: NHD = non-home discharge | PAC = post-acute care | SNF = skilled nursing facility | NH = nursing home | IRF = inpatient rehabilitation facility | ORF = outpatient rehabilitation facility | LTCH = long term care hospital | HHC = home health care | D = in-hospital death | IC = intermediate care | AMA = against medical advice

Predictors varied across models, and most only occurred 2 or less models. **Figure 2** and **Table 2** show variables present in 3 or more models. The average number of predictors per model was 10.2, median 8 (range 4-29). The most common were demographic variables, hospitalization, and comorbid conditions.

Figure 2: Summary of Variables Present in 3 or More Models



Legend: **= Charlson, Elixhauser, or other summary comorbidity measure | + = medication/dependency. Each color corresponds to their corresponding category. Blue = demographic variables, green = hospitalization variables, orange = functional status variables, yellow = comorbidities/diagnosis variabl*

39

Table 2: Variables Present in 3 or More Prediction Models by Study

Author	AlHilli, M. M. et al (2013)[111]	Bailey, E. A. et al (2017)[112]	Ballester, N. et al (2018)[106]	Barsoum, W. K. et al (2010)[107]	Bowles, K. H. et al (2009)[51]	Bowles, K. H. et al (2017)[105] - Step 1	Bowles, K. H. et al (2017)[105] - Step 2	Chang, D. C. et al (2007)[121]	Cho, J. S. et al (2017)[99] - Extended	Cho, J. S. et al (2017)[99] - Simplified	James, M. K et al (2018)[100]- Late	James, M. K et al (2018)[100]- Admission	Karhade, A. V. et al (2018)[113]	Karnuta, J. M. et al (2020)[114]	Kimmel L.A. et al (2011)[115]	Louis Simonet, M. et al (2008)[123]	McGirt, Matthew J. et al (2017)[116]	Nassour, I. et al (2017)[101]- Pre-op	Nassour, I. et al (2017)[101]- Post-op	Oldmeadow, L. B et al (2003)[124]	Pattakos, G. et al (2012)[108]	Penn, C. A. et al (2017)[117]	Ramkumar, P. N. et al (2019)[122]	Rondon, A. J. et al (2018)[102]- Pre-op	Rondon, A. J. et al (2018)[102]- Post-op	Sharma, B. S. et al (2018)[109]	Stineman, M. G. (2014)[118]	Stuebe, J. et al (2018)[103]- Pre-op	Stuebe, J. et al (2018)[103]- Post-op	Sultana, I. et al (2019)[119]	Tan, T. L. et al (2019)[104] - Pre-op	Tan, T. L. et al (2019)[104]- Post-op	Tapper, E. B. et al (2015)[110]	Wasfy, J. H. et al (2018)[120]	Zureik, M. et al (1997)[125]	**Total**
Demographic Variables	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	35
Age	1	1	1	1	1				1	1	1	1	1	1	1		1	1	1	1	1	1	1	1	1	1		1	1	1	1	1	1	1	1	30
Sex		1	1	1				1	1	1			1	1				1	1	1	1		1	1	1	1		1	1	1	1	1	1			22
Race									1	1				1									1	1	1					1	1	1				9
Marital Status																											1	1	1	1						4
Caregiver				1	1	1	1									1				1															1	7
Living Alone											1	1								1															1	4
Insurance									1		1	1			1		1				1			1	1									1		9
Hospitalization	1	1		1	1			1	1		1		1	1	1		1	1	1		1	1	1		1	1	1	1	1	1		1		1		24
Admission Source									1					1							1		1				1									5
ASA class	1												1				1	1	1			1														6
Weight				1				1					1					1	1		1					1		1	1							9
Complication																			1						1							1				3
Emergency Admission		1												1							1		1					1	1							6
Length of Stay					1						1								1		1								1	1						6
General Functional Status				1	1	1	1				1	1	1		1	1	1			1		1					1						1		1	15
Fall Risk score						1	1																										1			3
Ambulation				1	1	1	1				1	1					1			1																8
Bathing						1	1									1																				3
Transfer						1	1									1																				3
Comorbidities /Diagnosis		1	1	1	1	1	1		1	1	1	1	1			1					1	1	1	1	1		1	1	1	1	1	1	1		1	25
Composite Score*		1	1		1	1					1	1				1						1	1	1	1					1	1	1	1		1	16
Alcohol/Drug+																										1		1	1	1						4
Primary Diagnosis/Procedure			1	1										1	1			1			1							1	1	1						9

Diabetes				1					1	1			1				1													1						6
Dialysis																										1		1	1							3
Hyperlipidemia									1	1																				1						3
Hypertension			1	1					1	1																1				1						6
Psych/Mental Health					1				1	1														1	1	1		1	1		1	1				10
Anemia								1																		1				1						3
Cardiac Disease				1				1	1	1											1									1				1		7
PAD									1	1											1															3
Pulmonary disease				1				1													1							1	1	1						6

Legend: * = Charlson Index, Elixhauser or # diagnoses | + medication/dependency

Study Quality

Table 3 displays the summary quality evaluation.

Table 3: PROBAST Risk of Bias and Applicability Summary

Study		Risk of Bias				Applicability			Overall	
Author	Model	Participants	Predictors	Outcome	Analysis	Participants	Predictors	Outcome	ROB	Applicability
AlHilli, M. M. et al (2013) [111]		High	Low	Low	High	Low	Low	Low	High	Low
Bailey, E. A. et al (2017) [112]		High	Low	Low	High	Low	Low	Low	High	Low
Ballester, N. et al (2018) [106]		Low	Low	High	High	Low	Low	Low	High	Low
Barsoum, W. K. et al (2010) [107]		Low	Low	Low	High	Low	Low	High	High	High
Bowles, K. H. et al (2009) [51]		Low	Low	Low	High	Low	Low	Low	High	Low
Bowles, K. H. et al (2017) [105]	Step 1	Low	Low	High	Unclear	Low	Low	Low	Unclear	Low
	Step 2	Low	Low	High	Unclear	Low	Low	Low	Unclear	Low
Chang, D. C. et al (2007) [121]		High	High	High	High	High	Low	High	High	High
Cho, J. S. et al (2017) [99]	Full Model	High	High	Low	High	Low	Low	Low	High	Low
	Simplified Model	High	High	Low	High	Low	Low	Low	High	Low
James, M. K et al (2018) [100]	Early	Low	Low	Low	High	Low	Low	Low	High	Low
	Regular/Late	Low	Low	Low	High	Low	Low	Low	High	Low
Karhade, A. V. et al (2018) [113]		High	High	Low	High	Low	Low	Low	High	Low
Karnuta, J. M. et al (2020) [114]		High	Low	Low	Unclear	High	Low	Low	High	High
Kimmel L.A. et al (2011) [115]		Low	Low	Low	High	Low	Low	Low	High	Low
Louis Simonet, M. et al (2008) [123]		Low	Low	Low	High	Low	Low	Low	High	Low
McGirt, Matthew J. et al (2017) [116]		High	High	High	High	Low	Low	High	High	High
Nassour, I. et al (2017) [101]	Pre-operative	High	High	Low	High	High	Low	Low	High	High
	Post-operative	High	High	Low	High	High	Low	Low	High	High
Oldmeadow, L. B et al (2003) [124]		Low	Low	Low	High	Low	Low	Low	High	Low
Pattakos, G. et al (2012) [108]		Low	Low	Low	High	Low	High	Low	High	High
Penn, C. A. et al (2017) [117]		Low	High	Low	High	Low	Low	Low	High	Low
Ramkumar, P. N. et al (2019) [122]		Low	Low	Low	Unclear	Low	Low	Low	Unclear	Low
Rondon, A. J. et al (2018) [102]	Pre-operative	High	High	Low	High	High	High	Low	High	High
	Hospital Course	High	High	Low	High	High	High	Low	High	High
Sharma, B. S. et al (2018) [109]		High	Low	High	High	Low	Low	High	High	High
Stineman, M. G. (2014) [118]		Low	Low	Low	High	Low	Low	Low	High	Low
Stuebe, J. et al (2018) [103]	Pre-operative	Low	Low	Low	High	Low	Low	Low	High	Low
	Post-operative	Low	Low	Low	High	Low	Low	Low	High	Low
Sultana, I. et al (2019) [119]		High	High	Low	High	Low	Low	Low	High	Low
Tan, T. L. et al (2019) [104]	Pre-operative	Low	Low	Low	High	High	Low	Low	High	High
	Post-operative	Low	Low	Low	High	High	Low	Low	High	High
Tapper, E. B. et al (2015) [110]		Low	Low	High	High	Low	Low	High	High	High
Wasfy, J. H. et al (2018) [120]		High	Low	Low	High	Low	Low	Low	High	Low
Zureik, M. et al (1997) [125]		Low	Low	Low	High	High	Low	Low	High	High

Legend:
Low: Low risk of bias (no signaling questions in the domain answered as indicating risk of bias)
High: High risk of bias (1 or more signaling questions in the domain answered as indicating risk of bias)
Unclear: Unclear risk of bias (1 or more signaling questions in the domain answered as unclear)

Based on application of the PROBAST tool, nearly all models demonstrated high risk of bias and 14 models demonstrated high risk of applicability concerns.[101,102,104,107-110,114,116,121,125] Two models were high risk of bias in all 4 PROBAST domains,[116,121] 9 models were high risk in 3 domains,[99,101,102,109,113,119] 6 models were high risk in 2 domains,[106,110-112,117,120] 17 were high risk in one domain,[51,100,103-105,107,108,114,115,118,123-125] and 1 model was unclear risk in one domain.[122] Fifteen models demonstrated risk of bias in the participants' domain,[99,101,102,109,111-114,116,119-121] 11 models in the predictors' domain,[99,101,102,113,116,117,119,121] 7 models in the outcomes' domain,[105,106,109,110,116,121] and 31 models in the analysis domain (4 were unclear[105,114,122]). Participant issues included use of retrospective data without adjustment for original cohort or registry outcome frequency, and no description of data source. Predictor issues were related to unclear definitions. Outcome issues included creating composite outcomes after analysis or inappropriate time interval between predictor assessment and outcome determination. Analysis issues included not accounting for unbalanced samples, inadequate events per variable (EPV), unclear variable transformation, predictor selection based on univariable analysis, and inadequate reporting. Only 10 models explicitly stated variable transformation methodology.[51,106-108,110-112,118,120,125] Only 4 models used multiple imputation.[51,108,113,120] Eighteen models did not address missing data,[100-102,104-107,111,112,114,117,121,122] and 13 models used complete case analysis.[99,103,109,110,115,116,118,119,123-125]

For applicability, there were 9 models with participant concerns,[101,102,104,114,121,125] 3 models with predictor concerns,[102,108] and 5 models with outcome concerns[107,109,110,116,121] related to narrow inclusion criteria and unclear predictor or outcome definitions.

Model Performance

Table 4 reports model performance. Eighteen models performed internal validation only.[51,100,103,105,106,109-115,118-120,122,125] Seventeen models included discrimination and 8 models included calibration.[103,109,111-113,115,118,120] One model claimed to conduct validation, but did not report discrimination or calibration.[125] Sampling methods included random splits, bootstraps, cross-validation, and using the full sample for both training and testing.

Ten models performed external validation only,[99,101,104,116,117,121,124] of which 9 reported discrimination, and 2 reported calibration[99,117] (one model reported neither).[124] Sampling strategies included different years and/or geographic areas. Seven models performed internal and external validation,[102,103,107,108,122,123] all of which reported discrimination and 4 reported calibration.[103,107,108,123]

Table 4: Evaluation of Model Performance

Author/ Model	Candidate Variable Selection	Final Model Selection	Number of Predictors	Predictors	Validation Type	Discrim-ination (D)	Discrim-ination (IV)	Discrim-ination (EV)	Calibration (D)	Calibration (IV)	Calibration (EV)	Other Measures
AlHilli, M. M. et al (2013)[111]	UA	Stepwise and backward selection	4	Age, ECOG performance status, ASA score, CA-125	IV (300 BS)		0.88			Calibration Plot		Cut point 0.10, Sensitivity 86%, Specificity 71%
Bailey, E. A. et al (2017)[112]	UA	Variables in univariate analysis where p<0.2 included in multivariable model	7	Age, sex, # Elixhauser comorbidities, emergency admission, ≥1 admission in previous year, open operation, new ostomy	IV (k-fold CV, k=10)	0.83	0.83		H-L Statistic p=1.00	H-L Statistic p=0.99		
Ballester, N. et al (2018)[106]	AA within 24 hours of admission	Backward stepwise Selection	9	Age, sex, primary diagnosis (including neoplasms, diseases of nervous system, diseases of musculoskeletal system and connective tissue), # diagnoses, previous primary diagnoses (including diseases of circulatory system, external causes of injury and supplemental classification), previous discharge disposition, comorbidities	IV (70:30 RS)	0.75	0.74					D sensitivity: 83%, D specificity 46%, V sensitivity 82%, V sensitivity 48%

					(including hypertension, neurological disorders)								
Barsoum, W. K. et al (2010)[107]		EC	Full model	17	Procedure, age, gender, BMI, heart disease, diabetes, hypertension, pulmonary comorbidity, infection, projected weight bearing, arthritis, preoperative ambulatory status, number of entry steps, bedroom location, bathroom location, caregiver assistance, home location	IV (200 BS) and EV (T, G, 1000 BS)		0.867	0.861	Calibration plot	Calibration plot (intercept: 1.082, slope: 0.653)		
Bowles, K. H. et al (2009)[51]		TD, EC	Forward Selection	7	Frequency of available help, walking function, subjective health rating, LOS, depression score, age, number of comorbidities	IV (Monte Carlo CV with 500 replications & 20% validation set)		0.863					Cut point 0.69, sensitivity: 87.6%, specificity: 65.2%, overall PV 83.2%, cross validated PV 80.1%
Bowles, K. H. et al (2017)[105]	Yes/ No	TD, EC	Penalized Regression	16	Employment status, # hospital stays within past 6 months, fall risk score, equipment, home accessibility, wound present, ambulation current, ambulation change (level A to B),	IV (RS)		0.915					Sensitivity: 90.1%, specificity: 76.9%, PPV: 94.2%, NPV: 65%

					ambulation change (level B to C), transfer change (level A to B), transfer change (level B to C), bathing change, eating prior, number of comorbidities, caregiver presence, discharged on narcotics								
	Where to Refer	TD, EC	Penalized Regression	13	Braden score, fall risk score, ambulation current, ambulation change (level of decline from A to B by discharge), transfer current, transfer change (level of decline in transfer function in level A to B by discharge), toileting current, bathing current, bathing change (level of decline in bathing function from level A to B by discharge), eating prior, caregiver presence, caregiver availability, caregiver relationship	IV (RS)		0.897					Sensitivity: 89.2%, specificity: 68.0% , PPV: 91.6%, NPV: 61.8%

Chang, D. C. et al (2007)[121]		AA present in >5% of sample	Full model (except age)	9	Gender, osteoarthritis, CHF, atrial fibrillation, COPD, MI, anemia, obesity, renal disorder	EV (G)	0.635		0.644				D pseudo R^2: 0.0345, V pseudo R2: 0.0408, sensitivity: 58.7%, specificity: 62.1%
Cho, J. S. et al (2017)[99]	Full Model	UA	Full Model	8	Sex, age, race, stroke type, comorbidity (categories including diabetes, heart disease, hypertension, peripheral artery disease, chronic kidney disease, hyperlipidemia, arrythmia, depression), source of admission, primary payer, secondary payer	EV (T)	0.737		0.724	Predicted vs. observed probability plot		Predicted vs. observed probability plot	
	Simpli-fied Model	UA	Not specified	5	Sex, age, race, stroke type, comorbidity (categories including diabetes, heart disease, hypertension, peripheral artery disease, chronic kidney disease, hyperlipidemia, arrythmia, depression)	EV (T)	0.693		0.679				
James, M. K et al (2018)[100]	Early	UA	Created and merged different models based on	13	Age, face injury, chest injury, fracture present, intubation status, number of comorbidities,	IV (k-fold CV, k=10, 10		0.82					Sensitivity: 0.80, specificity: 0.70, PPV: 0.75, NPV:

			confusion matrix		Medicaid insurance type, Medicare insurance type, social security income status, living arrangements, mood affect status, ADLs and ambulation status	repetitions)							0.75, accuracy: 74.9%
	Late	UA	Created and merged different models based on confusion matrix	15	Age, face injury, chest injury, fracture present, intubation status, number of comorbidities, injury severity score, ICU length of stay, Medicaid insurance type, Medicare insurance type, social security income status, living arrangements, mood affect status, ADLs and ambulation status	IV (k-fold CV, k=10, 10 repetitions)		0.86					Sensitivity: 0.83, specificity: 0.74, PPV: 0.79, NPV: 0.79, accuracy: 78.5%
Karhade, A. V. et al (2018)[113]		AA missing in <30% of sample	Neural Network	9	Age, sex, BMI, fusion, functional status, ASA classification, diabetes, preoperative hematocrit, level of surgery	IV (80:20 RS) and 10-fold CV	0.815	0.823		Graph; slope: 0.896, intercept: 0.026	Calibration Plot; slope: 0.935; intercept: 0.026		Brier Score (D): 0.0725 Brier Score (V): 0.0713

Karnuta, J. M. et al (2020)[114]	CR	Machine learning	8	Age group, gender, race, ethnicity, type of admission, APR risk of mortality, APR severity of illness, CCS diagnosis code	IV (Stratified k-fold CV (k=10, 90% training, 10% validation), 100 adaptive boosting rounds)		0.906					Accuracy: 0.878
Kimmel L.A. et al (2011)[115]	UA	Backward Elimination	14	Age, compensable, private insurance, rural region, injury=shaft of femur, injury=distal femur, injury=proximal tibia, injury=shaft of tibia, injury=distal tibia, mechanism=transport related, mechanism=high fall, mechanism=other-not low fall, working prior, pre-injury disability	IV (RS)	0.92	0.86		H-L statistic: 11.6 (p= 0.17), calibration plot	H-L statistic: 38.0 (p=0.001), calibration plot		Sensitivity D: 69.9%, Specificity D: 94.4%, NPV D: 91.13%, PPV D: 79.1%, Sensitivity V: 59.8%, Specificity V: 90.3%, NPV V: 86.5%, PPV V: 68.5%

Louis Simonet, M. et al (2008)[123]		UA	Backward Elimination	5	Number of active medical problems, inability of patient's partner to provide home help, dependency for bathing, dependency for transfers, inability in medication self-management before admission	IV (CV) and EV (T)	0.82	0.81	0.77	H-L Statistic p=0.21			Cut point 8: Sensitivity: 87%, Specificity: 63%, PPV: 53%, NPV: 91%
McGirt, Matthew J. et al (2017)[116]		AA preoperatively	Stepwise multivariate regression	7	Fusion, ASA class IV/V, age ≥ 70, ODI ≥70, diabetes, ambulation-assisted, private insurance	EV (T)			0.731				Wald's chi-squared: 4.939, p=0.026
Nassour, I. et al (2017)[101]	Pre-Op	UA	Pre-operative variables where p<0.2	7	Age, sex , BMI, ASA class III/IV vs I/II, albumin, >10% weight loss, diagnosis group	EV (T)	0.77		0.75				
	Post-Op	UA	Forward Selection	9	Age, sex, BMI, ASA class III/IV vs I/II, albumin, >10% weight loss, return to OR, LOS ≥14 days, any inpatient complications	EV (T)	0.82		0.81				
Oldmeadow, L. B et al (2003)[124]		EC	Backward elimination	6	Age, gender, mobility, gait aid, community support, caregiver in the home	EV (T)							Predictive accuracy 75.2%, overall accuracy 74.6%

Pattakos, G. et al (2012)[108]	AA in databa se	Forward selection using bootstrap sampling	23	Age, gender, time to hospital drive, preoperative admission length of stay, payor, BMI, stroke, COPD, PAD, albumin, BUN, cholesterol, anticoagulation, steroids, lipid-lowering medication, ACE-I or ARB, emergency operation, preoperative IABP, NYHA class II, complete heart block, LVEF%, percutaneous, procedure (descending aortic graft, aortic root ascending arch, CABG)	IV (1000 BS) and EV (T)		0.88	0.87			Brier score: 0.1; Calibrati on Plot	
Penn, C. A. et al (2017)[117]	BA	Stepwise logistic regression	6	Age ≥ 70 years, dependent functional status, ASA class ≥3, open surgical approach, 1 comorbidity, ≥2 comorbidities	EV (G)	0.895		0.84	H-L statistic: 9.81 ($p=0.2$)			
Ramkumar, P. N. et al (2019)[122]	AA at admissi on	Deep learning artificial neural network	15	Age, gender, ethnicity, race, type of admission, emergency department, APR risk of mortality, APR risk of severity of illness, number of conditions, number of diagnoses, comorbidity status,	IV (k-fold CV, k=10) and EV (G)		0.794	0.701				Accuracy IV: 72.2%, Accuracy EV: 70.5%

					weekend admission, hospital type, income quartile, transferred from another hospital								
Rondon, A. J. et al (2018)[102]	Pre-Op	UA	Variables where p<0.2 in univariate analysis	6	Age ≥75, insurance status- Medicare, Charlson Comorbidity index, gender, race, history of depression	IV and EV (G*)		0.79					
	Hospital Course	UA	Full pre-op model with hospital/ operative variables	8	Age ≥ 75, insurance status- Medicare, Charlson comorbidity index, gender, race, history of depression, in hospital complication, procedure duration	IV and EV (G*)		0.8					
Sharma, B. S. et al (2018)[109]		UA	Combination of Backward and Forward Selection	10	General anesthesia, age ≥65, preoperative opioid use, preoperative METS <4, dialysis, gender, preoperative anemia, psychiatric history, hypertension, obesity (BMI >30 kg/m2)	IV (k-fold CV, k=10)	0.806	0.794		H-L Statistic p=0.08; calibration plot	H-L Statistic p=0.38; calibration plot		

Stineman, M. G. (2014)[118]		TD, BA	Backward selection	8	Marital status, location before hospitalization, discharge physical grade, discharge cognitive stage, comorbidities- liver disease, mechanical ventilation, nonoral feeding, ICU admission	IV (60:40 RS)	0.82	0.8		H-L Statistic p=0.23	H-L Statistic p=0.3		Probability ranged from 67.76% to 97.43%
Stuebe, J. et al (2018)[103]	Pre-Op	UA	Backward Selection	13	Age, sex, marital status, BMI, dialysis, PVD, prior cerebrovascular accident or TIA, pulmonary disease, alcohol or drug dependency, operation type, psychiatric disease, urgent operation, prior cardiac operation	IV (1000 BS) and EV (T)	0.82	0.817	0.814		Calibration plot; slope = 0.268	Calibration Plot	D Brier score: 0.149; V Brier score: 0.118; R^2: 0.349
	Post-Op	UA	Forward selection	16	Age, gender, marital status, BMI, preoperative dialysis, PVD, prior stroke, pulmonary disease, drug or alcohol dependency, psychiatric disease, prior cardiac surgery, urgent surgery, procedure, LOS, post-op neurological event, infection	IV (1000 BS)		0.86			Calibration plot; Slope = 0.343		R^2: 0.439, net reclassification index 6.5% when using a 0.3 cutoff probability for NHD
Sultana, I. et al (2019)[119]		BA	Multinomial Logistic Regression (p<0.1)	29	US region, hospital bed size, hospital status, teaching facility affiliation,	IV (10-fold CV)		Overall AUC: 0.685					IV accuracy: 62.5%

				gender, marital status, race, age, length of stay, Charlson index, coronary bypass of 2 coronary arteries, coronary bypass of four or more coronary arteries, coronary artery bypass of three coronary arteries, coronary artery bypass of one coronary artery, open replacement of aortic valve with tissue graft, open replacement of aortic valve, diabetes mellitus without complication, tobacco use disorder, atrial fibrillation, unspecified hypertension, coronary atherosclerosis, intermediate coronary syndrome, hyperlipidemia, posthemorrhagic anemia, acute myocardial infarction, congestive heart failure, anemia unspecified, pulmonary collapse, acute kidney failure			Home: 0.72 IRF: 0.53 LTCH: 0.52 SNF: 0.58 HHC: 0.72 Other: 0.46					

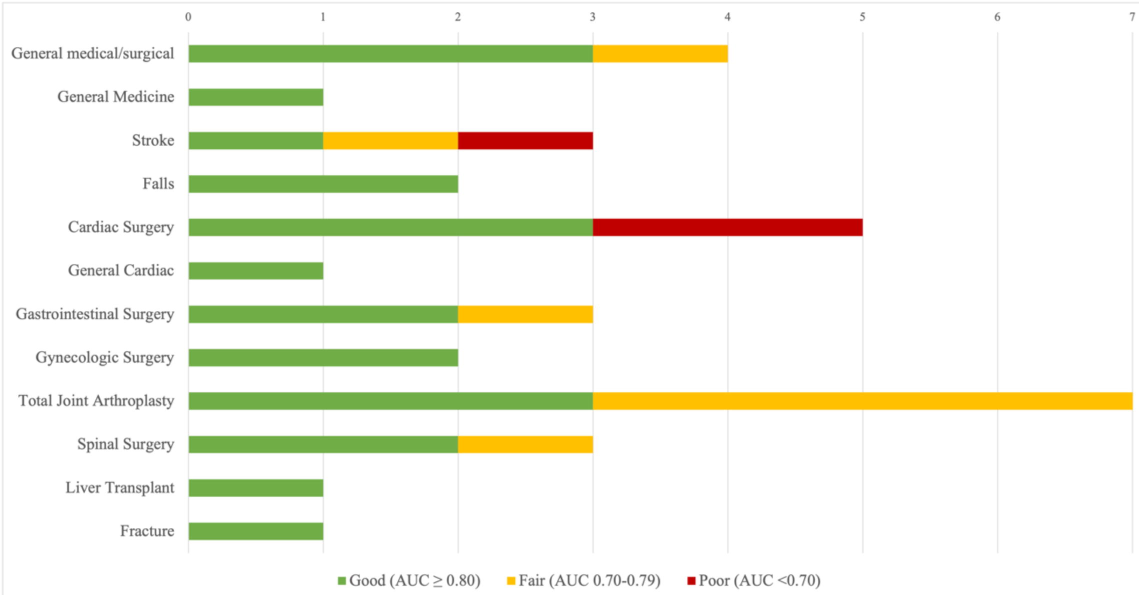
Tan, T. L. et al (2019)[104]	Pre-op	UA	Full Model	5	Age ≥65, sex, race, Charlson comorbidity index, history of depression	EV (G*)			0.77				
	Post-op	UA	Full Model	7	Age ≥ 65, sex, race, Charlson comorbidity index, history of depression, surgical time ≥90 min, in-hospital complication	EV (G*)			0.80				
Tapper, E. B. et al (2015)[110]		CR	Full model	8	Gender, age, Charlson comorbidity index, MELD score, admission sodium, ADL score, Braden score, Morse fall risk score	IV (2:1 RS)	0.85	0.77					
Wasfy, J. H. et al (2018)[120]		CR	Logistic Regression: Backward Selection	9	Age, Medicare, Medicaid, Uninsured, Heart failure, Heart rate at first contact, Shock at first contact, prior cerebrovascular disease, initial hemoglobin	IV (70:30 RS)		0.827			Calibration Plot		
Zureik, M. et al (1997)[125]		UA	Backward elimination	6	Principal carer's wishes about patient returning home, chronic condition, ability to perform toileting, age, ability to know name of place, living alone	IV (2:1 RS)				H-L statistic p = 0.476			Correct classification 67.4%, Sensitivity: 77.6%, specificity: 50%, PPV: 51.1%, NPV: 67.8%
Legend													

Candidate Selection: UA = univariate association | BA = bivariate association | CR = clinical relevance | TD = theory driven | EC = expert consensus | AA = all variables available
Validation Type: D = development | IV = internal validation | EV = external validation | BS = bootstrap samples | CV = Cross Validation | RS = Random Split |T = temporal | G = geographic | * = details not provided
Other Measures: D = Development | V = Validation | PV = predictive value | PPV = positive predictive value | NPV = Negative Predictive Value

Model Performance by Clinical Population

Figure 3 shows model discrimination by clinical population.

