ELECTRONIC HEALTH RECORD (EHR) DATA QUALITY AND TYPE 2 DIABETES MELLITUS CARE

Kevin Keith Wiley, Jr.

Submitted to the faculty of the University Graduate School
in partial fulfillment of the requirements
for the degree
Doctor of Philosophy
in the Richard M. Fairbanks School of Public Health,
Indiana University

June 2022

Accepted by the Graduate Faculty of Indiana University, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Doctoral Committee

Joshua Vest, PhD, MPH, Chair

Justin Blackburn, PhD

April 26, 2022

Mary De Groot, PhD

Nir Menachemi, PhD, MPH

Eneida Mendonca, MD, PhD

DEDICATION

I dedicate this work to my wonderful wife, Michele and daughter, Olivia. Without your encouragement, time, and the peace you have brought to my life, none of this work would have been possible. I also dedicate this work to my parents, Shelia and Kevin Wiley, Sr. May we continue to revel in the joys of our successes.

ACKNOWLEDGEMENTS

None of this work and the successes that I have enjoyed while completing it would have been possible without many mentors, friends, colleagues, coaches, and family supporting my journey.

First and foremost, I want to thank my many mentors spanning multiple institutions and organizations. Dr. Joshua Vest, you have served as a long-time mentor and who has been instrumental to my growth as a health services researcher, leader, and father. You have pushed me to pursue my scientific interests thoughtfully, critically, and more clearly for the past 10+ years. I am a better health services and informatics researcher, communicator, and instructor because of your training and mentorship. I am forever grateful for your accessibility and confidence throughout my career.

Dr. Justin Blackburn, working with you has been incredibly rewarding as a student and research mentee. I am grateful that you took time to extend your methodological expertise in a way that is empathetic and digestible. Additionally, your service as program director has been important in ensuring that my colleagues and I avoided common pitfalls throughout the PhD journey. Your perspective, approach, and dedication to service and research are invaluable.

Drs. Brian Dixon and Titus Schleyer, your mentorship through the T15 training program has been immensely critical to my development as a health services and population health informatics scholar. Being a part of the T15 training program is one of my proudest achievements. The approaches I have learned as a trainee have enabled me to explore important research questions that I will carry forward in my career.

Dr. Chris Harle, your honesty helped improve my own decision making as a burgeoning independent scholar. Although we spent a short time together in Indiana, I look forward to opportunities to learn from you and the impactful work you are doing.

Dr. Nir Menachemi, it would be difficult to capture the extent to which your mentorship has impacted my professional and academic life. I am grateful to have been able to learn and glean many important lessons from you: about research design, research writing, professional guidance, and managing the responsibility to create new knowledge.

Dr. Peter Jenkins, I have always been proud of the work we began and excited about the research we will conduct in the future. Our common experiences and research interests to improve care quality made our mentor/mentee relationship an immensely valuable one.

Dr. Lydia Isaac, there was never a time you denied your students help or support at moments of crisis. My time in the Health Policy Research Scholars (HPRS) program under your leadership made me reflect on racial and ethnic representation in the Health Policy and Management discipline. Our interactions were major factors in my decision to pursue a career in academia.

Dr. Keshia Pollack Porter, I am grateful and constantly awestruck by your accessibility and willingness to provide scholarly support to your mentees. You and Dr. Shannon Frattaroli are an inspiration to me and many scholars in the Health Policy and Management discipline. I enjoyed every opportunity I have had to learn from you and other mentors/leaders in HPRS.

Laurie Unruh, thank you for your professional support through the HPRS program. You were ever-present at various events on behalf of a team who ensured HPRS representation and support at critical moments.

Fellow Health Policy Research Scholars (HPRS), I have become a better researcher for knowing all of you and the important work you do in your respective fields of study. I am especially appreciative of my fellow Cohort 3 members who have been an integral part of my development. Demar, Sheridan, Jason, Tiana, those writing and reading accountability groups continue to be a staple in my preparation regime. Additionally, I am glad we were able to spend time learning and sharing our milestones throughout our doctoral journeys. Stephanie, Sireen, Angela, Cynthia, Kevin, Kathy, Kristy, Michael, and Ben, I am a better scholar, friend, and human for knowing you.

Colleagues at the National Committee for Quality Assurance, I am grateful for the work you have entrusted me to lead. My training as a Phyllis Torda Health Care Quality and Equity Fellow pushed me to be thoughtful about how I pose and explore research questions and to ensure results are accessible and communicated in a clear, effective way to the public.

Professors and mentors in the Indiana University Fairbanks School of Public Health and Center for Research on Inclusion and Social Policy (CRISP), I thank you for all your support and training. Ann, you challenged me to be more deliberate about aligning research questions with mechanics and methods that improved my approach to scientific inquiry in ways I could not imagine. Valerie, I have applied many of the qualitative research sessions to my current work and am grateful for your courses. Roxy

and Breanca, you showed me how to effectively conduct community-engaged research and evaluation.

Brittany, Joanna, Elizabeth D, Elizabeth G, Carol, Gary, Wei Wei, Abdulaziz, Katie, Robin, you have challenged me intellectually throughout the past few years. I have learned a lot from each one of you through laughs, tears, and many puns.

Nate, you have been an incomparable professional and academic peer mentor. I am grateful for the many chats, runs, puns, and other jokes we have shared. Your friendship means the world to me and my family. It has been a joy working with you.

Casey, it has been a pleasure learning from you. There is no one I would rather struggle through analysis or technical writing with, dear friend. I will make sure to maintain a consistent stream of puns and memes to keep our collective laughs going.

Heather, there is no chance that I could have completed doctoral studies without your support and friendship. When we started, we were both cheerful and ready to improve society's gravest ills. While we still have maintained that excitement, we are better prepared and more cautious to deploy it. Together, we realized the importance of tempering the eagerness of our scholarly pursuits and ambition with the realization that it takes time. Namely, time spent away from our families. I am grateful for that lesson and many others we learned through much frustration and anger. There is no one in this galaxy that I would rather be paired with if I had to do this again. I am proud of the work we have done and am excited about the next phase of your career.

It would have been impossible to complete my doctoral studies without the financial, professional, and academic support of the National Library of Medicine (NLM)/National Institutes of Health (NIH) Pre-Doctoral Training Program at Indiana

University and the Regenstrief Institute, Inc., the Robert Wood Johnson Foundation's (RWJF) Health Policy Research Scholars Program, Graduate Student Emissary Program, Center for Research on Inclusion and Social Policy (CRISP) at the Public Policy Institute, and National Committee for Quality Assurance (NCQA).

To my siblings and friends: I am who I am because of your love and support. You have all improved my life in real and substantive ways.

Finally, to Michele and Olivia: it takes more than words to describe my appreciation for both of you.

Kevin Keith Wiley, Jr.

ELECTRONIC HEALTH RECORD (EHR) DATA QUALITY AND TYPE 2 DIABETES MELLITUS CARE

Due to frequent utilization, high costs, high prevalence, and negative health outcomes, the care of patients managing type 2 diabetes mellitus (T2DM) remains an important focus for providers, payers, and policymakers. The challenges of care delivery, including care fragmentation, reliance on patient self-management behaviors, adherence to care management plans, and frequent medical visits are well-documented in the literature. T2DM management produces numerous clinical data points in the electronic health record (EHR) including laboratory test values and self-reported behaviors. Recency or absence of these data may limit providers' ability to make effective treatment decisions for care management. Increasingly, the context in which these data are being generated is changing. Specifically, telehealth usage is increasing. Adoption and use of telehealth for outpatient care is part of a broader trend to provide care at-a-distance, which was further accelerated by the COVID-19 pandemic. Despite unknown implications for patients managing T2DM, providers are increasingly using telehealth tools to complement traditional disease management programs and have adapted documentation practices for virtual care settings. Evidence suggests the quality of data documented during telehealth visits differs from that which is documented during traditional in-person visits. EHR data of differential quality could have cascading negative effects on patient healthcare outcomes.

The purpose of this dissertation is to examine whether and to what extent levels of EHR data quality are associated with healthcare outcomes and if EHR data quality is

improved by using health information technologies. This dissertation includes three studies: 1) a cross-sectional analysis that quantifies the extent to which EHR data are timely, complete, and uniform among patients managing T2DM with and without a history of telehealth use; 2) a panel analysis to examine associations between primary care laboratory test ages (timeliness) and subsequent inpatient hospitalizations and emergency department admissions; and 3) a panel analysis to examine associations between patient portal use and EHR data timeliness.

Joshua Vest, PhD, MPH, Chair

Justin Blackburn, PhD, MS

Mary De Groot, PhD

Nir Menachemi, PhD, MPH

Eneida Mendonca, MD, PhD

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ADA	American Diabetes Association
ANOVA	Analysis of Variance
CKD	Chronic Kidney Disease
CDC	Centers for Disease Control and Prevention
CDSS	Clinical Decision Support System
CMS	Centers for Medicare and Medicaid Services
COVID-19	Coronavirus Disease
EDW	Enterprise Data Warehouse
EHR	Electronic Health Record
EMR	Electronic Medical Record
F2F	Face-to-face
HHS	Department of Health and Human Services
HIE	Health Information Exchange
HIPAA	Health Information Portability and Accessibility Act of 1996
HITECH	Health Information Technology for Economic and Clinical Health Act
ICD	International Classification of Diseases
INPC	Indiana Network for Primary Care
IS	Information Systems
MU	Meaningful Use
NAM	National Academy of Medicine
NLP	Natural Language Processing
ONC	Office of the National Coordinator for Health Information Technology

T2DM	Type 2 Diabetes Mellitus

CHAPTER 1: INTRODUCTION

Assessing Electronic Health Record (EHR) Data Quality for Reuse

The Health Information Technology for Economic and Clinical Health (HITECH) Act of 2009 brought about widespread adoption and use of electronic health records (EHR) among US health care providers.[1] Since HITECH's passage, health care organizations and providers have documented and amassed volumes of patient health data in various forms, including prescription data, laboratory data, genetic information, billing, administrative, and demographic data.[2–4] These patient health data are primarily documented in electronic health records (EHR) or electronic medical records (EMR). The electronic medical record (EMR) historically includes digital patient information and is contained within the confines of a provider's office.[5] EHRs are similar, but they include a broader set of patient information that constitutes medical and non-medical data from a range of clinical providers.[6,7] Clinical providers document patient information in both the EMR and EHR which can be manually or electronically extracted. Data are collected to support clinical and administrative functions (e.g., billing) including data observed in clinical exams, data obtained from laboratory and diagnostic tests, data from monitoring devices, and patient-reported items.[8–10]

In addition to monitoring patient health over time, clinical documentation supports administrative claims processing and billing for care provided in medical and health care settings.[11–13] The Health Information Portability and Accessibility Act (HIPAA) of 1996 preceded HITECH and ensured that patient health information was securely maintained as a result of clinical documentation for record-keeping and billing.[14] These data were not commonly used to evaluate cost, access, or quality of care until

Meaningful Use (MU) provisions of the HITECH Act provided incentives to develop ways to streamline documentation and data collection to improve patient health outcomes.[15] As a result, hospitals and providers have adopted EHRs at increasing rates since the passage of HITECH.[1,10] Clinical data proliferated in recent years providing an opportunity to conduct empirical studies (observational, quasi-experimental, and population health research), program evaluations, and develop clinical quality improvement programs.[16] Despite the increasing use of clinical data for examining the organizational, patient, and larger health systems outcomes, data derived from the EHR have historically had varying levels of quality; namely, completeness, and timeliness.[17–19] Prior literature suggests that these three dimensions are fundamental to assessing the quality of clinical data and is supported by the National Academy of Medicine (NAM).[20–22]

Clinical data sources are commonly found in one of three data formats: semi-structured, structured, and unstructured.[23–27] Semi-structured and structured clinical data include simple dropdowns using menu-based, interactive and searchable forms. Whereas, unstructured data include documents constituting free text, namely clinical narrative like progress or discharge notes.[23,28,29] Notably, unstructured data is not organized in any specific manner, and it includes critical yet un-synthesized information to patient health.[26] Additionally, unstructured data are the most voluminous source comprising nearly 80 percent of all health care data.[26]

Researchers have relatively recently begun investigating the quality of EMR and EHR data quality to ensure sufficiency for reuse in health services research.[16,17,21,27,30] Clinical data have also been found to be important for quality improvement[16,31], program evaluation[32,33], and empirical research.[16,31–33] Clinical data quality varies by type and content, including laboratory, medication, or narrative reports, which may diminish usability for program quality improvement and evaluation.[27,31] More importantly, these data are of variable quality which can result in mismeasurement and erroneous outcomes should they be used in empirical research.[34]

As data captured in the EHR grows in volume and variety, the general quality of clinical data also varies and is largely dependent on provider and organizational documentation preferences and the complexity of patient conditions.[35] Data quality is described broadly as fitness for use.[20,36] In the context of the current study, we determine fitness for use as reusability of clinical data in empirical research, program evaluation, and quality improvement. Importantly, reuse of clinical data and the assessments thereof are dependent on each specific case.[20] For example, past research determined that more than half of all coded patient data were missing for serious cancer conditions and many of the variables were summarily excluded from analysis as a result.[17] In two survival analyses, researchers manually inferred missing values to model methodological approaches.[17,37] The content and quality of clinical data has been found to be associated with some health care outcomes.[37–40] Prior research has shown that available clinical data is associated with care quality[38], disrupted referral patterns[41–43], prescribing behavior[44], and diagnostic accuracy[37].

Generally, data quality assessments are conducted on data that were documented as a result of in-person or face-to-face (F2F) visits.[45] Few studies have examined the quality of patient data resulting from telehealth visits nor have they compared attendant outcomes to F2F visits.[45,46] Moreover, what constitutes clinical data and information among these modalities may differ slightly and include a range of data domains important for improving patient and population health outcomes. This is especially true for patients managing chronic disease. Furthermore, health care for patients at greater risk of chronic disease, like type 2 diabetes (T2DM), present complex challenges for effectively collecting important patient information for use in developing care management programs and medical decision-making.

Measuring EHR Data Quality

Standardized clinical documentation remains a challenge as health care organizations adapt documentation practices and standards to better address disease classification[47], treat more patients with complex conditions[48], and utilize new care modalities during a pandemic[49–51]. Recently, clinicians began providing most of their care virtually to avoid overwhelmed systems and prevent disease transmission due to COVID-19.[52,53] Additionally, providers are expected to begin transitioning from previous *International Classification of Disease (ICD)* version 10 to 11 by January 1, 2022.[47] For example, some health care organizations have experienced difficult transitions from ICD-9 to ICD-10 due to complications and cost which may compromise data integrity for both in-person and telehealth documentation.[54] This may become worse as telehealth access is expanded to care for COVID-19 patients and patients with delayed health care needs as a result of the pandemic. Moreover, an evolving body of research has examined provider burnout as a result of excess documentation and EHR screen viewing and fatigue.[55–58] The resulting challenges and variations in clinical documentation compromise the integrity of clinical data quality; namely, inaccuracies, incompleteness, loss of recency, and other errors. Additionally, disruptions to clinical workflows for providers and staff that may have worsened documentation procedures thereby affecting patient care coordination.[59] Varying quality of clinical data may have downstream impacts on clinical decisions and outcomes[38,60], as well as social and economic effects.[61,62]

Clinical data extracted for health services evaluation and research should be “fit for use” and require higher quality to withstand processing errors.[20,61] In the clinical setting, data are used in clinical decision support systems (CDSS) as a result of admit-discharge-transfer feeds[63,64], remote-patient monitoring and home health[46], and health information exchange[63]. How these data are entered into various EHR systems and the constraints posed by the care modality or format, virtual vs. in-person care, determine overall completeness, and timeliness.[20] Nahm (2012) and Sebastian-Coleman (2010) outlined three main phases of data recording, processing, and analysis that may, at each phase, impact information and data quality.[20,65,66] The current study is concerned with the latter phase of assessing and measuring EHR data quality. Chiefly, measurement of data quality dimensions. Additionally, this study will examine patient health care consequences of varying levels of clinical data quality measures and to what extent correcting data to be “fit for use” improves information quality and subsequent patient outcomes. Examining the integrity of clinical information and the difficulty measuring data quality dimensions require separate examination.

First, how data are recorded or documented is an important determinant of whether the data will be accurate, timely, and complete. Variation exists among clinical providers who subscribe to different documentation patterns brought about through organizational and professional association mandates, which may have legal and practical implications.[67,68] Second, data processing and management methods may be informed by data collection practices, complexity, the conceptualization of data use, and experience of the individual processor.[20,66] Importantly, data quality measures are affected by a high level of user subjectivity.[69–71] Lastly, challenges in the analysis are defined by misrepresentations of data for the final purpose which may include reuse in reporting, quality improvement, or empirical research.[31] Thus, researchers have developed methods to effectively audit clinical data quality prior to reuse, irrespective its final purpose.[19,21,39,62] Data quality dimensions described in past research to assess data quality largely overlap across existing concepts, measures, and methods.[21] These concepts, measures, and methods are applied across disciplines and industries but have recently gained importance in health services and clinical research. Namely, the adoption and use of EHR-based data collection methods underscore the importance of examining variation in clinical data and information quality in telehealth and in-person care for patients managing chronic disease.

To understand and operationalize measures of data quality, past research proposed using a data quality framework summarized in a systematic review by Weiskopf et al (2013) and, separately conceptual frameworks developed by Wang & Strong (1996) and Kahn et al. (2012).[21,61,72] Specific measures are quantified using quotients and metrics designed by Hinrichs (2002) and Heinrich et al. (2007) to assess data quality in information and business systems.[69,73] For example, researchers used the Levenshtein distance to compute correctness for word strings. (See Equation I) Simply, the Levenshtein distance is the total number of edits to an attribute value, most commonly a word string, required to match the real-world value, divided by the letter length of the real word value.

$$\text{"Patent"}, \text{"Patient"} = 1 - \frac{3}{7} = 57.2\%$$ [69] (Equation 1. Levenshtein's Distance)

Prior research used data quality dimensions to define, outline, and quantify measures of general information quality. Although these studies examined data quality generally, there have been few, if any, efforts to 1) compare EHR data quality measures of telehealth and in-person care for patients managing chronic disease, or 2) examine associations between patient-level data quality measures and health care outcomes. Specifically, this study relies on overlapping definitions of data quality identified by Weiskopf et al (2013) and the National Academy of Medicine (NAM) and includes the following (Table 1)[21,61]:

Data Quality Dimension	Dimension definition/measure
Availability of data	Overall presence of data for specific disease (T2DM, hypertension) indicators
Concordance	Represents agreement between elements in the EHR and another clinical dataset
Correctness	Represents elements in the EHR that are true
Completeness	Represents elements in the EHR that are true based on other patient data
Information Density	Information density scores are measures of completeness that account for the irregular nature of patient measurements taken over time
Timeliness	Represents EHR data for patient health status at a given point in time
Plausibility	Represents elements in the EHR that make sense in light of other knowledge about what that element is measuring

Table 1**.** Data Quality Dimensions [21,74,75]

Concordance, completeness, and timeliness represent measures for comparison to existing data sets that serve as relative gold standards (e.g., disease registries). Other studies have included the extent to which data for the quality audit use case exists. We define availability of data as the overall presence of data by T2DM phenotype. This study will stratify the extent to which patients are seen via telehealth vs. in-person, to classify visits among T2DM phenotypes, and to determine visits among specialists and facilities.

Data Quality Dimensions, Measurement, and the Information Systems Success Model

Data quality dimensions identified by Weiskopf et al. (2013) inform the overall framework used to classify and measure data quality as a result of clinical documentation. Measurement approaches were validated across studies in a systematic review conducted by Weiskopf et al. (2013) and NAM.[21] Final data quality dimensions mutually identified by investigators include completeness, concordance, and timeliness.[13,20,38,76] An additional measure, availability of data, was added to determine the differences in data quality stratified by care modalities (i.e., telehealth and in-person diabetes care), patient characteristics, clinical provider characteristics, organizational characteristics, and general T2DM phenotypes. The current study relies on these dimensions and their definitions (Table 1) to create measures using coded diagnoses, laboratory tests, and narrative texts. These data domains form the basis of measurement for the overall study.