Figure 3: Model Discrimination (AUC) Grouped by Clinical Population



Among the 4 medical population models, discrimination ranged from 0.679[99] to 0.81[123] and was good in 2 models,[118,123] fair in 1 model[99] and poor in 1 model[99] (1 model did not report discrimination[125]). Among the 22 surgical models, discrimination ranged from 0.52 in the Sultana et al. long term care hospital model[119] to 0.906[114] and was good in 12 models,[101-104,107,108,111-114,117] fair in 7 models,[101,102,104,109,110,116,122] and poor in 2 models[119,121] (one did not report discrimination[124]). Among 4 models of general medical and surgical patients,[51,105,106] discrimination ranged from to 0.75[106] to 0.915.[105] Discrimination was good in 3 models[51,105] while fair in 1 (calibration not reported).[106] Among the 2 general medicine models, one[123] had good discrimination (AUC 0.81) and good calibration (H-L statistic p=0.21), while another[125] did not measure discrimination but reported poor correct classification (67.2%). For 3 models of stroke patients, discrimination was split equally among 1 poor model,[99] 1 fair model,[99] and 1 good model[118] with AUC ranging from 0.679 and good calibration[99] to 0.80 and poor calibration. [118] Both of the models[100] for falls patients demonstrated good discrimination with AUC 0.82 for the early model and 0.86 for the late model (calibration not reported).

Among 5 cardiac surgery models,[103,108,119,121] discrimination was poor in 2 models[119,121] and good in 3[103,108] with AUC ranging from 0.644[121] to 0.87.[108] Only 2 models from one study[103] reported calibration demonstrating overfitting (calibration slope 0.268 for pre-operative model and 0.343 for post-operative model) with good discrimination (AUC 0.817 and 0.86 respectively). One model[108] graphically demonstrated good calibration. The cardiac model for medical and surgical patients demonstrated good discrimination (AUC 0.827).[120] For the gastrointestinal surgery models, the postoperative model had better discrimination (AUC 0.81) than the preoperative model (AUC 0.75).[101] Another model had the best discrimination (AUC 0.83) and good calibration (H-L statistic p=0.99).[112] Both gynecologic surgery models demonstrated good discrimination and calibration, with AUC 0.88 and a calibration plot demonstrating good calibration across the range in one[111] model, and AUC 0.84 and H-L statistic 9.81 (p=0.2) in the another.[117] For 8 joint arthroplasty models, discrimination was fair (4)[102,104,109,122] to good (3),[102,104,109] with AUC ranging from 0.701[122] to AUC 0.867.[107] Only 2 models reported calibration.[107,109] One model that did not report either statistic had predictive accuracy 74.6%.[124] Among the 3 spinal surgery models, one[116] had fair discrimination (AUC 0.731), one had good discrimination (AUC 0.906)[114] and one had good discrimination (AUC 0.823) and good calibration (slope 0.935, intercept 0.026).[113]

The liver transplant model[110] had fair discrimination (AUC 0.77), and the fracture model[115] had good discrimination (AUC 0.86) but poor calibration (reported graphically).

Model Performance in Admission versus Hospital Course Timing

Four models validated pre-and post-operative models. All models later in the hospitalization had better discrimination.[101-104] One study validated admission and hospital course models with the same pattern.[100] This trend also continued into the general sample. For discrimination across all studies, the 22 preoperative or admission models demonstrated 9 good models,[100,103,107,111,113-115,117,120] 9 fair models,[99,101,102,104,106,109,110,116,122] and 2 poor models[121] (2 models did not report discrimination[124,125]). Nearly all 13 models used post-operatively or throughout hospitalization had good discrimination,[51,100-105,108,112,118,123] with only one poor model.[119]

Historical Trends in Model Performance

Model discrimination has not necessarily improved over time. Among 5 models developed from 2000-2010, discrimination was good in 3 models,[51,107,123] and poor in one model;[121] (one model did not report it[124]). All 5 models developed from 2011-2015 had good discrimination. [108,110,111,115,118] Most models (24) were developed from 2016-2020 and discrimination was good in 14 models,[100-105,112-114,117,120] fair in 8 models,[99,101,102,104,106,109,116,122] and poor in 2 models.[99,119]

Model Discrimination by Method

Although machine learning and artificial intelligence modeling are becoming increasingly popular in healthcare, PAC models are still predominantly developed using regression. Among 32 regression models; 3 models reported poor,[99,119,121] 7 models reported fair,[99,101,102,104,106,109,110,116] and 20 models reported good discrimination.[51,100-105,107,108,111,112,115,117,118,120,123] Two models reported neither.[124,125] The Naïve Bayes machine learning model had good discrimination AUC 0.91 (no calibration).[114] The 2 artificial neural network models' performance ranged from fair to good, with AUC 0.701[122] and AUC 0.82 with good calibration (slope 0.935, intercept 0.026).[113]

Model Translation

Over half of models were translated into clinical tools (although not necessarily implemented) with 5 nomograms,[101,107,111,117] 6 online or EHR algorithms,[51,105,108,114,122] 12 score charts,[99,103,116,118,120,121,123-125] and 1 probability scoring tool.[109] Five tools are available for free online.[107,108,113,114,122] Most of the studies with translated models indicated intent to test them in clinical practice in future studies. Only a few teams have published about their models beyond the original studies to report quasi-experimental testing studies,[42,126] to externally validate using hospital EHR data[111] or to update with new variables.[108,127]

Discussion

This study identified 35 models across 28 studies predicting PAC use after hospital discharge in adults. The most common populations were orthopedic and cardiac surgery, and both had surprising variation in performance given the narrowness of the population. One possible explanation is that many surgeons or insurance companies have their own pre-determined PAC pathways,[11,128] and many elective surgeries are moving into ambulatory settings, leaving the sickest patients for in-hospital surgeries. [129] Four out of the 6 general patient models had good discrimination.[51,105,123] While other models demonstrated good discrimination, it is difficult to compare across clinical populations because each has unique PAC needs. Additionally, most models did not perform calibration. The general models' better performance might be attributed to the holistic set of predictors including caregivers, functional status, and comorbidities. Some of the better performing studies cited variable selection based on clinical relevance[107,116,120,124] and theory.[51,105,118] Instead of relying heavily on univariable analysis, future studies should consider incorporating clinically credible factors based on expert consensus, literature reviews,[43,130] and theory such as Orem's Self-Care Deficit Theory[46] which was used in 2 studies[51,105] to reduce bias and possibly improve performance.

Overall there is a growing trend to consider both physical and mental health issues in prediction models,[131,132] although only 10 models did so.[51,99,102-104,109] The most common conditions across all the models included diabetes, hypertension, and cardiac disease, making sense because these comorbidities are often seen as predictors of overall health status. A growing body of research suggests that functional status may be a better predictor than comorbidities of PAC and negative outcomes,[133,134] but comorbidities were more common in this review. Only fifteen models included functional status, and only 10 incorporated specific aspects of function beyond an overall functional status score.[51,100,105,107,110,116,123,124] It is possible that this information is not systematically documented in EHRs or included in registries. Moving forward, health systems and clinical data registries should consider including this critical data source.

All studies included demographic variables such as sex, race or insurance. If the goal is to improve decision making, this could sustain existing disparities in PAC referral practices.[11,135] One recent study revealed a 14.5% difference in PAC referral rates to skilled nursing facility versus inpatient rehabilitation facility between black and white patients.[135] Few models incorporated hospital characteristics like hospital size[119,122] and urban versus rural.[115,119] Recent research suggests that hospital factors can explain post-discharge location more than individual factors due to affiliation with PAC sites like cardiac rehabilitation, proximity to PAC in rural areas, and day of the week in hospitals with limited resources.[20,136,137]

Future studies should also be mindful of data sources. The majority of studies used retrospective discharge disposition data for all available patients in a database as outcomes to build the models. Unfortunately due to the large variation in discharge planning practices at the provider and hospital level, it is possible that these models were trained to predict common practices that are known to be biased[138] rather than best practices.[139] One possible solution is the use of discharge planning experts to build the models by identifying discharge disposition based on case studies similar to the approach used in two studies.[51,105] By using experts rather than the outcome generated from clinical practice, which is known to create disparities, experts may build models based on better decisions. Focusing on patient needs (clinical and functional) rather than demographic characteristics like insurance or race removes the bias where patients in need may not have gotten referred due to lack of insurance coverage, for example. However, experts are not always readily available, could be expensive and take longer than using outcomes within existing datasets.

Implementation of clinical decision support in practice and access to open source tools are growing trends.[140] Toward that end, a large majority of models created simplified tools for clinical practice (although not necessarily implemented),[51,99,101,103,105,107-109,111,114,116-118,120-125] and 5 provided open-access online tools.[107,108,113,114,122] The RAPT[124] model for orthopedic surgery has been updated and validated in different populations and cited in the literature over 130 times.[141,142] Its success is likely due to being one of the earliest tools, and ease of use with only 6 patient reported questions.

It is important to critically analyze all models in terms of missing data, sample size, validation technique, and misclassification error, but especially in machine learning where interpretability is more challenging than more traditional statistical methods. Although only 3 models in this review used machine learning, this type of evaluation will become more important as machine learning becomes more accessible and common. Patients may have fractured care or may not have the health literacy to access specialty care providers, contributing to missing data.[143] Subgroups of patients may not be present in sufficient numbers even in large sample sizes. Low-income patients may be seen in clinics with less documentation and clinicians may leave certain assessment elements blank if patients are cognitively impaired and cannot answer questions.[139,144] To lower the risk of bias, teams should incorporate expert clinicians and community members from vulnerable groups in the predictor selection process. Performing model validation is another statistical strategy to increase generalizability and lower the risk of bias by reducing the possibility of overfitting a model.[43] For these reasons, this review focused on studies that reported both development and validation. Future studies should be sure to include validation to enhance the rigor of their methods.

Historically, model discrimination has not necessarily improved over time despite new statistical methods and increased access to healthcare data. There were many quality issues across studies that call for recommendations for future research. Future studies should follow the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) guidelines.[48] For cohort selection, researchers should adjust for the original cohort or registry outcome frequency such as reweighting control and case samples by inverse sampling fraction in logistic regression to correctly estimate baseline risk.[145-147] Studies should address how they measure predictors and the quality of the data sources.[148] Performance is overestimated when the same dataset is used for development and validation, especially in small sample sizes.[149,150] With the wide availability of EHR and registry data, future studies might consider external validation of models prior to publication, or at least acknowledge it as a limitation.

Researchers should be mindful of sample size and candidate variable selection. Although 10 events per variable (EPV) was once considered acceptable, more recent research recommends 20.[43] Most registry studies had over 1,000 EPV, but smaller studies did not meet 20 EPV.[151,152] Future studies should be mindful about variable transformations in their procedures. Dichotomizing variables during analysis reduces power and predictive ability, and it is important to avoid data-driven transformations when possible unless using machine learning.[153]

Missing data was not mentioned or handled optimally in the majority of studies. Only 4 studies imputed missing data.[51,108,113,120] Multiple imputation reduces bias, creates correct p- values and standard errors, and studies have shown that it leads to better precision in validation studies.[43,154] As statistical software becomes more sophisticated, imputation is becoming more common. However, there are cases where imputation can create bias, so researchers should be cautious about using imputation for data that is missing not at random, such as for cognitively impaired patients.[155]

Finally, it is important to take these factors into consideration and be mindful about how to appropriately use prediction models in real-world clinical settings. If models are used to develop clinical decision support systems to help discharge planning teams determine PAC destination, it is crucial that these tools incorporate holistic data to identify where a patient would have the best probability of a positive outcome rather than emulate past (potentially biased) clinician behaviors. When deciding between retrospective datasets and expert-driven methods to develop models, researchers should consider the tension between time, cost, and data quality of each. During the implementation process, clinicians should be educated about how the tools were developed, clinical decision support systems' role in aiding decisions in addition to clinical judgment, and the limitations of the technology.

Limitations

This study is not without limitations. Nearly all studies demonstrated high risk of bias in the PROBAST assessment. PROBAST was published in 2019, which follows TRIPOD reporting guidelines (released in 2015 after some studies in the review were published), so it is possible that these tools will be updated and/or more studies will begin to follow standardized reporting criteria.[43]

There was great variability in predictors, outcomes, and methods across studies which made objective comparison challenging. For example, several models included comorbidity predictors but defined them as the Charlson Comorbidity Index,[102] Elixhauser Comorbidity Index,[112] or number of diagnoses.[105] Some studies reported model coefficients, while others included odds ratios, risk ratios, bootstraps, or predictor names only. Prediction modeling studies are encouraged to use standard outcome definitions, but none exist for combined PAC outcomes. Most studies created binary composite non-home discharge outcomes with a wide range of definitions. Only one model used a categorical outcome of 6 discharge disposition categories, which weakened performance.[119] Our model performance evaluation focused heavily on discrimination. It was difficult to compare other performance measures across studies with different validation methods (internal and/or external validation), and few studies met the minimum reporting criteria of discrimination and calibration.[43]

Conclusion

Prediction model development and validation studies have become more prevalent in the literature following advances in healthcare technology, especially in surgical populations. Model development and validation methodology differ across studies. Although models currently focus on demographic predictors, future models should consider using theory- and/or expert-driven approaches for variable selection and incorporate holistic variables like functional status, especially as these data become more widely available in EHRs. At this point, model performance within and across populations is variable. Future studies should ensure rigorous candidate variable selection and be sure to follow TRIPOD reporting guidelines for model development and validation. New models using artificial intelligence are becoming popular but should be mindful of their data sources and methodology to avoid prolonging biased clinical decision making. Finally, the majority of prediction models are waiting for implementation or translation to new populations.

CHAPTER 3: Paper 2

COMPARISON OF CLINICAL DECISION SUPPORT SYSTEM RECOMMENDATION FOR DISCHARGE DISPOSITION TO USUAL DECISION MAKING: EVALUATION OF 30-DAY READMISSIONS

Abstract

Objective: The goal of this study is to apply the Discharge Referral Expert System for Care Transitions (DIRECT) clinical decision support system (CDSS) for discharge planning in a new setting and determine differences in patient characteristics and 30-day readmission rates based on DIRECT's recommendation among older adults in a large urban academic health system who were discharged without post-acute care (PAC). This provides an opportunity to examine patient outcomes when those identified by CDSS as needing PAC do not get PAC.

Materials and Methods: A retrospective analysis of electronic health record data from adults aged 55 years or older hospitalized in two hospitals in one large urban health system was performed. Thirty-day readmission rates were examined using multiple logistic regression, with DIRECT PAC recommendation (yes/no) as the primary predictor variable, in patients discharged without referral. Subgroup analysis was performed to assess differences in patient characteristics and outcomes between surgical and non-surgical patients.

Results: Among 3,385 older adults discharged home without PAC, 2,776 (82%) patient encounters were flagged by DIRECT as needing PAC. The overall 30-day readmission rate was 15.2%, and those flagged experienced 0.5% lower rates of readmissions compared to those not flagged (15.1% vs. 15.6%, p=0.75). The sociodemographic characteristics and algorithm elements between those flagged vs. not flagged were significantly different. Subgroup analysis of surgical patient encounters (N=1,489) yielded an 8.6% higher 30-day readmission rate, and surgical patients flagged by DIRECT as needing PAC experienced an adjusted 51.8% higher odds of readmission (p=0.041).

Discussion/Conclusion: These findings suggest the transportability of DIRECT CDSS to new health systems and potential value in large urban hospitals especially for surgical patients. Real world challenges of transporting CDSS to new settings such as different clinical workflows and documentation practices, missing data, and lack of interoperability are discussed. Future directions include close work with stakeholder groups to assure the collection of algorithm data elements, use of natural language processing to extract reasons why patients did not receive PAC from clinical notes and installing DIRECT CDSS in the health system to examine the impact of sharing the algorithm with clinicians on clinical outcomes such as 30-day readmissions.

Introduction

Coordinated discharge planning is a key component of successful transitions from hospital to post-acute care (PAC), and has been associated with reductions in 30-day readmissions and improved patient outcomes.[4] A coordinated process includes an interdisciplinary assessment of patient needs throughout the hospitalization and collaborative, culturally competent planning with the patient and their caregiver(s) to identify the appropriate type of PAC after a hospitalization.[5] PAC includes destinations such as long term acute care hospitals, inpatient rehabilitation facilities, skilled nursing facilities, and home health care.[3] This is especially important for older adults, who experience more chronic complex conditions. Although 42% of Medicare discharges receive PAC referrals annually, discharge processes vary significantly at the patient, provider, and hospital level.[1,2]

Many barriers to high quality discharge planning and PAC referral exist. In general, discharge planning is complicated, involving complex decision making, coordination across inpatient and outpatient settings as well as communication between patients, multiple disciplines of healthcare providers, and insurance companies.[7] At the system level, hospital characteristics, geography, and insurance coverage are known barriers to PAC referrals.[19] Rural areas may have capacity constraints or limited PAC availability.[19,20] Insurance barriers include type of coverage, benefit limits for PAC, authorization requirements, narrow provider networks, ambiguity in medical necessity definitions, and lack of insurance.[19] At the provider level, communication issues include time constraints, inconsistent assessment, and variance in risk tolerance, which contributes to subjective decision making.[4] Some providers do not value PAC,[11] and racial and gender disparities in PAC referrals are well-documented in cardiology and orthopedics.[12,13,156-160] At the patient level, although nearly 75% of older adults will need formal care at some point, only 40% of Americans expect to need it.[14,15] Even when clinicians do recommend appropriate PAC, patients refuse up to 28% of the time, and these patients were readmitted at twice the rate of those who received PAC in one study.[86]

Together, these discharge planning challenges contribute to unplanned hospital readmissions, unnecessary treatments,[21] increased costs,[22] and decreased patient satisfaction.[19] Patients who receive coordinated discharge planning with evidence-based PAC referrals have better outcomes including reductions in errors[25] and hospital readmissions.[24] A recent systematic review of discharge communication practices found that well-designed technology solutions in discharge planning improve patient satisfaction and outcomes.[26]

One of the most successful and widely implemented technology solutions has been clinical decision support systems (CDSS), which leverage prediction models to improve clinical decision making. CDSS equip clinicians, patients, and other stakeholders with relevant and/or person-specific knowledge at appropriate times to improve decision making to ultimately enhance health and healthcare.[42,161] CDSS is frequently integrated into electronic health record (EHR) workflows to be used at the point of care. A recent study found that only 0.3% of published CDSS tools are replicated in the literature. Replication and implementation of CDSS in new settings has the potential to improve both efficiency and effectiveness of CDSS as well as minimize harms related to technology in healthcare delivery.[67] Furthermore, widespread implementation of CDSS aligns with the Office of the National Coordinator's Health Information Technology's interoperability goals.[68]

The goal of this study was to apply the Discharge Referral Expert System for Care Transitions (DIRECT) CDSS[41,42,86] in a new setting, a large urban academic health system with different leadership, resources, and with a more diverse patient population than the suburban community hospitals in the original study. DIRECT is an expert clinical decision support system[162] developed using consensus from interdisciplinary discharge planning experts.[41,42] The CDSS is a 2-step algorithm calculated from structured nursing and administrative EHR fields that identify (step 1) if a patient needs post-acute care (yes/no), and if yes, (step 2) the level of care as home health care or facility level care.

The specific aims of the study were: Among patients discharged home without PAC, (1) compare patient characteristics and (2) 30-day readmission rates between those identified by DIRECT as needing PAC and those not identified as needing a PAC referral. The hypotheses were that among patients discharged home without services, (1) patients identified by the algorithm as needing PAC would be older, with more limitations in activities of daily living, more comorbidities, and more hospitalizations in the 6 months prior to hospitalization compared to patients not flagged for PAC. Additionally (2) those flagged by the algorithm as needing PAC would also experience higher rates of 30-day readmissions compared to patients not flagged for PAC.

Methods

Design

This study was a retrospective analysis of existing clinical data on a cohort of inpatients in a large, urban, academic health system. The study was approved by the University of Pennsylvania Institutional Review Board (#843687).

Sample

The sample was drawn from clinical records of patients at two hospitals in a large regional academic medical system. One hospital (Site 1) is a large, urban, tertiary medical center and the other hospital (Site 2) is an urban community hospital. All patients admitted between December 1, 2018 and December 1, 2019, aged 55 years or older, admitted to medical or surgical service lines and units, with a hospitalization greater than or equal to 48 hours (to avoid observation stays), and discharged alive were eligible for this study. Study dates were selected in conjunction with the health system's data warehouse to ensure that discharge planning fields regarding patient preferences for discharge that were added to the EHR in November 2018 would be captured. Although the inpatient hospitalization data concluded on December 1, 2019, 30-day follow up data was captured on all patients to compute readmission rates. The period of the COVID-19 pandemic beginning in 2020 was avoided since hospital operations were disrupted. Age was limited to adults 55 years or older because the algorithm was developed and validated in this age group. Medical and surgical service lines were selected to represent a broad clinical population without limiting the study to a specific disease or procedure. Length of stay greater than or equal to 48 hours was selected to focus on hospitalizations rather than observational stays. The final cohort included patients discharged home without PAC services (discharge disposition was home to self-care).

DIRECT Algorithm and Study Data Elements

DIRECT CDSS identifies a patient's need for PAC services in general hospitalized adults aged 55 years or older. In model development and validation, the area under the receiver operating curve (AUC) was 95.1% for step 1 and 89.7% for step

2.[41] DIRECT was tested in a quasi-experimental pre-post study and its use was associated with significant reduction in readmissions at 7, 14, and 30 days. Patients experienced better outcomes when the discharge disposition matched the algorithm's recommendation, achieving a 22% relative reduction in readmission rates.[42] However, it recommended 25.6% more patients for PAC compared with their discharge disposition.[42] Details about algorithm development and testing have been described elsewhere.[41,42]

Structured sociodemographic, administrative, and clinical data were extracted from the EHR using EPIC Clarity to examine patient characteristics and compute the algorithm.[163,164] Sociodemographic variables including gender, race, ethnicity, age, employment status, marital status, and insurance type were extracted to describe the sample. Administrative variables included number of hospitalizations in the 6 months prior to index admission, encounter diagnosis (ICD-10-CM codes), comorbid conditions, length of stay, and 30-day readmission. Clinical data was extracted from nursing flowsheets including Morse fall risk, assistive devices, home accessibility, presence of a wound, changes in activities of daily living (ADLs), caregiver information, and discharge medications. Discharge disposition was evaluated as PAC (including all facility discharge destinations and home health care) versus home without services (home to self-care).

Statistical Analysis

We calculated step 1 of the DIRECT algorithm to determine patients' need for PAC (yes/no) for patients who were discharged home without PAC services (analytic sample). Statistical analyses compared two groups: those flagged by the algorithm as needing PAC, and those not flagged by the algorithm as needing PAC. Clinical and demographic data were described using frequencies and cross tabulations of proportions, means, standard deviations, and interquartile ranges, and compared using standard bivariate tests (Chi-square, Kruskal-Wallis, and t-test).

Multiple logistic regression examined the association between 30-day readmission rates and the primary predictor variable of the DIRECT PAC recommendation (yes/no). Additional covariates in the initial multiple logistic regression included 6 sociodemographic and clinical factors: sex (binary), age (continuous), insurance type (categorical), hospital (binary), length of stay (continuous), surgical encounter (binary). Backward elimination and Akaike Information Criterion (AIC) were used to determine the most parsimonious model for covariates. All analyses were conducted with type I error rate = 0.05 and R statistical software version 1.3.1093 was used for data analysis.

Subgroup Analysis

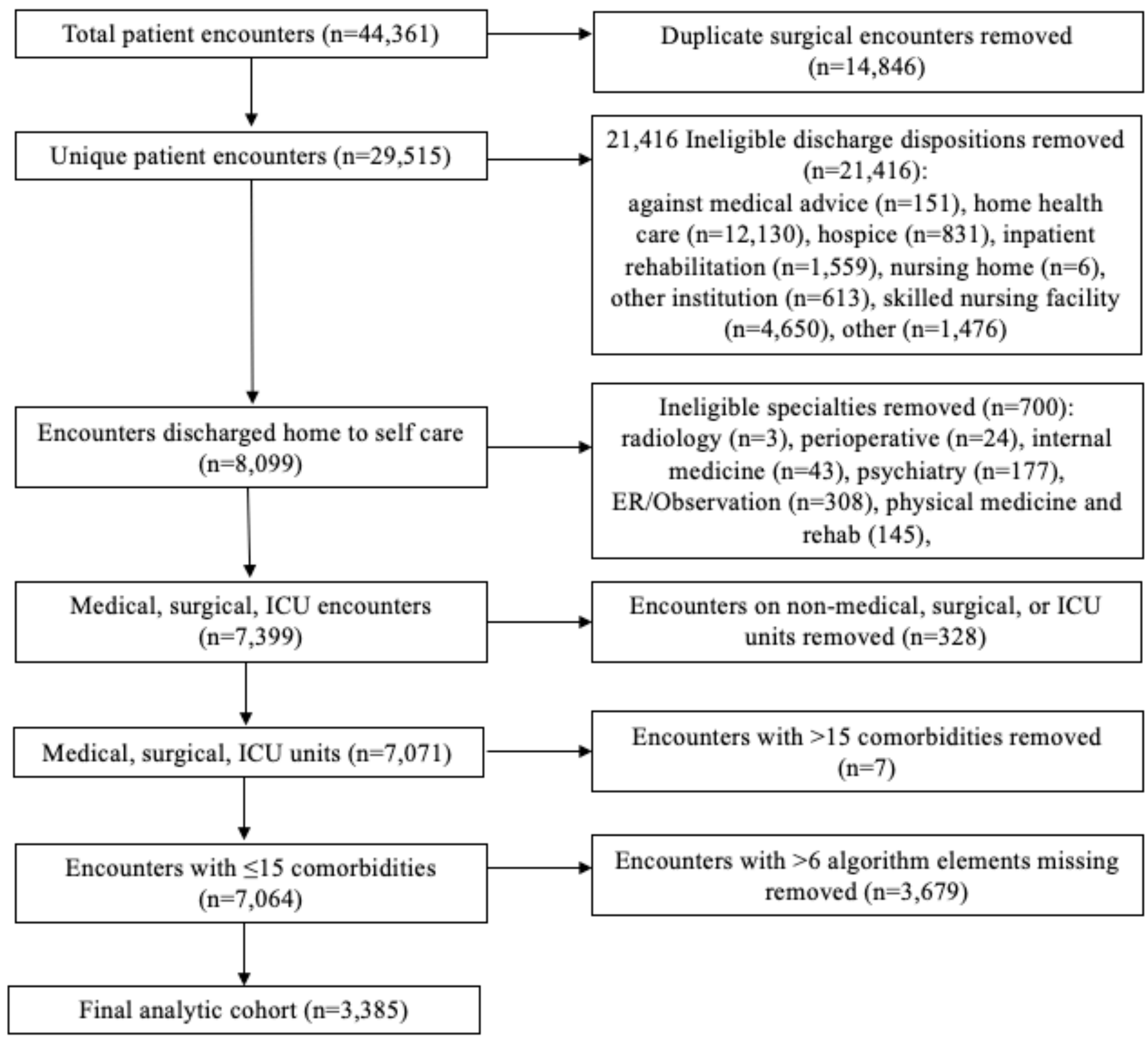
The subgroup analysis focused on the comparison of surgical and non-surgical patients for two reasons. First, some surgical patients were included in the original DIRECT study but were not the focus.[42] Additionally, several studies have identified disparities in PAC referral practices in surgical patients.[12,13,159] We compared patient characteristics and outcomes between surgical and non-surgical patients to identify any differences in clinical population that could impact implementation of DIRECT in real-world clinical settings. We compared logistic regression models with the outcome of 30-day readmission and all original predictors in surgical and non-surgical patient populations.

Results

Cohort selection is demonstrated in **Figure 4**. The initial cohort contained 29,515 unique encounters. After removing discharge dispositions other than home to self-care

(N=21,416) and specialties or units outside of medical, surgical and ICU patients (N=1,028), 7,071 patient encounters remained. Assuming data entry errors, encounters with five times greater than the interquartile range of comorbid conditions were excluded to eliminate outliers (N=7), yielding 7,064 patient encounters. The DIRECT algorithm requires fewer than 7 data elements missing, so patient encounters with greater than 6 missing elements were excluded (N=3,679). The final analytic sample contained 3,385 patient encounters from adults aged 55 years or older discharged home to self-care.

Figure 4: Cohort Selection



2,776 (82%) of patient encounters were flagged by the algorithm as needing PAC, while 609 (18%) were not flagged by the algorithm. The overall readmission rate was 15.2%, and the median length of stay was 4 days. Biological sex was 58% male and 42% female. Median age was 67 years. Although all patients in the sample went home without PAC, there were statistically significant differences between sociodemographic patient characteristics and algorithm elements for those flagged and not flagged by the DIRECT algorithm, demonstrated in **Table 5**.

Table 5: Comparison of Patient and Algorithm Characteristics between Patients Flagged vs. Not Flagged as Needing PAC by DIRECT CDSS

Characteristic	**Overall (N=3385)**		**Not Flagged (N=609)**		**Flagged (N=2776)**		**p-value**
Clinical Variables	**Median [Range] or N (%)**						
30-Day Readmission	515	(15.2%)	95	(15.6%)	420	(15.1%)	0.75
Age	67	[55-101]	64	[55-93]	68	[55-101]	**<0.001**
Length of Stay	4	[2-101]	4	[2-26]	4	[2-101]	0.451
Sex							**<0.001**
Male	1962	(58%)	392	(64.4%)	1570	(56.6%)	
Female	1423	(42%)	217	(35.6%)	1206	(43.4%)	
Encounter Type							**<0.001**
Surgical	1489	(44%)	220	(36.1%)	1269	(45.7%)	
Non-Surgical	1896	(56%)	389	(63.9%)	1507	(54.3%)	
Marital Status							**<0.001**
Married/partnered	1877	(55.5%)	420	(69%)	1457	(52.5%)	
Divorced/separated	370	(10.9%)	41	(6.7%)	329	(11.9%)	
Widowed/single	1103	(32.6%)	145	(23.8%)	958	(34.5%)	
Other	35	(1%)	3	(0.5%)	32	(1.2%)	
Race							**<0.001**
White	2052	(60.6%)	428	(70.2%)	1624	(58.5%)	
Black	1087	(32.1%)	133	(21.8%)	954	(34.4%)	
Other	167	(4.9%)	37	(6.1%)	130	(4.7%)	
Missing*	79	(2.3%)	11	(1.8%)	68	(2.4%)	
Ethnicity							0.86
Hispanic/Latino	58	(1.7%)	11	(1.8%)	47	(1.7%)	
Not Hispanic/Latino	3263	(96.4%)	588	(96.6%)	2675	(96.4%)	
Missing*	64	(1.8%)	10	(1.6%)	54	(1.9%)	
Employment status							**<0.001**

Employed part-time, full-time, or per diem	723	(21.4%)	373	(61.2%)	350	(12.6%)	
Retired/Disabled/ Unemployed	2416	(71.4%)	211	(34.6%)	2205	(79.4%)	
Missing*	246	(7.3%)	25	(4.1%)	221	(8%)	
Insurance							**<0.001**
Medicaid/Managed Medicaid	302	(8.9%)	33	(5.4%)	269	(9.7%)	
Medicare/Managed Medicare	1998	(59%)	248	(40.7%)	1750	(63%)	
Private/Commercial or Managed Care/ Self-Pay	1083	(32%)	328	(53.9%)	755	(27.2%)	
Missing*	2	(<0.1%)	0	(0%)	2	(0.1%)	
Hospital							**<0.001**
Tertiary Hospital (Site 1)	2055	(60.7%)	439	(72.1%)	1616	(58.2%)	
Community Hospital (Site 2)	1330	(39.3%)	170	(27.9%)	1160	(41.8%)	
Algorithm Variables							
Employment							**<0.001**
Employed	723	(21.4%)	373	(61.2%)	350	(12.6%)	
Not currently employed	2662	(78.6%)	236	(38.8%)	2426	(87.4%)	
Hospitalization in the 6 months prior to admission							**0.03**
No hospitalization	2441	(72.1%)	461	(75.7%)	1980	(71.3%)	
Hospitalization	944	(27.9%)	148	(24.3%)	796	(28.7%)	
Morse Fall Risk Score (0-125)							**<0.001**
Fall risk ≤20	224	(6.6%)	103	(16.9%)	121	(4.4%)	
Fall risk >20	3161	(93.4%)	506	(83.1%)	2655	(95.6%)	
Use of Equipment/Assistive Devices at Home							**<0.001**
No equipment used	2990	(88.3%)	599	(98.4%)	2391	(86.1%)	
Equipment used	395	(11.7%)	10	(1.6%)	385	(13.9%)	
Home Accessibility Concerns							0.06
No concerns	179	(5.3%)	42	(6.9%)	137	(4.9%)	
Concerns	3206	(94.7%)	567	(93.1%)	2639	(95.1%)	
Presence of Wound							**<0.001**
No Wounds	3216	(95%)	602	(98.9%)	2614	(94.2%)	
Wound present	169	(5%)	7	(1.1%)	162	(5.8%)	

Ambulation							**<0.001**
Improved	629	(18.6%)	86	(14.1%)	543	(19.6%)	
No change	2361	(69.7%)	517	(84.9%)	1844	(66.4%)	
Declined	391	(11.5%)	5	(0.8%)	386	(13.9%)	
Missing*	4	(0.1%)	1	(0.1%)	3	(0.1%)	
Transfer							**<0.001**
Improved	384	(11.3%)	23	(3.8%)	361	(13%)	
No change	1400	(41.4%)	410	(67.3%)	990	(35.7%)	
Declined	179	(5.3%)	1	(0.1%)	178	(6.4%)	
Missing*	1422	(42%)	175	(28.7%)	1247	(44.9%)	
Bathing							**<0.001**
Improved	177	(5.2%)	26	(4.3%)	151	(5.4%)	
No change	1359	(40.1%)	311	(51.1%)	1048	(37.8%)	
Declined	347	(10.3%)	8	(1.3%)	339	(12.2%)	
Missing*	1502	(44.4%)	264	(43.3%)	1238	(44.6%)	
Eating							**<0.001**
Improved	94	(2.8%)	12	(1.9%)	82	(3%)	
No change	863	(25.5%)	216	(35.5%)	647	(23.3%)	
Declined	40	(1.2%)	1	(0.1%)	39	(1.4%)	
Missing*	2388	(70.5%)	380	(62.4%)	2008	(72.3%)	
Number of Comorbid Conditions	2	[0-14]	2	[0-11]	2	[0-14]	**<0.001**
Caregiver							**<0.001**
Caregiver	939	(27.7%)	325	(53.4%)	614	(22.1%)	
No caregiver/Unknown	2446	(72.3%)	284	(46.6%)	2162	(77.9%)	
Spousal Caregiver							**<0.001**
Spousal caregiver	547	(16.2%)	236	(38.8%)	311	(11.2%)	
Non-spousal caregiver or unknown	2838	(83.8%)	373	(61.2%)	2465	(88.8%)	
Discharged with an opioid							**<0.001**
Discharged with an opioid	627	(18.5%)	38	(6.2%)	589	(21.2%)	
Not discharged with an opioid	2758	(81.5%)	571	(93.8%)	2187	(78.8%)	

Legend: Tests include Chi-squared, Kruskal-Wallis, or Fisher's Exact as appropriate | Bold indicates p<0.05 |* indicates missing values were excluded from descriptive statistics tests

Comparison of 30-Day Readmissions by Algorithm Flag

Overall, 515 (15.2%) encounters experienced a 30-day readmission at one of the health system hospitals. Patient encounters flagged by the algorithm experienced an unadjusted 0.5% lower 30-day readmission rate compared to those not flagged (15.1% vs. 15.6%, p=0.75). The original logistic regression model, presented in **Table 6**, included all original variables. The final logistic regression model, presented in **Table 7**, was selected based upon having the lowest AIC and included all initial covariates except sex. Adjusted 30-day odds of readmission were 1.3% lower in the flagged groups (OR 0.987, 95% CI 0.77-1.277; p=0.922). The AIC of the final model was 2825.5, while the AIC of the model containing all initial variables was 2827.1.

Table 6: Original Logistic Regression Model with the Outcome of 30-Day Readmission (AIC 2827.1)

Factor	**OR (95% CI)** Unadjusted	Adjusted	**P-Value** Unadjusted	Adjusted
Flagged by algorithm (yes)	0.965 (0.76-1.234)	0.992 (0.773-1.283)	0.77	0.951
Sex (female)	0.88 (0.726-1.065)	0.94 (0.773-1.141)	0.191	0.534
Age (years)	0.994 (0.983-1.004)	0.992 (0.979-1.005)	0.24	0.24
Insurance				
Medicare/managed Medicare (REF)				
Medicaid/managed Medicaid	0.64 (0.425-0.932)	0.652 (0.422-0.979)	**0.025**	**0.046**
Private, Commercial, Managed Care, Self-Pay	1.123 (0.917-1.371)	0.997 (0.787-1.261)	0.26	0.981
Hospital (Site 1)	1.511 (1.239-1.85)	1.568 (1.276-1.933)	**<0.001****	**<0.001****
Length of Stay	1.009 (0.993-1.024)	1.007 (0.99-1.023)	0.224	0.414
Encounter Type (Surgical)	1.936 (1.602-2.343)	2.006 (1.653-2.437)	**<0.001****	**<0.001****

Legend: Bold indicates p<0.05, * indicates p<0.01, ** indicates p<0.001

Table 7: Final Logistic Regression Model with the Outcome of 30-Day Readmission – No Sex Variable (AIC = 2825.5)

Factor	**OR (95% CI)**		**P-Value**	
	Unadjusted	Adjusted	Unadjusted	Adjusted
Flagged by algorithm (yes)	0.965 (0.76-1.234)	0.987 (0.77-1.277)	0.77	0.922
Age (years)	0.994 (0.983-1.004)	0.992 (0.979-1.005)	0.191	0.238
Insurance				
Medicare/managed Medicare (REF)				
Medicaid/managed Medicaid	0.64 (0.425-0.932)	0.652 (0.422-0.98)	**0.025**	**0.046**
Private, Commercial, Managed Care, Self-Pay	1.123 (0.917-1.371)	0.999 (0.788-1.263)	0.26	0.993
Hospital (Site 1)	1.511 (1.239-1.85)	1.571 (1.278-1.937)	**<0.001****	**<0.001****
Length of Stay	1.009 (0.993-1.024)	1.007 (0.99-1.023)	0.224	0.413
Encounter Type (Surgical)	1.936 (1.602-2.343)	2.015 (1.662-2.447)	**<0.001****	**<0.001****

Legend: Bold indicates p<0.05, * indicates p<0.01, ** indicates p<0.001

Subgroup Analysis

Differences in patient characteristics between surgical and non-surgical patients are illustrated in **Appendix C**. Surgical patients experienced an 8.6% higher readmission rate than non-surgical patients (20% vs. 11.4%), and this difference was statistically significant (p<0.001). On average, surgical patients were older, male, white, married/partnered, and employed compared with non-surgical patients, and these differences were statistically significant. A higher proportion of surgical patients were flagged by DIRECT as needing PAC than non-surgical patients (85.2% vs. 79.5%, p<0.001).

There were also notable differences in DIRECT algorithm variables. Surgical patients had higher rates of hospitalization in the 6 months prior to admission (39.8% vs. 18.6%, p<0.001), fewer surgical patients had caregivers (24.1% vs. 30.6%, p<0.001), and surgical patients were discharged with an opioid more frequently (28.5% vs. 10.7%, p<0.001). Surgical patients also experienced greater decline across all activities of daily living (ambulation, transfer, bathing, eating), and these differences were statistically significant for transfer (p<0.001) and bathing (p<0.001).

Table 8 shows the logistic regression model for surgical patients, while **Table 9** shows the logistic regression model for non-surgical patients. Surgical patients flagged by DIRECT as needing PAC experienced an adjusted 51.8% higher rate of 30-day readmission, and this difference was statistically significant (p=0.041). This was the only statistically significant variable in both unadjusted and adjusted analysis for surgical patients. In **Table 9**, non-surgical patients flagged by DIRECT as needing PAC experienced an adjusted 23.8% lower rate of 30-day readmission. This difference was not statistically significant (p=0.116). The only statistically significant covariate in the logistic regression model for non-surgical patients was hospital (p<0.001), with patients from Site 1 (tertiary hospital) experiencing 3.66 times higher odds of 30-day readmission compared with Site 2 (community hospital).

Table 8: Logistic Regression Model with the Outcome of 30-Day Readmission – Surgical Patients (AIC = 1497.3)

Factor	**OR (95% CI)**		**P-Value**	
	Unadjusted	Adjusted	Unadjusted	Adjusted
Flagged by algorithm (yes)	1.496 (1.022-2.252)	1.518 (1.029-2.3)	**0.045**	**0.041**
Sex (female)	0.865 (0.663-1.124)	0.851 (0.651-1.108)	0.282	0.233
Age (years)	1.001 (0.986-1.016)	0.999 (0.982-1.018)	0.875	0.968
Insurance				
Medicare/managed Medicare (REF)				
Medicaid/managed Medicaid	0.717 (0.39-1.239)	0.692 (0.366-1.236)	0.257	0.234
Private, Commercial, Managed Care, Self-Pay	0.969 (0.736-1.269)	1.016 (0.737-1.397)	0.819	0.922
Hospital (Site 1)	1.019 (0.789-1.316)	0.99 (0.763-1.287)	0.887	0.942
Length of Stay	1.015 (0.992-1.037)	1.016 (0.992-1.039)	0.197	0.176

Legend: Bold indicates p<0.05, * indicates p<0.01, ** indicates p<0.001

Table 9: Logistic Regression Model with the Outcome of 30-Day Readmission – Non-Surgical Patients (AIC=1298.6)

Factor	**OR (95% CI)**		**P-Value**	
	Unadjusted	Adjusted	Unadjusted	Adjusted
Flagged by algorithm (yes)	0.605 (0.442-0.836)	0.762 (0.545-1.075)	**0.002***	0.116
Sex (female)	0.987 (0.742-1.31)	1.048 (0.783-1.4)	0.928	0.753
Age (years)	0.981 (0.964-0.997)	0.986 (0.965-1.006)	**0.023**	0.17
Insurance				
Medicare/managed Medicare (REF)				
Medicaid/managed Medicaid	0.692 (0.388-1.159)	0.676 (0.363-1.195)	0.148	0.196
Private, Commercial, Managed Care, Self-Pay	1.326 (0.978-1.789)	1.005 (0.704-1.43)	0.067	0.978
Hospital (Site 1)	3.953 (2.684-6.03)	3.656 (2.464-5.611)	**<0.001****	**<0.001****
Length of Stay	1.008 (0.984-1.028)	0.996 (0.967-1.02)	0.477	0.76
Legend: Bold indicates p<0.05, * indicates p<0.01, ** indicates p<0.001				

Discussion

We found significant differences between patient characteristics of those flagged and not flagged for PAC referral in nearly all sociodemographic factors and DIRECT CDSS variables. This is consistent with a prior quasi-experimental study using DIRECT CDSS for patients at a community hospital. In that study, patients flagged by DIRECT CDSS as needing PAC tended to be older, with poor self-rated health, more comorbid conditions, more hospitalizations in the 6 months prior, higher fall risk, greater decline in activities of daily living, female, single or widowed than those not flagged.[86] After the CDSS was implemented in the community hospital, patients with poor self-rated health and greater than 4 hospitalizations in the 6 months prior to admission were less likely to be discharged home to self-care.[42]

A higher percentage of females were flagged as needing PAC than males. This is consistent with prior literature demonstrating that women have longer life expectancies and frequently less in-home caregivers compared with men.[165,166] Black patients were flagged by DIRECT CDSS as needing PAC at 12.6% higher rates compared to white and other races, which is supported by literature demonstrating racial disparities in PAC referrals. Although this study focuses on patients discharged home without PAC, recent literature comparing referrals to home health care vs. facility level care reported non-white patients experienced lower rates of home health care referrals compared with white patients, increasing their risk for negative outcomes. A recent study of total knee arthroplasty patients reported that African American patients were more likely to be discharged to inpatient rehabilitation or skilled nursing compared to white patients ($p<0.001$), and these PAC destinations were associated with higher odds of 90-day readmission ($p<0.001$).[167] Similar results were presented in a study of brain tumor surgery patients where Black patients had 7% higher rates of non-home discharge.[168] Reducing racial bias in predictive models and clinical decision making is a key focus in data science,[169,170] and tools like the DIRECT CDSS which identify a patient's need for PAC services based on clinical factors alone could be particularly useful in reducing biased decision making in these populations.

Patient encounters flagged by DIRECT CDSS in this study experienced slightly lower 30-day readmission rates compared to those not flagged. In the prior community hospital study, patients discharged home to self-care but flagged by DIRECT had a 67.8% higher risk of 30-day readmission than patients not flagged ($p=0.006$).[86] This difference may be partly explained by differences in patient population, health system resources, missing data, and changes in PAC referrals over time. Compared with that study, the sample in this study had fewer females (42% vs. 54.6%), was more racially diverse (60.6% white vs. 85.4% white) and was younger (mean age 67.6 vs. 75.9 years).