In addition to using Weiskopf et al. (2013) to develop relative data quality measurements, we adapt the Information System Success Models (Figure 1) to operationalize patient- and encounter-level EHR data quality dimensions.[65,77] Specifically, we orient the IS Success Model to include its constructs as independent variables that explain IS Success as organizational and individual impact (Figure 2).[78] Petter et al. (2013) defined independent variables that were related to tasks, general system characteristics, and user and social factors.[65] Information and system quality indicators inform system usage which are modeled as determinants of organizational and individual impact.[65,78] Information quality is synonymous with data quality in this flowchart (Figure 1). The organizational and individual impact is defined as inpatient hospitalizations and emergency department visits. These outcomes affect healthcare organizations and the patients they serve by creating care inefficiencies and compromising the overall quality and coordination of critical care.

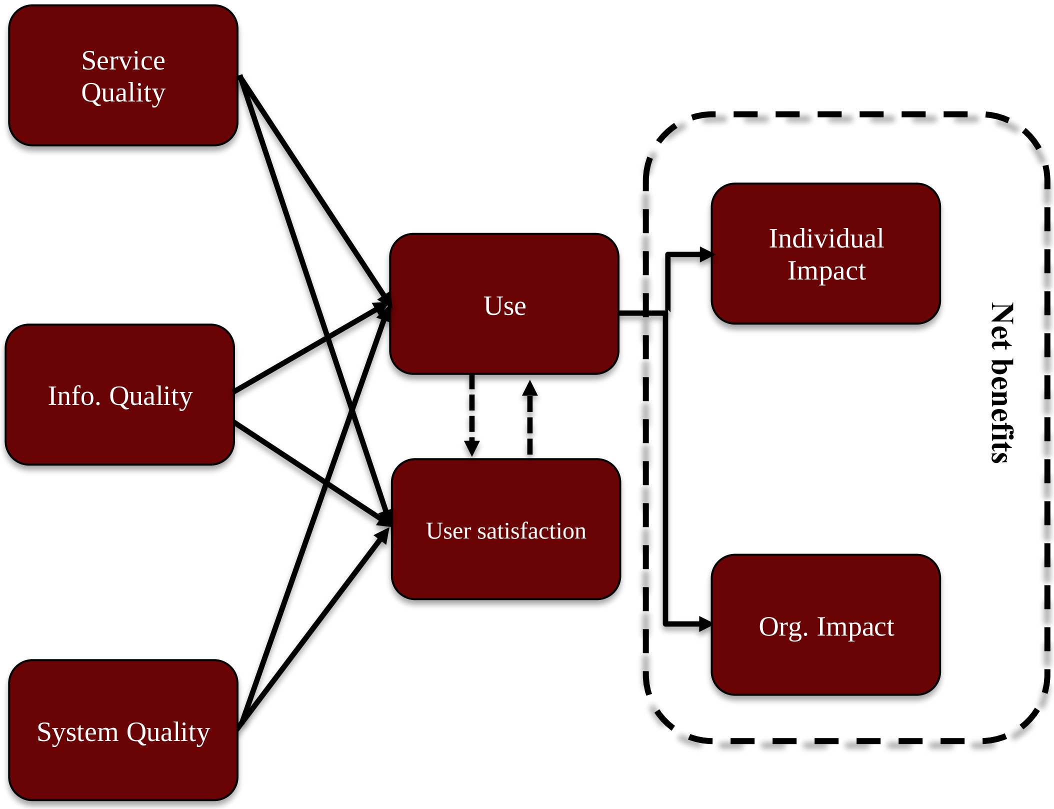


Figure 1. Updated Petter, DeLone, and McLean Information Systems Success Model

Relative Gold Standards for Comparing the Quality of Clinical Datasets & Sources

Comparisons derived in data quality audits are often incomplete without using a second institutional dataset as a relative gold standard for validation purposes. Relative gold standards are critical data sources that allow the use of an internal or external data source to assess clinical data quality in another data source.[19] Gold standard datasets are considered relative because there are no truly accurate or perfect datasets to measure against for data quality comparisons. The primary data source may be derived externally but are generally internal and linkable based on patient or other identifiers. In the current study, we use EHR data that is sourced for chronic disease registry databases, namely, T2DM. Natural language processing (NLP) is commonly used to extract narrative text from clinical documents and subsequently compare data quality dimensions and data availability across datasets. However, validating NLP is limited by its reliance on rare gold standard corpora derived from clinical notes (i.e., discharge and progress notes).[79] Registry data is considered a relative gold standard due to its function in administrative and operational processes that go beyond clinical use (e.g., billing).[80] Importantly, key coded diagnoses are highly reliable in these datasets because of the nature of prescribed treatments for conditions like cancer and chronic kidney disease.[19,40]

Type 2 Diabetes Mellitus Management

Diabetes Mellitus is a chronic condition that represents the sixth leading cause of death in the US. Persons living with diabetes currently constitute approximately 9 percent of the US population, but 25 percent of all hospitalizations.[81] Type 2 Diabetes Mellitus is the most common type of diabetes in the US and can be treated with diet, exercise, prescription medication, and insulin.[82] Costs associated with diabetes hospitalizations total approximately $124 billion, where $25 billion were attributable to readmissions.[83] The current estimate of age-gender weighted average lifetime medical costs for T2DM was $85,200. Approximately 53% was due to treating diabetes complications.[84] The cost of managing macrovascular complications accounted for 57% of the total complication cost. Health care providers have undergone quality improvement and program development efforts to optimize T2DM care, yet diabetes management is still subpar.[85]

Patients with T2DM require care management plans that incorporate multiple provider specialties across diverse clinical settings to effectively treat comorbid conditions, including hypertension, hyperlipidemia, chronic kidney disease (CKD), and obesity.[86,87] Additionally, care management for persons living with diabetes relies on data-rich environments from clinical data sources collected by documentation, remote monitoring and patient-reported health.[85,88] Moreover, diabetes care management programs are supported by telehealth technologies and patient health portals.[9,85,89] Although prior research suggests that patients managing T2DM utilize care at high rates, resulting in expensive, fragmented management plans, telehealth programs serve to improve care coordination and patient health outcomes.[83,87,90] Telehealth programs targeting persons living with diabetes seek to provide education and wellness classes, peer-to-peer support, and manage care through both synchronous and asynchronous telehealth technologies (i.e., remote patient monitoring, follow-up care).[85,91]

Patient Portal Use and Type 2 Diabetes Mellitus Care

Patient portals provide important linkages to health information and clinical providers.[92] These connections enable patients to manage their own care by having access to their own information including laboratory test results, visit summaries, secure messaging, and medication orders as well as a clinician to aid in interpretation.[92] Accessible and shareable patient information are important features of T2DM care management. Successful T2DM management relies on team-based care and shared information to effectively coordinate care processes among health care providers. This is especially important given patients with T2DM struggle with other comorbid conditions including hypertension, hyperlipidemia, cardiovascular disease, and obesity.[93–96] Diabetes is also accompanied by mental health conditions (e.g., major depression, depressive symptoms, anxiety, and diabetes distress) that may make care management more difficult.[10] Informatics researchers have developed tools within the patient health portal to accommodate chronic disease management and other concomitant conditions.[97–99]

Researchers have acknowledged that patient engagement issues and clinical inertia inhibit patient portal access and use.[100–102] These deficiencies cause downstream process of care outcomes which may disproportionately affect patients managing T2DM and other comorbid conditions. For example, persons living with diabetes who are more active and engaged in portal use have the greatest likelihood of glycemic control.[80,103,104] Similarly, the use of the EHRs, often where patient portal information is derived, among providers and persons living with diabetes is associated with fewer hospitalizations and ED visits.[10] Care coordination is informed by patient data which is used by both patients and providers to effectively manage diabetes care. Increasing the use of patient portals is critical for improving T2DM care management and outcomes.

Gaps in the Literature

Past research has struggled to identify and operationalize measures of data quality using electronic health record (EHR) data.[17,39] Rather, investigators rely on a broad set of data quality dimensions that require available indicators that are task or research specific.[21,61,69] Complications arise in the variety of quality issues apparent across multisite, system, and nationwide EHR samples.[37,105] Quality issues are found in all clinical data sources due to the nature of how these data are collected.[106,107] Data are populated by providers relying on electronic systems and subject to professional and organizational guidelines which can be formulaic.[108] Guidelines may be more stringent in an emergency like the COVID-19 pandemic.[109] Fewer still are those studies that examine the effect of patient- and encounter-level data quality dimensions or measures on patient healthcare outcomes.[37,38] The difficulty of measuring and examining relationships between patient-level data quality measures lies in expected missingness in clinical data and the methodological adjustments that can be made to improve data integrity and subsequent outcomes[44,45].

Missing and older data can be attributed to provider documentation attributes, patient demographic data availability, and dissimilar patient characteristics.[55,106] This is one of a few studies to examine whether patient-level data quality indicators influence healthcare outcomes for patients managingT2DM. Similar research used predictive survival models to determine the relationship between clinical data quality and patient care outcomes.[40] Lastly, patient portal use mitigates communication challenges by empowering patient activation and engagement to manage their own health.[97,110–114] These features of self-management are entirely dependent on the accessibility and quality of patient data for providers across the care spectrum.[92,103,104]

Dissertation Overview

The proposed dissertation combines three related approaches to examine whether a relationship exists between patient and encounter-level measures of data quality and healthcare outcomes. Data from the electronic health record (EHR) are being used nearly exclusively to conduct health services and informatics research regardless of whether data has been audited for integrity or general quality.[17,20,21] Few studies have examined the influence EHR data has on care coordination and processes.[115,116] My overall research questions for this dissertation are 1) to what extent does digital health tools including patient portal and telehealth care improve measures of EHR data quality and 2) how does EHR data quality effect processes of care for patients with T2DM?

This dissertation and the studies therein rely on two related frameworks. First, we used data quality dimensions to determine how EHR data are measured based on a systematic review conducted by Weiskopf et al. (2013) and metrics frameworks in previous research.[17,39,69,73] These dimensions include a range of published data quality dimensions used to audit the quality of business, information technology, and other industry data sources.[20,61] Additionally, we operationalize data quality dimensions using quotients and computation developed by Heinrichs et al (2007) whereby timeliness had clearly defined and clinically interpretable equations for contextualizing results. Our analyses use EHR data to quantify measures of patient data quality generally, and in T2DM where care was provided via telehealth and in-person. We performed and validated a natural language processing (NLP) approach by creating a reference/gold standard using T2DM clinical narrative derived from patient charts (Appendix A-G).[102] Specifically, we construct a reference/gold standard for NLP evaluation by manually prescreening a randomly selected set of 400 clinical notes. Of these, approximately 80% will contain T2DM diagnoses, labs, and medications; approximately 20% will not contain any T2DM references and will serve as a test set in preliminary analyses.

The second framework is the Information Systems (IS) Success Model.[78,117] The IS Success model was originally developed to identify and measure dependent variables to determine whether an information system was successful in its intended outcomes.[78] Six variables were identified as measurements for information system success outcomes including system quality, information quality, system use, user satisfaction, organizational impact, and service quality. Hebert (2001) modified the IS Success Framework and its dependent variables for telehealth evaluation using Donabedian's structure-process-outcome framework.[77] Conversely, Petter et al. (2013) sought to identify interdependent variables that might bidirectionally influence information success outcomes.[65] Based on a systematic review, researchers found that tasks, people, technology, and structure explain socio-technical relationships between an IS and other aspects of the working environment.[65,118]

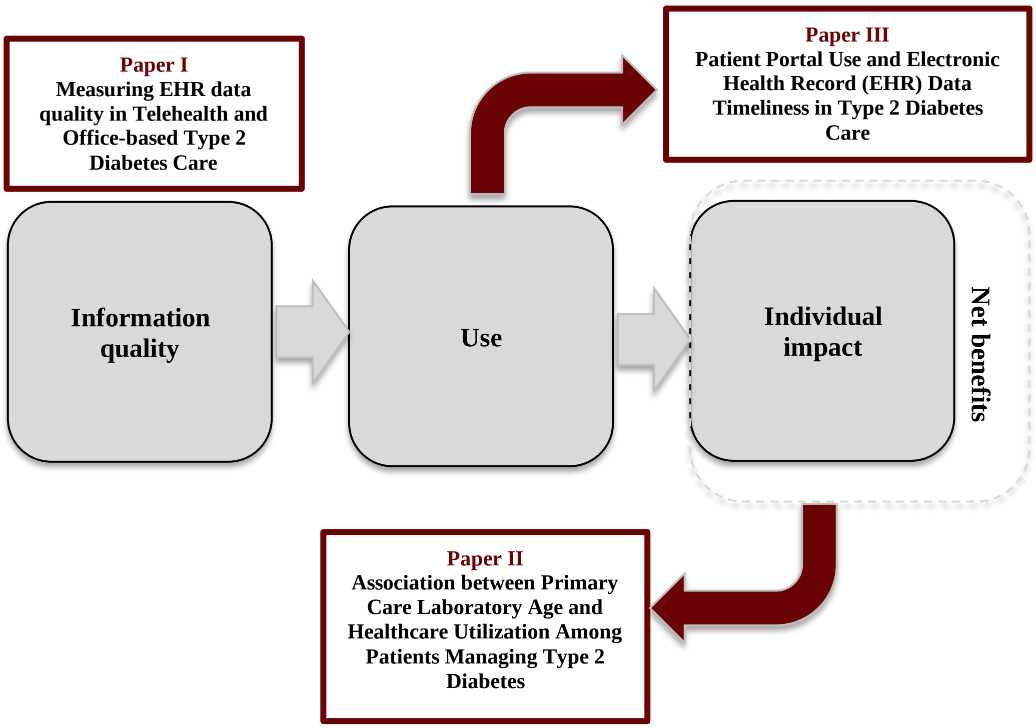


Figure 2. Modified Information Systems Success Model

This study relies on a modified version of the IS Success Model (Figure 2) to orient this overall dissertation. Data quality measurement in this study does not explicitly rely on empirical framing to determine antecedents of information quality. Rather, we posit that use of delivery modality (in-person vs. telehealth) influences or modifies the general quality of documented patient information, which we measure to examine its association with process of care outcomes.

CHAPTER 2: MEASURING ELECTRONIC HEALTH RECORD (EHR) DATA QUALITY IN TYPE 2 DIABETES MELLITUS CARE

INTRODUCTION

Electronic health records (EHR) are often a source of secondary data for use in clinical and health services research.[27,62] Nevertheless, such data may face numerous data quality challenges[119] from manual entry, lags between the patient visits and actual documentation, differential documentation guidelines, and changes in standards.[56,58] As a result, EHR data are commonly discordant and incomplete.[17,39] Extant research on EHR data quality is based primarily on traditional office-based, face-to-face visits between patients and providers. However, the extent to which EHR data quality is associated with other types of visits is not well known.[17,120,121]

Specifically, the proportion of remote or telehealth visits has steadily increased over the past 20 years[50,122,123] and became even more common during the COVID-19 pandemic.[124,125] While limited, evidence suggests the use of telehealth may have further effects on EHR data quality. For example, patients have expressed concerns about the accuracy and quality of data gathered during the encounter.[126] Likewise, providers and researchers note that telehealth visits may not include sufficient diagnostic data for providers to effectively manage patient symptoms.[127,128] Additionally, prior attempts to leverage data collected via telehealth for decision support systems were unsuccessful due to data quality issues.[46] In general, examining whether and to what extent new health care tools have improved or worsened EHR data quality is an understudied phenomenon.[129] Thus, a better understanding of different care modalities effects on data quality is needed.

Type 2 Diabetes Mellitus (T2DM) is an appropriate focal condition for such a comparison. First, opportunities for comparisons exist as T2DM self-management, nutrition consultations, and wellness programs are amenable to telehealth.[130,131] Second, T2DM is associated with a wide array of EHR data elements and types: laboratory measurements, medication regimens, self-management education, and nutrition consultations.[132] Patients routinely interact with multispecialty team-based clinicians during care management and care coordination as the standard of medical care for patients managing T2DM. Data collection across care settings and providers may increase data incompleteness because of a failure to consistently record relevant T2DM measurements.[133]

Objective

This study compared EHR data quality among patients managing T2DM with and without a history of telehealth use. We assessed structured EHR data quality in terms of timeliness, completeness, and patient information density. As telehealth tools become commonly used across care settings, the quality of EHR data is highly relevant to both the clinical care teams who rely on comprehensive patient information and those secondary data users seeking to inform research, quality measurement, population health management, and clinical guideline and policy development.

METHODS

Study Population

The study sample included patients managing T2DM aged ≥18 years who were seen between 2016 and 2021 at two health systems in central Indiana. The study population was defined as any patient who received a T2DM diagnosis between their index and latest encounter date. The study period includes uniform telehealth adoption, use, and documentation among both in health systems which began in 2020.

Data

The primary dataset used in the current study was derived from each health systems' enterprise data warehouse (EDW) and the Indiana Network for Patient Care (INPC). The INPC was established in 1994 as a repository for cross-institutional patient data including EHR data.[134] We obtained data representing patient demographics and clinical diagnoses inputted by treating clinicians regardless of specialty.[106]

Telehealth Status

Telehealth status was defined as any encounter where remote use of audio or video services were provided. Because of the potential selection bias associated with use of telehealth, we frequency matched controls on age, sex, and total visit count.

Clinical Visit Type

We identified patients with a history of telehealth use as any outpatient visit linkable to patients by common identifiers and was coded as having used audio or video technologies during a visit. Patients with no history of telehealth use were identified where there was no available indication of telehealth use or virtual care technologies present during or as a result of the visit.

EHR Data Quality Measures

We quantified measures of timeliness, completeness, and information density to assess the quality of EHR data for patients managing T2DM with and without a history of telehealth use. Each measure of quality was computed for data representing the following patient characteristics and relevant T2DM structured data elements and measurements: body mass index (BMI), body weight (lbs.), serum creatinine, glycated hemoglobin A1c, cholesterol, blood pressure, and smoking status.

Timeliness

Timeliness was defined as data elements that represent a patient's health state at a desired time of interest.[21,69] We operationalized timeliness as the number of days between a patient encounter and the most recent T2DM laboratory test and measurement dates.[135] Timeliness is measured at the encounter level for each patient in the study.

Completeness

Completeness is defined as the presence (the opposite of absence or missingness) of a measurement in the EHR.[21] We measured completeness using elements from the structured EHR data elements. T2DM measurements were identified and flagged as complete where a laboratory value or indicator was available at its first indication for sequential patient encounters. Completeness values ranged from 0 to 1 where a higher score indicated that data were more complete.

Information Density Score

Chronic disease care relies on repeated health care interactions and measurements of multiple indicators. Information density scores are measures of completeness that account for the irregular nature of patient measurements taken over time.[74,75] The information score, measured at the patient level, is expressed as the following:

$$I = \frac{2}{n} + \frac{n-2}{n}\left[1 - \sqrt{(n-1)\mathrm{Var}\{g_t; i = 1, \ldots, n-1\}}\right],$$

$$\text{where n is the number of } g_t = \frac{x_{i+1} - x_i}{x_n - x_1}$$

I is the average amount of information each observation provides for a patient observed n times. The information density score is a number between 0 and 1 where higher scores indicate that patient measurements were more equally distributed across patient visits.

Analysis

Frequencies, percentages, and means described EHR data quality measures by patient characteristics and over time. We compared data quality measures by patient history of telehealth use using two-tailed t-tests, χ^2 tests, and ANOVA statistical tests to examine the equality of means and relationships between patient characteristics. Correlation analyses examined the relationship between EHR data quality measures (Appendix 8).

RESULTS

Our final matched study population included 5,027 patients. Most clinical visits were for patients who were documented as African American (52.1%) and female (66.1%) (Table 3). The mean patient age in the sample was approximately 57.6 years.

Telehealth Use

Nearly 40% of patients in this sample had a history of telehealth visits (Table 3). Most telehealth visits constituted patients who were female, non-Hispanic Black, middle-aged (56-65), had Charlson scores of 1, and whose insurer was Medicare (Table 3). Clinical encounters in this sample mostly took place prior to the COVID-19 pandemic due to partial data availability for 2021.

EHR Data Timeliness

In terms of timeliness, across all patient encounters and data elements, the average data element was 77.8 (SD=63.8) days old. On average, data were timelier, i.e., fewer days between EHR data attribute updates, for males, patients who were documented as non-Hispanic White, elderly patients, patients with severe comorbidities, patients on Medicare, and during pre-COVID-19 years (Table 4). There were differences in timeliness among some patient characteristics. For example, non-Hispanic White patients had lower timeliness (in days) compared to the remaining racial categories.

In general, T2DM measurements were timelier among patients who had a history of telehealth use compared to patients with no history of telehealth use (Table 5). That is, for all T2DM clinical measurements examined in this study, the average number of days between EHR data attribute updates was shorter for patients with a history of telehealth use.

EHR Data Completeness

The mean completeness for all relevant patient data was 0.891 (SD=0.04) indicating moderate-to-high completeness for data recorded in T2DM measurements (Table 4). Completeness scores were, on average, higher among patients with severe comorbidities and patients who were seen in years preceding the COVID-19 pandemic (Table 4). These scores for T2DM measurements were generally higher among patients who had a history of telehealth use compared to patients with no history of telehealth use (Table 5).

EHR Data Information Density

The mean information density score for T2DM EHR data was 0.787 (SD=0.14). This demonstrates that patient visits and relevant T2DM measurements were somewhat uniform during the period in which they were identified in the EHR dataset. Information density scores for EHR data were higher among females, patients who were documented as Asian, older patients, and patients whose insurance status was Medicare (Table 3). T2DM measurements became more uniform across patient visits in the year following the COVID-19 pandemic. Information density scores, which account for the irregular nature of T2DM physiological measurements, were, on average higher for patients who had a history of telehealth use. However, blood pressure, cholesterol, and serum creatinine information density scores were higher among patients who did not have a history of telehealth use compared to patients who used telehealth during the study period.

DISCUSSION

EHR data quality for adult patients managing T2DM was generally similar or more improved among patients with a history of telehealth use than for those with no history of telehealth use. Establishing the quality of EHR data across modalities and patient populations is critical for health care organizations and researchers as the delivery of care evolves in the US.

The management of T2DM requires comprehensive, high-quality data.[136] While patients with a history of telehealth tended towards higher quality, all measures across both patient groups had room for improvement. For example, clinical guidelines recommend measuring HbA1c at least twice a year, but the average HbA1c measure was nearly a year old.[22] Options for improving data collection and documentation exist. For example, health portals that allow for patient input and self-reporting may help bridging incomplete or erroneous data.[137–139] Similarly, health portals also give patients opportunities to view, verify, and potentially correct information.[139] Likewise, organizations could consider the use of remote monitoring tools. Specifically, the improved accuracy of blood glucose monitors enable for remote data sharing and access.[19] During the patient encounters, the use of in-built structured forms and support personnel, like scribes, has been linked to reductions in workflow disruptions, improvements in documentation completion time, and efficient and high-quality documentation.[140]

The positive association between telehealth visits and data quality for most relevant clinical is reassuring given the increase in telehealth nationwide. An increasing number of payers and policy-makers had already been encouraging the wider use of telehealth services for chronic disease management.[141] The Centers for Medicare and Medicaid Services (CMS) finalized two policy changes for remote therapeutic and physiologic monitoring services extending beyond the Public Health Emergency.[142] These new rules seek to improve care delivery, cost management, and health outcomes for chronic disease patients who rely on remote monitoring as part of disease management.[142,143] As noted above, opportunities and strategies exist to improve data quality for T2DM patients, but these findings at least suggest telehealth visits may not be detrimental to data quality.

Overall, the levels of data quality in this study population were mostly consistent with the literature examining gaps in complete data for patients managing diabetes.[23] This study did not determine the reasons for less than timely and incomplete data, but determinants of poor EHR data quality in chronic disease care domains can include lack of standards, incomplete or inaccurate data entry, spelling and coding errors, non-compliant data protocols, and errors in extraction, to name a few.[133,144,145] Direct comparisons with other studies on timeliness and completeness are challenging. While timeliness and completeness are well-defined data quality constructs,[26-29] many studies of EHR data quality do not always account for the longitudinal nature of patient clinical interactions. We adapted our approach to examine the completeness of relevant chronic disease and demographic data elements longitudinally to examine the full historical account of data at the patient level.[75]

Additionally, this study highlighted additional notable differences in data quality. EHR data for underrepresented racial minority patients were marginally less timely than for non-Hispanic White patients. The magnitude of these differences may reflect distributions in patient characteristics that were not examined in this study. Given our sample is majority African American/Black and T2DM affects a large share of African American/Black patients in the US, the effects of incomplete and less timely data on care quality and care coordination should be explored further.[146] In addition, this study illustrates the very disruptive nature of the COVID-19 pandemic on EHR data quality. EHR data completeness scores decreased during the pandemic period. This indicates future EHR-based research may have to account for lower quality data during this period.

Limitations

These findings on data quality are limited in terms of the generalizability of the measurements and the patient population. First, these findings solely included structured data elements. It is possible that relevant data were documented in clinical notes or other text documents, and therefore would have been available at either encounter type. While potentially available for clinical care, any text-based data would be less accessible for secondary uses. Second, due to the COVID-19 pandemic, timeliness measures may have been affected by delays in care experienced by all primary care settings. Third, this study did not assess where the observed levels of data quality were sufficient for effective decision-making and disease management. The levels of data quality may not be generalizable to other settings as with different documentation practices, scheduling practices, and workflows.

CONCLUSION

EHR data for patients managing T2DM with a history of telehealth use were generally more timely and more complete than data for patients with no history of telehealth use. Differences in data quality among visit types and across patient characteristics may limit care delivery and secondary data uses. Improvements to data collection and quality will be needed as telehealth and hybrid delivery models become more common.

Table 2. Patient and Encounter Descriptive Characteristics by History of Telehealth Use

	History of patient telehealth use by demographic characteristics			
	Total, n (%)	History of Telehealth Use, n (%)	No History of Telehealth Use, n (%)	P value
Patient Sex				0.412
Male	1705 (33.9)	638 (33.2)	1067 (34.4)	
Female	3322 (66.1)	1284 (66.8)	2038 (65.6)	
Patient Race				<0.001
White	1480 (29.4)	620 (32.3)	820 (27.7)	
Black	2617 (52.1)	1029 (53.5)	1588 (51.1)	
Hispanic	579 (11.5)	161 (8.4)	418 (13.5)	
Asian	48 (0.9)	16 (0.8)	32 (1.0)	
American Indiana/Alaska Native	5 (0.1)	19 (0.9)	5 (0.2)	
Native Hawaiian/Other Pacific Islander	59 (1.2)	35 (1.8)	40 (1.3)	
More than one Race	143 (2.8)	42 (2.2)	54 (1.7)	
Age				<0.001
18-25	66 (1.3)	28 (1.5)	38 (1.2)	
26-45	781 (15.5)	284 (14.8)	497 (16.0)	

46-55	1130 (22.5)	421 (21.9)	709 (22.8)	
56-65	1592 (31.7)	561 (29.2)	1031 (33.2)	
66-75	966 (19.2)	425 (22.1)	541 (17.4)	
≥75	492 (9.8)	102 (10.6)	289 (9.3)	
Charlson Comorbidity Score				<0.001
0	549 (10.9)	188 (9.8)	361 (11.6)	
1	1919 (38.2)	674 (35.1)	1245 (40.1)	
2	1554 (30.9)	644 (33.5)	910 (29.3)	
3	726 (14.4)	305 (15.9)	421 (13.6)	
4	221 (4.4)	86 (4.5)	135 (5.4)	
5	48 (0.9)	20 (1.0)	28 (0.9)	
6	10 (0.2)	5 (0.3)	5 (0.2)	
Insurance				<0.001
Commercial	574 (11.4)	140 (7.3)	434 (13.9)	
Medicare	2756 (54.8)	1245 (64.8)	1511 (48.7)	
Medicaid	1395 (27.6)	449 (23.4)	946 (30.5)	
Self-pay	205 (4.1)	47 (2.5)	158 (5.1)	
Other	96 (1.9)	41 (2.1)	55 (1.8)	
COVID-19				0.359
Pre-COVID	4787 (95.2)	1823 (94.9)	2964 (95.5)	
Post-COVID	240 (4.8)	99 (5.2)	141 (4.5)	
Year				0.002

2016	2120 (42.2)	857 (44.6)	1263 (40.7)	
2017	1414 (28.1)	516 (26.9)	898 (28.9)	
2018	628 (12.5)	219 (11.4)	409 (13.2)	
2019	518 (10.3)	181 (9.4)	337 (10.9)	
2020	335 (6.7)	147 (7.7)	188 (6.1)	
2021	12 (0.2)	2 (0.1)	10 (0.32)	

Table 3. Patient and Encounter Characteristics by Timeliness and Completeness Measures for T2DM EHR Data

	Measures of Timeliness, Completeness, and Information Density Score of T2DM EHR Data					
	Timeliness (in days)[a]		Completeness[b]		Sperrin's I (Information Density Score)[b]	
	Mean (SD)	P value	Mean (SD)	P value	Mean (SD)	P value
Telehealth Use		0.045		0.901		<0.001
No	79.8 (65)		0.892 (0.05)		0.784 (0.11)	
Yes	73.3 (61.6)		0.892 (0.04)		0.795 (0.08)	
Patient Sex		0.021		0.882		0.033
Male	62.8 (63.6)		0.896 (0.05)		0.784 (0.10)	
Female	101 (163)		0.892 (0.04)		0.790 (0.09)	
Patient Race/Ethnicity		0.003		<0.001		<0.001
White	71.7 (63.4)		0.889 (0.06)		0.790 (0.09)	
Black	77.1 (62.4)		0.890 (0.04)		0.792 (0.09)	
Hispanic	88.8 (68.3)		0.903 (0.04)		0.775 (0.11)	
Asian	99.3 (79.1)		0.891 (0.04)		0.807 (0.09)	
American Indian/Alaska Native	85 (33.5)		0.882 (0.04)		0.736 (0.19)	
Native Hawaiian/Other Pacific Islander	83.5 (47.8)		0.906 (0.03)		0.768 (0.12)	

More than one Race	73.3 (63.3)		0.906 (0.04)		0.751 (0.11)	
Age		<0.001		0.172		<0.001
18-25	86.4 (60.2)		0.887 (0.07)		0.775 (0.13)	
26-45	88.3 (73.3)		0.891 (0.04)		0.776 (0.11)	
46-55	78.8 (65.7)		0.892 (0.04)		0.785 (0.09)	
56-65	77.5 (65.6)		0.891 (0.04)		0.789 (0.09)	
66-75	71.2 (53.1)		0.892 (0.05)		0.795 (0.09)	
≥75	66.7 (53)		0.895 (0.05)		0.798 (0.10)	
Charlson Comorbidity Score		<0.001		<0.001		0.188
0	87.9 (81.3)		0.876 (0.04)		0.798 (0.13)	
1	92.3 (75.5)		0.890 (0.05)		0.786 (0.11)	
2	72.2 (50.6)		0.895 (0.04)		0.781 (0.08)	
3	53.4 (26.8)		0.897 (0.03)		0.793 (0.07)	
4	45.9 (23.2)		0.904 (0.04)		0.804 (0.06)	
5	37.7 (17.4)		0.893 (0.03)		0.825 (0.04)	
6	42.2 (19.2)		0.900 (0.03)		0.806 (0.04)	
Insurance		0.008		<0.001		0.018

Commercial	86.3 (60.8)		0.886 (0.05)		0.776 (0.11)	
Medicare	72 (58)		0.892 (0.05)		0.796 (0.09)	
Medicaid	80.9 (70.5)		0.893 (0.04)		0.779 (0.11)	
Self-pay	96 (87.9)		0.896 (0.04)		0.788 (0.15)	
Other	84.7 (58.9)		0.899 (0.03)		0.762 (0.11)	
COVID-19		0.203		<0.001		<0.001
Pre-COVID	72.2 (68.5)		0.893 (0.04)		0.786 (0.09)	
Post-COVID	143 (278)		0.862 (0.09)		0.822 (0.15)	
Year		0.052		<0.001		<0.001
2016	69.7 (54)		0.896 (0.04)		0.799 (0.08)	
2017	94.8 (73.8)		0.894 (0.04)		0.782 (0.09)	
2018	78.7 (59.8)		0.896 (0.04)		0.772 (0.11)	
2019	75.3 (66.7)		0.884 (0.05)		0.766 (0.12)	
2020	54.5 (61.2)		0.867 (0.08)		0.803 (0.14)	
2021	15.4 (10.8)		0.826 (0.12)		0.905 (0.16)	

Note:
[a]Timeliness is measured at the encounter level.
[b]Completeness is measured at the patient level.

Table 4. Data Quality Measurements for T2DM EHR Data by Patients' Use of Telehealth

	Quality Measures for T2DM EHR Data		
	History of Telehealth Use	No History of Telehealth Use	P value
	Mean (SD)	Mean (SD)	
Timeliness (in days)[a]			
Body Mass Index (BMI)	50.6 (112)	61.2 (120)	0.021
Body Weight	52 (114)	63 (131)	0.006
Blood pressure	42.6 (91)	52.1 (99)	0.041
HbA1c	324 (386)	412 (488)	<0.001
Cholesterol	326 (322)	432 (481)	<0.001
Serum Creatinine	123 (219)	135 (212)	0.284
Smoking Status	29 (76.6)	31.9 (88.8)	0.712
Completeness[b]			
Body Mass Index (BMI)	0.991 (0.07)	0.990 (0.11)	0.032
Body Weight	0.992 (0.04)	0.992 (0.06)	0.882
Blood pressure	0.993 (0.05)	0.993 (0.09)	0.701
HbA1c	0.886 (0.22)	0.836 (0.32)	<0.001
Cholesterol	0.903 (0.35)	0.814 (0.31)	0.001
Serum Creatinine	0.989 (0.12)	0.983 (0.18)	0.044
Smoking Status	0.993 (0.09)	0.994 (0.10)	0.844

Sperrin's I (Information Density Score)[b]			
Body Mass Index (BMI)	0.831 (0.12)	0.806 (0.13)	0.012
Body Weight	0.877 (0.10)	0.807 (0.09)	0.442
Blood pressure	0.820 (0.09)	0.866 (0.13)	<0.001
HbA1c	0.881 (0.12)	0.827 (0.12)	<0.001
Cholesterol	0.822 (0.12)	0.838 (0.08)	0.011
Serum Creatinine	0.791 (0.11)	0.814 (0.12)	<0.001
Smoking Status	0.889 (0.12)	0.880 (0.10)	0.061

Note:
[a]Timeliness is measured at the encounter level. Lower timeliness, in days, indicates higher levels of data quality.
[b]Completeness is measured at the patient level. Higher levels of completeness indicate that more data are available for each patient.

CHAPTER 3: ASSOCIATIONS BETWEEN PRIMARY CARE LABORATORY TEST AGE AND HEALTHCARE UTILIZATION AMONG PATIENTS MANAGING TYPE 2 DIABETES MELLITUS

INTRODUCTION

Patient data generated by clinicians across settings, including laboratory tests, play a key role in monitoring type 2 diabetes mellitus (T2DM) disease progression and care management.[150,151] For example, T2DM management relies on glycemic control and is an important indicator of care quality that requires regular laboratory testing.[151] Cholesterol and serum creatinine laboratory tests also inform T2DM management as important indicators of disease progression.[151,153,155] Moreover, laboratory tests provide important data attributes that support patient self-management, provider clinical decision-making, and are intended to monitor disease progression to prevent unplanned hospitalizations.[88,98] However, laboratory test data are not often captured uniformly across settings which may create data gaps and affect care quality.[159]

The US Preventive Services Task Force and the American Diabetes Association (ADA) developed guidelines recommending clinicians conduct glucose test based on T2DM risk factors including age, weight, and comorbid conditions.[154,165] There are still well-known screening and testing process deficiencies that create patient data gaps which may have downstream implications for T2DM care quality and care coordination. For example, evidence that determines the optimal frequency of glucose and HbA1c tests is still limited.[154] Additionally, glucose test management practices are suboptimal even as follow-up reporting and care are supported by EHRs and other digital healthcare tools that enable timely recognition of test results.[148] One study reported a follow-up failure rate between 50%-62% for reporting abnormal test values.[148] Timely access to laboratory test features including reported values and test frequency inform the extent to which T2DM has advanced.[136,148,155] These challenges are indicative of system-level factors driving laboratory tests gaps that are critical to T2DM management.[148,150] Chronic disease management and integrated care are improved by timely, shareable laboratory test data across providers and clinical settings.[116,152]

However, the quality of laboratory data available to providers may not be sufficient to support care and reduce unnecessary utilization. For example, older laboratory tests may impede the ability of providers and patients to timely and effectively reduce T2DM progression and unnecessary hospital admissions.[150] Additionally, older laboratory tests may indicate poor compliance with standards of care, including increasing testing frequencies for concerning laboratory test results and advanced disease management.[151] These issues may limit access to relevant patient information and timely treatment to avoid costly care.[156–157] Prior research found that patients with Diabetes and concomitant clinical and demographic risk factors have an increased risk of unplanned hospitalizations and longer lengths of stay than patients without Diabetes. [155, 157] Identifying means to improve care for those with T2DM is critical as patients with T2DM account for more costs, hospitalizations, and overall utilization[151,154-162]

Objective

While timely laboratory test data are clearly important, we lack clear evidence on its association with subsequent healthcare utilization. In addition, numerous features of our health care system inhibit access to timely laboratory information such as poor access to care, patient compliance, cost barriers, and non-interoperable health information technology.[165-168] The purpose of this study is to examine the association between laboratory test ages and subsequent inpatient hospitalizations and emergency department (ED) visits.

METHODS

Study Design

We used a panel design to examine associations between T2DM primary care laboratory test ages and subsequent inpatient hospitalizations and ED visits.

Study Population

The study sample included patients managing T2DM aged ≥18 years who were seen between 2016 and 2021 at two health systems in central Indiana. The sample was defined as any patient who received a T2DM diagnosis between their index and latest encounter date and had a subsequent inpatient hospitalization or ED visit.

Data

The EHR dataset used in the current study was derived from the Indiana Network for Patient Care (INPC). The INPC is a statewide health information exchange that was established in 1994 as a repository for cross-institutional patient data.[134]

Primary Care Laboratory Test Age

We defined a measure of timeliness as the prior year (i.e., lagged) average age of primary care laboratory tests relevant to T2DM care. We quantified laboratory age as the number of days between laboratory test dates and subsequent encounter dates for three tests: serum creatinine, glycated hemoglobin A1c, and cholesterol tests. We scaled the resulting days into months to improve interpretability of findings (i.e., divided days by 30.44). Fewer months indicate that laboratory tests are newer.[135]

Study Outcome

We identified patient encounters that were coded as inpatient hospitalizations and ED visits. We determined the frequency of each visit type to create count variables for each year a patient was included in the sample.

Analysis

We described the sample using frequencies, percentages, and means. To examine relationships between average laboratory test age and the frequency of inpatient hospitalizations and ED visits, we estimated multivariable Negative Binomial regression models with patient fixed effects and year dummies to control for linear trends. Average primary care laboratory test age (timeliness measure) was lagged by a year. All regression analyses controlled for prior-year primary care visits. We conducted stratified analyses to explore differences in effects by patient sex, age, race, and insurance status. Regression estimates were reported using marginal effects. All analyses were conducted in STATA 16.1. Results were considered significant at the p=0.05 level.

Sensitivity Analyses

First, we fit separate fixed-effects Poisson Regression models as a check on the Negative Binomial regressions. Next, some laboratory test dates were unmatchable by proximate dates and patient identifiers which produced missing ages for relevant T2DM laboratory tests. We determined that data were missing at random based on a mean imputation approach.[44] Results from this approach yielded little change to standard errors which indicated that overall laboratory test ages were only partially sensitive to mean imputation. Nevertheless, we examined whether the missing laboratory age data had influenced model estimates by imputing values. We imputed the 90th percentiles of average laboratory test ages to obtain estimates under the assumption that missing data were even older. Additionally, we imputed zeros in place of missing data as a check to determine changes in results if missing data were as recent as possible. We created a categorical measure of average laboratory test age to check for consistency of results against the continuous measure. Lastly, we used a binary indicator to determine pre- and post-COVID-19 status to examine whether the pandemic affected the recency of relevant laboratory tests. Post-COVID-19 status was determined as any encounter and laboratory date after March 13, 2020, when the Public Health Emergency Declaration was announced.