Greater proportions of the sample in this study were employed (21.4% vs. 15.1%) and married (55.5% vs. 49.8%).[42]

In this study, the average readmission rate was 15.2%, and statistically significant covariates in our final model included hospital, insurance (Medicaid or Managed Medicaid), and surgical encounter type. Recent research exploring the relationship between structural characteristics of hospitals and readmissions found important differences in size, region, teaching or safety-net status, nurse-patient ratios, and patient mix across the United States.[171] This larger, urban, academic health system has a more diverse patient population, culture of research, and several readmission reduction interventions already implemented compared to the community hospital in the prior quasi-experimental study conducted several years earlier, which may have contributed to these results. For example, we anticipated “home to self-care” (did not receive any PAC referral) would be the largest discharge disposition category among adults aged 55 years or older in the two hospitals, but we found that the largest category was home health care. 39.2% were discharged with home health care compared with 27.4% discharged home to self-care, which demonstrates that the health system was already accomplishing a high volume of PAC referrals. This change aligns with national trends in increased use of PAC after hospital discharge among Medicare beneficiaries from 2000 to 2015.[172] However, it should be noted that the COVID-19 pandemic has further impacted PAC usage with a shift towards greater use of home health care compared to other facility PAC.[173] It is also possible that there were differences in insurance status in this cohort, which can significantly impact access to PAC referrals and outcomes like 30-day readmission if the patient does not receive adequate care.[174,175]

Subgroup analysis revealed important differences between surgical and non-surgical patients. Non-surgical patients discharged home to self-care had significantly lower rates of 30-day readmissions and fewer declines in functional status. This could mean that existing discharge planning initiatives in the health system for medical patients are successfully identifying the right patients for PAC. However, surgical patients were more likely to be flagged for PAC, and those flagged experienced significantly higher rates of 30-day readmission. One of the interesting aspects of including surgical patients is that many of those admissions are planned with pre-determined discharge dispositions, which may reflect physician preferences prior to surgery rather than patient needs after the procedure.[176,177] A recent qualitative study of patients with planned surgical procedures reported that more than half of patients demonstrated lower coping scores in the early post-discharge period despite having preoperative teaching sessions, which highlights the risk of pre-determined discharge dispositions.[178] Another quantitative study of older surgical patients reported higher rates of PAC referrals among patients with pre-operative risk factors like age and lower functional status and/or one or more post-surgical complication.[179]

Surgical patients in this study experienced greater declines in activities of daily living (ambulation, transfer, bathing, eating) over the course of the hospitalization, fewer caregivers, and a higher percentage of opioid prescriptions at discharge compared with non-surgical patients, which warrants the need for patient education and/or consideration of PAC referral for skilled nursing care at discharge. The higher readmission rate among surgical patients holds clinical importance from both a patient outcomes and cost perspective. The average cost of a readmission for Medicare patients is $15,500,[180] which is higher than the average cost of an index hospitalization.[181] Preventing negative outcomes like 30-day readmission among even a small fraction of patients can reduce suffering, improve the quality of care, reduce readmission-related penalties for hospitals, and reduce costs for insurers. In the prior study after the intervention period where DIRECT CDSS was implemented, discharge disposition agreement with the algorithm was associated with a 4% decrease in 30-day readmissions (22% relative reduction).[42]

Missing data is a common challenge in research using EHR data and our study was no exception. 52% of eligible older adults discharged home to self-care were excluded from the study because they had greater than 6 DIRECT elements missing. Less than half of the analytic sample had complete data for all DIRECT elements, and most of the missing data occurred in the measures of activities of daily living. The algorithm was trained to be highly sensitive to missing data, and partially explains why over 80% of patients in the analytic sample were flagged as needing PAC in this study. Therefore, patients with more missing algorithm variables have higher odds of being flagged for PAC referral so that the clinician can be alerted and make a judgment call based on additional information that might not be present in the structured EHR data. This is consistent with other hospital CDSS tools which are trained to be highly sensitive.[182,183]

Another challenge regarding data quality was the inclusion a broad sample across two hospitals because different units have different nursing documentation flowsheets and required documentation elements, contributing to missing data. For example, it is likely that most nurses assess patients' ambulation consistently, but it is documented inconsistently at a health system level. We analyzed data from four different structured nursing documentation sources and found wide variation across units (for example, the “eating” functional status variable had 70.2% missing cases), which is consistent with other large studies.[184,185] Further complicating this situation in an older adult population is the possibility of missing data if clinicians are biased against certain patient populations or have difficulty completing assessments due to cognitive impairment[186,187] or other limitations, especially for activities of daily living. We agree with the recent recommendation from Holmes et al.[188] to prospectively advocate for standardized, required nursing documentation to reduce incompleteness, inaccuracy, and variation in data collection. Ideally, these standardized assessments could be mapped to ontologies like the Systematized Nomenclature of Medicine (SNOMED) and Logical Observation Identifiers Names and Codes (LOINC) to improve access and interoperability.[189]

A recent study found that only 0.3% of CDSS studies are replicated in the literature, and posited that this could be due to both a research culture that values novelty over replication research and that replication research may only have a minimal impact on implications for real-world settings.[190] We hoped to address this gap by generating initial evidence of external validity of the DIRECT algorithm in a new setting through a retrospective study of EHR data, but still struggled with practical issues relevant to implementation including EHR data quality and interoperability which likely impacted algorithm performance. CDSS is designed to utilize data from a central, standardized data repository, which is rarely possible in real-world hospital settings. One recent study even argued that unstandardized data collection leads to corrupt CDSS data.[183] Although some variables in this study like employment status, caregiver, fall risk, and medications were documented in a standardized assessment with very little missing data, other variables like activities of daily living were not consistently documented in a central nursing flowsheet, leading to data quality issues. This is consistent our recent systematic review which found very few discharge planning prediction models incorporate activities of daily living, which suggests that health systems may not prioritize its documentation.[87]

These data quality problems are closely related to standardization and interoperability challenges. The EHR in this health system used similar but not identical nursing assessment documentation flowsheets as was used during the algorithm development study. There was also variation in assessment definitions within the health system. The four ambulation documentation sources mentioned earlier defined ambulation differently and assessed it on unique numeric scales. This diversity in documentation required us to recode variables to develop a close but not perfect match to DIRECT CDSS variable definitions. This unstandardized documentation may also have contributed to the unexpected results for the 30-day logistic regression model, but this is a frequent challenge in CDSS studies as a whole.[183] The 2014 Improving Medicare Post-Acute Care Transformation (IMPACT) Act requires the standardization and interoperability of categories of patient assessment content.[191] Although currently focused on post-acute settings, expansion of this work to acute care would greatly benefit implementation of CDSS systems. In addition, Sutton et al. recommends that health systems and CDSS utilize standards like Health Level 7 (HL7), Fast Healthcare Interoperability Resources (FHIR), and other cloud based solutions to improve operability and portability to new health systems.[183]

Finally, there are likely other factors outside of the DIRECT CDSS that may prevent flagged patients from getting the care they need, such as social determinants of health and patient refusal. Social determinants of health such as socio-economic status, insurance coverage, healthcare access, education, neighborhood, and social context contribute to discharge planning decision making and have become a recent focus area in research.[192] One study found that up to 28% of patients refuse PAC, and those patients experienced higher rates of 30-day readmissions compared to those who did not refuse.[86] Future studies should assess the impact of these factors on PAC destination during the discharge planning process using natural language processing, qualitative, or mixed-methods to uncover barriers to PAC.

Limitations

Study limitations included the use of retrospective EHR data from one health system. Challenges of using retrospective EHR data in research include incompleteness, inaccuracy, and inconsistency.[193] We lost over half of eligible older adult patient encounters due to missing data and data quality issues. To address these issues and minimize bias, the team interviewed nurse informaticians during the data extraction process to ensure that we captured all possible data sources. We were unable to capture the full scope of readmission rates and patient mortality if patients experienced these negative outcomes outside of our health system, but since the study was conducted in a large academic health system with six hospitals in the region we were able to capture outcomes data in five of the six hospitals. The last hospital's outcomes were not captured due to EHR incompatibility, but it is unlikely that patients would be readmitted there because it is located a significant distance from the index admission hospitals. Since we simulated the DIRECT algorithm using retrospective data, the algorithm has not been installed in the health system and therefore clinicians were not able to view the DIRECT recommendation. Future directions might include external validation with model updating, use of natural language processing to understand barriers to receiving a PAC referral among patients flagged, and/or implementation studies.

Conclusion

These results demonstrate the potential applicability of DIRECT CDSS in large urban health systems, especially for surgical patients, and highlight the real-world challenges of translating CDSS to new settings. The DIRECT algorithm identifies patients who need post-acute care to prevent poor outcomes using a set of holistic clinical and administrative variables. Implementing the CDSS in practice and assuring the collection of algorithm variables could improve its value in identifying those who need PAC services and potentially leading to reductions in negative outcomes like 30-day readmissions. Future directions include using natural language processing on discharge planning notes to identify reasons why patients did not receive PAC and implementing DIRECT CDSS in the health system to understand how discharge planning teams viewing the algorithm impacts PAC referral rates and patient outcomes.

CHAPTER 4: Paper 3

IDENTIFYING BARRIERS TO POST-ACUTE CARE REFERRAL AND CHARACTERIZING NEGATIVE PATIENT PREFERENCES AMONG HOSPITALIZED OLDER ADULTS USING NATURAL LANGUAGE PROCESSING

*This paper is under review with the American Medical Informatics Association Annual Symposium Student Paper Competition.

Abstract

Our objective was to detect common barriers to post-acute care (B2PAC) among hospitalized older adults using natural language processing (NLP) of clinical notes from patients discharged home when a clinical decision support system recommended post-acute care. We annotated B2PAC sentences from discharge planning notes and developed an NLP classifier to identify the highest-value B2PAC class (negative patient preferences). Eight machine learning models were compared with Amazon's AutoGluon deep learning model. The study included 594 acute care notes from 100 patient encounters (1156 sentences contained 11 B2PAC) in a large academic health system. The most frequent and modifiable B2PAC class was negative patient preferences (18.3%). The best supervised model was XGBoost (F1: 0.859), but the deep learning model performed better (F1: 0.916). Alerting clinicians about negative patient preferences early in the hospitalization can prompt interventions like patient education to ensure patients are appropriately informed about PAC, participate in shared decision making, and ultimately avoid negative outcomes.

Introduction

Older adults experience more chronic, complex health conditions with greater hospitalization rates and higher acuity than younger adults, making them susceptible to negative health outcomes like hospital readmissions.[194-197] Coordinated discharge planning is an evidence-based strategy to reduce negative outcomes like readmission.[4] The goal of discharge planning is to determine a patient's discharge disposition, which is usually either home or post-acute care (PAC). PAC is defined as skilled home health care (as opposed to no care or a home health aide), long-term acute care hospitals, inpatient rehabilitation, and skilled nursing facilities.[3] Discharge planning teams that are most effective in reducing 30-day readmissions are multidisciplinary (medicine, nursing, physical therapy, social work, case management, and others) and begin the process early in the hospitalization.[198] However, these teams increase the complexity of an already difficult decision due to the involvement and communication between multidisciplinary inpatient and outpatient healthcare providers or PAC organizations, insurance companies, patients, and families.[7] This process is individualized to each patient, but as acuity increases and clinicians face more time constraints, standardization through informatics solutions can be used to ensure effective discharge planning.

Although many discharge disposition clinical decision support systems (CDSS) and 30-day readmission risk prediction models have been developed to aid clinician decision making processes for discharge disposition,[175] to our knowledge, no studies have comprehensively explored the barriers to actually receiving these services. This study builds on the team's prior research involving the Discharge Referral Expert System for Care Transitions (DIRECT) CDSS that supports patient-centered discharge planning for older adults. The DIRECT algorithm identifies a patient's need for PAC referral (discharge home to self-care vs. PAC including home health care). Algorithm development[41,199] and quasi-experimental testing in a community hospital[42,86] are described in prior research. In the testing study, discharge disposition agreement with DIRECT was associated with a statistically significant reduction in 30-day readmission rates (4% overall, 22% relative reduction). However, it recommended 25.6% more patients for PAC than actually got services.[42] The readmission risk was 68% higher among those patients who were recommended PAC, but were discharged home without services.[86] DIRECT identifies a patient's clinical need for post-acute care, but does not identify real-world barriers that could prevent a patient from getting PAC. This study illuminates those barriers.

Understanding the reasons why patients do not receive PAC referrals requires identifying and quantifying complex clinical, social, and economic factors that impact the patient throughout the hospitalization. Standardized terminologies like the Systematized Nomenclature of Medicine Clinical Terms (SNOMED-CT) and Unified Medical Language System (UMLS) are still in the early stages of developing standards for these concepts.[200] Because each discharge planning process is uniquely centered around the patient, most of this information is documented in unstructured clinical notes. Natural language processing (NLP) algorithms unlock this important source of data by automating the process of reading and extracting information from clinical notes, enabling thousands of notes to be processed systematically by a computer much more efficiently than a human.[201] This data can be used to improve CDSS, public health surveillance, and other applications.[60] A recent call to action from CDSS experts recommended the use of NLP to improve CDSS.[61]

Preliminary barriers to PAC have been identified in qualitative studies with patients and clinicians,[11] including patient preferences, clinical reasoning, and psychosocial factors.[4,11,18,202,203] NLP systems have been developed to extract related psychosocial factors such as chronic stress, social isolation, financial insecurity, housing insecurity, and criminal justice.[62,200,204] Although researchers have developed solutions for extracting some high-value barriers to PAC including social determinants more broadly; few have extracted factors related to patient and family preferences.

The aim of the study was to address this gap by identifying common barriers to post-acute care (B2PAC), then developing and evaluating an NLP system to encode sentences containing negative preferences among hospitalized older adults. The aim was accomplished with two main objectives:

1. Develop a representation and reference standard of known (literature review) and observed (data-driven chart review) barriers to post-acute care (B2PAC) among hospitalized older adults using retrospective clinical notes.
2. Develop and validate an automatic negative preferences classifier for sentences from clinical notes.

Methods

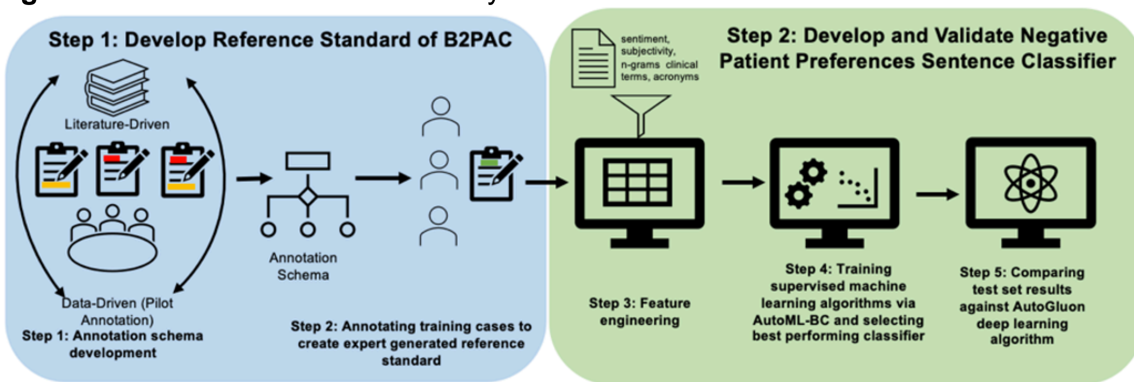
Study Dataset

The NLP study was approved by the University of Pennsylvania Institutional Review Board (#843687). The sample includes case management, social work, and discharge summary notes of patient encounters at two large urban hospitals in a large academic health system in Philadelphia, Pennsylvania. Stakeholder interviews with health system leaders informed the selection of these particular clinical notes. The clinical notes were de-identified of protected health information using a text deidentification system called De-ID.[205] The sample includes patients admitted between December 1, 2018 and December 1, 2019, aged 55 years or older, admitted to medical or surgical service lines, with a hospitalization ≥ 48 hours (to avoid observation stays), and discharged alive. The DIRECT algorithm was applied to the retrospective data. The sample in this study is a randomly sampled subset of 100 patient encounters whose discharge disposition was “home to self-care” although the DIRECT algorithm recommended PAC. For objective 1, results are reported at both the sentence level (unit of analysis for the NLP study) and the patient level in order to report clinical insights. For objective 2, the sample was randomly split into training (70%) and testing (30%) sets at the sentence level, which is typical for an NLP study.

B2PAC Schema Development and Annotation Study

Multi-class B2PAC definitions were constructed through literature review as common reasons why patients may not get PAC to inform the annotation schema. Three broad classes were identified from the literature search: preferences, psychosocial factors, and clinical reasoning. Preferences were defined as individuals’ evaluation of dimensions of health based on cognition, experience and values.[202] Patient and family preferences influence shared decision making for discharge disposition.[16,17,203,206] Psychosocial factors were defined as social circumstances that shape health risks and outcomes and were included because they impact access to PAC.[62,204,207] Clinical reasoning was defined as a recommendation based on clinician assessment, and was included because clinical evaluation may reveal needs beyond structured data from CDSS.[62,208,209] **Figure 5** illustrates the study workflow.

Figure 5: Process Workflow and Study Methods



The annotation schema was refined through an annotation study, where additional novel data-driven categories for inclusion were considered, and an expert-generated reference standard was annotated in the extensible Human Oracle Suite of Tools (eHOST) software.[76] To create the reference standard, the 100 patient encounters were divided into 11 batches. Three annotators (EK, KB, and AD) applied the initial schema to the first batch of clinical notes using eHOST.[76] The team computed inter-annotator agreement (IAA) using match criterion defined by the NLP sub-class type. After each iteration, IAA was assessed using F1-score.[210] The team discussed and resolved disagreements using consensus review and updated the guidelines accordingly. The team iteratively refined, applied, and updated the annotation schema and guidelines over batches 0 through 2. Once the schema was finalized, two annotators completed annotations for batches 3 through 6 (EK and AD), and two annotators completed annotations for batches 7 through 10 (EK and KB). All three annotators (EK, KB, AD) met to resolve any remaining disagreements and finalize the reference standard.

After the annotation study, the final schema included 11 classes in **Table 10**. To address small sub-class sizes, similar sub-classes were combined and classes appearing in fewer than 10 sentences were removed. B2PAC distributions are reported at both the sentence and patient level. Pearson correlation coefficients were computed for pairwise combinations of all 11 classes at the patient encounter level to determine potential correlations between observed B2PAC classes and is visualized as a heatmap. Correlation coefficients range from -1 to 1, with larger values indicating a higher likelihood of observing both classes for a patient. The highest-value B2PAC class from both a data-driven and clinical standpoint was selected for development into an NLP sentence classifier.

Feature Engineering

All feature engineering, model development, and analysis was conducted in Python 3 using various NLP libraries and open-source tools.[211] Lexical, sentiment, and semantic features were encoded. Lexical features included creating n-grams and applying Porter stemming using the natural language toolkit (NLTK).[212] Sentiment features like negative, positive, fear, worry, and happy were extracted using Empath.[213] Subjectivity features including direction (positive, negative, neutral) and magnitude (weak, strong) were encoded using the Multi-Perspective Question Answering (MPQA) Subjectivity Lexicon.[214] Clinical concepts were extracted using the UMLS entity linker via scispaCy.[215,216] An abbreviations feature set was created to identify common discharge planning abbreviations such as "HHC" for home health care.

Experiments with Supervised Machine Learning Classifiers

Using the annotated training set, a newly developed automated machine learning (ML) system called AutoMLPipe-BC[217] (available online) was trained to learn prediction models that accurately classify sentences from the clinical notes according to instances of negative preferences. AutoMLPipe-BC is a rigorous ML analysis pipeline that applies scikit-learn[218] ML modeling algorithms with automated pre-processing, feature selection, hyperparameter optimization, feature importance evaluation, statistical analysis, and data/evaluation visualizations. Eight ML algorithms including Naïve Bayes, Logistic Regression, Decision Tree, Random Forest, XGBoost, Support Vector Machines (SVM), Artificial Neural Network (ANN), and K-Nearest Neighbors (K Neighbors) were applied to train respective models with 10-fold cross validation on the 70% training set, then applied to the 30% hold out test set for evaluation. The team evaluated which textual features and algorithms best predicted negative preferences from clinical notes. Feature importance was evaluated uniformly across all models using a permutation-based estimator. To evaluate how well the classifiers identified negative preferences, we focused on area under the receiver operating curve (AUROC) as well as precision-recall curve (AUPRC) and average precision score (APS) as the data had imbalanced class counts.[77,82] Significant differences in ML performance, for each metric, were evaluated with non-parametric Kruskall-Wallis and subsequent pairwise Mann-Whitney U testing.

Experiment with Amazon's AutoGluon Deep Learning Classifier

A deep learning model was developed using Amazon AutoGluon's TextPredictor for comparison.[219] The model was developed with 9 epochs and 46 iterations on the 70% training set, then applied to the 30% test set. The only pre-processing step included reducing text case; no feature engineering was performed. TextPredictor relies on pretrained NLP models including ELECTRA for transfer learning to fit a transformer neural network model.[219-221] Evaluation metrics included AUROC, average precision, precision, recall, and F1-score.

Results

Sample

The final study sample contained 594 notes from 100 encounters of older adult patients who were discharged home to self-care when the DIRECT algorithm identified them as needing B2PAC. The sample was 58% male, median age was 66, and median length of stay was 7 days. 56% of patient encounters were white, 35% were black, and 9% were another race or unknown. Ethnicity was 5% Hispanic/Latino, 93% not Hispanic/Latino, and 2% unknown. Insurance was 60% Medicare or Managed Medicare, 15% Medicaid or Managed Medicaid, 14% managed care, and 11% private or commercial insurance. Most patient encounters were retired or disabled (64%); in contrast, 16% were employed, and 16% were not currently employed (4% unknown).

Annotation Study

The final annotation schema is described in **Table 10** including definitions, examples, and frequencies. Patients experienced a mean of 3.68 B2PAC classes (median 3, range 1-7). Average IAA from early batches to the final round of annotations between EK and AD improved by 12.8% (64.9% in batches 0-2 vs. 77.7% in batches 3-6), and IAA between EK and KB improved by 8.3% (51.3% in batches 0-1 vs. 59.6% in batches 7-10). During the annotation study, the "Received PAC Referral" class was added because the team discovered that some patients were transferred to PAC units within the hospitalization, or outside facilities at discharge and their discharge disposition was incorrect.

Table 10: Final B2PAC Annotation Schema and Sub-Class Frequency at the Sentence and Encounter Level

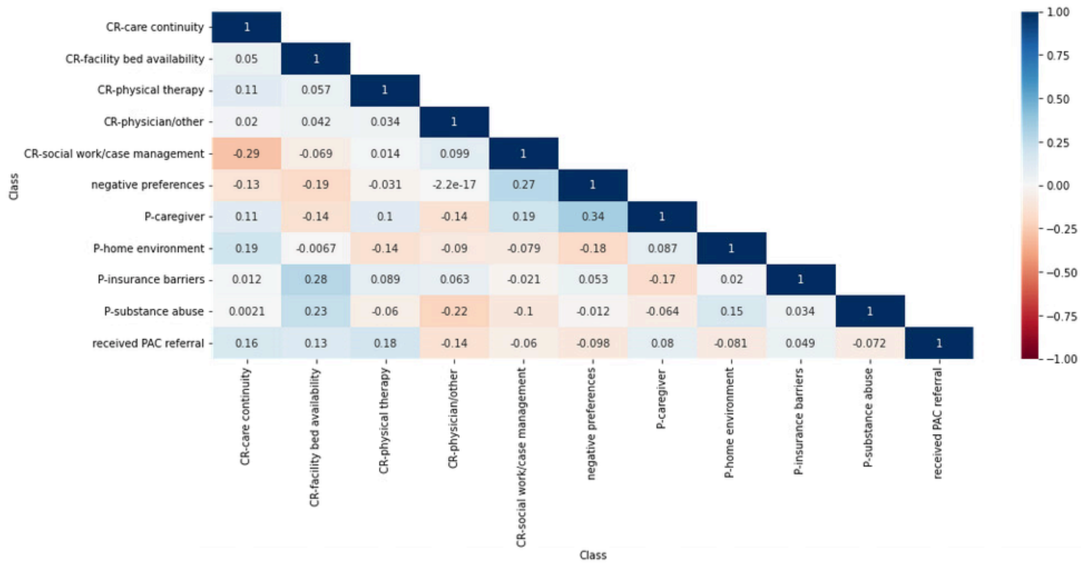
Class.	Sub-Class	N+	N*(%)	IAA %~	Definition	Example Sentence
Received PAC Referral		13	25 (2.2%)	48.3%	Patients who received a PAC referral but their discharge disposition was incorrectly coded as "home to self-care"	"On [**DATE**], the patient was discharged to [Skilled Nursing Facility Unit]"
Negative Prefer-ences	Patient or family	72	211 (18.3%)	88.7%	Statements from the patient or family indicating they prefer not to have post-acute care or are unsure	"Patient and wife [**NAME**] made an informed decision to refuse home care"
Psycho-social Factors	Caregiver	95	474 (41%)	76%	Patient has a full-time caregiver living in the home or patient has caregiver support outside the home	"His daughter is a nurse and will help out with care during most days"
	Substance Abuse	14	75 (6.5%)	66%	Evidence of current substance abuse issues that could impact the patient's recovery or cause concern for PAC agency	"The patient has approximately 30-year history of alcohol abuse"
	Home Environment	15	17 (1.5%)	36.4%	Indicators of a safe environment for an older adult to recover at home (e.g. handicap accessible)	"Patient lives in a single-story home with a ramp to enter and an accessible bath on the 1st floor"
	Insurance Barriers	18	49 (4.2%)	88.9%	Evidence of insurance-related barriers including lack of coverage, non-covered services, or missing eligibility criteria for certain PAC	"Pt denied by insurance due to pt improving and no longer requiring skilled level of care"
Clinical Reasoning	Physical/ Occupational Therapy	20	28 (2.4%)	50%	Recommendations from PT or OT that the patient is ready/recommends to be discharged home	"He was evaluated by physical and occupational therapy and deemed stable for discharge to home once medically ready"
	Social Work/Case Management	68	140 (12.1%)	56.5%	Recommendation from SW/CM that patient can return home without PAC	"No home care needs identified by the Case Manager at this time"
	Physician Reasoning/Other	22	32 (2.8%)	11.1%	Recommendation from physician or general team for the patient to discharge home to self-care	"Per Medical Team, Patient can discharge tomorrow with no additional needs post discharge"
	Care Continuity	21	78 (6.7%)	50.8%	An indication that the patient received skilled PAC before hospitalization (not including unskilled aids)	"She has home care services through [Agency Name] Health"
	No Facility Bed Availability	7	27 (2.3%)	50%	Facility has no beds available at the time of discharge or cannot accommodate a patient due to level of acuity.	"the facility does not have any bed availability; they do not anticipate any availability until possibly next week"

Legend: B2PAC are divided into 4 main categories (received PAC referral, negative preferences, psychosocial factors, and clinical reasoning) | + = Encounter level (N=100, 99 encounters had annotations), patients could experience more than one class | * = Sentence level. Sentences could contain multiple sub-classes | ~ = Average IAA across batches 3-10

The correlation heatmap is demonstrated in **Figure 6**. Class pairs with the highest correlation were *negative preferences* and *psychosocial caregiver* (0.34), followed by *psychosocial insurance barriers* and *clinical reasoning facility bed availability*

(0.28). The class pairs with the lowest correlation were *clinical reasoning social work or case management* and *clinical reasoning care continuity* (-0.29), followed by *clinical reasoning physician/other* and *psychosocial substance abuse* (-0.22).