RESULTS

The final sample included a total of 15,033 person-year observations. The mean age of patients in the sample was 52.3 years. The majority of our sample included female (60.6%), non-Hispanic Black (51.6%) patients who were primarily Medicare recipients (42.4%) (Table 6). The average laboratory test age for the sample was 7.6 (SD=7.8) months. The average age of primary care laboratory tests identified among ED visits were generally older compared to inpatient admissions.

Associations between Primary Care Laboratory Test Age and Inpatient and ED Admissions

Older laboratory tests were associated with an increase in the expected counts of subsequent inpatient hospitalizations (ME=0.047; $p<0.001$) and ED Visits (ME=0.034; $p<0.001$) controlling for prior-year primary care visits and linear trends (Table 6). Results from stratified analyses were consistent with estimates from the main findings for the inpatient hospitalization outcome (Figures 3 & 4). Patient demographic subgroups with older laboratory tests were separately associated with statistically significant increases in the expected counts of inpatient hospitalizations except for patients whose insurance status was "Self-pay". Similarly, older laboratory tests for patient demographic subgroups were associated with statistically significant increases in the expected counts of ED visits. Notably, the estimate for patients in age groups 26-45, 46-55, and 56-65 indicated that as laboratory tests aged, the expected counts of inpatient hospitalizations gradually increased compared to the main effect (Figure 3).

Sensitivity Analyses

Sensitivity analyses aligned with the results of our primary analysis (Tables 7-15). The fixed-effects Poisson regression models were consistent (Table 7). Laboratory test age categories indicated that tests older than 12 months increased the total expected counts of inpatient hospitalizations by approximately 81 percentage points ($p<0.001$) (Table 8) and ED visits by 120 percentage points per patient per year (Table 9). In fixed-effects Negative Binomial regressions that modeled the independent variable with imputed zeros, older laboratory tests were associated with an increase in the expected counts of both inpatient hospitalizations and ED Visits (Table 10 & 11). Imputing the 90th

percentiles of laboratory test ages determined no change and were supportive of main model marginal effects estimates. Imputing 12 months (equivalent of a year) into missing laboratory test ages produced effect estimates that reflect results from our primary analysis (Tables 14 & 15). Patients who were seen in inpatient settings after the Public Health Emergency Declaration for the COVID-19 pandemic were predicted to, on average, have more recent laboratory tests by 0.12 months (Table 16). Additionally, patients who were seen in the emergency department after the Public Health Emergency Declaration for the COVID-19 pandemic were predicted to, on average, have newer laboratory tests by 1.17 months.

DISCUSSION

In a two-health system sample, older laboratory tests were associated with an increase in subsequent inpatient hospitalizations and ED visits among adult patients managing T2DM. This study contributes to the clinical data quality literature by operationalizing a clinically relevant measure of timeliness and its relationship with subsequent healthcare utilization. Prior research examined how diabetes progression is associated with higher healthcare utilization.[155] While our study did not examine factors related to T2DM progression, i.e. values of laboratory tests, older laboratory tests produced in primary care settings are possibly linked to unnoticed disease advancement which may result in inpatient hospitalizations and/or ED visits. This pattern was also present in stratified analyses where the age of prior-year laboratory tests was associated with increases in inpatient hospitalizations in parallel with ascending patient age groups.

Healthcare organizations could consider several proven interventions to improve access to timely laboratory test data in primary care settings. First, engagement in health information exchange has been demonstrated to improve EHR data quality and data specific to Diabetes.[186] Second, organizations could encourage patients to use digital healthcare tools that generate data to improve test monitoring. For example, patients managing T2DM have access to continuous glucose monitoring (CGM) digital technologies that provide data on the percentage of time spent in the target range of glycemic control.[147] Use of these tools may remove barriers to timely, automated data entry which is critically important to monitoring disease progression. In addition, patient portals enable patients with the ability to review and verify information important to managing their own health, including laboratory test values and the presence of recent tests.[139]

In considerations of data quality, prior research has primarily focused on quantifying measures of completeness and accuracy.[17,39,133] These are important aspects of data quality; however, this study was focused on operationalizing a measure of timeliness. Timeliness has been variously defined in the literature as data elements that represent a patient health status at a given point in time and reflected using quotients that account for average intervals between data attribute updates.[69-71] Such measures are very useful from an informatics perspective, however, we opted for a measure of timeliness, i.e., age in months, with straightforward interpretability.

The American Diabetes Association (ADA), the US Preventive Services Task Force, and the Centers for Disease Control and Prevention (CDC) recommend regular testing to improve Diabetes self-management to avoid inpatient hospitalizations and ED admissions.[154,161,163] However, patients managing T2DM may be subject to irregular laboratory testing and disparate collection of other physiological measurements.[161]

Importantly, T2DM care management is improved when patients maintain a regular testing regime to avoid duplicate or unnecessary laboratory tests. Findings from this study suggest that T2DM testing was generally uniform among older patients with a history of comorbid conditions, which is reflective of ADA and CDC guidelines. However, these guidelines do not consider scenarios or effective approaches when laboratory tests are out of date or too old to support meaningful clinical decision-making. Uniform testing provides clinicians with recent laboratory test results to better inform care.[37] Clinicians are expected to ensure timely monitoring of patients with laboratory tests that are clinically concerning to avoid unnecessary hospital admissions.[159] Thus, timeliness is a focal data quality dimension for examining clinical domains where regular pathophysiological measurements are required to manage and treat chronic disease.

Limitations

This study had several limitations. We used months as the scale for laboratory test ages computed using patient encounter and laboratory dates. This approach facilitates interpretability but may not be the most precise method of describing the age of relevant T2DM laboratory tests. We only examined timeliness and did not operationalize other data quality dimensions (e.g., completeness, correctness, and accuracy). Other factors besides laboratory test age could be driving the increase in inpatient hospitalizations and ED visits that are not addressed in our fixed-effects models. The COVID-19 pandemic may have led to critical care delays for patients in our sample, particularly in 2020 and 2021. As a result, patients in our sample may have forgone laboratory-based blood draws to avoid disease exposure and summarily contributed to higher-than-average laboratory tests ages. Patients may also be subject to missed opportunities and self-management stigmas that may impede improved care coordination and laboratory test sequencing that that are critical to better health outcomes. Lastly, data used in this study were specific to this sample of patients at two health systems. Thus, these results cannot be generalized to other health systems or similar patient populations.

CONCLUSION

This study found that older laboratory tests were associated with increases in subsequent inpatient hospitalizations and ED visits among patients managing T2DM. Chronic disease management relies on uniform, timely, and accessible patient data. Improving the uniformity and timeliness of laboratory testing data may aid in reducing exacerbations of T2DM and unplanned hospital visits.

Table 5. Patient Characteristics

	Total Sample (n= 15,033)	Average Prior Year Laboratory Test age (in months)
Inpatient Hospitalizations, mean (SD)	0.3 (1.2)	7.7
Emergency Department Visits, mean (SD)	0.5 (1.1)	7.9
Patient Sex (%)		
Male	39.4	7.9
Female	60.6	8.3
Patient Race/Ethnicity (%)		
White	26.7	7.5
African American/Black	51.6	8.1
Hispanic	15.3	9.2
Asian	1.0	7.9
American Indian/Alaska Native	0.2	9.3
Native Hawaiian/Other Pacific Islander	1.5	7.5
More than one Race	3.7	7.9
Age (%)		
18-25	2.9	9.7
26-45	24.8	9.3
46-55	27.4	8.3
56-65	27.8	7.4
66-75	11.3	6.8
≥75	5.8	5.9
Charlson Comorbidity Score (%)		

0	17.3	10.4
1	42.9	8.8
2	24.9	7.0
3	8.8	4.9
4	2.1	3.7
5	0.5	3.1
6	3.5	2.5
Insurance (%)		
Commercial	13.5	9.0
Medicare	42.4	7.3
Medicaid	35.5	8.5
Self-pay	5.8	9.0
Other	2.8	8.3
COVID-19 Status		
Pre-COVID	68.6	6.8
Post-COVID	31.4	8.7
Year (%)		
2016	26.8	4.1
2017	23.1	6.5
2018	17.4	8.7
2019	15.6	9.3
2020	11.4	9.5
2021	5.7	8.0

Table 6. Fixed-effects Negative Binomial Regression Models Marginal Effects for Inpatient Hospitalizations and Emergency Department Visits – Primary Care Laboratory Test Age in Months

	Dependent Variable: Inpatient Hospitalizations Independent Variable: Laboratory Test Age (in Months) Model 1: Fixed-effects Negative Binomial Regression					Dependent Variable: ED Visits Independent Variable: Laboratory Test Age (in Months) Model 2: Fixed-effects Negative Binomial Regression				
	ME	se	95% CI		P value	ME	se	95% CI		P value
Prior Year Laboratory Test age (timeliness)	0.047	0.007	0.033	0.062	<0.001	0.034	0.007	0.021	0.047	<0.001
Prior Year Outpatient Visits	-0.025	0.009	-0.042	-0.008	0.004	-0.049	0.011	-0.071	-0.029	<0.001
Year (lagged)										
2018	0.048	0.086	-0.121	0.217	0.578	0.270	0.115	0.044	0.496	0.019
2019	0.336	0.096	0.148	0.523	<0.001	0.169	0.118	-0.062	0.402	0.152
2020	0.657	0.112	0.436	0.877	<0.001	-0.614	0.122	-0.853	-0.375	<0.001
2021	-1.071	0.11	-1.288	-0.855	<0.001	-3.086	0.192	-3.462	-2.709	<0.001

Table 7. Fixed-effects Poisson Model Marginal Effects for Inpatient Hospitalizations and Emergency Department Visits – Primary Care Laboratory Test Age in Months

	Dependent Variable: Inpatient Hospitalizations Independent Variable: Laboratory Test Age (in Months) Model 1: Fixed-effects Poisson Regression					Dependent Variable: ED Visits Independent Variable: Laboratory Test Age (in Months) Model 2: Fixed-effects Poisson Regression				
	ME	se	95% CI		P value	ME	se	95% CI		P value
Prior Year Laboratory Test age (timeliness)	0.032	0.004	0.024	0.041	<0.001	0.018	0.001	0.005	0.021	<0.001
Prior Year Outpatient Visits	-0.000	0.005	-0.009	0.009	0.996	-0.011	0.002	-0.015	-0.007	<0.001
Year (lagged)										
2018	0.101	0.050	0.003	0.199	0.044	0.082	0.026	0.031	0.132	0.001
2019	0.265	0.056	0.155	0.375	<0.001	0.046	0.026	-0.005	0.097	0.077
2020	0.434	0.066	0.306	0.562	<0.001	-0.150	0.024	-0.198	-0.103	<0.001
2021	-0.781	0.038	-0.856	-0.707	<0.001	-0.804	0.018	-0.839	-0.768	<0.001

Table 8. Fixed-effects Negative Binomial Model Marginal Effects for Inpatient Hospitalizations – Primary Care Laboratory Test Age Categories

	Fixed-effects Negative Binomial Regression for Inpatient Hospitalizations				
	ME	se	95% CI		P value
Marginal Effects for Prior Year Laboratory Test Age Categories					
≤3 months	0.407	0.073	0.264	0.549	<0.001
4-7 months	0.543	0.082	0.391	0.704	<0.001
8-11 months	0.591	0.090	0.415	0.768	<0.001
≥12 months	0.810	0.094	0.625	0.995	<0.001

Table 9. Fixed-effects Negative Binomial Model Marginal Effects for Emergency Department Visits – Primary Care Laboratory Test Age Categories

	Fixed-effects Negative Binomial Regression for ED Visits				
	ME	se	95% CI		P value
Marginal effects for Prior Year Laboratory Test Age Categories					
≤3 months	1.041	0.058	0.928	1.153	<0.001
4-7 months	1.089	0.060	0.971	1.206	<0.001
8-11 months	1.086	0.063	0.962	1.210	<0.001
≥12 months	1.204	0.064	1.079	1.328	<0.001

Figure 3. Fixed-effects Negative Binomial Regression Marginal Effects Estimates: Stratified by Patient Race, Age, Insurance Status, and Sex – Inpatient Hospitalizations

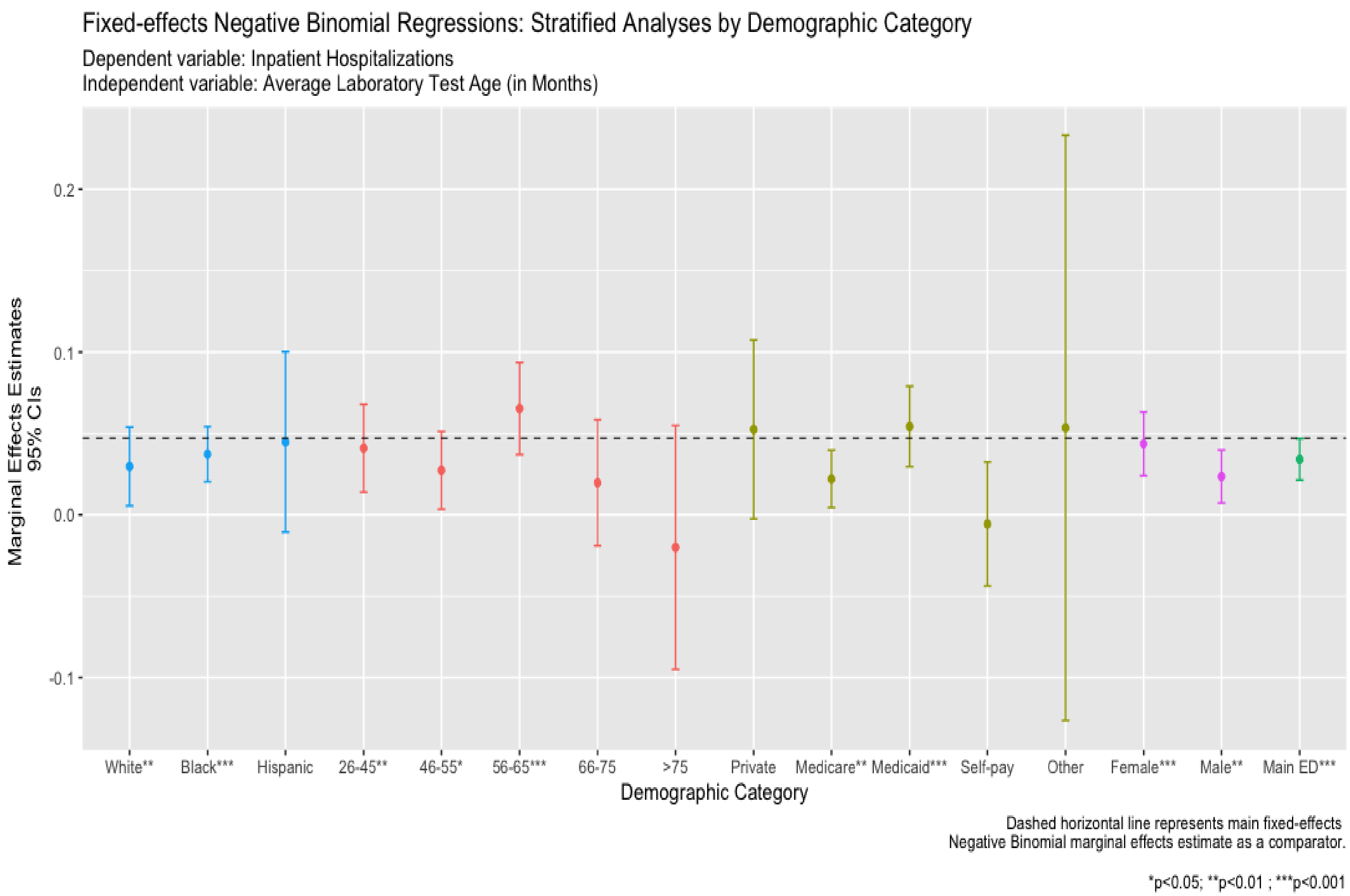


Figure 4. Fixed-effects Negative Binomial Regression Marginal Effect Estimates: Stratified by Patient Race, Age, Insurance Status, and Sex – Emergency Department Visits

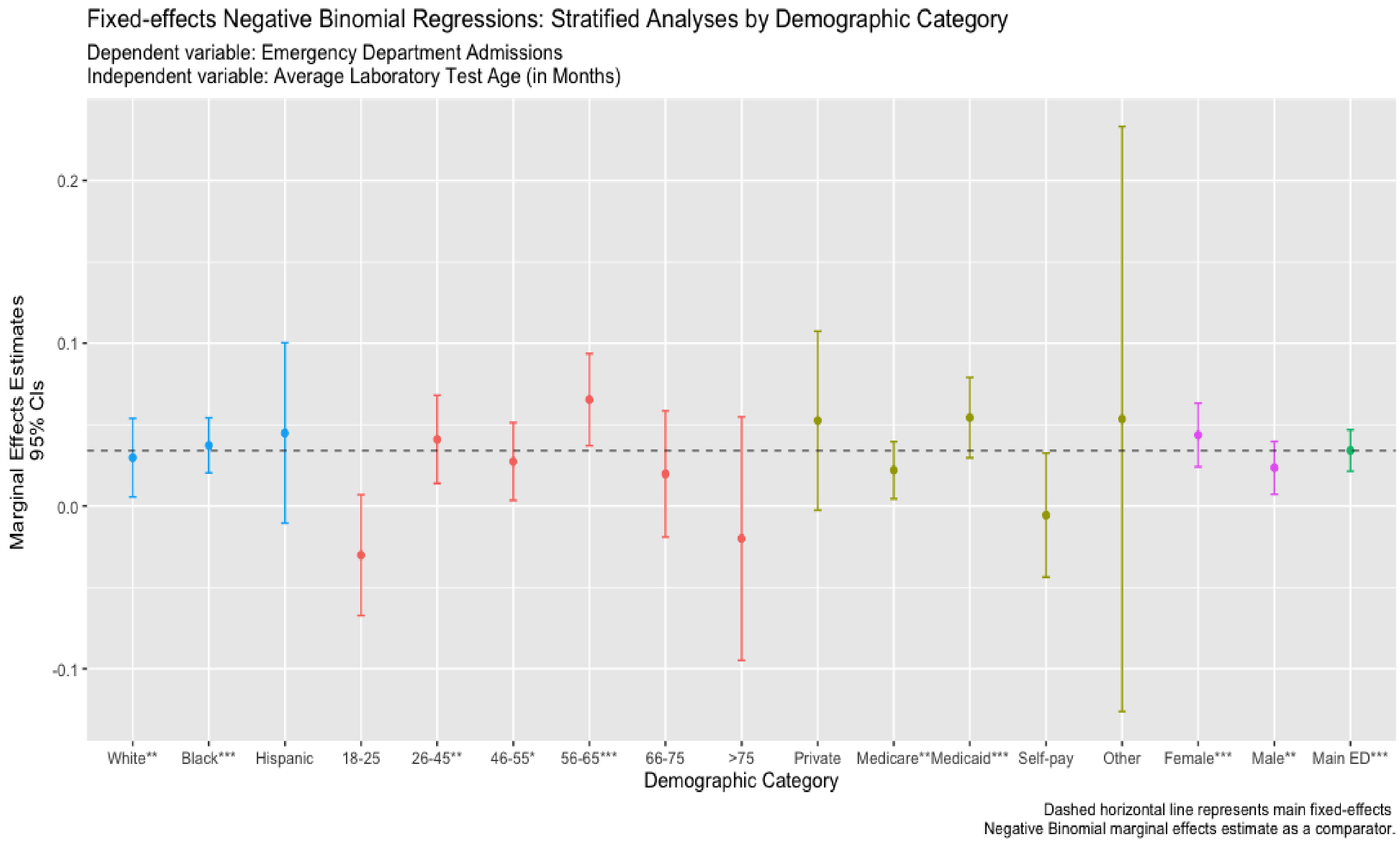


Table 10. Sensitivity Analyses: Primary Care Laboratory Age modeled with Zeros imputed in Laboratory Test Age Variable in *Months* – Inpatient Hospitalizations (Fixed-effects Negative Binomial Regression Models)