Figure 6: Pearson Correlation Heatmap of B2PAC Classes at the Patient Encounter Level



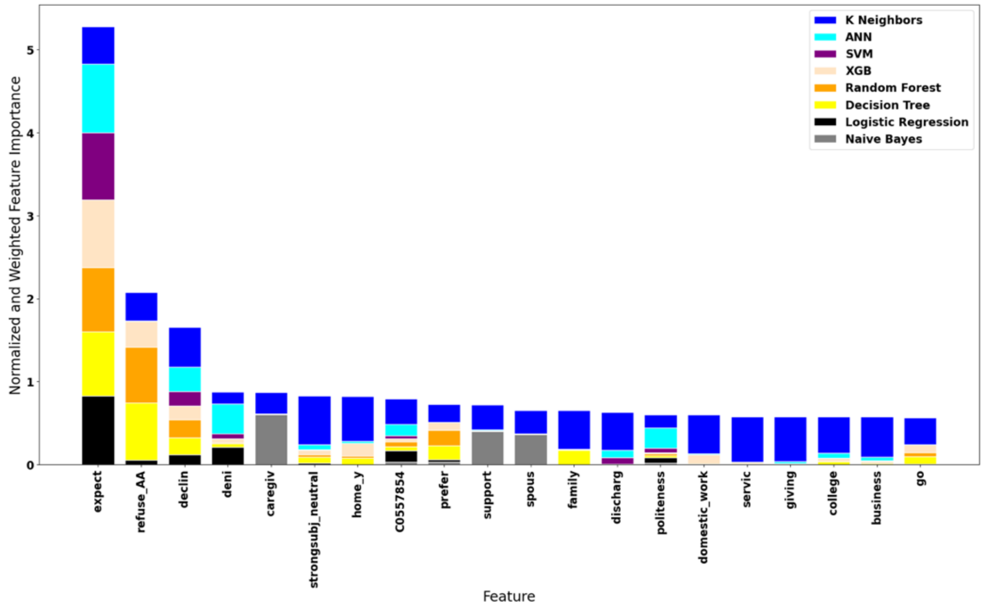
Legend: *Correlation is measured from -1 to 1. Higher scores and darker blue colors represent stronger correlations, while lower scores and darker red numbers represent weaker correlations.CR = clinical reasoning, P = psychosocial.*

Supervised Machine Learning Results

The *negative patient preference* class was selected for automation from a clinical standpoint because it is potentially modifiable, and from a data-driven standpoint for its high frequency of training instances (N=211 sentences), high IAA (88.7%), and high prevalence at the patient level (72%). The full dataset contained 211 (18.3%) negative preferences sentences; The training set contained 146 (18%) and the testing set contained 65 (18.7%). A total of 1297 features were encoded.

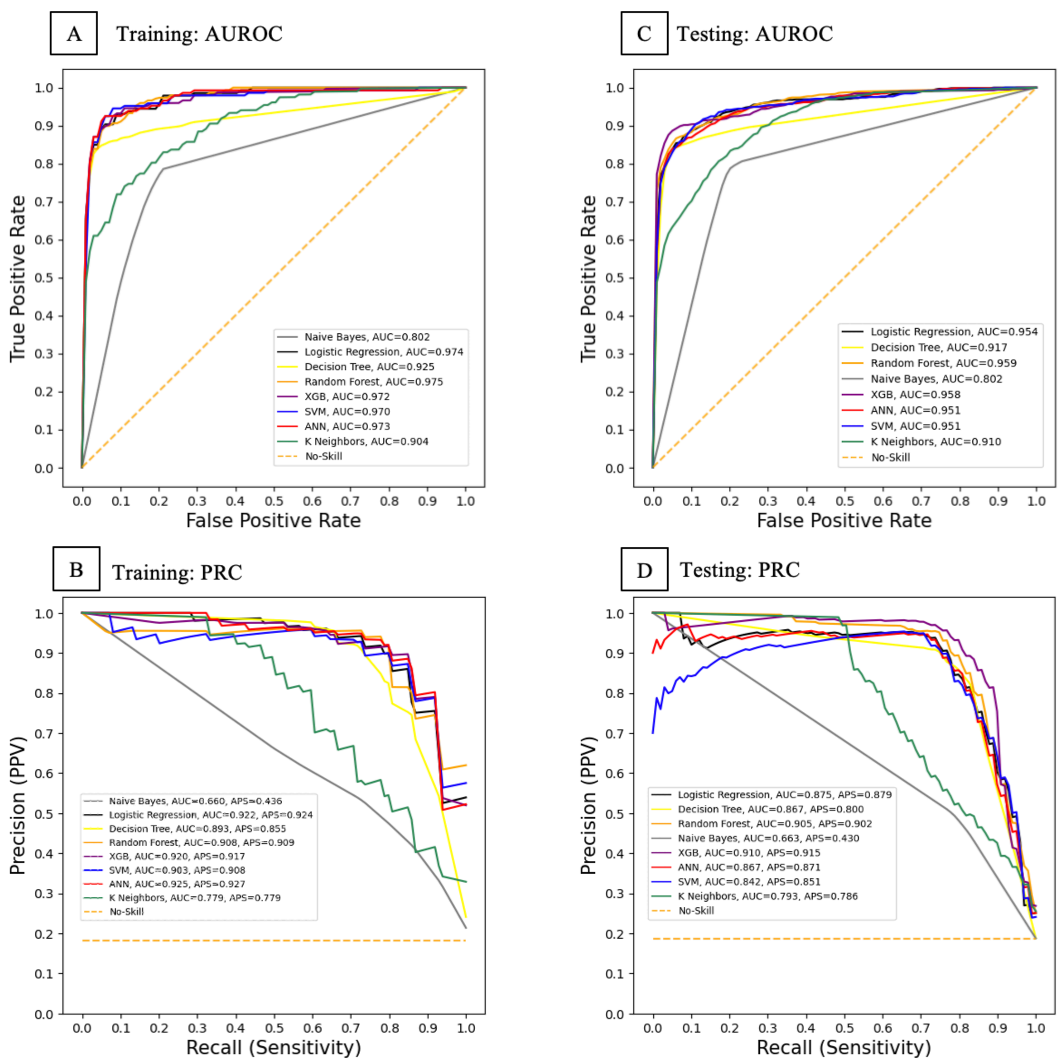
Figure 7 shows normalized weighted feature importance of the top 20 features across all 8 models in the training set. The top 5 important features are “expect,” "refuse," "declin," "deni," and "caregiv." Most of the top 20 features were n-grams, but UMLS, Empath, abbreviation, and subjectivity lexicon features sets were represented.

Figure 7: Normalized and Weighted Importance of Top 20 Features Across 8 Supervised Learning Classifiers



Model performance is illustrated in **Figure 8**. Although all models demonstrated strong AUROC performance ≥ 0.80 in both training and testing sets, AUPRC and APS performance was weaker for some models including Naïve Bayes and K-Nearest Neighbors. Across all performance metrics, the best performing models were Random Forest, XGBoost, and Artificial Neural Network. In terms of AUROC, random forest had the best performance in both the training (0.975) and testing sets (0.959). This difference was statistically significant for Naïve Bayes, Decision Trees, and K-Nearest Neighbors in the training set. In the testing set, it was statistically significant for all classifiers except XGBoost and Logistic Regression. For AUPRC, the Artificial Neural Network model had the highest performance in the training set (0.925). This difference was statistically significant for Naïve Bayes and K-Nearest Neighbors. The XGBoost model had the highest performance in the testing set (0.91). This difference was statistically significant for all classifiers except Random Forest. For APS, the Artificial Neural Network model had the best performance in the training set (0.927). This difference was statistically significant for Naïve Bayes and K-Nearest Neighbors. XGBoost had the best performance in the test set (0.915). This difference was statistically significant for all classifiers except Random Forest.

Figure 8: Supervised Machine Learning Model Performance in Training and Testing Datasets



Legend: AUROC = Area under the receiver operating curve, PRC = precision-recall curve

Deep Learning Results

AutoGluon's TextPredictor training was completed in 45.45 minutes. For the training set, AUROC was 0.999, precision was 0.953, average precision was 0.994, recall was 0.966, and F1-score was 0.959. For the testing set, AUROC was 0.991, precision was 0.909, average precision was 0.969, recall was 0.923, and F1-score was 0.916. Comparison of precision, recall and F1-score to supervised machine learning models is demonstrated in **Table 11**. AutoGluon out-performed all supervised machine learning models.

Table 11: Comparison of Supervised Algorithm Performance to AutoGluon

	Training			Testing		
	Precision	Recall	F1-Score	Precision	Recall	F1-Score
Naive Bayes	0.508	0.771	0.607	0.492	0.795	0.608
K-Nearest Neighbors	0.463	0.801	0.579	0.486	0.851	0.612
Logistic Regression	0.768	0.890	0.821	0.738	0.863	0.795
Artificial Neural Network	0.833	0.871	0.847	0.758	0.840	0.796
Support Vector Machines	0.865	0.843	0.850	0.812	0.815	0.813
Decision Tree	0.892	0.801	0.835	0.844	0.802	0.822
Random Forest	0.931	0.788	0.849	0.906	0.786	0.841
XGBoost	0.874	0.842	0.854	0.872	0.848	0.859
AutoGluon TextPredictor	0.953	0.966	0.959	0.909	0.923	0.916

Discussion

Our study revealed that most patients who DIRECT identified as needing PAC but were discharged home without it experienced multiple B2PAC, illustrating the complex interaction between social determinants of health and access to care, clinical judgment, and patient preferences in discharge planning. The social work and case management notes contained the majority of B2PAC annotations, while this information was often left out of the discharge summary. It is important to consider note types from all disciplines involved in discharge planning when developing future NLP systems for this purpose since it is a team process. Some barriers identified were more modifiable than others. For example, evidence-based interventions have been developed to address *negative patient preferences* through patient education[222] and *substance abuse* through intervention teams.[223] Other barriers such as insurance and no facility bed availability would need to be addressed through policy change to improve access to insurance, coverage of outpatient social services,[224] and access to home health or facility-level care, especially in rural or underserved areas.[225] The *Received PAC Referral* class was unexpected because within-hospital transfers to skilled nursing or inpatient rehabilitation units and errors in coding discharge disposition were not anticipated. However, this finding highlights common data quality issues in structured EHR data.[226] In the future, NLP could be a valuable tool to verify or correct structured information such as discharge disposition.

The Pearson correlation heatmap in **Figure 6** highlighted interesting patterns in B2PAC documentation and co-occurring classes. *Negative preferences* and *caregiver* had the highest correlation (0.34). In the qualitative literature it is common for patients who live with or near a caregiver to prefer that their loved-one helps them at home without realizing the need for skilled nursing to perform tasks like medication administration, wound care, etc. and further highlights the need for patient education.[203] *Insurance barriers* and *no facility bed availability* also had a high correlation (0.28), highlighting healthcare access issues at multiple levels. *Care continuity* and *social work or case management clinical reasoning* had the lowest correlation (-0.29). Because patients containing *care continuity* sentences were using skilled nursing care prior to the hospitalization, it is unlikely that the social worker or case manager would document reasons why the patient should return home without those services.

Case management and social work notes contained a combination of structured (e.g. “Patient Expects to be Discharged To: Home”) and unstructured (e.g. “I informed patient of the recommendations, And he again declined all services,” followed by “I can take care of myself”) sentences containing *negative patient preferences*. The structured field occurs at the beginning of the note and serves as a helpful initial indicator of *negative preference*, but the unstructured explanations added important context which could guide future interventions. These documentation patterns are reflected in the feature importance visualization in **Figure 7**, where the “expect” n-gram is the most important feature across most machine learning models. N-grams can be a useful tool for information retrieval because they detect common documentation patterns in clinical notes, which may explain why they encompassed the majority of the top 20 features.

UMLS features were incorporated to map concepts to a standardized vocabulary, a growing trend in NLP research.[227] “C0557854” is a UMLS concept unique identifier that maps to “services,” and was positive in sentences such as “He was previously seen in [department name], and declined home care.” Because *negative patient preferences* involve opinions, subjectivity, and emotion, other feature types like Empath and the subjectivity lexicon were incorporated to extract these higher-level meanings from words. “Domestic_work” is an Empath feature that appeared in the top 20 features, and an example was “Mr. [**NAME**] lives at home with his caretaker [**NAME**] and her daughters, who also help take care of him.” These lexicons were trained on non-biomedical data, so one of the shortcomings of using them for clinical NLP is that context can change the meaning of words. For example, “spoke to patient’s wife who denied the need for home services,” contained “home” which maps to “strongsub_neutral” (strong magnitude of neutral subjectivity) from the subjectivity lexicon. In a discharge planning context, declining home health services would be considered negative. Future research is needed to develop lexicons detecting negative opinions and emotions from clinical notes.

In **Table 11**, Random Forest, XGBoost, and Artificial Network models consistently had the highest performance across AUROC, AUPRC, and APS in both the training and testing sets. However, the deep learning model from AutoGluon's TextPredictor outperformed all supervised models from a classification standpoint. The model trained in less than an hour and relied on only sentences and labels without feature engineering, which highlights the growing trend of user friendly AutoML tools to make machine learning more accessible.[228] Despite the strong performance and efficient development time compared to supervised approaches, it is important to consider the tradeoffs of using deep learning for clinical problems. Deep learning approaches lack the explainability of feature importance, which can be crucial for implementation in clinical settings if the goal is to incorporate the negative patient preferences NLP CDSS at the point of care.[229] Prior research shows that clinicians are more likely to trust CDSS where they can understand what is driving the prediction model.[230] Knowing which features are contributing to an alert can help the clinical team select appropriate interventions. Prior to implementation of any model in a clinical setting, an external validation study is warranted to determine if the performance holds in other health systems with potentially different documentation patterns and determine a meaningful cutoff score for an alert by balancing sensitivity, specificity, and the resources available at that hospital to address negative preferences.

Limitations and Future Work

Study limitations include small sample size of the annotation study, limited note types, only developing an NLP classifier for one sub-class, and the 13% mis-coding of discharge disposition. Some features between feature sets may have overlapped, leading to confounding. The supervised machine learning models cannot be perfectly compared to the deep learning model because there were notable differences in training methodology (10-fold cross validation in the supervised approaches vs. 9 epochs in the deep learning model) and outcomes because we did not measure AUPRC or APC for the deep learning model. Future directions include the creation of a larger reference standard to train classifiers for the other 10 sub-classes, and integrating the NLP system into CDSS systems like DIRECT to improve performance and prompt interventions such as patient education.

Conclusion

The annotation study revealed a wide variety of B2PAC and the majority of patients experienced more than one barrier, which demonstrates the complex interaction of social determinants of health, patient preferences, and clinical judgement in discharge planning. The top 20 features across all models included all feature types, highlighting the importance of incorporating a holistic set of text, sentiment, and standardized vocabulary features for clinical NLP studies. The deep learning model outperformed all supervised machine learning algorithms without any feature engineering but lacks feature importance. This highlights the tradeoffs between time, computational burden, and model explainability when developing and implementing prediction models at the point of care.

CHAPTER 5: CONCLUSION

This three-paper dissertation produced new knowledge on the development, simulation, and further advancement of clinical decision support for discharge referral decision making. This chapter will present a brief conclusion and implications for future research, practice, and policy for each paper.

Chapter 2 – Paper 1

The objective of this paper was to conduct a systematic review of studies reporting development and validation of models predicting post-acute care (PAC) destinations, summarize areas of model development and variables in the final models, and assess risk of bias and applicability using the Prediction Model Risk of Bias Assessment Tool (PROBAST). Our goal was to evaluate the state of the science of models that could be used for discharge planning clinical decision support systems (CDSS) and understand how the Discharge Expert Referral System for Care Transitions (DIRECT) CDSS compares. Several systematic reviews of models predicting 30-day readmissions have been published, but to our knowledge, there were no comprehensive systematic reviews of models predicting post-acute care destination, which is a distinct outcome with important implications for discharge planning and can help *prevent* readmission. Findings from this paper informed some of the design and methods of Papers 2 and 3. In the broader transitions in care literature, this information could be useful for future model development, updating, and implementation in clinical practice.

Although 35 models were identified in the literature and met inclusion criteria for the study, it proved difficult to objectively compare model development methods and performance across the studies due to differing definitions of variables and outcomes, as well as lack of transparent reporting performance. Most of the models were developed for specific clinical populations, especially orthopedic and cardiac surgery, and defined a binary outcome tailored to that population. For example, an orthopedic surgery model may be developed to predict inpatient rehabilitation facility vs. home discharge, while a general medicine model may be developed to predict home vs. non-home discharge. Predictor variables like comorbidities were also defined in unique ways such as number of diagnoses, a published comorbidity index score, or did not include definitions. Reporting was inconsistent as a whole, with very few studies following the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis Or Diagnosis (TRIPOD) guidelines.[48] As a result, our statistical evaluation was mostly focused on discrimination (area under the receiver operating curve), which can be biased in unbalanced samples.[77,78]

Despite these challenges, we were able to make generalized comparisons across models and identify interesting patterns. Most of the models were developed for specific surgical populations and used retrospective EHR, data warehouse, or registry datasets. The most common types of variables included in models were demographic (i.e. age, sex, race), comorbidities/diagnoses (i.e. comorbidity score, diabetes, hypertension), and hospitalization (i.e. admission source, weight, length of stay). 29% of models included a mental health diagnosis, and this may increase in future models as there is a growing trend to incorporate mental and physical health.[131,132] Less than half of models incorporated measures of functional status, despite recent research suggesting that functional status may be a better predictor of negative health outcomes than comorbidities.[133,134] Only 3 models were developed using machine learning (Naïve Bayes, Artificial Neural Network), but these methods will likely increase as machine learning becomes more accessible through tools like Auto-Machine Learning[228] and new methods emerge.

Model performance varied within and across populations, but models developed for medical patients had a higher proportion of good discrimination (AUC ≥ 0.80) than those developed for surgical patients. Model discrimination has not improved over time despite new statistical methods and increased access to healthcare data. The only substantial trend identified for performance was that models designed to run later in the hospitalization (when the patient has more data in the EHR) had higher performance than models at admission or early in the hospitalization. Over half of the studies presented clinical tools developed from their models to be used at the bedside, and many of the studies published within the last 5 years included open access versions online. However, only 5 teams have published about their models beyond the original studies to report quasi-experimental testing, external validation, or update the model with new variables.

The Discharge Referral Expert System for Care Transitions (DIRECT) CDSS was unique from other models identified in the systematic review in 4 key ways. Although several models incorporated clinician feedback for variable selection, it was the only model that also cited a specific theory (Orem's Self-Care Deficit Theory) to guide their process. It was the only two-step model to first identify a patient's need for *any* PAC, then classify the need as facility-level care or home care. Traditionally, prediction models are developed using retrospective data from EHRs or registries, meaning that the decisions of less experienced or biased clinicians are weighed equal to experts in the field. DIRECT was one of the only expert-developed systems, meaning the data used to build the model came from Delphi rounds of case studies with interdisciplinary discharge planning experts determining a patient's discharge disposition. Finally, it was one of the only models tested in a real-world clinical setting in a quasi-experimental study and had a statistically significant impact on the reduction of 30-day readmissions.

Implications for future research include emphasizing the importance of following TRIPOD guidelines and a growing need for replication and/or implementation studies. According to the Prediction Model Risk of Bias Assessment Tool (PROBAST), nearly all models demonstrated high risk of bias.[43] The Cochrane Prognosis Methods Group has published several guides for conducting rigorous modeling studies including the TRIPOD statement. Key recommendations include being mindful of data sources and variables used for model development to reduce bias, incorporating clinical stakeholders in the development process, and transparent reporting of missing data and model performance, especially for calibration. Existing models incorporate structured EHR data, so future studies should consider incorporating natural language processing where appropriate.

From a clinical standpoint, health systems should carefully evaluate implementation considerations when deciding to develop a new model or implement an existing one. A recent study from our team identified five components to consider, including alert timing (e.g. at admission vs. throughout hospitalization), user (e.g. physician, nurse), CDSS design (e.g. EHR alert, email), and outcomes studied (e.g. adoption, 30-day readmission rate).[231] If implementing an existing model, health systems should evaluate the methods and compare context in which the model was originally developed. Clinical stakeholders are needed to evaluate workflow considerations and ensure that new technology is not adding to documentation burden.[232]

The policy implications of this paper involve data standards and interoperability. A recent study found that only 0.3% of CDSS studies are replicated,[67] which may be due to problems with interoperability, data governance policies, the proprietary nature of EHRs, and the higher value placed on generation of new knowledge over confirmation of previous findings. The rise of ontologies like SNOMED-CT and UMLS have been instrumental in standardizing definitions and documentation of common data elements,[189] but some areas such as functional status are still under-developed or lack consensus. More work is needed to develop new documentation standards as well as maintain and streamline existing ones. Initiatives from the Office of the National Coordinator for Health IT like the Interoperability Roadmap[233] and the 10-Year Vision to Achieve an Interoperable Health IT Infrastructure[234] will be crucial in the effort to close the gap in interoperability issues.

Chapter 3 – Paper 2

The goal of Paper 2 was to apply DIRECT CDSS for discharge planning in a new setting and determine differences in patient characteristics and 30-day readmission rates based on DIRECT's recommendation among older adults in a large urban academic health system who were discharged home without post-acute care (PAC). This provided an opportunity to examine patient outcomes when those identified by CDS as needing PAC do not get PAC. We hypothesized that (1) among patients discharged home without services, patient characteristics of those identified by the algorithm as needing PAC would be older, with more limitations in activities of daily living, more comorbidities, and more hospitalizations in the 6 months prior compared to patients not flagged for PAC. Additionally, (2) those flagged by the algorithm as needing PAC would also experience higher rates of 30-day readmissions compared to those not flagged for PAC.

82% of patients were flagged by DIRECT as needing PAC, and the overall readmission rate was 15.2%. Hypothesis 1 was partially supported. Patients flagged by DIRECT as needing PAC had a higher median age, with more limitations across all measures of activities of daily living (use of assistive device, ambulation, transfer, bathing, eating), and had a higher proportion of hospitalizations in the 6 months prior to admission compared to patients not flagged. All of these differences were statistically significant. Although we expected flagged patients to have more comorbidities, both groups had a median of 2. However, the range for comorbidities was larger for flagged patients (0-14) than not flagged patients (0-11). Notable findings included that among females, 84.8% were flagged (compared with 80% of males); among Black patients, 87.8% were flagged (compared with 79.1% of white patients); and 89.1% of Medicaid or Managed Medicaid and 87.6% of Medicare or Managed Medicare beneficiaries were flagged (compared with 69.7% of private/commercial insurance or self-pay). The gender finding aligns with prior research showing that women tend to have longer life expectancies and less in-home caregivers, indicating a greater need for PAC.[165,166] Similar racial disparities have been identified about PAC referral practice in other studies of orthopedic patients[167] and brain tumor surgery.[168] Use of standardized discharge planning tools like DIRECT that identify patients for care based on clinical need can aid in the reduction of biased decision making. The insurance findings may be explained by age for Medicare beneficiaries and socioeconomic status among Medicaid beneficiaries. Privately insured patients tend to be younger and of higher socioeconomic status. However, more research is needed to understand specific policies that might be impacting access for these groups.

Hypothesis 2 was not supported. The readmission rate among flagged patients was actually 0.5% lower than patients not flagged (15.1% flagged vs. 15.6% not flagged), and the adjusted odds of 30-day readmission in the final logistic regression model were 1.3% lower among flagged patients (p=0.922). In the prior study, patients flagged by DIRECT had a 67.8% higher risk of readmission than patients not flagged (p=0.006).[42] Changes in referral patterns over time and hospital differences from the prior study may partially explain the surprising results. Nationally, PAC referrals have increased over the last two decades.[172] Data from this study was collected in 2019, three years after completion of the original study in 2016. Although we expected the most common discharge disposition to be home to self-care among older adults before starting the study, we found that it was actually 11.8% lower than home health care (27.4% vs. 39.2%), so this health system was already accomplishing a significant volume of PAC referrals. Additional differences in patient characteristics included that this study had fewer females, was more racially diverse, and was younger. Greater proportions of the sample in this study were employed and married.[42]

Seven initial variables were considered for logistic regression models (DIRECT flag, sex, age, insurance, hospital, length of stay, and surgical encounter type) to predict 30-day readmission, and the final model contained all variables except for sex. Statistically significant variables were Medicaid or Managed Medicaid insurance (OR 0.652, p=0.046), Hospital Site 1 tertiary medical center (OR 1.571, p<0.001), and surgical encounter type (OR 2.015, p<0.001). The Medicaid finding may be explained by several transitions in care interventions that have been implemented in the health system to reduce negative outcomes in this population.

Although some surgical patients were included in the original study, they were not the focus. Therefore, one of the secondary goals of this study was to evaluate its potential impact on this population. Although DIRECT was not a statistically significant predictor for 30-day readmission in the full sample, the subgroup analysis revealed its potential value among surgical patients in this health system. We found that surgical patients were experiencing 8.6% higher rates of 30-day readmission compared with non-surgical patients. In the logistic regression model of surgical patients only, patients flagged by DIRECT had adjusted 51.8% higher odds of 30-day readmission compared with patients not flagged, and this difference was statistically significant (p=0.041). Surgical patients experienced greater declines in activities of daily living (ambulation, transfer, bathing, eating) over the course of the hospitalization, fewer had caregivers, and these patients had a higher percentage of opioid prescriptions at discharge compared with non-surgical patients.