	Dependent Variable: Inpatient Hospitalizations				
	Independent Variable: Laboratory test age (zero imputation)				
	Model 1: Fixed-effects Negative Binomial Regression				
	ME	se	95% CI		P value
Prior Year Laboratory Year Test Age (timeliness)	0.001	0.000	0.001	0.002	<0.001
Prior Year Outpatient Visits	-0.017	0.004	-0.026	-0.001	<0.001
Year (lagged)					
2018	-0.014	0.044	-0.099	0.071	0.747
2019	0.131	0.045	0.043	0.219	0.004
2020	0.243	0.047	0.150	0.335	<0.001
2021	-1.013	0.067	-1.146	-0.880	<0.001

Table 11. Sensitivity Analyses: Primary Care Laboratory Age modeled with Zeros imputed in Laboratory Test Age Variable in Months – Emergency Department Visits (Fixed-effects Negative Binomial Regression Models)

	Dependent Variable: ED Visits Independent Variable: Laboratory test age (zero imputation)) Model 1: Fixed-effects Negative Binomial Regression				
	ME	se	95% CI		P value
Prior year Laboratory age (timeliness)	0.001	0.000	0.001	0.001	<0.001
Prior year Outpatient Visits	-0.061	0.010	-0.081	-0.041	<0.001
Year (lagged)					
2018	0.170	0.107	-0.039	0.379	0.111
2019	0.019	0.110	-0.197	0.235	0.863
2020	-0.761	0.117	-0.991	-0.532	<0.001
2021	-3.341	0.187	-3.706	-2.976	<0.001

Table 12. Sensitivity Analyses: Primary Care Laboratory Age modeled with 90th percentile Laboratory Test Age (in Months) Imputation – Inpatient Hospitalizations (Fixed-effects Negative Binomial Regression Models)

	Dependent Variable: Inpatient Hospitalizations Independent Variable: Laboratory Test age (90th percentile imputation) Model 1: Fixed-effects Negative Binomial Regression				
	ME	se	95% CI		P value
Prior Year Laboratory Test Age (timeliness)	0.010	0.002	0.007	0.014	<0.001
Prior Year Outpatient Visits	-0.023	0.007	-0.037	-0.008	0.002
Year (lagged)					
2018	0.038	0.069	-0.098	0.175	0.584
2019	0.319	0.079	0.165	0.473	<0.001
2020	0.533	0.089	0.358	0.709	<0.001
2021	-0.993	0.089	-1.168	-0.818	<0.001

Table 13. Sensitivity Analyses: Primary Care Laboratory Age modeled with 90th percentile Laboratory Test Age (in Months) Imputation – Emergency Department Visits (Fixed-effects Negative Binomial Regression Models)

	Dependent Variable: ED Visits Independent Variable: Laboratory test age (90th percentile imputation) Model 1: Fixed-effects Negative Binomial Regression				
	ME	se	95% CI		P value
Prior Year Laboratory Test Age (timeliness)	0.017	0.002	0.003	0.011	<0.001
Prior Year Outpatient Visits	-0.054	0.010	-0.074	-0.034	<0.001
Year (lagged)					
2018	0.227	0.102	0.028	0.426	0.025
2019	0.116	0.104	-0.087	0.319	0.263
2020	-0.637	0.108	-0.849	-0.424	<0.001
2021	-3.154	0.172	-3.493	-2.814	<0.001

Table 14. Sensitivity Analyses: Primary Care Laboratory Age modeled with 12 months (a Year) imputed in Laboratory Test Age Variable – Inpatient Hospitalizations (Fixed-effects Negative Binomial Regression Models)

	Dependent Variable: Inpatient Hospitalizations Independent Variable: Laboratory Test Age (12-month imputation) Model 1: Fixed-effects Negative Binomial Regression				
	ME	se	95% CI		P value
Prior Year Laboratory Test Age (timeliness)	0.049	0.006	0.037	0.061	<0.001
Prior Year Outpatient Visits	-0.022	0.008	-0.037	-0.006	0.005
Year (lagged)					
2018	-0.019	0.076	-0.169	0.129	0.795
2019	0.256	0.084	0.090	0.421	0.002
2020	0.496	0.096	0.301	0.684	<0.001
2021	-1.119	0.100	-1.317	-0.923	<0.001

Table 15. Sensitivity Analyses: Primary Care Laboratory Age modeled with 12 months (a Year) imputed in Laboratory Test Age Variable – Emergency Department Visits (Fixed-effects Negative Binomial Regression Models)

	Dependent variable: ED Visits Independent Variable: Laboratory Test Age (12-month imputation) Model 1: Fixed-effects Negative Binomial Regression				
	ME	se	95% CI		P value
Prior Year Laboratory Test Age (timeliness)	0.037	0.006	0.025	0.048	<0.001
Prior Year Outpatient Visits	-0.052	0.010	-0.072	-0.032	<0.001
Year (lagged)					
2018	0.172	0.105	-0.035	0.378	0.103
2019	0.020	0.108	-0.192	0.232	0.851
2020	-0.747	0.114	-0.971	-0.523	<0.001
2021	-3.302	0.182	-3.658	-2.946	<0.001

Table 16. Primary Laboratory Test Age Pre- and Post-COVID-19 Emergency Declaration
Dependent Variables: Inpatient Hospitalizations and Emergency Department Visits (Fixed-effects Negative Binomial Regressions)

	Fixed-effects Negative Binomial Regression for Inpatient Hospitalizations				Fixed-effects Negative Binomial Regression for ED Visits			
	ME	95% CI		P value	ME	95% CI		P value
COVID-19 Status (ref: Pre-COVID-19)								
Post-COVID-19	-0.124	-0.034	-0.012	0.004	-1.169	-1.288	-1.051	<0.001

CHAPTER 4: PATIENT PORTAL USE AND ELECTRONIC HEALTH RECORD (EHR) DATA TIMELINESS IN TYPE 2 DIABETES MELLITUS CARE

INTRODUCTION

Type 2 Diabetes (T2DM) care relies on a wide array of electronic health record (EHR) data elements including laboratory results, medication regimens, and behavioral indicators (e.g., smoking status).[132] These data are collected by multispecialty care teams during various interactions with patients. The quality of EHR data may be subject to challenges because of numerous touchpoints and differential documentation practices across care settings and providers.[133] As a result, patient EHR data are subject to varying levels of data quality. Errors and missing patient information can have costly care coordination ramifications if uncorrected.[164] Adverse outcomes from missing patient information, particularly in the primary care setting, include duplicate medications, missed or delayed diagnoses, missed immunizations, and repeat testing and procedures.[165,166]

Patient portals enable bidirectional communication between providers and patients to share and receive health information for critical care processes.[137] Approximately 44 million patients have access to ambulatory care health information documented and maintained in their patient portal.[139] With this technology, patient portals serve as a means to improve data quality by allowing patients to view, verify, and correct their records where information may be missing, misrepresented, or repeated.[139,167] The opportunity to improve EHR data quality via patient portals may optimize care coordination and self-management, which are critical features of improving chronic disease care outcomes.[92,100,137] Likewise, there may be potential outcome improvements among older and underrepresented patient populations due to historically low digital literacy, low patient portal use, and lower levels of EHR data quality.[168–170]

Although patient portals enable data verification and correction, there are barriers that may impede any tangible improvements in overall data quality. Namely, prior research has identified challenges limiting patients’ access and consumption of their health information as a result of disparate documentation patterns, health system capacity, and inaccessible patient portals.[171,172] Despite these challenges, patients have still become more perceptive to mistakes in diagnoses, medical history, medications, and test results.[10-13] Patient portal use has also been found to increase overall patient engagement by enabling access to their own health information to manage various aspects of their health.[92,173] Thus, patient portals reduce unnecessary care utilization and care fragmentation for patients managing T2DM.[4] While prior research has investigated implications of digital health tool use on data quality dimensions in clinical settings[174], studies have not examined associations between patient portal use and EHR data timeliness in T2DM care.

Objective

This study examined the relationship between patient portal use and EHR data timeliness for patients managing T2DM. We quantified a measure of timeliness using structured EHR data. Although patient portal use is becoming more common, there are still access barriers that limit registration and use among some disadvantaged patient groups who may benefit most.[11] Therefore, we specifically examined associations between patient portal use and EHR data quality stratified by patient race, age, sex, and insurance status.

METHODS

Study Design

We examined the association between patient portal usage and EHR data timeliness in a panel of patients aged ≥18 years between 2017 and 2021.

Sample & Setting

Our study sample was derived from two large health systems in central Indiana. These systems operate more than 30 outpatient and specialty facilities. Approximately a third of outpatient care is provided by specialists at these health systems. Patients were included in the sample if they were diagnosed with T2DM by a primary or specialty care provider between their index and final encounter.

Data

We extracted patient demographic, encounter, and laboratory data from the Indiana Network for Primary Care (INPC), a statewide health information exchange data repository, for patients seen at any facility affiliated with two large hospital systems in central Indiana between 2017-2021. The INPC was established in 1994 as a repository for 38 health systems, 19,095 practices, and 19 million patients.[134]

Outcome Variable: EHR Data Timeliness

Timeliness was defined as the age of data elements that represent a patient's health state at a desired time of interest.[21,69] We quantified EHR data timeliness as the number of days between patient encounters where available EHR attribute updates for T2DM measurements were present including body mass index, body weight, glycated hemoglobin A1c, cholesterol, blood pressure, serum creatinine, and smoking status. Fewer days between EHR attribute updates indicate better timeliness at the time of the patient encounter.[135]

History of Patient Portal Use

We created a dichotomous measure of active patient portal use determined as patients having received and opened a secure message sent by a health care provider or provider organization. Patients without a history of active patient portal use did not receive or open a secure message sent by a health care provider or provider organization. Under this definition, patients who had been issued a patient portal account, but had never used the technology, were non-users.

Demographic Variables

We included patient characteristics identified in EHR data for subgroup analyses. The patient characteristics included patient age, patient race, patient sex, and insurance status. Age categories were measured as (1) 18-25; (2) 26-45; (3) 46-55; (4) 56-65; and (5) >65. Race was categorized as (1) non-Hispanic White; (2) African American/Black; (3) Hispanic; and (4) Other. We modeled patient sex as (1) Female; and (2) Male. Lastly, insurance status is categorized as (1) Commercial; (2) Medicare; (3) Medicaid; (4) Self-Pay; and (5) Other.

Analytic Methods

We computed descriptive statistics and cross-tabulations using frequencies, percentages, means, and standard deviations. One-way ANOVA analyses and chi-square tests examined bivariate relationships between independent and dependent variables. Negative binomial regressions with fixed effects estimated the association between patient portal use and EHR data timeliness. We used negative binomial models as appropriate for the over-dispersed count-based timeliness outcome (i.e., total number of days). Analyses accounted for linear time trends for 2017-2021 using yearly time dummy variables. Hausman tests were performed to determine proper model fit.[175] We fit separate regressions to examine variations in patient portal use and EHR data timeliness by patient age, patient race, patient sex, and insurance status. For all models, we reported marginal effects estimates. Statistical analyses were performed using Stata 16.0 (Stata Corp., College Station, TX, USA). The institutional review board (IRB) at Indiana University reviewed and approved this research protocol.

Sensitivity Analyses

To check the robustness of our findings, we undertook several sensitivity analyses. First, we repeated analyses by fitting fixed effects Poisson models with time dummies. Second, to check that results were not the product of extreme values, we fit separate fixed-effects regression where we truncated the dependent variable, EHR data timeliness, at the 90th and 95th percentiles. We tested a different measure of timeliness using a quotient that accounts for the mean attribute update times between patient encounters (Appendix 9).[21] Lastly, we adjusted our main analysis to account for pre- and post-COVID-19 status using a binary indicator.

RESULTS

Descriptive Analyses

The study sample included 35,759 patient-encounter date observations. The average age was 53.6 years. Most patients in the study were documented as non-Hispanic Black (47.4%), female (65.8%), and had a Charlson comorbidity score of 1 or higher (50.8%) (Table 18).

Nearly a third (31.3%) of the sample used the patient portal (Table 18). Mean EHR data timeliness, i.e., days between EHR attribute updates, was lower among patients who actively used patient portals (111.9 days) compared to patients who did not use patient portals at all (136.7 days; $p<0.001$) (Table 18). Timeliness was lower among patients who were documented as female, non-Hispanic White, >65 years of age, had a Charlson score of 3, and were insured by Medicare. EHR data timeliness improved during the study period. Use of patient portals steadily increased during the study period (2017-2021) (Figure 6).

Association between Patient Portal Use and Data Timeliness

Patient portal use was associated with an expected decrease in the EHR data timeliness (in days) of -0.036 ($p<0.001$) (Table 19). That is, patient portal usage was associated with more timely data. In stratified analyses, patient portal usage was generally associated with more time data for all patient groups. Specifically, patient portal use was associated with more timely data among female (ME= -0.041; $p<0.001$) and male (ME= -0.029; $p=0.017$) patients, and among patients who were enrolled in Medicare (ME= -0.033; $p=0.003$) and Medicaid (ME= -0.044; $p<0.001$) (Figure 7).

Patient portal use was negatively associated with the timeliness measure, i.e., data were timelier among non-Hispanic White patients (ME= -0.061; $p<0.001$) and non-Hispanic Black patients (ME= -0.035; $p<0.001$). Lastly, patient portal use was associated with expected decreases in the EHR data timeliness measure among patients aged 26-45 (ME= -0.040; $p=0.001$) and 56-65 (ME=-0.063; $p<0.001$).

Sensitivity Analyses

Sensitivity analyses were largely supportive of our primary analysis. In Poisson and negative binomial regressions that examined outliers in the dependent variable at the 90th and 95th percentiles, patient portal use was associated with expected statistically significant decreases in the EHR data timeliness measures (i.e., more timely data) and was comparable to the main model (Tables 18 & 19). We modeled the dependent variable using a quotient to account for mean time between updates to EHR data attributes including body weight, body mass index (BMI), blood pressure, cholesterol, serum creatinine, glycated HbA1c, and smoking status (Table 20). The quotient computes a timeliness value between 0 and 1 (Appendix 9). These results showed that patient portal use decreased EHR data timeliness albeit with smaller effect sizes due to the construction of the measure (Table 21). Patients who used portals were predicted to have fewer days between EHR attribute updates following the Public Health Emergency Declaration for the COVID-19 pandemic.

DISCUSSION

In examining the relationship between patient portal use and EHR data timeliness, we found that patient portal usage was associated with shorter time intervals between relevant EHR attribute updates. Further, findings from this study underscore the important role health information access plays in improving overall data access and quality, which is critical for patient engagement and managing T2DM.[92,100,176] A considerable number of health information technology (HIT) policies have been developed to ensure that patients have access to their patient information to improve engagement and chronic disease management. Prior research has examined how patient portal use improves engagement and self-management[92,173,177], and this work adds insights to the potential association with data quality.

Prior research has suggested that patient portal access and use enables data correction and verification.[139] Patient portal usage represents additional contact between the patient and the health care system for additional data collection.[35] Either or both of these mechanisms (i.e., verification or additional collection opportunities) could explain the association between usage and timeliness observed in this study. Regardless of the mechanism, improvements in EHR data timeliness is an important goal for the health care system, as timely data are important diagnostic indicators in the management of chronic disease. Additionally, the association between patient portal usage and timelier EHR data was consistent across numerous patient groups. Prior research that examined differential patient portal use outcomes among patient demographics, raising concerns that such technological interventions may not benefit all patients.[7,178,179] For example, research suggests lower usage of patient portals among the elderly, and the uninsured.[13,180]

Our findings suggest, that once using the patient portal, all patients may benefit in terms of higher data quality. In general, benefits only lagged among the older age groups. Generally, lower levels of health and computer literacy are concerning among older age groups.[13] Also, chronic conditions are prevalent among these age groups, and older patients tend to be higher utilizers of services. Health care organizations might consider specifically increasing patient portal education among this population and others that may not have experience or confidence in using digital health tools.

Health systems have increasingly adopted and used EHR technologies that support patient portal functionality to improve patient engagement.[26-29] The Meaningful Use (MU) program sponsored by the federal government incentivized the use of EHRs and patient portals to improve patient information access.[181,182] These efforts were refined in the Centers for Medicare and Medicaid Services (CMS) through its Promoting Interoperability reports. Additionally, the 21st Century Cures Act established guidelines for improving patient portal access to optimize chronic disease management and reduce unnecessary health care interactions.[183] The effects of greater patient portal access and use on EHR data quality remains underexamined. Our findings provide evidence that patient portal use does reduce time intervals between updates to relevant data. However, these results may be unique to patients who are closely monitored as part of chronic disease management programs and shifts from in-person care and test reporting to virtual delivery formats to limit disease exposure.[184] We observed effects during, before, and after the COVID-19 pandemic indicating that for any change in patient portal use, the time between EHR attribute updates decreased slightly in the months following the Public Health Emergency Declaration.

Past research examined the relationship between a single measure of EHR data timeliness and care quality.[133] We sought to operationalize EHR data timeliness using multiple metrics and approaches to demonstrate measure applicability in future research that examines the quality of EHR data and its fitness and subsequent use to examine critical associations. However, our methodology is not exhaustive nor is it conclusive to our chronic disease domain of interest. Thus, further research is needed to examine other factors that might contribute to variations in attribute update times and how they might be mitigated by the adoption and use of health information technologies like patient portals and EHRs. For example, active use of patient portals may also improve EHR data completeness in chronic disease settings, but improvements may vary in non-chronic disease settings and across provider types where the informed presence of disease indicators is less discernable.

Limitations

There are several limitations in this study. First, we measured timeliness in terms of days between updates. The biomedical informatics literature has other definitions of timeliness, and findings from this study may not have been similar with timeliness. In addition, timeliness is only one dimension of quality, and we cannot speak to other quality aspects. Second, we may not be able to generalize to other conditions and data types. We specifically selected T2DM patients because patients managing chronic disease have more frequent visits and physiological measurements. Patient groups relying on fewer diagnostic measures or those that are of shorter duration may not see the same associations in timeliness from patient portal use.

CONCLUSION

This study found that patient portal use was associated with EHR data timeliness in T2DM care. Reducing EHR data timeliness is an important health care outcome that may indicate that improvements in overall data quality are bolstered by the active use of patient portals and other digital health solutions.

Table 17. Descriptive Patient Characteristics and EHR Data Timeliness

	EHR Data Timeliness for All Patients			
	(%)	Mean Timeliness (in days) (SE)	95% CI	P value
Patient portal use				<0.001
Yes	31.3	111.9 (19.0)	73.9-149.8	
No	68.7	136.7 (27.6)	81.8-191.6	
Patient Sex				<0.001
Female	65.8	120.9 (23.8)	73.5-168.3	
Male	34.2	128 (23.7)	80.9-175.1	
Patient Race/Ethnicity				0.004
White	28.3	112.9 (25.1)	63.1-162.8	
African American/Black	47.4	129.5 (22.3)	85.0-173.9	
Hispanic	16.6	158.5 (56.4)	46.2-270.8	
Other	3.9	42.8 (16.6)	9.6-75.9	
Age				0.007
18-25	5.7	105.3 (36.5)	32.5-178.1	
26-45	39.3	155.4 (33.6)	88.4-222.3	
46-55	25.8	107.9 (27.9)	52.2-163.5	
56-65	20.1	101 (26.9)	47.5-154.5	
>65	6.7	113 (52.6)	8.5-217.9	
Charlson Comorbidity Score				0.021
0	26.9	165 (47.2)	71.3-259.2	
1	50.8	115.5 (19.4)	86.9-154.1	
2	17.3	121.8 (42.0)	38.2-205.5	
3	4.1	71.2 (22.3)	26.9-115.4	

Insurance				<0.001
Commercial	25.2	136.6 (42.6)	51.7-221.5	
Medicare	24.1	81.9 (16.5)	49.1-114.8	
Medicaid	41.4	134.4 (23.9)	86.6-182.1	
Self-pay	6.0	151.6 (48.7)	54.6-248.6	
Other	3.2	15.5 (5.5)	4.5-26.5	
COVID-19 Status				<0.001
Pre-COVID-19	67.2	66.5 (0.3)	65.9-67.1	
Post-COVID-19	32.8	71.2 (0.5)	70.2-72.1	
Year				<0.001
2017	42.3	112.2 (15.4)	81.5-142.9	
2018	21.9	152.4 (45.9)	60.9-243.8	
2019	21.4	128.9 (49.9)	29.3-228.4	
2020	13.6	123.8 (41.6)	41.1-206.6	
2021	0.7	43 (8.3)	15.2-68.3	

Figure 5. Mean Timeliness by Year and Patient Portal Use (2017-2021)

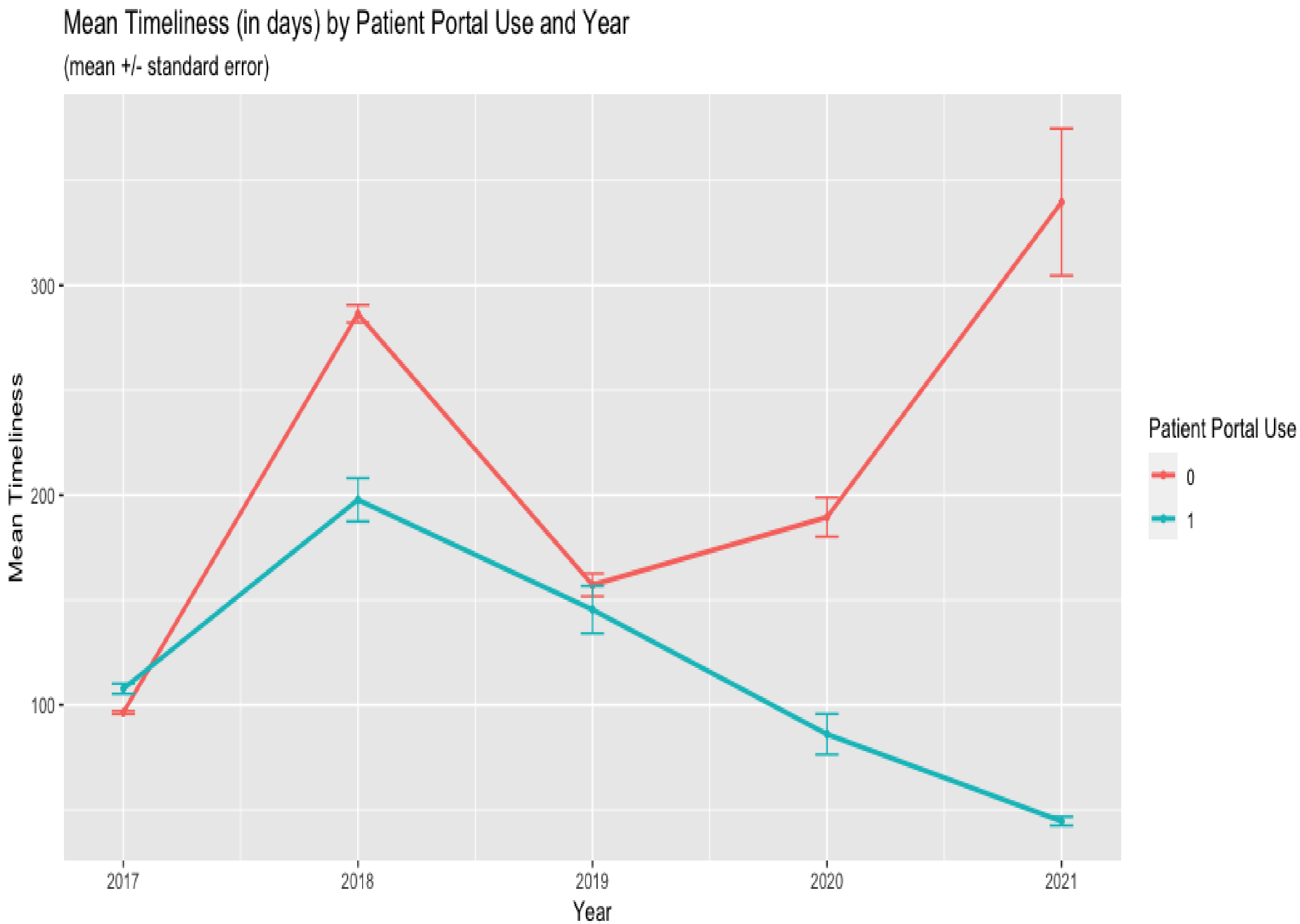


Table 18. Association between Patient Portal Use and EHR Data Quality – Conditional Negative Binomial Regression with Fixed-effects (reported as marginal effects estimates)

	EHR Data Timeliness Model 1. Main Negative Binomial Regression with Fixed effects				
	ME	se	95% CI		P value
Patient Portal Use	-0.036	0.007	-0.049	-0.022	<0.001
Year (ref: 2017)					
2018	0.088	0.009	0.071	0.105	<0.001
2019	0.038	0.008	0.021	0.054	<0.001
2020	0.029	0.009	0.012	0.047	0.001
2021	0.027	0.014	-0.007	0.055	0.056

Figure 6. Stratified Fixed-effects Negative Binomial Regression Analyses: Race, Age, Insurance Status, Sex, and the Main Regression Model

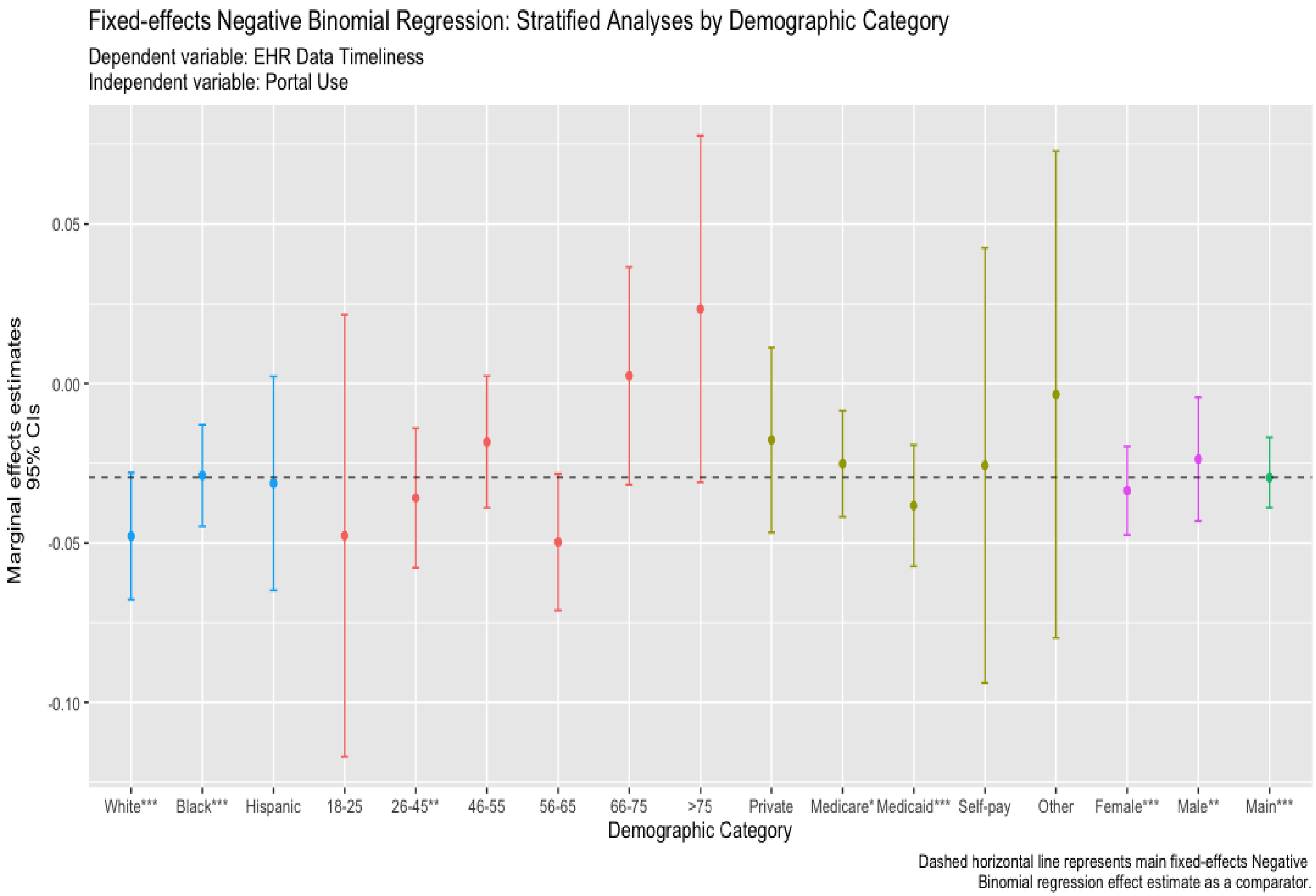


Table 19. Sensitivity Analyses for Examining the Relationship between Patient Portal Use and EHR Data Timeliness - Conditional Negative Binomial Regression with Fixed-effects Truncated at the 90th Percentile

	EHR Data Timeliness Truncated at the 90th Percentile Model 1. Negative Binomial Regression with Fixed effects			
	ME	95% CI		P value
Patient Portal Use	-0.025	-0.036	-0.015	<0.001
Year (ref: 2017)				
2018	0.439	0.031	0.057	<0.001
2019	0.015	0.003	0.028	0.017
2020	0.014	0.001	0.027	0.043
2021	0.034	0.015	0.055	0.002
Constant	0.561	0.549	0.549	0.574

Table 20. Sensitivity Analyses for Examining the Relationship between Patient Portal Use and EHR Data Timeliness - Conditional Negative Binomial Regression with Fixed-effects Truncated at the 95th Percentile

	EHR Data Timeliness Truncated at the 95th Percentile Model 1. Negative Binomial Regression with Fixed effects			
	ME	95% CI		P value
Patient Portal Use	-0.059	-0.069	-0.049	<0.001
Year (ref: 2017)				
2018	0.054	0.041	0.067	<0.001
2019	0.012	-0.002	0.025	0.077
2020	-0.004	-0.018	0.009	0.542
2021	-0.019	-0.042	0.004	0.101
Constant	0.245	0.232	0.257	<0.001

Table 21. Sensitivity Analyses for Examining the Relationship between Patient Portal Use and EHR Data Timeliness - Conditional Negative Binomial Regression with Fixed-effects using Timeliness Quotient (See Appendix 9)

	EHR Data Timeliness Quotient Accounting for Mean Attribute Update Time Model 1. Fixed-effects Regression model			
	me	95% CI		P value
Patient Portal Use	-0.000	-0.000	-0.000	0.800
Year (ref: 2017)				
2018	-0.000	-0.000	-0.000	<0.001
2019	-0.000	-0.000	-0.000	<0.001
2020	0.000	-0.000	-0.000	<0.001
2021	0.000	-0.000	-0.000	<0.001
Constant	0.001	0.001	0.001	0.001

Table 22. Patient Portal Use Pre- and Post-COVID-19 Emergency Declaration
Dependent Variable: EHR Data Timeliness

	Fixed-effects Negative Binomial Regression for EHR Data Timeliness			
	ME	95% CI		P value
COVID-19 Status (ref: Pre-COVID-19)				
Post-COVID-19	-0.016	-0.028	-0.004	0.011

CHAPTER 5: CONCLUSIONS

Electronic health record (EHR) data are consistently used in program evaluation, quality improvement, and secondary empirical research, yet the "fitness for reuse" of elements that comprise these data are rarely and rigorously examined.[17,19-21] More data are being generated from new care delivery formats and increasing use of digital healthcare tools, including patient portals, remote patient care and virtual modalities.[53,80,138] However, research streams examining the quality of such data are still developing. The purpose of this dissertation was to examine whether and to what extent levels of EHR data quality are associated with healthcare outcomes and if EHR data quality is improved by using health information technologies (HIT). Chapter 2 derived EHR data quality dimensions and metrics to operationalize timeliness, completeness, and information density measures among patients managing T2DM with and without a history of telehealth use. I sought to quantify EHR data quality dimensions in a regimented clinical domain by applying existing frameworks informed by the information systems (IS) and health services and informatics literature.[21,65,75] In Chapter 3, we used a panel design to examine the effect of primary care laboratory test ages, a measure of timeliness, on subsequent inpatient hospitalizations and ED Visits. Lastly, Chapter 4 used a panel analysis to examine the effect of patient portal use on EHR data timeliness. Timeliness was quantified as the EHR attribute update time for relevant T2DM measurements and laboratory tests including body mass index (BMI), body weight, cholesterol, serum creatinine, glycated hemoglobin, blood pressure, and smoking status.

Metrics used to quantify EHR data quality have generally been applied in chronic disease settings where patient health data are routinely collected: T2DM, chronic obstructive pulmonary disease, and cardiovascular disease.[37,39-40] The focal clinical domain in this study was T2DM. Importantly, metrics and quotients used to operationalize EHR data quality measures must account for how data are collected, the setting in which these data are collected, and for what purpose.[21,69] Research indicates that more investigators are examining the quality of clinical data prior to reuse, however, these methods and approaches remain underused.[39,135] Moreover, the mechanics to audit clinical and EHR data quality are improving in their reliability, validity, and generalizability which enables accounting for task dependencies; chiefly, secondary health services and informatics research.[30,65] We tested this assumption by modeling timeliness, a critical EHR data quality dimension for chronic disease management, interdependently as an independent variable and dependent variable in two separate analyses.

Summary of Key Findings

In Chapter 2, entitled "Measuring EHR data quality in telehealth and office-based Diabetes care", I operationalized three EHR data quality dimensions for patients managing type 2 diabetes to determine fitness for use in secondary research. I specifically operationalized clinically relevant common measures of completeness, information density, and timeliness among telehealth and office-based primary care visits to determine differences between care delivery models and formats. In a cross-sectional analysis of EHR data, we found that data quality for adult patients managing T2DM was generally similar or more improved among patients with a history of telehealth use. We specifically found that EHR data were generally more timely, more complete, and T2DM measurements were generally more uniform where patients had a history of telehealth use compared to patients with no history of telehealth use during their medical visits.

Examining EHR data quality across care delivery modalities and complex patients is critically important to improving comprehensive information access and some healthcare outcomes. Although EHR data quality measures more improved for patients who had a history of telehealth care, there is still the possibility for improving EHR data quality across care delivery formats. Clinical guidelines and workflow optimization have been determined to improve the quality of data documented by healthcare providers. Care delivery is transitioning from in-person and telehealth delivery to hybrid formats based on self-reported data from a national survey.[188]

Chronic disease settings are appropriate for examining the quality of data documented and produced to inform care. Primary care and specialty providers who treat patients managing chronic diseases like T2DM have specific clinical guidelines that structure information collection practices.[185] For example, some providers may intentionally exclude some data and collect meaningful patient health data meant to aid in monitoring health status which is described as "informed presence".[186,187] We considered this description in how we constructed measures, specifically for EHR data completeness.

Completeness was operationalized using structured data elements from the EHR as any T2DM measurement identified and flagged as complete where a laboratory value or indicator was available at its first indication for sequential patient encounters. Constructing the completeness measure using this approach was especially relevant given data included in this study captured COVID-19 pandemic care delivery transitions from in-person to mostly remote and virtual delivery formats. We computed a measure of timeliness that was clinically interpretable as the number of days between a patient encounter and the most recent measurement dates. Lastly, we used a metric to quantify information density described in the methods of Aim I. Simply, information density is a measure of completeness that accounts for the irregular nature of patient measurements taken over time. Because these measures were operationalized for data that covered the COVID-19 pandemic, we could analyze important changes in chronic disease care. Thus, we were able to capture trends in EHR data quality among patients with and without a history of telehealth use before and after the Public Health Emergency Declaration signed into a presidential Executive Order in March 2020. Using these measures, we were able to linearly determine the quality of EHR data in two care modalities over time.

Chapter 3, entitled "Effects of Primary Care Laboratory Age on Healthcare Utilization in Type 2 Diabetes Care," modeled EHR data timeliness as an independent variable to determine its relationship with subsequent expected counts of inpatient hospitalizations and emergency department Visits. We defined timeliness as the most recent T2DM laboratory test date preceding an encounter date to produce a total number of days or age of the laboratory test at the next encounter. The information systems success model determined individual impact and data quality variables are interdependent which enabled us to examine bidirectional relationships between EHR data timeliness and healthcare outcomes, like total expected counts of hospitalizations in inpatient and ED settings. In this analysis, we used a panel design that controlled for the linear trends, patient, and encounter characteristics. Our findings from this study indicate that older laboratory tests are associated with an increase in the expected total counts of inpatient hospitalizations and ED Visits per patient per year. The effect size was largest among ED

Visits, and the marginal changes between laboratory test age categories indicated that tests older than a year increased the expected ED Visits by 120 percentage points per patient per year. These results emphasize that the age of laboratory tests performed in primary care or outpatient settings serve as important indicators of future healthcare utilization. However, these findings may be confounded by other unobserved factors given healthcare utilization, particularly among patients managing chronic disease, are subject to variations that cannot be controlled for in the study design. Thus, further research is needed to understand utilization patterns among patients managing chronic diseases, especially where repeated measurements and close monitoring are needed.

In Chapter 4, entitled “Patient portal use and EHR data timeliness in T2DM care”, we examined associations between patient portal use and changes in the level of EHR data timeliness. That is, we determined whether use of patient portal technologies improved the time between EHR updates for relevant T2DM measurements and laboratory tests. These measurements include body mass index, body weight, serum creatinine, glycated HbA1c, cholesterol, blood pressure, and smoking status. We oriented analyses and modeling considering the structure of the Information Systems Success Model. We modeled EHR timeliness as an outcome variable in this study. A panel design was used to conduct longitudinal analyses of patient portal use among patients managing T2DM. Given there are differential patient portal use patterns among patients by age, sex, insurance status, and race categories, we conducted stratified analyses among these subgroups using separate fixed effects negative binomial regression models. We found that patient portal use decreases the average timeliness of EHR attribute updates among patients managing T2DM. These findings suggest that increased use of patient portals among the full T2DM population improves time between EHR attribute updates. Results from stratified analyses demonstrates that patient portal use is more effective among some patient demographic categories. For example, patients who were documented as Female, non-Hispanic White, 26-45 years of age, and were Medicaid recipients had lower levels of EHR data timeliness compared to the overall sample. These results were statistically significant at the $p = 0.05$ level. Although we observed improvements in this definition of EHR data timeliness, organizations should deploy strategies to continue improving adoption and use of digital healthcare tools among patients managing chronic disease.

Summary of Contributions

Current health services research is utilizing more approaches to assess the quality of clinical data reused to conduct such tasks like empirical research, quality improvement, program evaluation, and clinical guideline development. This has been especially true for EHR data where unobserved context is difficult to parse.[186] Our findings serve to inform data quality assessments and approaches by first measuring data quality dimensions and modeling them in secondary research. This has enabled us to specify tasks and research questions to determine whether data quality measures are truly interdependent as the Information System Success Model suggests. In Chapter 2, we applied metrics to create measures in a chronic disease domain that are applicable to similar complex disease domains. Past research conducted similar assessments of data quality, but primarily focused on completeness.[74-76] Conversely, timeliness is an important indicator of data access for patients who need their information to make important decisions about their health. To my knowledge, there are no studies that quantify timeliness, completeness, and information density using EHR data inputted on behalf of patients managing T2DM. Results from Chapter 2 provide researchers and clinical practitioners with three separate approaches to quantifying EHR data quality in a chronic disease domain. Additionally, these results contribute to evolving frameworks and guidelines developed to improve care management in virtual and in-person settings.[132,163,185] Systematically assessing the quality of EHR data in chronic disease domains serves to improve clinical decision making, which relies on time-sensitive patient information, and secondary reuse to refine inferences or conclusions made from using such data.[21,75]

Chapter 3 demonstrates how EHR data quality dimensions can be modeled to determine whether these measures, indeed, are fit for use in secondary research largely dependent on the source. We derived our data from a state-wide health information exchange that conducts its own preprocessing. Thus, we are confident that our results reflect near-truths about patient health care outcomes and quantifiable measures of EHR data quality. The age of a laboratory test produced in primary care or outpatient settings is an important timeliness measure that is underexplored in the literature. This is a valuable addition to the health services and informatics literature supporting few past studies that examined the influence of EHR data quality on care quality.[133] However, we found a statistically significant relationship between EHR data timeliness and two health care outcomes. To our knowledge this was the first study to rigorously examine these relationships. Findings from this study inform the formulation of clinical guidelines that direct providers to increase monitoring for patients who receive concerning laboratory test results. Namely, ensuring uniformity in testing both in frequency and setting given primary care and outpatient laboratory test are associated with subsequent inpatient hospitalizations and ED Visits, two outcomes clinical providers work to avoid.