Throughout the study, we faced many real-world challenges common in EHR and CDSS replication research, especially around data quality, missingness, and variable definitions. The prior study was conducted in a smaller health system with a different EHR, and all DIRECT algorithm variables were required daily nursing documentation elements. We worked with nursing and social work stakeholders across both hospitals for months to attempt to map this health system's EHR elements to the DIRECT variables, and many were not a perfect match. For example, we identified 3 separate nursing flowsheets documenting activities of daily living on 3 different ordinal scales and used consensus to map them to DIRECT's variables, which were measured on a different ordinal scale. Several of the EHR elements that we identified are not required nursing documentation elements, so some variables like "eating" had 70.2% missing values. We were unable to run Step 2 of the algorithm because one of the required variables (caregiver availability) does not exist in this health system's EHR. The DIRECT algorithm was trained to be highly sensitive to missing data, and partially explains why over 80% of patients were flagged as needing PAC in this study. Therefore, patients with more missing algorithm variables have higher odds of being flagged for PAC referral so that the clinician can be alerted and make a judgment call based on additional information that might not be present in the structured EHR data.

More research is needed to better understand the outcome differences in surgical patients and identify whether any important differences exist within subgroups of surgical patients. Potential areas of exploration might include severity or invasiveness of the procedure, gaps in insurance coverage of services or within specific patient populations, patient refusal of PAC, and whether PAC has been determined prior to hospitalization for planned surgical procedures. It is common practice for surgeons to recommend a PAC destination before surgery such as outpatient rehabilitation,[11,176] which may not reflect a patient's clinical needs after the procedure. Once a target population has been identified, a quasi-experimental study of DIRECT is warranted to evaluate its potential impact on patient outcomes in real-world inpatient surgical settings. This study should additionally measure implementation outcomes,[235,236] since this is one of the key gaps identified in Paper 1. If implemented, the model might require re-calibration or updating to better reflect the patient population or resource availability.[237]

Clinical implications include emphasizing the importance of required documentation of functional status and considering where to implement CDSS in existing workflows without increasing clinician documentation burden or alert fatigue. We agree with the recent recommendation[188] to advocate for standardized, required nursing documentation to reduce incompleteness, inaccuracy, and variation in data collection. In addition to our team's suggested considerations for implementation of transitions in care CDSS,[231] several general guides for CDSS implementation have been published and focus on settings, policy considerations, and strategies.[238-240] Several systematic reviews and strategy papers have been published with recommendations and frameworks to reduce documentation burden[241] and alert fatigue for different types of CDSS.[242-245] Many include recommendations for redesigning the EHR user interface. We also advocate for implementing CDSS early in the hospitalization to promote early discharge planning, since it is associated with lower rates of 30-day readmission and leaves more time for patient education and securing resources after discharge.[246]

From a policy standpoint, CDSS has been an important component of Meaningful Use legislation since it was passed in 2009.[247] Unfortunately in real world clinical settings, it can be challenging to implement CDSS in a timely and affordable manner.[183] Many of the barriers that led to the initial policy development still exist today. More legislation is needed to support increased interoperability and reduced cost of CDSS. Future CDSS agendas should emphasize the importance of incorporating nursing data, and develop better standards for documentation of activities of daily living and caregiver status in the EHR. Ideally, these standardized assessments could be mapped to ontologies like the Systematized Nomenclature of Medicine (SNOMED) and Logical Observation Identifiers Names and Codes (LOINC) to improve access and interoperability.[189]

Chapter 4 – Paper 3

The objective of Paper 3 was to identify common barriers to post-acute care (B2PAC), then develop and evaluate an NLP system to encode sentences containing negative preferences among hospitalized older adults. Our goal was to facilitate a deeper understanding of the reasons why so many patients flagged by DIRECT as needing PAC do not receive it, using an untapped data source: discharge planning notes and discharge summaries. Was there important information not captured by DIRECT but somewhere else in the EHR informing the clinicians' recommendations, or were these patients recommended PAC by the discharge planning team but experienced some other type of barrier? We hoped that the findings could identify novel variables to include in CDSS and/or areas for future discharge planning intervention development. We hypothesized that patient refusal and insurance would be the most frequent reasons for not achieving recommended PAC.

We used a combination of literature- and data-driven methods to complete the annotation study. We started by examining the literature for other NLP systems or quantitative studies of B2PAC, but only found a few recent studies related to social determinants of health. Our original annotation schema was mainly derived from qualitative studies of patients and clinicians about their experiences with discharge planning and care transitions, where we identified 3 main classes: patient/family preferences, social determinants of health, and clinical reasoning. We used the first 3 batches of the annotation study to adapt the schema based on the real-world data, including modifying definitions, identifying novel classes, and removing classes that we weren't able to find in our dataset. The task of each annotation was to ask, "why did this patient not receive PAC?" One of the most challenging parts of the annotation study for all annotators and one of the most common themes in the notes from our meetings was that we *wanted* to ask, "why did this patient *actually need* PAC?" These discussions were carefully tracked, and we found that 44% of patient encounters in the annotation study were identified by clinicians as needing PAC or expressed interest in it at some point during the hospitalization. Another surprising finding was that the discharge summary rarely contained information about the reason for a particular discharge disposition. Most of the annotations came from the social work or case management notes. This process took several months, and 15 iterations of annotation schemas.

Over the course of the annotation study, our inter annotator agreement (IAA) F1-score improved by an average of 10.6% from early batches to the final batch of annotations. Due to the small sample size of the annotation study (594 notes from 100 encounters), some classes had to be collapsed or removed because there were too few instances (<10 sentences) to draw conclusions or automate using NLP. Our final annotation schema contained 11 classes grouped into 4 categories: negative preferences, clinical reasoning, psychosocial factors, and received PAC referral. The *received PAC referral* class was our only data-driven class, identified because we discovered that some patients were actually transferred to PAC units within the hospital, or to an outside facility at inpatient discharge and their discharge disposition was inaccurate. A recent study found that approximately 9% of discharge dispositions in Medicare claims for hip and knee replacements are inaccurate,[248] which is close to our finding of 13%.

Patients experienced a mean of 3.68 B2PAC. Our hypothesis was partially supported, as *negative preferences* was the second most prevalent class (72% of encounters). The most prevalent class was patient has a *caregiver* (95%). *Insurance* barriers were the 6th most prevalent class (18%). Patients experienced barriers across different classes, and some patients experienced as many as 7, which highlights the complex interaction between clinical judgment, socioeconomic factors, healthcare access, and patient decision making in discharge planning.

Although our eventual goal is to develop NLP systems for all 11 classes, it was not possible in this study due to very small sample sizes of most classes. Supervised and unsupervised machine learning algorithms achieve better performance with larger sample sizes for training and testing, and our threshold for developing an NLP system was 200 sentences. Unfortunately only 2 classes (*caregiver* and negative *preferences*) met that threshold. We chose to develop an NLP system for the highest-value class from both a clinical and data-driven perspective. While both classes met the data-driven criteria at both the sentence and patient level, we believed that *negative preferences* held higher clinical value because it is modifiable through patient education. For example, one study found that up to 28% of patients refuse PAC.[18] Having a *caregiver* at home to support the recovery process is a positive finding that we did not feel required intervention, other than the need to better assess availability as seen in Paper 2.

Since we were automating an NLP system for *negative preferences*, we tailored our feature engineering process to fit the task by incorporating sentiment and subjectivity features in addition to traditional NLP features. One of the limitations of using sentiment and subjectivity lexicons for biomedical research is that they were trained on non-biomedical data, so some words can change meanings in a different context. For example, “patient” is a positive adjective in a non-medical setting, but sometimes a neutral noun in medicine. This may explain why only one subjectivity feature (strong magnitude, neutral subjectivity) was present in our top 20 features. However, the sentiment features performed better and appeared in 6 of the top 20 features. Future research is needed to tailor these lexicons to the biomedical domain. We also incorporated UMLS features to ensure our research was mapped to standardized terminology.

After feature engineering, we trained and tested 8 supervised machine learning algorithms. All of the final models achieved AUROC performance > 0.80 in both the training and testing sets. However, the deep learning model developed without any feature engineering achieved the highest performance across all measures on the testing set (AUROC 0.991, precision 0.909, recall 0.923, F1- score 0.916). Despite the strong performance and efficient development time compared to supervised approaches, it is important to consider the tradeoffs of using deep learning for clinical problems. Deep learning approaches lack the explainability of feature importance, which can be crucial for implementation in clinical settings.[229]

Future research directions include running the system on a larger sample of patients to estimate the prevalence of negative patient preferences in this health system and conducting a larger annotation study to generate enough data to develop future NLP systems for the remaining classes. Depending on the system performance and prevalence of B2PAC in a wider patient population, some or all of these features could be incorporated into future CDSS tools or aid in quality improvement research. These variables have been traditionally inaccessible in quantitative studies but could be incorporated into future models in health services research to explore their potential associations with patient outcomes like 30-day readmission. Patient education interventions about PAC can be developed and tested for patients with negative preferences as a component of early discharge planning.

The clinical implications occur at the patient and the health system level. Although clinicians are aware that these barriers exist, they have not traditionally been tracked at a system level. The ability to systematically evaluate barriers such as negative preferences can enable health systems to better optimize resource allocation and reduce negative outcomes. The average cost of a readmission for Medicare patients is $15,500,[180] which is higher than the average cost of an index hospitalization.[181] Preventing negative outcomes like 30-day readmission among even a small fraction of patients can improve the quality of care and reduce costs for hospitals. Potential strategies to address barriers might include hiring more case managers[249] or incorporating interventions to promote early, multidisciplinary discharge planning including care integration with community settings, patient education, and specialist follow-up.[198] Using this NLP system presents a systematic approach to identifying patient needs, but specific strategies can be tailored to the patient.

Health equity has become a major policy focus area,[250,251] and future work is needed to continue to emphasize the importance of developing standards for and documenting social determinants of health and functional status, as well as insurance and payment reform to improve healthcare access. Although at least 4 standardized vocabularies include social determinants of health codes, there is little consensus across vocabularies and gaps exist in screening, diagnosis, and intervention.[252] Future work is needed to streamline these efforts. *Insurance barriers* and *no facility bed availability* had the second highest correlation of all classes at the patient level (0.28), indicating different types of healthcare access issues. Policy change is needed at multiple levels to improve access to insurance, coverage of outpatient social services,[224] and access to home health or facility-level care, especially in rural or underserved areas.[225]

Summary

This dissertation began by examining the literature for models predicting PAC destinations after hospitalization to evaluate the state of the science of this research. Important study quality and reporting issues were identified, as well as gaps in replication research and use of natural language processing. These findings led to the research questions explored in the second and third papers.

The second paper aimed to simulate the application of DIRECT in a large urban academic health system and evaluate its potential impact on 30-day readmissions in a new setting. In the overall sample, patients flagged by DIRECT did not experience higher rates of 30-day readmission than those not flagged, most likely due to temporal differences, hospital characteristics, and missing data. However, surgical patients flagged by DIRECT experienced significantly higher 30-day readmissions than those not flagged, and this difference was not seen in non-surgical patients. These patients experienced greater declines in activities of daily living over the course of the hospitalization, fewer had caregivers, and they had a higher percentage of opioid prescriptions at discharge. Use of DIRECT in a surgical population could help clinicians systematically identify patients at risk of needing PAC.

The third paper aimed to identify common barriers to PAC among patients flagged by DIRECT who were discharged home and develop an NLP system to detect sentences containing negative preferences. Eleven barriers were identified, and most patients experienced several, reinforcing the complex nature of discharge planning. An NLP system was developed to detect sentences from discharge planning notes containing negative patient preferences and achieved high performance after internal validation. Future research is needed to understand the higher readmission rate among surgical patients, develop NLP systems for the other barriers, and integrate these systems into CDSS to drive future health system interventions and policy.

APPENDICES

Appendix A: PRISMA Reporting Checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	20
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	20
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	21
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	22
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	N/A
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	23
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	23
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	Appendix B
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	23
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	23
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	23

Section/topic	#	Checklist item	Reported on page #
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	24
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	24
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	24
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	N/A
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	N/A
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	25
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	26, Table 1
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome-level assessment (see Item 12).	40, Table 3
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group and (b) effect estimates and confidence intervals, ideally with a forest plot.	N/A
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	N/A
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	N/A
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	N/A
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., health care providers, users, and policy makers).	60

Section/topic	#	Checklist item	Reported on page #
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review level (e.g., incomplete retrieval of identified research, reporting bias).	65
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	66
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	

Appendix B: Final Search Strategies in Pubmed, CINAHL, and Embase from Inception to June 5, 2020

Pubmed Final Search

((((("Hospitalization"[Mesh] OR "Inpatients"[Mesh] OR inpatient OR hospitaliz*))
AND
((“algorithms”[Mesh] OR algorithm[tiab] OR “machine learning”[tiab] OR “Medical informatics”[Mesh] OR “clinical decision support”[tiab] OR “clinical decision support systems”[tiab] OR “clinical decision rules” OR “decision support techniques”[Mesh] OR “decision aid”[tiab] OR “decision tool”[tiab] OR “Models, statistical”[Mesh] OR “logistic model”[tiab] OR “multivariate”[tiab] OR “risk score”[tiab])))
AND
((“Risk”[Mesh] OR “clinical decision-making”[Mesh] OR referral[title] OR use[title] OR utilization[title] OR “predictive value of tests”[Mesh])))
AND
((“Patient Discharge”[Mesh] OR “discharge disposition”[tiab] OR “discharge location”[tiab] OR “subacute care”[Mesh] OR “post-acute care“[tiab] OR “home care services”[Mesh] OR “home nursing”[Mesh] OR “home care”[tiab] OR “Hospitals, rehabilitation”[Mesh] OR “Nursing homes”[Mesh])))
NOT
((child OR child* OR infant OR infan* OR newborn* OR neonat* OR toddler* OR adolescen* OR teen* OR pediatric* OR paediatric*))

<u>2785 Results</u>

CINAHL

Monday, June 08, 2020 12:19:24 PM

#	Query	Limiters/Expanders	Last Run Via	Results
S6	S1 AND S2 AND S3 AND S4 AND S5	Expanders - Apply related words; Also search within the full text of the articles; Apply equivalent subjects Search modes - Boolean/Phrase	Interface - EBSCOhost Research Databases Search Screen - Advanced Search Database - CINAHL	609
S5	(MH "Patient Discharge+") OR (MH "Discharge Planning+") OR (discharge disposition) OR (discharge location) OR (MH "Subacute Care") OR (post-acute care) OR (MH "Home Health Care+") OR (MH "Home Nursing, Professional") OR (MH "Nursing Homes+") OR (MH "Rehabilitation Centers+")	Limiters - English Language; Peer Reviewed; Age Groups: All Adult Expanders - Apply related words; Also search within the full text of the articles; Apply equivalent subjects Search modes - Boolean/Phrase	Interface - EBSCOhost Research Databases Search Screen - Advanced Search Database - CINAHL	39,603
S4	(risk) OR (MH "Decision Making, Clinical") OR (MH "Referral and Consultation+") OR (TI referral) OR (TI use) OR (TI utilization) OR (MH "Predictive Value of Tests") OR (TI predict*)	Limiters - English Language; Peer Reviewed; Age Groups: All Adult Expanders - Apply related words; Also search within the full text of the articles; Apply equivalent subjects Search modes - Boolean/Phrase	Interface - EBSCOhost Research Databases Search Screen - Advanced Search Database - CINAHL	496,767
S3	(MH "Decision Making, Computer Assisted+") OR (MH "Medical Informatics") OR (MH "Algorithms") OR (MH "Decision Trees") OR (“decision support tool”) (“decision aid”) OR (MH "Decision Support	Limiters - English Language; Peer Reviewed; Age Groups: All Adult Expanders - Apply related words; Also search within the full text of the articles; Apply equivalent subjects	Interface - EBSCOhost Research Databases Search Screen - Advanced Search Database - CINAHL	123,774

Monday, June 08, 2020 12:19:24 PM

#	Query	Limiters/Expanders	Last Run Via	Results
S6	S1 AND S2 AND S3 AND S4 AND S5	Expanders - Apply related words; Also search within the full text of the articles; Apply equivalent subjects Search modes - Boolean/Phrase	Interface - EBSCOhost Research Databases Search Screen - Advanced Search Database - CINAHL	609
S5	(MH "Patient Discharge+") OR (MH "Discharge Planning+") OR (discharge disposition) OR (discharge location) OR (MH "Subacute Care") OR (post-acute care) OR (MH "Home Health Care+") OR (MH "Home Nursing, Professional") OR (MH "Nursing Homes+") OR (MH "Rehabilitation Centers+")	Limiters - English Language; Peer Reviewed; Age Groups: All Adult Expanders - Apply related words; Also search within the full text of the articles; Apply equivalent subjects Search modes - Boolean/Phrase	Interface - EBSCOhost Research Databases Search Screen - Advanced Search Database - CINAHL	39,603
S4	(risk) OR (MH "Decision Making, Clinical") OR (MH "Referral and Consultation+") OR (TI referral) OR (TI use) OR (TI utilization) OR (MH "Predictive Value of Tests") OR (TI predict*)	Limiters - English Language; Peer Reviewed; Age Groups: All Adult Expanders - Apply related words; Also search within the full text of the articles; Apply equivalent subjects Search modes - Boolean/Phrase	Interface - EBSCOhost Research Databases Search Screen - Advanced Search Database - CINAHL	496,767
S3	(MH "Decision Making, Computer Assisted+") OR (MH "Medical Informatics") OR (MH "Algorithms") OR (MH "Decision Trees") OR (“decision support tool”) (“decision aid”) OR (MH "Decision Support	Limiters - English Language; Peer Reviewed; Age Groups: All Adult Expanders - Apply related words; Also search within the full text of the articles; Apply equivalent subjects	Interface - EBSCOhost Research Databases Search Screen - Advanced Search Database - CINAHL	123,774

609 Results

Embase

....

No.	Query Results	Results	Date
#6.	#1 AND #2 AND #3 AND #4 AND [english]/lim AND ([young adult]/lim OR [adult]/lim OR [middle aged]/lim OR [aged]/lim OR [very elderly]/lim)	1,993	5 Jun 2020
#5.	#1 AND #2 AND #3 AND #4	2,938	5 Jun 2020
#4.	'hospital discharge'/exp OR 'hospital discharge' OR disposition.tw OR 'discharge location' OR 'subacute care'/exp OR 'subacute care' OR 'home care'/exp OR 'home care' OR 'rehabilitation center'/exp OR 'rehabilitation center' OR 'nursing home'/exp OR 'nursing home'	299,055	5 Jun 2020
#3.	'risk'/exp OR 'risk' OR 'clinical decision making'/exp OR 'clinical decision making' OR 'patient referral'/exp OR 'patient referral' OR use.ti OR 'utilization'/exp OR 'utilization' OR 'predictive value'/exp OR 'predictive value'	4,449,089	5 Jun 2020
#2.	'medical informatics'/exp OR 'medical informatics' OR 'algorithm'/exp OR 'algorithm' OR 'machine learning'/exp OR 'machine learning' OR 'decision support system'/exp OR 'decision support system' OR 'decision aid'/exp OR	1,287,639	5 Jun 2020

	'decision aid' OR 'statistical model'/exp OR 'statistical model' OR 'logistic model'/exp OR 'logistic model' OR 'multivariate analysis'/exp OR 'multivariate analysis' OR 'risk score'/exp OR 'risk score'		
#1.	'hospitalization'/exp OR 'hospitalization' OR 'hospital patient'/exp OR 'hospital patient'	596,095	5 Jun 2020

....

1993 Results

Appendix C: Comparison of Patient Characteristics Between Surgical and Non-Surgical Patients

Characteristic	Overall (N=3385)		Non-Surgical (N=1896)		Surgical (N=1489)		p-value
Clinical Variables	Median [Range] or N (%)						
30-Day Readmission	515	(15.2%)	217	(11.4%)	298	(20%)	**<0.001****
Age	67	[55-101]	66	[55-101]	68	[55-94]	**0.029**
Length of Stay	4	[2-101]	4	[2-101]	4	[2-70]	0.133
Sex							**<0.001****
Male	1962	(58%)	1043	(55%)	919	(61.7%)	
Female	1423	(42%)	853	(45%)	570	(38.3%)	
DIRECT Algorithm							**<0.001****
Flagged	2776	(82%)	1507	(79.5%)	1269	(85.2%)	
Not Flagged	609	(18%)	389	(20.5%)	220	(14.8%)	
Marital Status							**<0.001****
Married/partnered	1877	(55.5%)	978	(51.6%)	899	(60.4%)	
Divorced/separated	370	(10.9%)	216	(11.4%)	154	(10.3%)	
Widowed/single	1103	(32.6%)	689	(36.3%)	414	(27.8%)	
Other	35	(1%)	13	(0.7%)	22	(1.5%)	
Race							**<0.001****
White	2052	(60.6%)	1018	(53.7%)	1034	(69.4%)	
Black	1087	(32.1%)	742	(39.1%)	345	(23.2%)	
Other	167	(5.1%)	101	(5.3%)	66	(4.4%)	
Missing*	79	(2.3%)	35	(1.8%)	44	(3%)	
Ethnicity							0.594
Hispanic/Latino	58	(1.7%)	35	(1.8%)	23	(1.5%)	
Not Hispanic/Latino	3263	(96.4%)	1834	(96.7%)	1429	(96%)	
Missing*	64	(1.9%)	27	(1.4%)	37	(2.5%)	
Employment status							**0.016***
Employed part-time, full-time, or per diem	723	(21.4%)	381	(20.1%)	342	(23%)	
Retired/Disabled/ Unemployed	2416	(71.4%)	1397	(73.7%)	1019	(68.4%)	
Missing*	246	(7.3%)	118	(6.2%)	128	(8.6%)	
Insurance							**<0.001****
Medicaid/Managed Medicaid	302	(8.9%)	206	(10.9%)	96	(6.4%)	
Medicare/Managed Medicare	1998	(59%)	1106	(58.3%)	892	(59.9%)	
Private/Commercial or Managed Care/ Self-Pay	1083	(32%)	583	(30.7%)	500	(33.6%)	
Missing*	2	(0.1%)	1	(0.1%)	1	(0.1%)	
Hospital							**<0.001****
Tertiary Hospital (Site 1)	2055	(60.7%)	1231	(64.9%)	824	(55.3%)	

Community Hospital (Site 2)	1330	(39.3%)	665	(35.1%)	665	(44.7%)	
Algorithm Variables							
Employment							**0.047**
Employed	723	(21.4%)	381	(20.1%)	342	(23%)	
Not currently employed	2662	(78.6%)	1515	(79.9%)	1147	(77%)	
Hospitalization in the 6 months prior to admission							**<0.001****
No hospitalization	2441	(72.1%)	1544	(81.4%)	897	(60.2%)	
Hospitalization	944	(27.9%)	352	(18.6%)	592	(39.8%)	
Morse Fall Risk Score (0-125)							0.403
Fall risk ≤20	224	(6.6%)	132	(7%)	92	(6.2%)	
Fall risk >20	3161	(93.4%)	1764	(93%)	1397	(93.8%)	
Use of Equipment/Assistive Devices at Home							**0.008***
No equipment used	2990	(88.3%)	1650	(87%)	1340	(90%)	
Equipment used	395	(11.7%)	246	(13%)	149	(10%)	
Home Accessibility Concerns							0.189
No concerns	179	(5.3%)	109	(5.7%)	70	(4.7%)	
Concerns	3206	(94.7%)	1787	(94.3%)	1419	(95.3%)	
Presence of Wound							0.812
No Wounds	3216	(95%)	1803	(95.1%)	1413	(94.9%)	
Wound present	169	(5%)	93	(4.9%)	76	(5.1%)	
Ambulation							0.326
Improved	629	(18.6%)	338	(17.8%)	291	(19.5%)	
No change	2361	(69.7%)	1342	(70.8%)	1019	(68.4%)	
Declined	391	(11.6%)	214	(11.3%)	177	(11.9%)	
Missing*	4	(0.1%	2	(0.1%)	2	(0.1%)	
Transfer							**<0.001****
Improved	384	(11.3%)	214	(11.3%)	170	(11.4%)	
No change	1400	(41.4%)	884	(46.6%)	516	(34.7%)	
Declined	179	(5.3%)	90	(4.7%)	89	(6%)	
Missing*	1422	(42%)	708	(37.3%)	714	(48%)	
Bathing							**<0.001****
Improved	177	(5.2%)	82	(4.3%)	95	(6.4%)	
No change	1359	(40.1%)	564	(29.7%)	795	(53.4%)	
Declined	347	(10.3%)	71	(3.7%)	276	(18.5%)	
Missing*	1502	(44.4%)	1179	(62.2%)	323	(21.7%)	
Eating							0.058
Improved	94	(2.8%)	55	(2.9%)	39	(2.6%)	