Chapter 4 examined how patient digital health use improved timeliness as the sole outcome. We ascertained patient portal use data to determine if this digital healthcare tool improved EHR data timeliness. Our results are a major contribution to the EHR data quality and broader health services and informatics literature because it helps inform how the intermediary digital healthcare tools used to share patient information improve upon patient information. Chiefly, information in the patient portal and EHR are improved by enabling patients with viewing, verification, and correction capabilities that shortened the time between EHR attribute updates examined at the patient-encounter date level.[139,172] Although past descriptive studies have enumerated and described the types of patients who are most likely to use patient portals and correct their information, this is the first study to examine the extent to which patient portals improve an EHR data quality dimension like timeliness. Findings show there are secondary benefits of adopting and using advance functions of EHRs beyond improve care delivery pursuant to meeting Meaningful Use recommendations. By improving access and use of patient portals, healthcare organizations enable optimize care processes and information sharing that aids in refining chronic disease management as is promoted by federal agencies like CMS and the Office of the National Coordinator for Health Information Technology.

Future Directions

This study quantified measures of completeness, timeliness, and information density or uniformity. We then modeled EHR data timeliness as an independent and dependent variable in separate analyses. However, this dissertation did not extensively examine other dimensions of data quality. Additionally, we did not model operationalized measures of completeness in these studies. However, EHR data are well-suited to examining other data quality dimensions, especially in chronic disease domains that rely on uniform laboratory testing and patient visits. For example, vital measurements like blood pressure and laboratory values may specifically be recorded at specific times depending on whether a patient is being monitoring for exacerbations of chronic disease.[186] How missingness manifest in EHR data are also dependent on where and how measurements are collected. [187] Future studies should explore the extent to which other dimensions of data quality are present in EHR data and the extent to which measures influence health and healthcare outcomes. The three studies that constitute this dissertation focused solely on structured data elements. Future research might consider operationalizing data quality measures using unstructured notes documented by T2DM providers. Measures, including concordance and correctness, are commonly developed from clinical notes. Moreover, clinical notes may serve to supplement structured EHR data through imputation methods where structured data are unavailable. In preliminary analyses, we were able to link only a small percentage of clinical notes to the EHR dataset. Improving data inclusion criteria for unstructured and structured datasets will enable application of advanced informatics methods to text-mine and analyze both data formats. Likewise, correctness requires extensive data preprocessing that is informed by clinical domain expertise.[69] Few, if any, studies have used measures of correctness due to the highly computational nature of assessing correction, which includes Levenshtein's Distance; a method for examining word construction for errors.[69]

APPENDIX

Appendix 1. T2DM Term Frequency-Inverse Document Frequency Dictionary (Not included in final analyses)

T2DM TERM FREQUENCY-INVERSE DOCUMENT (tf-idf) FREQUENCY TERM DICTIONARY		
Term type	Term group	Existing NLP terms
T2DM medication terms	A	furosemide, epixaban, avastin, atenolol, entresto, losartan (high blood pressure), clopidogrel, latanoprost, minoxidil, humalog, gabapenton, acarbose, keratopathy
	B	keratopathy
T2DM laboratory terms	C	BMI, bmi, hemoglobin A1C, HbA1c, serum cholesterol, glucose, glycemic, venipuncture
T2DM diagnostic terms	D	hydrodistention, presbyopia, ambylopia, oropeza (spanish for family history), diabetes, hypokalemia, carotid, epigastric, epigastirc, discolored, foot, orthostasis, glaucomatous, macular, gastroparesis, degenerative, rheumatology, uveitis, hyperparathyroidism, thyroglossal, pterygium, cornea, morgagnian, hypermetabolic, nephrolithiasis, prediabetes

Telehealth/Telemedicine terms	E	virtual, virtual-phone, phone, video visits, remote monitoring, remote patient monitoring
T2DM Care Process terms	F	dictation, dic, dict, endocrinology, behavior, collaboration, referral, coordination
T2DM Treatment terms	G	administered, resolved, dialysis, attempted, informed, amputated, hemodialysis
	H	treated with, receiving treatment with, treated by, receiving treatment by, tx with, receiving tx with, tx by, receiving tx by, examined
	I	receiving
Guidance on term combinations	Rule Group	
[T2DM Treatment term] + [T2DM laboratory term]	J	[G+C], [H+C], [I+C]
[T2DM Treatment term] + [T2DM medication term]	K	[G+A], [H+A], [I+A], [G+B], [H+B], [I+B]
[T2DM Treatment term] + [T2DM diagnostic term]	L	[G+D], [H+D], [I+D]

[T2DM Care Process term] + [T2DM diagnostic term]	M	[F+D]
Terms	Group	
T2DM medication terms	A, B	
T2DM laboratory terms	C	
T2DM diagnostic terms	D	
Telehealth/Telemedicine terms	E	
T2DM Care Process terms	F	
T2DM Treatment terms	G, H, I	

Natural Language Processing (NLP) Reference Standard Development Framework

Data Quality in Type 2 Diabetes Mellitus Study

Disclaimer: Please do not copy, download, or discuss any patient information in the clinical notes. Violating confidentiality and privacy of patient information is subject to penalties under the Health Insurance Portability and Accountability Act (HIPAA) of 1996 and related regulations ensuring protection of patient health information. Please feel free to contact me if you have any questions or concerns.

What is natural language processing and why is it important?

Natural language processing (NLP) “refers to the branch of computer science—and more specifically, the branch of artificial intelligence or AI—concerned with giving computers the ability to understand text and spoken words in much the same way human beings can.

NLP combines computational linguistics—rule-based modeling of human language—with statistical, machine learning, and deep learning models. Together, these technologies enable computers to process human language in the form of text or voice data and to ‘understand’ its full meaning, complete with the speaker or writer’s intent and sentiment”[1]

Why is NLP important in health care? NLP provides computational techniques to generate structured information from clinical notes that most often constitute free-text data in electronic health records (EHRs) and other clinical reports (lab tests, medications, summary reports)[2]

What is a reference standard?

A reference standard for NLP provides a real-world data comparison for extracted clinical data. If no clear reference or gold standard exists, researchers/investigators can develop one through manual chart review. In manual chart review, one or more independent and blinded abstractors manually review patient clinical notes from a randomly sampled set of patients. Similarly, a gold standard is a real-world accurate reflection of patient data (e.g., patient registry data). In the absence of a gold standard, we use a composite reference standard as mentioned above. [3]

Reference standards are used across domains to measure variability in different data sources (clinical narrative vs. coded diagnoses) for consistency, accuracy, and prominence. What is generally observed in manual medical chart review may not reflect computational queries or text extraction of free-text. Creating a reference standard provides a near real-world comparison.

Chart Review Procedures

The following describe the chart review procedures for the current study examining quality of clinical data among patients with Type 2 Diabetes Mellitus. You are one of two abstractors who have chosen to participate in this study. Your responsibility is to ensure confidentiality of patient information as outlined in CITI training and in the IRB, which you have been added to for participation in this study.

Disclaimer: Please do not copy, download, or discuss any patient information in the clinical notes. Violating confidentiality and privacy of patient information is subject to penalties under the Health Insurance Portability and Accountability Act (HIPAA) of 1996 and related regulations ensuring protection of patient health information. Please feel free to contact me if you have any questions or concerns.

How to Access Patient Charts

Upon completion of CITI and HIPAA training you will be given access to a link to a OneDrive file whereby you will be able to review notes and input specific patient information required to sufficiently assess the notes.

I. Open Excel .csv file that is available in the OneDrive folder

   A. Only two individuals will be given permission to access this folder. Sharing information with anyone that is not IRB approved or part of this project will be subjected to HIPAA violations.

II. In this T2DM notes file, you will find the following eight variables and their description listed from left to right:

A. sid: unique patient identifier

B. encounter_id: number/identifier of encounter associated with the patient unique identifier (sid)

C. contact_date: date of provider/patient contact

D. note_id: identifier for each patient note that is associated with both the encounter_id and sid variable

E. Note_text: full text note for the patient based on unique patient identifier, note_id, and encounter_id

F. Note_year: year the note was written, which is derived from the contact_date variable

G. Note_month: month the note was written, which is derived from the contact_date variable

III. There are column headers with no empty cells that will require you to complete upon reading each chart. Lists of expected selections are already programmed into each cell/row. The name of these column headers/questions are as follows:

A. Is this a type 2 diabetes encounter (includes labs/tests, medications, symptoms)?

1. "yes" if encounter is T2DM related
2. "no" if encounter is <u>NOT</u> T2DM related

B. Is this a telehealth (telephone, video, audio, or patient portal) encounter?

1. "yes" if note indicates telehealth encounter

2. “no” if note DOES NOT indicate telehealth encounter

C. Was family history mentioned in note?

1. “yes” if family history is mentioned
2. “no” if family history is NOT mentioned

D. Was smoking mentioned in note?

1. “yes” if smoking is mentioned
2. “no” if smoking is NOT mentioned

E. Was body weight mentioned in note?

1. “yes” if body weight is mentioned
2. “no” if body weight is NOT mentioned

F. Was BMI mentioned in note?

1. “yes” if BMI is mentioned
2. “no” if BMI is NOT mentioned

G. Was blood pressure mentioned in note?

1. “yes” if blood pressure is mentioned
2. “no” if blood pressure is NOT mentioned

H. Was HbA1c mentioned in note?

1. “yes” if HbA1c is mentioned
2. “no” if HbA1c is NOT mentioned

I. Was cholesterol mentioned in note?

1. “yes” if cholesterol is mentioned
2. “no” if cholesterol is NOT mentioned

J. Was serum creatinine mentioned in note?

1. “yes” if Serum creatinine is mentioned
2. “no” if Serum creatinine is NOT mentioned

K. What is the note type?

1. “Admission”
2. “Discharge”
3. “Progress note”
4. “None” if note type is unavailable

L. What type of clinician/provider wrote the note?

1. “Physician”
2. “Nurse”
3. “Physician Assistant”
4. “Undetermined” if there is no mention of clinician/provider type

M. What was the patient’s disposition?

1. “Alive” if note indicates patient is alive
2. “Deceased” if note indicates patient is deceased
3. “Patient Transferred” if patient was transferred to another location/facility
4. “Undetermined” if disposition is undetermined

N. Please list any uncommon terms/words in the note that are not commonly used to describe a Type 2 Diabetes patient encounter. ________________________

Term Dictionary, Explained

A Term dictionary is a set of terms or words that are commonly found in a corpus of condition specific text(s). Term dictionaries are often referred to as “lexicons” or “term lists” in the context of NLP research and developed as a result of reviewing patient notes/charts.

Building a term dictionary allows investigators to compile commonly used words/terms and identify lexical patterns for use in both rule-based classification and using machine learning approaches. Further, identified terms are critical to refining NLP procedures. This study will use both manual and automated term dictionary development. Manual requires abstractors maintain a list of terms that appear in the notes that are important to the overall study. This would include manually noting terms common to type 2 diabetes care (e.g., Metformin, HbA1c). Automated term development is conducted using term frequency-inverse document frequency (tf-idf). This process is enabled by statistical software to determine high frequency and low frequency terms using a mathematical formula.

Term Dictionary Development Procedures

I. In a Google Sheets sheet titled, Term Dictionary, a predefined list of common T2DM terms will be listed in the second column. As you are reviewing notes, please document any lexical (how terms are used or T2DM conditions are described) variations and spelling variations. This will allow us to refine the rule-based NLP approach to include differently spelled T2DM-related terms and to determine lexical variations within the notes.

II. Make sure to include partial term spellings, where available (e.g., gluc for glucose).

Appendix 3. T2DM Data Agreement and Cohen's Kappa Coefficients

Question	First agreement	First agreement due to chance	First kappa	Second agreement	Second agreement due to chance	Second kappa	Agreement diff	Agreement due to chance diff	Kappa difference	Kappa - 3rd reviewer
Is this a Diabetes encounter? (Include mentions of Rx, diagnoses, or labs)	0.88	0.59	0.70	0.89	0.57	0.75	0.02	-0.01	0.05	0.92
Is this a T2DM encounter?	0.88	0.59	0.70	0.89	0.57	0.74	0.02	-0.01	0.05	0.86

Is this a telehealth (telephone, video, audio, or patient portal) encounter?	0.93	0.93	-0.00376	0.97	0.87	0.80	0.04	-0.06	0.80	0.88
Was family history mentioned in the note?	0.97	0.92	0.57	0.97	0.92	0.59	0.00	0.00	0.02	0.96
Was smoking mentioned in the note?	0.96	0.87	0.69	0.95	0.87	0.65	-0.01	-0.01	-0.04	0.92

Was BMI mentioned in the note?	0.97	0.91	0.68	0.97	0.89	0.68	-0.01	-0.02	0.00	0.91
Was blood pressure mentioned in the note?	0.88	0.75	0.52	0.90	0.72	0.63	0.02	-0.03	0.11	0.82
Was HbA1c mentioned in the note?	0.95	0.85	0.67	0.95	0.84	0.71	0.00	-0.01	0.04	0.97
Was cholesterol mentioned in the note?	0.98	0.93	0.69	0.98	0.92	0.68	0.00	-0.01	0.00	0.95
Was serum creatinine	0.98	0.93	0.62	0.98	0.93	0.64	0.00	0.00	0.02	0.87

mentioned in the note?										
Was body weight mentioned in the note?	0.94	0.78	0.72	0.94	0.73	0.77	0.00	-0.05	0.06	0.94
What was the note type?	0.84	0.70	0.46	0.89	0.74	0.59	0.05	0.04	0.12	0.91
Is this an outpatient or inpatient encounter?	1.00	1.00	1	1.00	1.00	1	0.00	0.00	0.00	1
What type of clinician/provider wrote the note?	0.85	0.55	0.66	0.86	0.55	0.68	0.01	0.00	0.02	0.8

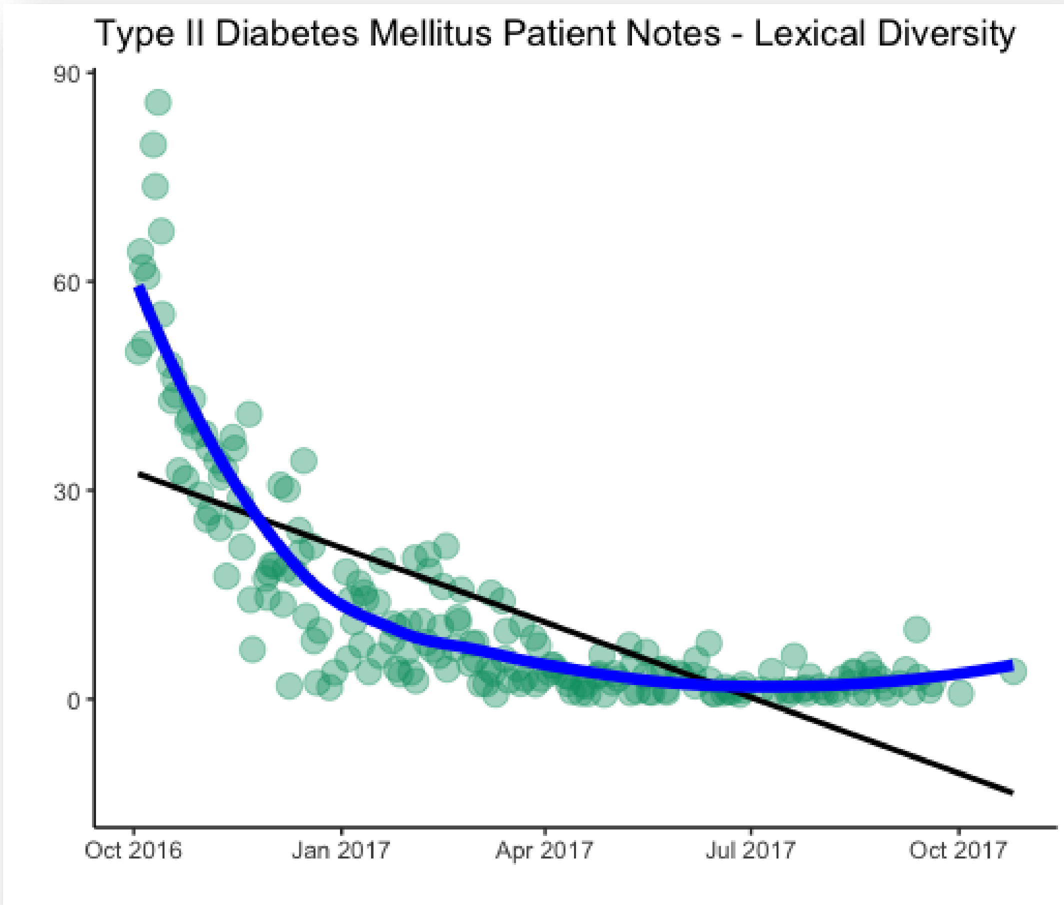
What was the patient's disposition?	0.89	0.89	0.89	0.96	0.96	0.89	0.07	0.07	0.00	0.89

Appendix 4. T2DM NLP Performance Metrics (Not included in final analyses)

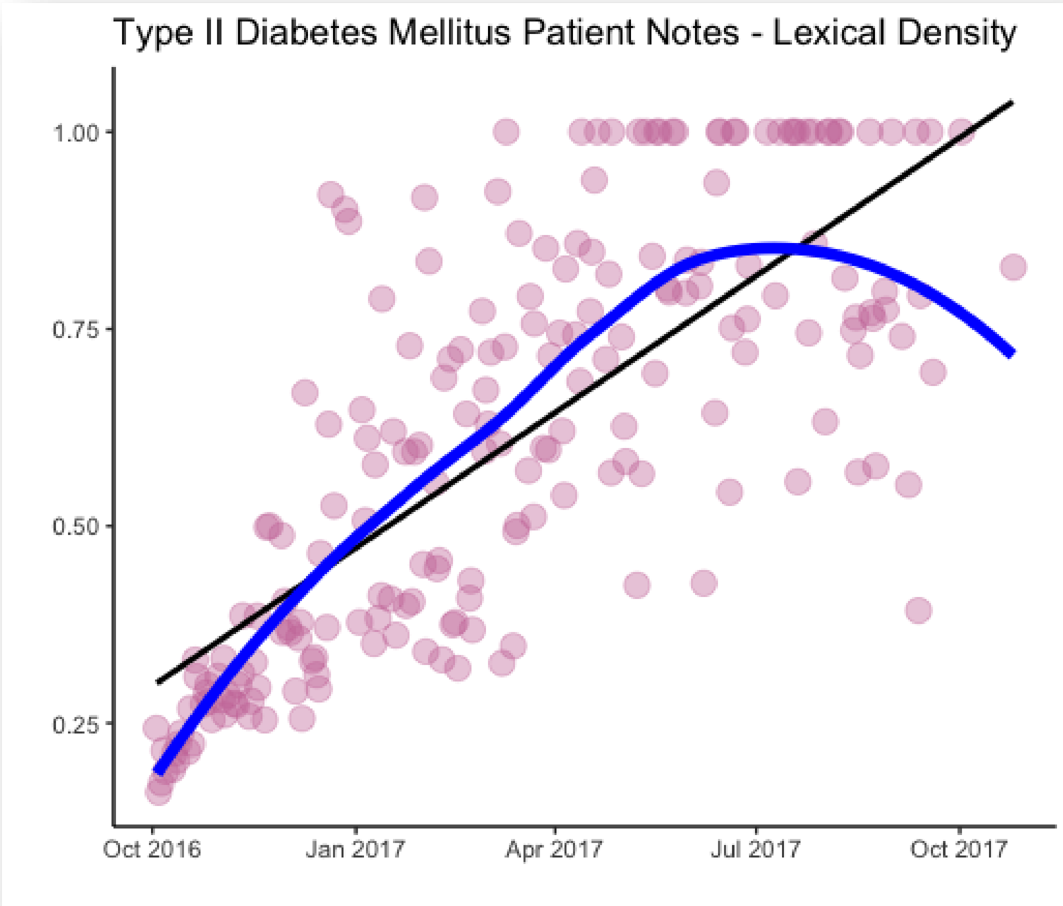
	Rule-based				Support Vector Machine				Naive Bayes			
Type of mention in note	Accuracy	Precision	Recall	F1 Score	Accuracy	Precision	Recall	F1 Score	Accuracy	Precision	Recall	F1 Score
T2DM encounter	0.94	0.88	0.97	0.95	0.96	0.96	0.99	0.98	1.00	0.99	1.00	1.00
Telehealth (patient portal, video, audio)	0.98	0.95	0.83	0.86	0.98	0.98	0.98	0.98	0.92	0.95	0.99	0.97
Family history	1.00	1.00	0.92	0.96	0.99	0.99	0.99	0.99	0.92	0.99	0.97	0.98
Smoking	0.99	1.00	0.86	0.93	0.99	0.99	0.99	0.99	0.97	1.00	1.00	1.00
Body Mass Index	0.99	0.96	0.84	0.89	1.00	0.97	0.99	0.98	0.98	0.99	0.99	0.99
Blood pressure	0.95	0.97	0.73	0.84	0.94	0.95	0.95	0.95	0.96	0.98	0.98	0.98
HbA1c	0.99	0.98	0.94	0.96	0.99	0.99	0.98	0.98	0.95	0.99	0.95	0.99
Cholesterol	1.00	1.00	0.91	0.95	1.00	0.99	0.99	0.99	0.97	0.98	0.98	0.98
Serum Creatinine	0.99	0.90	0.77	0.83	0.99	0.97	0.98	0.98	0.91	0.93	0.95	0.95
Body Weight	0.98	0.96	0.94	0.95	0.98	0.98	0.99	0.98	0.96	0.99	0.97	0.98