No change	863	(25.5%)	592	(31.2%)	271	(18.2%)	
Declined	40	(1.2%)	23	(1.2%)	17	(1.1%)	
Missing*	2388	(70.5%	1226	(64.7%)	1162	(78%)	
Number of Comorbid Conditions	2	[0-14]	2	[0-14]	2	[0-14]	**<0.001****
Caregiver							**<0.001****
Caregiver	939	(27.7%)	580	(30.6%)	359	(24.1%)	
No caregiver/Unknown	2446	(72.3%)	1316	(69.4%)	1130	(75.9%)	
Spousal Caregiver							0.672
Spousal caregiver	547	(16.2%)	311	(16.4%)	236	(15.8%)	
Non-spousal caregiver or unknown	2838	(83.8%)	1585	(83.6%)	1253	(84.2%)	
Discharged with an opioid							**<0.001****
Discharged with an opioid	627	(18.5%)	203	(10.7%)	424	(28.5%)	
Not discharged with an opioid	2758	(81.5%)	1693	(89.3%)	1065	(71.5%)	

Legend: Tests include Chi-squared, Kruskal-Wallis, or Fisher's Exact as appropriate | Bold indicates p<0.05 |* indicates missing values were excluded from descriptive statistics tests

BIBLIOGRAPHY

1. Centers for Medicare and Medicaid Services. Medicare and Medicaid Programs; Revisions to Requirements for Discharge Planning for Hospitals, Critical Access to Hospitals, and Home Health Agencies. https://www.federalregister.gov/documents/2019/09/30/2019-20732/medicare-and-medicaid-programs-revisions-to-requirements-for-discharge-planning-for-hospitals (Accessed June 2020).
2. Tian W. *Statistical Brief #205. Healthcare Cost and Utilization Project (HCUP).* Agency for Healthcare Research and Quality, Rockville, MD; 2016.
3. Medicare Payment Advisory Commission. Post-Acute Care. 2019. http://www.medpac.gov/-research-areas-/post-acute-care (Accessed June 2020).
4. Holland DE, Bowles KH. Standardized discharge planning assessments: impact on patient outcomes. *J Nurs Care Qual.* 2012;27(3):200-208.
5. Phillips LR, Salem BE, Skrine Jeffers K, et al. Developing and proposing the ethno-cultural gerontological nursing model. *J Transcult Nurs.* 2015;26(2):118-128.
6. Bradley EH, Sipsma H, Horwitz LI, Curry L, Krumholz HM. Contemporary data about hospital strategies to reduce unplanned readmissions: what has changed? *JAMA Intern Med.* 2014;174(1):154-156.
7. Benson KJ, Weech-Maldonado R, Gamm LD. The role of value in provider-patient-payer relationships. *Hosp Top.* 2003;81(3):21-28.
8. Prusaczyk B, Kripalani S, Dhand A. Summary discharge planning network for older adults discharged from the hospital. In: Networks of hospital discharge planning teams and readmissions, ed: *Journal of Interprofessional Care*; 2019.
9. Meiboom A, Diedrich C, Vries HD, Hertogh C, Scheele F. The hidden curriculum of the medical care for elderly patients in medical education: a qualitative study. *Gerontology & Geriatrics Education.* 2015;36(1):30-44.
10. Goldman J, Reeves S, Wu R, Silver I, MacMillan K, Kitto S. A sociological exploration of the tensions related to interprofessional collaboration in acute-care discharge planning. *J Interprof Care.* 2016;30(2):217-225.
11. Mead H, Ramos C, Grantham SC. Drivers of racial and ethnic disparities in cardiac rehabilitation use: patient and provider Perspectives. *Med Care Res Rev.* 2016;73(3):251-282.
12. Borkhoff CM, Hawker GA, Kreder HJ, Glazier RH, Mahomed NN, Wright JG. Influence of patients' gender on informed decision making regarding total knee arthroplasty. *Arthritis Care Res (Hoboken).* 2013;65(8):1281-1290.
13. Blum MA, Ibrahim SA. Race/ethnicity and use of elective joint replacement in the management of end-stage knee/hip osteoarthritis: a review of the literature. *Clin Geriatr Med.* 2012;28(3):521-532.
14. Kemper P, Komisar HL, Alecxih L. Long-term care over an uncertain future: what can current retirees expect? *Inquiry.* 2005;42(4):335-350.
15. Henning-Smith CE, Shippee TP. Expectations about future use of long-term services and supports vary by current living arrangement. *Health Aff (Millwood).* 2015;34(1):39-47.
16. Zupanc S, Lee K, Moseley E, Yeh I, Lindvall C. Finding what matters: using natural language processing to identify patient care preferences within clinical notes. *BMJ Supp Palliat Care*; 2019; 9.

17. Islam MS, Hasan MM, Wang X, Germack HD, Noor-E-Alam M. A systematic review on healthcare analytics: application and theoretical perspective of data mining. *Healthcare (Basel).* 2018;6(2):54.
18. Topaz M, Kang Y, Holland DE, Ohta B, Rickard K, Bowles KH. Higher 30-day and 60-day readmissions among patients who refuse post acute care services. *Am J Manag Care.* 2015;21(6):424-433.
19. Rogut L, Kothari P. *Difficult Decisions About Post-Acute Care and Why They Matter.* United Hospital Fund; 2018.
20. Henning-Smith C, Kozhimannil KB, Casey MM, Prasad S. Beyond clinical complexity: nonmedical barriers to nursing home care for rural residents. *J Aging Soc Policy.* 2018;30(2):109-126.
21. Cheah LP, Amott DH, Pollard J, Watters DA. Electronic medical handover: towards safer medical care. *Med J Aust.* 2005;183(7):369-372.
22. Friedman B, Basu J. The rate and cost of hospital readmissions for preventable conditions. *Med Care Res Rev.* 2004;61(2):225-240.
23. Pollack R. Perspective: Post-acute care is a vital part of health care delivery. 2019; https://www.aha.org/news/perspective/2019-07-12-perspective-post-acute-care-vital-part-health-care-delivery.
24. Kripalani S, Theobald CN, Anctil B, Vasilevskis EE. Reducing hospital readmission rates: current strategies and future directions. *Annu Rev Med.* 2014;65:471-485.
25. Bajorek SA, McElroy V. Discharge Planning and Transitions of Care. *Patient Safety Primer* 2020; https://psnet.ahrq.gov/primer/discharge-planning-and-transitions-care.
26. Newnham H, Barker A, Ritchie E, Hitchcock K, Gibbs H, Holton S. Discharge communication practices and healthcare provider and patient preferences, satisfaction and comprehension: a systematic review. *International Journal for Quality in Health Care.* 2017;29(6):752-768.
27. Office of the National Coordinator for Health Information Technology. Health IT Legislation. 2020; https://www.healthit.gov/topic/laws-regulation-and-policy/health-it-legislation (accessed 28 January, 2020).
28. Archie RR, Boren SA. Opportunities for informatics to improve discharge planning: a systematic review of the literature. *AMIA Annual Symposium proceedings AMIA Symposium.* 2009:16-20.
29. Zolbanin HM, Delen D. Processing electronic medical records to improve predictive analytics outcomes for hospital readmissions. *Decision Support Systems.* 2018;112:98-110.
30. HealthIT.gov. Clinical Decision Support. 2018; https://www.healthit.gov/topic/safety/clinical-decision-support (accessed March 11, 2021).
31. Osheroff JA, Teich JM, Middleton B, Steen EB, Wright A, Detmer DE. A Roadmap for national action on clinical decision support. *Journal of the American Medical Informatics Association.* 2007;14(2):141-145.
32. Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *BMJ.* 2005;330(7494):765.

33. Epstein D, Barak-Corren Y, Isenberg Y, Berger G. Clinical decision support system: a pragmatic tool to improve acute exacerbation of COPD discharge recommendations. *COPD.* 2019;16(1):18-24.
34. Bahm A, Freedman SB, Guan J, Guttmann A. Evaluating the impact of clinical decision tools in pediatric acute gastroenteritis: a population-based cohort study. *Society for Academic Emergency Medicine (US).* 2016;23(5):599-609.
35. Mahler SA, Riley RF, Hiestand BC, et al. The HEART Pathway randomized trial: identifying emergency department patients with acute chest pain for early discharge. *Circ Cardiovasc Qual Outcomes.* 2015;8(2):195-203.
36. Topaz M, Trifilio M, Maloney D, Bar-Bachar O, Bowles KH. Improving patient prioritization during hospital-homecare transition: a pilot study of a clinical decision support tool. *Res Nurs Health.* 2018;41(5):440-447.
37. Pawloski PA, Brooks GA, Nielsen ME, Olson-Bullis BA. A systematic review of clinical decision support systems for clinical oncology practice. *J Natl Compr Canc Netw.* 2019;17(4):331-338.
38. Knoery CR, Heaton J, Polson R, et al. Systematic review of clinical decision support systems for prehospital acute coronary syndrome identification. *Crit Pathw Cardiol.* 2020;19(3):119-125.
39. Rawson TM, Moore LSP, Hernandez B, et al. A systematic review of clinical decision support systems for antimicrobial management: are we failing to investigate these interventions appropriately? *Clin Microbiol Infect.* 2017;23(8):524-532.
40. Kennedy EE, Bowles KH, Aryal S. Systematic review of prediciton models for postacute care destination decision making. *Journal of the American Medical Informatics Association.* 2022;29(1):176-186.
41. Bowles KH, Ratcliffe SJ, Naylor MD, Holmes JH, Keim SK, Flores EJ. Nurse generated EHR data supports post-acute care referral decision making: development and validation of a two-step algorithm. *AMIA Annual Symposium Proceedings.* 2017:465-474.
42. Bowles KH, Ratcliffe SJ, Holmes JH, et al. Using a decision support algorithm for referrals to post-acute care. *Journal of the American Medical Directors Association.* 2019;20(4):408-413.
43. Moons KGM, Wolff RF, Riley RD, et al. PROBAST: A tool to assess risk of bias and applicability of prediction model studies: explanation and elaboration. *Ann Intern Med.* 2019;170(1):W1-w33.
44. Orem DE. *Nursing: Concepts of Practice.* 6th Ed. ed. St. Louis: Mosby; 2001.
45. Hartweg DL, Fleck LM, Parker ME, Smith MC. Dorothea Orem's Self-Care Deficit Theory In: *Nursing Theories & Nursing Practice.* 3rd ed. Philadelphia, PA: F. A. Davis Company 2010.
46. Orem DE. *Nursing: Concepts of Practice.* 4 ed. St. Louis, MO: Mosby-Year Book Inc; 1991.
47. Moons KGM, Altman DG, Reitsma JB, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): Explanation and Elaboration. *Ann of Intern Med.* 2015;162(1):W1-W73.
48. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med.* 2015;162(1):55-63.

49. Miotto R, Wang F, Wang S, Jiang X, Dudley JT. Deep learning for healthcare: review, opportunities and challenges. *Brief Bioinform.* 2018;19(6):1236-1246.
50. Bowles KH, Chittams J, Heil E, et al. Successful electronic implementation of discharge referral decision support has a positive impact on 30- and 60-day readmissions. *Res Nurs Health.* 2015;38(2):102-114.
51. Bowles KH, Holmes JH, Ratcliffe SJ, Liberatore M, Nydick R, Naylor MD. Factors identified by experts to support decision making for post acute referral. *Nursing Research.* 2009;58(2):115-122.
52. Bowles KH, Ratcliffe SJ, Naylor MD, Holmes JH, Keim SK, Flores EJ. Nurse Generated EHR Data Supports Post-Acute Care Referral Decision Making: Development and Validation of a Two-step Algorithm. *AMIA Annual Symposium proceedings.* 2017:465-474.
53. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology.* 1982;143(1):29-36.
54. Vittinghoff E, Glidden DV, Shiboski SC, CMcCulloch CE. *Regression Methods in Biostatistics.* Vol 2. New York, NY: Springer New York Dordrecht Heidelberg London; 2012.
55. Leeflang MMG, Moons KGM, Reitsma JB, Zwinderman AH. Bias in sensitivity and specificity caused by data-driven selection of optimal cutoff values: mechanisms, magnitude, and solutions. *Clinical Chemistry.* 2008:729+.
56. Parikh R, Mathai A, Parikh S, Chandra Sekhar G, Thomas R. Understanding and using sensitivity, specificity and predictive values. *Indian J Ophthalmol.* 2008;56(1):45-50.
57. Bowles KH, Heil E. From unmet clinical need to entrepreneurship: taking your informatics solution to market. *Stud Health Technol Inform.* 2014;201:315-320.
58. Sheikhalishahi S, Miotto R, Dudley JT, Lavelli A, Rinaldi F, Osmani V. Natural language processing of clinical notes on chronic diseases: systematic review. *JMIR Med Inform.* 2019;7(2):e12239-e12239.
59. Kwan JL, Lo L, Ferguson J, et al. Computerised clinical decision support systems and absolute improvements in care: meta-analysis of controlled clinical trials. *BMJ.* 2020;370:m3216.
60. Doan S, Conway M, Phuong TM, Ohno-Machado L. Natural language processing in biomedicine: a unified system architecture overview. *Methods Mol Biol.* 2014;1168:275-294.
61. Kawamoto K, McDonald CJ. Designing, conducting, and reporting clinical decision support studies: recommendations and call to action. *Ann Intern Med.* 2020;172(11_Supplement):S101-S109.
62. Conway M, Keyhani S, Christensen L, et al. Moonstone: a novel natural language processing system for inferring social risk from clinical narratives. *J Biomed Semantics.* 2019;10(1):6.
63. Rumshisky A, Ghassemi M, Naumann T, et al. Predicting early psychiatric readmission with natural language processing of narrative discharge summaries. *Transl Psychiatry.* 2016;6(10):e921.
64. Tang Y, Yang J, Ang PS, et al. Detecting adverse drug reactions in discharge summaries of electronic medical records using Readpeer. *Int J Med Inform.* 2019;128:62-70.

65. Popejoy LL, Khalilia MA, Popescu M, et al. Quantifying care coordination using natural language processing and domain-specific ontology. *Journal of the American Medical Informatics Association.* 2015;22(e1):e93-103.
66. Nawab K, Ramsey G, Schreiber R. Natural language processing to extract meaningful information from patient experience feedback. *Appl Clin Inform.* 2020;11(2):242-252.
67. Coiera E, Tong HL. Replication studies in the clinical decision support literature–frequency, fidelity, and impact. *Journal of the American Medical Informatics Association.* 2021.
68. Office of the National Coordinator for Health Information Technology. Connecting Health and Care for the Nation: A Shared Nationwide Interoperability Roadmap. 2019; https://www.healthit.gov/sites/default/files/hie-interoperability/nationwide-interoperability-roadmap-final-version-1.0.pdf (accessed 7 October, 2020).
69. Chan KS, Fowles JB, Weiner JP. Review: Electronic health records and the reliability and validity of quality measures: a review of the literature. *Med Care Res Rev.* 2010;67(5):503-527.
70. Elsevier. Care Planning with Epic. 2020; https://www.elsevier.com/clinical-solutions/about/partners-and-advisors/epic/care-planning-and-epic-ehr-integration#:~:text=Serving%20more%20than%20350%20Epic%20hospitals%2C%20Else Else%20Care,efficiently%20document%20and%20monitor%20care%20for%20every%20 patient (accessed June 2020).
71. Cheng HG, Phillips MR. Secondary analysis of existing data: opportunities and implementation. *Shanghai Arch Psychiatry.* 2014;26(6):371-375.
72. Smith AK, Ayanian JZ, Covinsky KE, et al. Conducting high-value secondary dataset analysis: an introductory guide and resources. *J Gen Intern Med.* 2011;26(8):920-929.
73. Bird S, Klein E, Loper E. *Natural Language Processing with Python.* Sebastopol, CA: O'Reilly Media, Inc; 2009.
74. Lin C-J, Cheng S-J, Shih S-C, Chu C-H, Tjung J-J. Discharge planning. *Int J Gerontol.* 2012;6(4):237-240.
75. Office of the National Coordinator for Health Information Technology. What is an electronic health record (EHR)? 2019; https://www.healthit.gov/faq/what-electronic-health-record-ehr (accessed February 2, 2021).
76. South BR, Mowery D, Suo Y, et al. Evaluating the effects of machine pre-annotation and an interactive annotation interface on manual de-identification of clinical text. *J Biomed Inform.* 2014;50:162-172.
77. Goutte C, Gaussier E. A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation. 2005; Berlin, Heidelberg.
78. Brownlee J. How to Use ROC Curves and Precision-Recall Curves for Classification in Python. 2018; https://machinelearningmastery.com/roc-curves-and-precision-recall-curves-for-classification-in-python/ (accessed April 21, 2021).
79. Di Pietro M. Text Analysis & Feature Engineering with NLP. In. Vol 2021: Towards Data Science; 2020.
80. Chowdhury GG. Natural language processing. *Annual Review of Information Science and Technology.* 2003;37(1):51-89.

81. Patel S. NLP Pipeline: Building an NLP Pipeline, Step-by-Step. 2020; https://medium.com/@suneelpatel.in/nlp-pipeline-building-an-nlp-pipeline-step-by-step-7f0576e11d08 (accessed April 21, 2021).
82. Davis J, Goadrich M. The relationship between Precision-Recall and ROC curves. 2006.
83. Pickell D. Structured vs. Unstructured Data- What's the Difference? 2018. https://learn.g2.com/structured-vs-unstructured-data#:~:text=What%20is%20the%20difference%20between,collect%2C%20process%2C%20and%20analyze (accessed April 21, 2021).
84. Supervised vs Unsupervised Learning: Key Differences. https://learn.g2.com/structured-vs-unstructured-data#:~:text=What%20is%20the%20difference%20between,collect%2C%20process%2C%20and%20analyze. Accessed April 21, 2021.
85. Centers for Medicare and Medicaid Services. Hospital Readmissions Reduction Program (HRRP). 2020; https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/HRRP/Hospital-Readmission-Reduction-Program. Accessed February 2, 2021.
86. Keim SK, Bowles KH. Comparison of algorithm advice for post-acute care referral to usual clinical decision-making: examination of 30-day acute healthcare utilization. *AMIA Annual Symposium proceedings* 2017:1051-1059.
87. Kennedy EE, Bowles KH, Aryal S. Systematic review of prediction models for postacute care destination decision-making. *Journal of the American Medical Informatics Association.* 2021.
88. Centers for Disease Control and Prevention. Number and Rate of Hospital Discharges, 2010. *National Hospital Discharge Survey* 2015; https://www.cdc.gov/nchs/nhds/nhds_tables.htm (accessed 11 August, 2020).
89. Prusaczyk B, Kripalani S, Dhand A. Networks of hospital discharge planning teams and readmissions. *J InterprofCare.* 2019;33(1):85-92.
90. American Hospital Association. *The Role of Post-Acute Care in New Care Delivery Models.* American Hospital Association: American Hospital Association;2015.
91. Vogenberg FR. Predictive and prognostic models: implications for healthcare decision-making in a modern recession. *Am Health Drug Benefits.* 2009;2(6):218-222.
92. Hansen LO, Young RS, Hinami K, Leung A, Williams MV. Interventions to reduce 30-day rehospitalization: a systematic review. *Ann Intern Med.* 2011;155(8):520-528.
93. Wang W, Kiik M, Peek N, et al. A systematic review of machine learning models for predicting outcomes of stroke with structured data. *PLOS ONE.* 2020;15(6):e0234722.
94. Kobewka DM, Mulpuru S, Chassé M, et al. Predicting the need for supportive services after discharged from hospital: a systematic review. *BMC Health Serv Res.* 2020;20(1):161.
95. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ.* 2021;372:n71.
96. Cochrane Prognosis Methods Group. Systematic Review Tools. 2020; https://methods.cochrane.org/prognosis/tools (accessed June 2020).
97. Moons KG, de Groot JA, Bouwmeester W, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLoS Med.* 2014;11(10):e1001744.

98. Evidence Partners. DistillerSR. 2020; https://www.evidencepartners.com/products/distillersr-systematic-review-software/ (accessed June 2020).
99. Cho JS, Hu Z, Fell N, Heath GW, Qayyum R, Sartipi M. Hospital discharge disposition of stroke patients in tennessee. *South Med J.* 2017;110(9):594-600.
100. James MK, Robitsek RJ, Saghir SM, Gentile PA, Ramos M, Perez F. Clinical and non-clinical factors that predict discharge disposition after a fall. *Injury.* 2018;49(5):975-982.
101. Nassour I, Wang SC, Christie A, et al. Nomogram to predict non-home discharge following pancreaticoduodenectomy in a national cohort of patients. *HPB (Oxford).* 2017;19(12):1037-1045.
102. Rondon AJ, Tan TL, Greenky MR, et al. Who goes to inpatient rehabilitation or skilled nursing facilities unexpectedly following total knee arthroplasty? *J Arthroplasty.* 2018;33(5):1348-1351.e1341.
103. Stuebe J, Rydingsward J, Lander H, et al. A Pragmatic preoperative prediction score for nonhome discharge after cardiac operations. *Ann Thorac Surg.* 2018;105(5):1384-1391.
104. Tan TL, Rondon AJ, Greenky MR, Shohat N, Goswami K, Purtill JJ. Which patients require unexpected admission to postacute care facilities after total hip arthroplasty? *J Am Acad Orthop Surg.* 2019.
105. Bowles KH, Ratcliffe SJ, Naylor MD, Holmes JH, Keim SK, Flores EJ. Nurse generated EHR data supports post-acute care referral decision making: development and validation of a two-step algorithm. *AMIA Annual Symposium Proceedings*;2017:465-474.
106. Ballester N, Parikh PJ, Donlin M. An early warning tool for predicting at admission the discharge disposition of a hospitalized patient. *Am J Manag Care.* 2018;24(10):e325-e331.
107. Barsoum WKMD, Murray TGMD, Klika AKMS, et al. Predicting patient discharge disposition after total joint arthroplasty in the United States. *J Arthroplasty.* 2010;25(6):885-892.
108. Pattakos GMDMS, Johnston DRMD, Houghtaling PLMS, Nowicki ERMDMS, Blackstone EHMD. Preoperative prediction of non-home discharge: a strategy to reduce resource use after cardiac surgery. *J Am Coll Surg.* 2012;214(2):140-147.
109. Sharma BS, Swisher MW, Doan CN, Khatibi B, Gabriel RA. Predicting patients requiring discharge to post-acute care facilities following primary total hip replacement: Does anesthesia type play a role? *J Clin Anesth.* 2018;51:32-36.
110. Tapper EB, Finkelstein D, Mittleman MA, Piatkowski G, Lai M. Standard assessments of frailty are validated predictors of mortality in hospitalized patients with cirrhosis. *Hepatology.* 2015;62(2):584-590.
111. AlHilli MM, Tran CW, Langstraat CL, et al. Risk-scoring model for prediction of non-home discharge in epithelial ovarian cancer patients. *J Am Coll Surg.* 2013;217(3):507-515.
112. Bailey EA, Hoffman RL, Wirtalla C, Karakousis G, Kelz RR. Development and validation of a prediction model for patients discharged to post-acute care after colorectal cancer surgery. *Surgery.* 2017;161(4):1049-1057.
113. Karhade AV, Ogink P, Thio Q, et al. Development of machine learning algorithms for prediction of discharge disposition after elective inpatient surgery for lumbar degenerative disc disorders. *Neurosurg Focus.* 2018;45(5):E6.

114. Karnuta JM, Golubovsky JL, Haeberle HS, et al. Can a machine learning model accurately predict patient resource utilization following lumbar spinal fusion? *Spine J.* 2020;20(3):329-336.
115. Kimmel LA, Holland AE, Edwards ER, et al. Discharge destination following lower limb fracture: Development of a prediction model to assist with decision making. *Injury.* 2011;43(6):829-834.
116. McGirt MJ, Parker SL, Chotai S, et al. Predictors of extended length of stay, discharge to inpatient rehab, and hospital readmission following elective lumbar spine surgery: introduction of the Carolina-Semmes Grading Scale. *J Spine Neurosurgery.* 2017:1-9.
117. Penn CA, Kamdar NS, Morgan DM, Uppal S. Preoperatively predicting non-home discharge after surgery for gynecologic malignancy. *Gynecol Oncol.* 2017;145:15-16.
118. Stineman MG, Kwong PL, Bates BE, Kurichi JE, Ripley DC, Xie D. Development and validation of a discharge planning index for achieving home discharge after hospitalization for acute stroke among those who received rehabilitation services. *Am J Phys Med Rehabil.* 2014;93(3):217-230.
119. Sultana I, Erraguntla M, Kum HC, Delen D, Lawley M. Post-acute care referral in United States of America: a multiregional study of factors associated with referral destination in a cohort of patients with coronary artery bypass graft or valve replacement. *BMC med Inform and Decis Mak.* 2019;19(1):223.
120. Wasfy JH, Kennedy KF, Masoudi FA, et al. Predicting length of stay and the need for postacute care after acute myocardial infarction to improve healthcare efficiency. *Circ Cardiovasc Qual Outcomes.* 2018;11(9):e004635.
121. Chang DC, Joyce DL, Shoher A, Yuh DD. Simple index to predict likelihood of skilled nursing facility admission after coronary artery bypass grafting among older patients. *Ann Thorac Surg.* 2007;84(3):829-834; discussion 834-825.
122. Ramkumar PN, Karnuta JM, Navarro SM, et al. Preoperative prediction of value metrics and a patient-specific payment model for primary total hip arthroplasty: development and validation of a deep learning model. *J Arthroplasty.* 2019;34(10):2228-2234.e2221.
123. Louis Simonet M, Kossovsky MP, Chopard P, Sigaud P, Perneger TV, Gaspoz JM. A predictive score to identify hospitalized patients' risk of discharge to a post-acute care facility. *BMC Health Serv Res.* 2008;8:154.
124. Oldmeadow LB, McBurney H, Robertson VJ. Predicting risk of extended inpatient rehabilitation after hip or knee arthroplasty. *J Arthroplasty.* 2003;18(6):775-779.
125. Zureik M, Lombrail P, Davido A, et al. Predicting the outcome in elderly patients of hospital admission for acute care in Paris, France: construction and initial validation of a simple index. *J Epidemiol Community Health.* 1997;51(2):192-198.
126. Bowles KH, Chittams J, Heil E, et al. Successful electronic implementation of discharge referral decision support Has a positive impact on 30- and 60-day readmissions. *Res Nurs Health.* 2015;38(2):102-114.
127. Tong MZ, Pattakos G, He J, et al. Sequentially updated discharge model for optimizing hospital resource use and surgical patients' satisfaction. *Ann Thorac Surg.* 2015;100(6):2174-2181.
128. Wadhera RK, Yeh RW, Joynt Maddox KE. The rise and fall of mandatory cardiac bundled payments. *JAMA.* 2018;319(4):335-336.

129. Van Horne A, Van Horne J. Patient-optimizing enhanced recovery pathways for total knee and hip arthroplasty in Medicare patients: implication for transition to ambulatory surgery centers. *Arthroplasty Today.* 2019;5(4):497-509.
130. Harrell FE, Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med.* 1996;15(4):361-387.
131. American Hospital Association. *Trendwatch: Increasing Access to Behavioral Health Care Advances Value for Patients, Providers, and Communities.* 2019. https://www.aha.org/system/files/media/file/2019/05/aha-trendwatch-behavioral-health-2019.pdf (accessed June 2020).
132. Ohrnberger J, Fichera E, Sutton M. The relationship between physical and mental health: A mediation analysis. *Social Science & Medicine.* 2017;195:42-49.
133. Shih SL, Gerrard P, Goldstein R, et al. Functional status outperforms comorbidities in predicting acute care readmissions in medically complex patients. *J Gen Intern Med.* 2015;30(11):1688-1695.
134. Fortinsky RH, Covinsky KE, Palmer RM, Landefeld CS. Effects of functional status changes before and during hospitalization on nursing home admission of older adults. *J Gerontol A Biol Sci Med Sci.* 1999;54(10):M521-526.
135. Lan RH, Kamath AF. Post-acute care disparities in total joint arthroplasty. *Arthroplast Today.* 2017;3(3):187-191.
136. Claflin J, Dimick JB, Campbell DA, Englesbe MJ, Sheetz KH. Understanding disparities in surgical outcomes for medicaid beneficiaries. *World J Surg.* 2019;43(4):981-987.
137. Beatty AL, Bradley SM, Maynard C, McCabe JM. Referral to cardiac rehabilitation after percutaneous coronary intervention, coronary artery bypass surgery, and valve surgery: data From the clinical outcomes assessment program. *Circ Cardiovasc Qual Outcomes.* 2017;10(6).
138. Steyerberg EW, Moons KGM, van der Windt DA, et al. Prognosis research strategy (PROGRESS) 3: prognostic model research. *PLoS medicine.* 2013;10(2):e1001381-e1001381.
139. Gianfrancesco MA, Tamang S, Yazdany J, Schmajuk G. Potential biases in machine learning algorithms using electronic health record data. *JAMA Intern Med.* 2018;178(11):1544-1547.
140. Nick J. Open access Part I: The movement, the issues, and the benefits. *Online J issues Nurs.* 2011;17:8.
141. Hansen VJ, Gromov K, Lebrun LM, Rubash HE, Malchau H, Freiberg AA. Does the risk assessment and prediction tool predict discharge disposition after joint replacement? *Clin Orthop Relat Res.* 2015;473(2):597-601.
142. Tan C, Loo G, Pua YH, et al. Predicting discharge outcomes after total knee replacement using the Risk Assessment and Predictor Tool. *Physiotherapy.* 2014;100(2):176-181.
143. Firmino RT, Fraiz FC, Montes GR, Paiva SM, Granville-Garcia AF, Ferreira FM. Impact of oral health literacy on self-reported missing data in epidemiological research. *Community Dent Oral Epidemiol.* 2018;46(6):624-630.
144. Parikh RB, Teeple S, Navathe AS. Addressing bias in artificial intelligence in health care. *JAMA.* 2019;322(24):2377-2378.

145. Moons KG, Kengne AP, Woodward M, et al. Risk prediction models: Development, internal validation, and assessing the incremental value of a new (bio)marker. *Heart.* 2012;98(9):683-690.
146. Ganna A, Reilly M, de Faire U, Pedersen N, Magnusson P, Ingelsson E. Risk prediction measures for case-cohort and nested case-control designs: an application to cardiovascular disease. *Am J Epidemiol.* 2012;175(7):715-724.
147. Sanderson J, Thompson SG, White IR, Aspelund T, Pennells L. Derivation and assessment of risk prediction models using case-cohort data. *BMC Med Res Methodol.* 2013;13:113.
148. Moons KG, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? *BMJ.* 2009;338:b375.
149. Steyerberg EW, Bleeker SE, Moll HA, Grobbee DE, Moons KG. Internal and external validation of predictive models: a simulation study of bias and precision in small samples. *J Clin Epidemiol.* 2003;56(5):441-447.
150. Steyerberg EW, Eijkemans MJ, Harrell FE, Jr., Habbema JD. Prognostic modeling with logistic regression analysis: in search of a sensible strategy in small data sets. *Med Decis Making.* 2001;21(1):45-56.
151. van Smeden M, de Groot JA, Moons KG, et al. No rationale for 1 variable per 10 events criterion for binary logistic regression analysis. *BMC Med Res Methodol.* 2016;16(1):163.
152. Ogundimu EO, Altman DG, Collins GS. Adequate sample size for developing prediction models is not simply related to events per variable. *J Clin Epidemiol.* 2016;76:175-182.
153. Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. *Stat Med.* 2006;25(1):127-141.
154. Vergouwe Y, Royston P, Moons KG, Altman DG. Development and validation of a prediction model with missing predictor data: a practical approach. *J Clin Epidemiol.* 2010;63(2):205-214.
155. Sterne JAC, White IR, Carlin JB, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ.* 2009;338:b2393.
156. Young BA. Health disparities in advanced heart failure treatment: the intersection of race and sex. *JAMA Network Open.* 2020;3(7):e2011034-e2011034.
157. Graham G. Disparities in cardiovascular disease risk in the United States. *Curr Cardiol Rev.* 2015;11(3):238-245.
158. Davis AM, Vinci LM, Okwuosa TM, Chase AR, Huang ES. Cardiovascular health disparities: a systematic review of health care interventions. *Medical Care Res Review.* 2007;64(5 Suppl):29S-100S.
159. Adelani MA, O'Connor MI. Perspectives of orthopedic surgeons on racial/ethnic disparities in care. *J Racial Ethn Health Disparities.* 2017;4(4):758-762.
160. Schoenfeld AJ, Tipirneni R, Nelson JH, Carpenter JE, Iwashyna TJ. The influence of race and ethnicity on complications and mortality after orthopedic surgery: a systematic review of the literature. *Medical Care.* 2014;52(9):842-851.
161. Osheroff JA, Teich JM, Middleton B, Steen EB, Wright A, Detmer DE. A roadmap for national action on clinical decision support. *Journal of the American Medical Informatics Association.* 2007;14(141-145).
162. Duda RO, Shortliffe EH. Expert Systems Research. *Science.* 1983;220(4594):261-268.
163. Clarity. Epic EHR Solution. 2020; https://www.clarity-ventures.com/ecommerce-integrations/emr-ehr/epic-integration (accessed June 2020).

164. Penn Medicine Data Analytics Center. Epic Clarity: Overview. https://www.med.upenn.edu/dac/epic-clarity-data-warehousing.html (Accessed June 2020).
165. Thomas T, Coombs NC, Mullette EJ, Bick T, Brant JM. bilateral total knee arthroplasty. *Orthop Nurs (Online).* 2020;39(2):92-96.
166. Werner RM, Coe NB, Qi M, Konetzka RT. Patient outcomes after hospital discharge to home with home health care vs to a skilled nursing facility. *JAMA Intern Med.* 2019;179(5):617-623.
167. Singh JA, Kallan MJ, Chen Y, Parks ML, Ibrahim SA. Association of race/ethnicity with hospital discharge disposition after elective total knee arthroplasty. *JAMA Network Open.* 2019;2(10):e1914259.
168. Muhlestein WE, Akagi DS, Chotai S, Chambless LB. The impact of race on discharge disposition and length of hospitalization after craniotomy for brain tumor. *World Neurosurg.* 2017;104:24-38.
169. Ioannidis JPA, Powe NR, Yancy C. Recalibrating the use of race in medical research. *JAMA.* 2021;325(7):623-624.
170. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science.* 2019;366(6464):447-453.
171. Horwitz LI, Bernheim SM, Ross JS, et al. Hospital characteristics associated with risk-standardized readmission rates. *Medical care.* 2017;55(5):528-534.
172. Werner RM, Konetzka RT. Trends in post–acute care use among medicare beneficiaries: 2000 to 2015. *JAMA.* 2018;319(15):1616-1617.
173. Werner RM, Bressman E. Trends in post-acute care utilization during the COVID-19 pandemic. *J Am Med Dir Assoc.* 2021;22(12):2496-2499.
174. Wang S, Zhu X. Predictive modeling of hospital readmission: Challenges and Solutions. *IEEE/ACM Transactions on Computational Biology and Bioinformatics.* 2021.
175. Zhou H, Della PR, Roberts P, Goh L, Dhaliwal SS. Utility of models to predict 28-day or 30-day unplanned hospital readmissions: an updated systematic review. *BMJ Open.* 2016;6(6):e011060.
176. Burnett RA, 3rd, Serino J, Yang J, Della Valle CJ, Courtney PM. National trends in post-acute care costs following total knee arthroplasty from 2007 to 2016. *J Arthroplasty.* 2021;36(7):2268-2275.
177. Mead H, Ramos C, Grantham SC. Drivers of racial and ethnic disparities in cardiac rehabilitation use: patient and provider perspectives. *Med Care Res Review.* 2015;73(3):251-282.
178. Hodgins MJ, Filiatreault S, Keeping-Burke L, Logan SM. Patterns of patient coping following hospital discharge from medical and surgical units: A pilot study. *Nurs Health Sci.* 2020;22(1):118-125.
179. Sacks GD, Lawson EH, Dawes AJ, Gibbons MM, Zingmond DS, Ko CY. Which patients require more care after hospital discharge? an analysis of post-acute care use among elderly patients undergoing elective surgery. *J Am Coll Surg.* 2015;220(6):1113-1121.e1112.
180. Weiss AJ, Jiang HJ. *Overview of Clinical Conditions with Frequent and Costly Hospital Readmissions by Payer, 2018.* 2021. https://www.hcup-

us.ahrq.gov/reports/statbriefs/sb278-Conditions-Frequent-Readmissions-By-Payer-2018.jsp.
181. Liang L, Moore B, Soni A. *National Inpatient Hospital Costs: The Most Expensive Conditions by Payer, 2017.* Healthcare Cost and Utilization Project; 2020.
182. Carli D, Fahrni G, Bonnabry P, Lovis C. Quality of decision support in computerized provider order entry: systematic literature review. *JMIR Med Inform.* 2018;6(1):e3-e3.
183. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. *NPJ Digital Medicine.* 2020;3(1):17.
184. Wells BJ, Chagin KM, Nowacki AS, Kattan MW. Strategies for handling missing data in electronic health record derived data. *EGEMS (Washington, DC).* 2013;1(3):1035-1035.
185. Beaulieu-Jones BK, Lavage DR, Snyder JW, Moore JH, Pendergrass SA, Bauer CR. Characterizing and managing missing structured data in electronic health records: data analysis. *JMIR Med Inform.* 2018;6(1):e11-e11.
186. McCombe N, Liu S, Ding X, et al. Practical strategies for extreme missing data Imputation in dementia diagnosis. *medRxiv.* 2021:2020.2007.2013.20146118.
187. Wong A, Young AT, Liang AS, Gonzales R, Douglas VC, Hadley D. Development and validation of an electronic health record–based machine learning model to estimate delirium risk in newly hospitalized patients without known cognitive impairment. *JAMA Network Open.* 2018;1(4):e181018-e181018.
188. Holmes JH, Beinlich J, Boland MR, et al. Why is the electronic health record so challenging for research and clinical care? *Methods Inf Med.* 2021;60(1-02):32-48.
189. Kim J, Macieira TGR, Meyer SL, et al. Towards implementing SNOMED CT in nursing practice: A scoping review. *Int J Med Inform.* 2020;134:104035.
190. Coiera E, Tong HL. Replication studies in the clinical decision support literature-frequency, fidelity, and impact. *Journal of the American Medical Informatics Association.* 2021;28(9):1815-1825.
191. Centers For Medicare and Medicaid Services. IMPACT Act of 2014 Data Standardization & Cross Setting Measures. 2021; https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Post-Acute-Care-Quality-Initiatives/IMPACT-Act-of-2014/IMPACT-Act-of-2014-Data-Standardization-and-Cross-Setting-Measures (accessed January 2022).
192. Mosher HJ, Lyckholm LJ. A Problem of Capacity, but Whose? The hospitalists' discharge dilemma and social determinants of health. *J Hosp Med.* 2017;12(1):57-58.
193. Botsis T, Hartvigsen G, Chen F, Weng C. Secondary use of EHR: data quality issues and informatics opportunities. *Summit Transl Bioinform.* 2010;2010:1-5.
194. Covinsky KE, Palmer RM, Fortinsky RH, et al. Loss of independence in activities of daily living in older adults hospitalized with medical illnesses: increased vulnerability with age. *J Am Geriatr Soc.* 2003;51(4):451-458.
195. Marcantonio ER, McKean S, Goldfinger M, Kleefield S, Yurkofsky M, Brennan TA. Factors associated with unplanned hospital readmission among patients 65 years of age and older in a Medicare managed care plan. *Am J Med.* 1999;107(1):13-17.
196. Musich S, Wang SS, Ruiz J, Hawkins K, Wicker E. The impact of mobility limitations on health outcomes among older adults. *Geriatric Nursing.* 2018;39(2):162-169.

197. Golden AG, Tewary S, Dang S, Roos BA. Care management's challenges and opportunities to reduce the rapid rehospitalization of frail community-dwelling older adults. *The Gerontologist.* 2010;50(4):451-458.
198. Coffey A, Leahy-Warren P, Savage E, et al. Interventions to promote early discharge and avoid inappropriate hospital (re)admission: a systematic review. *International J Environ Res Public Health.* 2019;16(14):2457.
199. Bowles KH, Ratcliffe S, Potashnik S, et al. Using electronic case summaries to elicit multi-disciplinary expert knowledge about referrals to post-acute care. *Appl Clin Inform.* 2016;7(2):368-379.
200. Bettencourt-Silva JH, Mulligan N, Sbodio M, et al. Discovering new social determinants of health concepts from unstructured data: framework and evaluation. *Stud Health Technol Inform.* 2020;270:173-177.
201. Demner-Fushman D, Chapman WW, McDonald CJ. What can natural language processing do for clinical decision support? *J Biomed Inform.* 2009;42(5):760-772.
202. Brennan PF, Strombom I. Improving health care by understanding patient preferences: the role of computer technology. *Journal of the American Medical Informatics Association.* 1998;5(3):257-262.
203. Sefcik JS, Ritter AZ, Flores EJ, et al. Why older adults may decline offers of post-acute care services: a qualitative descriptive study. *Geriatr Nurs.* 2017;38(3):238-243.
204. Dorr D, Bejan CA, Pizzimenti C, Singh S, Storer M, Quinones A. Identifying patients with significant problems related to social determinants of health with natural language processing. *Stud Health Technol Inform.* 2019;264:1456-1457.
205. Gupta D, Saul M, Gilbertson J. Evaluation of a deidentification (De-Id) software engine to share pathology reports and clinical documents for research. *Am J Clin Pathol.* 2004;121(2):176-186.
206. Dinescu A, Korc-Grodzicki B, Farber J, Ross JS. Discharge disposition disagreements and re-admission risk among older adults: a retrospective cohort study. *BMJ open.* 2012;2(6):e001646.
207. Healthy People 2020. Access to Health Services. 2020; https://www.healthypeople.gov/2020/topics-objectives/topic/Access-to-Health-Services (accessed June 2021).
208. Bowles KH, McCorkle R, Nuamah IF. Homecare referrals and 12-week outcomes following surgery for cancer. *Oncol Nurs Forum.* 2008;35(3):377-383.
209. Topaz M, Murga L, Gaddis KM, et al. Mining fall-related information in clinical notes: Comparison of rule-based and novel word embedding-based machine learning approaches. *J Biomed Inform.* 2019;90:103103.
210. Hripcsak G, Rothschild AS. Agreement, the f-measure, and reliability in information retrieval. *Journal of the American Medical Informatics Association.* 2005;12(3):296-298.
211. Van Rossum C, Drake FL. *Python 3 Reference Manual.* Scotts Valley, CA: CreateSpace; 2009.
212. Bird S, Loper E, Klein E. *Natural Language Processing with Python.* O'Reilly Media Inc.; 2009.
213. Fast E, Chen B, Bernstein MS. Empath: Understanding topic signals in large-scale text. *ACM Conference on Human Factors in Computing Systems 2016.* 2016.

214. Wilson T, Wiebe J, Hoffman P. Recognizing contextual polarity in phrase-level sentiment analysis. Paper presented at: *Proc. of HLT-EMNLP-20052005*.
215. Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. *Nucleic Acids Res.* 2004;1(D267-70).
216. Neumann M, King DR, Beltagy I, Ammar W. ScispaCy: Fast and robust models for biomedical natural langauge processing Paper presented at: *Proceedings of the 18th BioNLP Workshop and Shared Task*; Aug 2019, 2019; Florence, Italy.
217. Urbanowicz RJ, Suri P, Cui Y, et al. A rigorous machine learning analysis pipeline for biomedical binary classification: application in pancreatic cancer nested case-control studies with implications for bias assessments. *arXIV Preprint.* 2020;arXiv:2008.12829.
218. Pedregosa F, Varoquaux G, Gramfort A, al. e. Scikit-learn: machine learning in python. *J Mach Learn Res.* 2011;12:2825-2830.
219. Shi X, Mueller J, Erickson N, Li M, Smola A. Multimodal AutoML on structured tables with text fields. Paper presented at: 8th ICML Workshop on Automated Machine Learning (AutoML)2021.
220. Erickson N, al. e. AutoGluon-Tabular: robust and accurate AutoML for structured data. *arXiv preprint arXiv:200306505.* 2020.
221. Clark K, Luong MT, Le QV, Manning CD. ELECTRA: Pre-training text encoders as discriminators rather than generators. *arXIV Preprint.* 2020.
222. Carroll A, Dowling M. Discharge planning: communication, education and patient participation. *Br J Nurs.* 2007;16(14):882-886.
223. Tran TH, Swoboda H, Perticone K, et al. The substance use intervention team: a hospital-based intervention and outpatient clinic to improve care for patients with substance use disorders. *Am J Health Syst Pharm.* 2021;78(4):345-353.
224. Chen AH, Chin MH. What if the role of healthcare was to maximize health? *J Gen Intern Med.* 2020;35(6):1884-1886.
225. Siconolfi D, Shih RA, Friedman EM, et al. Rural-urban disparities in access to home- and community-based services and supports: stakeholder perspectives from 14 states. *J Am Med Dir Assoc.* 2019;20(4):503-508.e501.
226. Bowman S. Impact of electronic health record systems on information integrity: quality and safety implications. *Perspect Health Inf Manag.* 2013;10(Fall):1c-1c.
227. Kim MC, Nam S, Wang F, Zhu Y. Mapping scientific landscapes in UMLS research: a scientometric review. *Journal of the American Medical Informatics Association.* 2020;27(10):1612-1624.
228. Li Y, Wang Z, Ding B, Zhang C. AutoML: A Perspective where industry meets academy. proceedings of the 27th *ACM SIGKDD conference on knowledge discovery; Data Mining*; 2021; Virtual Event, Singapore.
229. Amann J, Blasimme A, Vayena E, Frey D, Madai VI. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. *BMC Med Inform Decis Mak.* 2020;20(1):310.
230. Khairat S, Marc D, Crosby W, Al Sanousi A. Reasons For physicians not adopting clinical decision support systems: critical analysis. *JMIR Med Inform.* 2018;6(2):e24.
231. Kennedy EE, Bowles KH. Human factors considerations in transitions in care clinical decision support system implementation studies. *AMIA Annual Symposium Proceedings*;2021.

232. Collins S, Couture B, Kang MJ, et al. Quantifying and visualizing nursing flowsheet documentation burden in acute and critical care. Paper presented at: *AMIA Annual Symposium Proceedings*; 2018.
233. Office of the National Coordinator for Health Information Technology. Shared Nationwide Interoperability Roadmap: The Journey to Better Health and Care. 2019; https://www.healthit.gov/infographic/shared-nationwide-interoperability-roadmap-journey-better-health-and-care (accessed March 17, 2022).
234. Office of the National Coordinator for Health Information Technology. 10-Year Vision to Achieve an Interoperable Health IT Infrastructure. https://www.healthit.gov/sites/default/files/ONC10yearInteroperabilityConceptPaper.ppd (accessed March 17, 2022).
235. Glasgow RE, Harden SM, Gaglio B, et al. RE-AIM planning and evaluation framework: adapting to new science and practice with a 20-Year review. *Front in public health.* 2019;7:64-64.
236. Feldstein AC, Glasgow RE. A practical, robust implementation and sustainability model (PRISM) for integrating research findings into practice. *Jt Comm J Qual Patient Saf.* 2008;34(4):228-243.
237. Mishra A, McClelland RL, Inoue LYT, Kerr KF. Recalibration methods for improved clinical utility of risk scores. *Med Decis Making.* 2021:272989x211044697.
238. Byrne C, Sherry D, Mercincavage L, Johnston D, Pan E, Schiff G. *Advancing Clinical Decision Support: Key Lessons in Clinical Decision Support Implementation.*
239. Centers Disease Control and Prevention. Implementing Clinical Decision Support Systems. *Best Practice Guides* 2021; https://www.cdc.gov/dhdsp/pubs/guides/best-practices/clinical-decision-support.htm (accessed March 17, 2022).
240. Tcheng JE, Bakken S, Bates DW, et al. Optimizing Strategies for Clinical Decision Support: Summary of a Meeting Series. *Learning Health System Series* 2017; https://www.healthit.gov/sites/default/files/page/2018-04/Optimizing_Strategies_508.pdf (Accessed March 17, 2022).
241. Gesner E, Gazarian P, Dykes P. The burden and burnout in documenting patient care: an integrative literature review. *MEDINFO 2019: Health and Wellbeing e-Networks for All.* 2019:1194-1198.
242. Wan PK, Satybaldy A, Huang L, Holtskog H, Nowostawski M. Reducing alert fatigue by sharing low-level alerts with patients and enhancing collaborative decision making using blockchain technology: scoping review and proposed framework (MedAlert). *JMIR.* 2020;22(10):e22013.
243. Joseph AL, Borycki EM, Kushniruk AW. Alert fatigue and errors caused by technology: A scoping review and introduction to the flow of cognitive processing model. *Knowledge Management & E-Learning: An International Journal.* 2021;13(4):500-521.
244. Kesselheim AS, Cresswell K, Phansalkar S, Bates DW, Sheikh A. Clinical decision support systems could be modified to reduce 'alert fatigue'while still minimizing the risk of litigation. *Health Affairs.* 2011;30(12):2310-2317.
245. Khalifa M, Zabani I. Improving utilization of clinical decision support systems by reducing alert fatigue: strategies and recommendations. Paper presented at: *ICIMTH2016*.

246. Fox MT, Persaud M, Maimets I, Brooks D, O'Brien K, Tregunno D. Effectiveness of early discharge planning in acutely ill or injured hospitalized older adults: a systematic review and meta-analysis. *BMC Geriatr.* 2013;13(1):70.
247. Murphy EV. Clinical decision support: effectiveness in improving quality processes and clinical outcomes and factors that may influence success. *Yale J Biol Med.* 2014;87(2):187-197.
248. Kim H, Grunditz JI, Meath THA, Quiñones AR, Ibrahim SA, McConnell KJ. Accuracy of hospital discharge codes in medicare claims for knee and hip replacement patients. *Medical care.* 2020;58(5):491-495.
249. Joo JY, Huber DL. Case management effectiveness on health care utilization outcomes: A systematic review of reviews. *West J Nurs Res.* 2018;41(1):111-133.
250. Health Resources and Services Administration. Office of Health Equity. 2020; https://www.hrsa.gov/about/organization/bureaus/ohe/index.html (Accessed March 18, 2022).
251. Centers Disease Control and Prevention. Health Equity. 2022; https://www.cdc.gov/healthequity/index.html (accessed March 18, 2022).
252. Arons A, DeSilvey S, Fichtenberg C, Gottlieb L. Documenting social determinants of health-related clinical activities using standardized medical vocabularies. *JAMIA Open.* 2018;2(1):81-88.