Clinical note type (progress, admission)	0.98	0.99	0.86	0.91	0.98	0.95	0.99	0.95	0.95	0.97	0.98	0.98
Note author (physician, nurse, NP)	0.91	1.00	1.00	1.00	0.93	0.95	0.96	0.96	0.77	0.84	0.90	0.87
Patient disposition	0.96	0.96	1.00	0.97	0.97	0.98	0.97	0.97	0.98	0.99	0.96	0.98

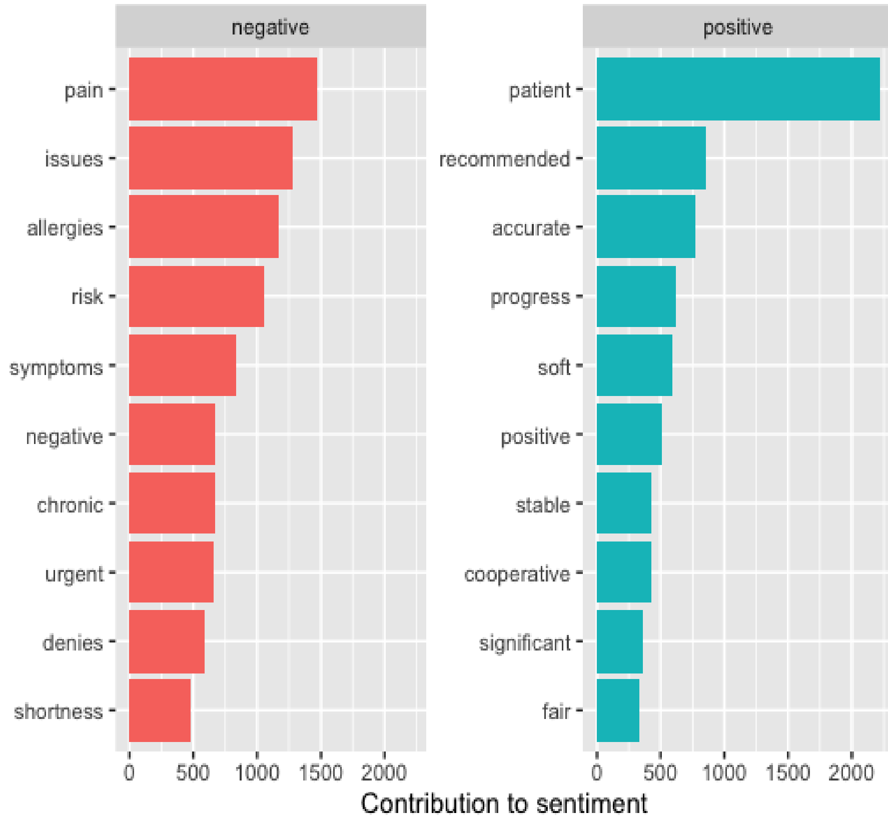
Appendix 5. NLP Lexical Analysis for Concordance Measure – Lexical Diversity



Appendix 6. NLP Lexical Analysis for Concordance Measure – Lexical Density



Appendix 7. Clinical Note Sentiment Analysis



Appendix 8. T2DM Data Quality Measure Correlation Analyses

	Timeliness	Completeness	Sperrin’s I
Timeliness	1.00	-0.008**	0.417***
Completeness	-0.008**	1.00	-0.053***
Sperrin’s I	0.417***	-0.053***	1.00

p<0.05*; p<0.01**;p<0.001***

Appendix 9. Timeliness Quotient

$$\text{Timeliness} = \frac{1}{(\text{mean attribute update time}) \times (\text{attribute age}) + 1}$$

REFERENCES

1. Adler-Milstein J, Jha AK. HITECH Act Drove Large Gains In Hospital Electronic Health Record Adoption. Health Aff . 2017;36(8):1416-1422.

2. What information does an electronic health record (EHR) contain? Accessed November 4, 2020. https://www.healthit.gov/faq/what-information-does-electronic-health-record-ehr-contain

3. Pendergrass SA, Crawford DC. Using Electronic Health Records To Generate Phenotypes For Research. Curr Protoc Hum Genet. 2019;100(1):e80.

4. Leveille SG, Mejilla R, Ngo L, et al. Do Patients Who Access Clinical Information on Patient Internet Portals Have More Primary Care Visits? Med Care. 2016;54(1):17-23.

5. Herwehe J, Wilbright W, Abrams A, et al. Implementation of an innovative, integrated electronic medical record (EMR) and public health information exchange for HIV/AIDS. J Am Med Inform Assoc. 2012;19(3):448-452.

6. Non-federal Acute Care Hospital Electronic Health Record Adoption. Accessed July 2, 2020. https://dashboard.healthit.gov/quickstats/pages/FIG-Hospital-EHR-Adoption.php

7. Graber ML, Byrne C, Johnston D. The impact of electronic health records on diagnosis. Diagnosis (Berl). 2017;4(4):211-223.

8. Ronda MCM, Dijkhorst-Oei LT, Rutten GEHM. Reasons and barriers for using a patient portal: survey among patients with diabetes mellitus. J Med Internet Res. 2014;16(11):e263.

9. Walker RC, Tong A, Howard K, Palmer SC. Patient expectations and experiences

of remote monitoring for chronic diseases: Systematic review and thematic synthesis of qualitative studies. Int J Med Inform. 2019;124:78-85.

10. Reed M, Huang J, Brand R, et al. Implementation of an outpatient electronic health record and emergency department visits, hospitalizations, and office visits among patients with diabetes. JAMA. 2013;310(10):1060-1065.

11. Swigert NB. Clinical documentation, coding, and billing. Semin Speech Lang. 2006;27(2):101-118.

12. Sidorov J. It Ain't Necessarily So: The Electronic Health Record And The Unlikely Prospect Of Reducing Health Care Costs. Health Aff . 2006;25(4):1079-1085.

13. Horth RZ, Wagstaff S, Jeppson T, et al. Use of electronic health records from a statewide health information exchange to support public health surveillance of diabetes and hypertension. BMC Public Health. 2019;19(1):1106.

14. Cohen IG, Mello MM. HIPAA and Protecting Health Information in the 21st Century. JAMA. 2018;320(3):231-232.

15. Gold M, McLAUGHLIN C. Assessing HITECH Implementation and Lessons: 5 Years Later. Milbank Q. 2016;94(3):654-687.

16. Weiner MG, Embi PJ. Toward reuse of clinical data for research and quality improvement: the end of the beginning? Ann Intern Med. 2009;151(5):359-360.

17. Botsis T, Hartvigsen G, Chen F, Weng C. Secondary Use of EHR: Data Quality Issues and Informatics Opportunities. AMIA Jt Summits Transl Sci Proc. 2010;2010:1-5.

18. Rowe C, Vittinghoff E, Santos GM, Behar E, Turner C, Coffin PO. Performance Measures of Diagnostic Codes for Detecting Opioid Overdose in the Emergency Department. Acad Emerg Med. 2017;24(4):475-483.

19. Kahn MG, Eliason BB, Bathurst J. Quantifying clinical data quality using relative gold standards. AMIA Annu Symp Proc. 2010;2010:356-360.

20. Nahm M. Data Quality in Clinical Research. In: Richesson RL, Andrews JE, eds. Clinical Research Informatics. Springer London; 2012:175-201.

21. Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. J Am Med Inform Assoc. 2013;20(1):144-151.

22. Committee on Improving the Patient Record, Institute of Medicine. The Computer-Based Patient Record: An Essential Technology for Health Care, Revised Edition. National Academies Press; 1997.

23. Weber GM, Mandl KD, Kohane IS. Finding the missing link for big biomedical data. JAMA. 2014;311(24):2479-2480.

24. Pak HS. Unstructured data in healthcare [Internet]. Fremont (CA): Healthcare Tech Outlook; c2018. cited at 2019 Jan 15.

25. Consultant HIT. Why unstructured data holds the key to intelligent healthcare systems [Internet]. Atlanta (GA): HIT Consultant; 2015. cited at 2019 Jan 15.

26. Kong HJ. Managing Unstructured Big Data in Healthcare System. Healthc Inform Res. 2019;25(1):1-2.

27. Meystre SM, Lovis C, Bürkle T, Tognola G, Budrionis A, Lehmann CU. Clinical Data Reuse or Secondary Use: Current Status and Potential Future Progress. Yearb Med Inform. 2017;26(1):38-52.

28. Zhang Y, Jiang M, Wang J, Xu H. Semantic Role Labeling of Clinical Text: Comparing Syntactic Parsers and Features. AMIA Annu Symp Proc. 2016;2016:1283-

1292.

29. Popejoy LL, Khalilia MA, Popescu M, et al. Quantifying care coordination using natural language processing and domain-specific ontology. J Am Med Inform Assoc. 2015;22(e1):e93-e103.

30. Ancker JS, Shih S, Singh MP, et al. Root causes underlying challenges to secondary use of data. AMIA Annu Symp Proc. 2011;2011:57-62.

31. Capurro D, Yetisgen M, van Eaton E, Black R, Tarczy-Hornoch P. Availability of structured and unstructured clinical data for comparative effectiveness research and quality improvement: a multisite assessment. EGEMS (Wash DC). 2014;2(1):1079.

32. Fagan PJ, Schuster AB, Boyd C, et al. Chronic care improvement in primary care: evaluation of an integrated pay-for-performance and practice-based care coordination program among elderly patients with diabetes. Health Serv Res. 2010;45(6 Pt 1):1763-1782.

33. Player M, O'Bryan E, Sederstrom E, Pinckney J, Diaz V. Electronic Visits For Common Acute Conditions: Evaluation Of A Recently Established Program. Health Aff . 2018;37(12):2024-2030.

34. Mullainathan S, Obermeyer Z. Does Machine Learning Automate Moral Hazard and Error? Am Econ Rev. 2017;107(5):476-480.

35. Sáez C, Zurriaga O, Pérez-Panadés J, Melchor I, Robles M, García-Gómez JM. Applying probabilistic temporal and multisite data quality control methods to a public health mortality registry in Spain: a systematic approach to quality control of repositories. Journal of the American Medical Informatics Association. 2016;23(6):1085-1095. doi:10.1093/jamia/ocw010

36. Juran JM, Gryna FM, Others. Juran's Quality Control Handbook. Vol 4. McGraw-Hill New York; 1988.

37. Li Y, Sperrin M, Martin GP, Ashcroft DM, van Staa TP. Examining the impact of data quality and completeness of electronic health records on predictions of patients' risks of cardiovascular disease. Int J Med Inform. 2020;133:104033.

38. Yates M, Bechman K, Dennison EM, et al. Data quality predicts care quality: findings from a national clinical audit. Arthritis Res Ther. 2020;22(1):87.

39. Faulconer ER, de Lusignan S. An eight-step method for assessing diagnostic data quality in practice: chronic obstructive pulmonary disease as an exemplar. Inform Prim Care. 2004;12(4):243-254.

40. Danckert B, Horsbøl TA, Andersen O, et al. Registrations of patients with renal cell carcinoma in the nationwide Danish renal cancer database versus the Danish Cancer Registry: Data quality, completeness and survival (DaRenCa study-3). Clin Epidemiol. 2020;12:807-814.

41. Graetz I, Reed M, Shortell SM, Rundall TG, Bellows J, Hsu J. The association between EHRs and care coordination varies by team cohesion. Health Serv Res. 2014;49(1 Pt 2):438-452.

42. Miller AM, Swartwout KD, Schoeny ME, Vail M, McClenton R. Care coordination to target patient complexity and reduce disparities in primary care. Public Health Nurs. 2019;36(4):451-460.

43. Dixon BE, Embi PJ, Haggstrom DA. Information technologies that facilitate care coordination: provider and patient perspectives. Transl Behav Med. 2018;8(3):522-525.

44. Hu Z, Melton GB, Arsoniadis EG, Wang Y, Kwaan MR, Simon GJ. Strategies for

handling missing clinical data for automated surgical site infection detection from the electronic health record. J Biomed Inform. 2017;68:112-120.

45. Mahmood T, Wittenberg P, Zwetsloot IM, Wang H, Tsui KL. Monitoring data quality for telehealth systems in the presence of missing data. Int J Med Inform. 2019;126:156-163.

46. Mohktar MS, Sukor JA, Redmond SJ, Basilakis J, Lovell NH. Effect of Home Telehealth Data Quality on Decision Support System Performance. Procedia Comput Sci. 2015;64:352-359.

47. National Committee on Vital and Health Statistics. Recommendation Letter-Preparting for Adoption of ICD-11 as a Mandated US Health Data Standard. Published online November 25, 2019. https://ncvhs.hhs.gov/wp-content/uploads/2019/12/Recommendation-Letter-Preparing-for-Adoption-of-ICD-11-as-a-Mandated-US-Health-Data-Standard-final.pdf

48. Jue JS, Spector SA, Spector SA. Telemedicine broadening access to care for complex cases. J Surg Res. 2017;220:164-170.

49. Kane CK, Gillis K. The Use Of Telemedicine By Physicians: Still The Exception Rather Than The Rule. Health Aff . 2018;37(12):1923-1930.

50. Adler-Milstein J, Kvedar J, Bates DW. Telehealth Among US Hospitals: Several Factors, Including State Reimbursement And Licensure Policies, Influence Adoption. Health Aff. 2014;33(2):207-215.

51. Harvey JB, Valenta S, Simpson K, Lyles M, McElligott J. Utilization of Outpatient Telehealth Services in Parity and Nonparity States 2010-2015. Telemed J E Health. Published online May 30, 2018. doi:10.1089/tmj.2017.0265

52. Nouri Sarah, Khoong Elaine C., Lyles Courtney R., Karliner Leah. Addressing Equity in Telemedicine for Chronic Disease Management During the Covid-19 Pandemic. Catalyst non-issue content. 1(3). doi:10.1056/CAT.20.0123

53. Mann DM, Chen J, Chunara R, Testa PA, Nov O. COVID-19 transforms health care through telemedicine: evidence from the field. J Am Med Inform Assoc. Published online April 23, 2020. doi:10.1093/jamia/ocaa072

54. James J. Health Policy Brief: Transitioning to ICD-10. Health Affairs. Published online March 10, 2014. doi:10.1377/hpb2014.6

55. Sinsky CA, Rule A, Cohen G, et al. Metrics for assessing physician activity using electronic health record log data. J Am Med Inform Assoc. 2020;27(4):639-643.

56. Adler-Milstein J, Zhao W, Willard-Grace R, Knox M, Grumbach K. Electronic health records and burnout: Time spent on the electronic health record after hours and message volume associated with exhaustion but not with cynicism among primary care clinicians. J Am Med Inform Assoc. 2020;27(4):531-538.

57. Arnold EL, Sikes DH, Harrington JT. Clinician Burnout and Professional Well-being. JAMA. 2020;323(13):1317-1318.

58. Khairat S, Coleman C, Ottmar P, Jayachander DI, Bice T, Carson SS. Association of Electronic Health Record Use With Physician Fatigue and Efficiency. JAMA Netw Open. 2020;3(6):e207385.

59. Cherry BJ, Ford EW, Peterson LT. Experiences with electronic health records: early adopters in long-term care facilities. Health Care Manage Rev. 2011;36(3):265-274.

60. Hasan S, Padman R. Analyzing the effect of data quality on the accuracy of clinical decision support systems: a computer simulation approach. AMIA Annu Symp

Proc. Published online 2006:324-328.

61. Wang RY, Strong DM. Beyond Accuracy: What Data Quality Means to Data Consumers. Journal of Management Information Systems. 1996;12(4):5-33.

62. Hersh WR. Adding value to the electronic health record through secondary use of data for quality assurance, research, and surveillance. Am J Manag Care. 2007;13(6 Part 1):277-278.

63. Unruh MA, Jung HY, Kaushal R, Vest JR. Hospitalization event notifications and reductions in readmissions of Medicare fee-for-service beneficiaries in the Bronx, New York. J Am Med Inform Assoc. 2017;24(e1):e150-e156.

64. Jakobsen AS, Laursen LC, Østergaard B, Rydahl-Hansen S, Phanareth KV. Hospital-admitted COPD patients treated at home using telemedicine technology in The Virtual Hospital Trial: methods of a randomized effectiveness trial. Trials. 2013;14:280.

65. Petter S, DeLone W, McLean ER. Information Systems Success: The Quest for the Independent Variables. Journal of Management Information Systems. 2013;29(4):7-62.

66. Sebastian-Coleman L, Ingenix – United Health Group. MIT Information Quality Industry Symposium July 14-16, 2010. MIT Information Quality Industry Symposium. Published online July 2020.
http://mitiq.mit.edu/IQIS/Documents/CDOIQS_201077/Papers/03_08_4B-1.pdf

67. Li J, Yazdany J, Trupin L, et al. Capturing a Patient-Reported Measure of Physical Function Through an Online Electronic Health Record Patient Portal in an Ambulatory Clinic: Implementation Study. JMIR Med Inform. 2018;6(2):e31.

68. Hatef E, Rouhizadeh M, Tia I, et al. Assessing the Availability of Data on Social

and Behavioral Determinants in Structured and Unstructured Electronic Health Records: A Retrospective Analysis of a Multilevel Health Care System. JMIR Med Inform. 2019;7(3):e13802.

69. Heinrich B, Kaiser M. HOW TO MEASURE DATA QUALITY? A METRIC BASED APPROACH. Twenty Eighth International Conference on Information Systems,. Published online 2007. https://epub.uni-regensburg.de/23633/1/heinrich.pdf

70. Cappiello C, Francalanci C, Pernici B. Data quality assessment from the user's perspective. In: Proceedings of the 2004 International Workshop on Information Quality in Information Systems. IQIS '04. Association for Computing Machinery; 2004:68-73.

71. Cappiello C, Francalanci C, Pernici B. Time-Related Factors of Data Quality in Multichannel Information Systems. Journal of Management Information Systems. 2003;20(3):71-92.

72. Kahn MG, Raebel MA, Glanz JM, Riedlinger K, Steiner JF. A pragmatic framework for single-site and multisite data quality assessment in electronic health record-based clinical research. Med Care. 2012;50 Suppl:S21-S29.

73. Hinrichs H. Datenqualitätsmanagement in Data Warehouse-Systemen. Published online 2002. Accessed December 11, 2020. http://oops.uni-oldenburg.de/id/eprint/279

74. Sperrin M, Thew S, Weatherall J, Dixon W, Buchan I. Quantifying the longitudinal value of healthcare record collections for pharmacoepidemiology. AMIA Annu Symp Proc. 2011;2011:1318-1325.

75. Weiskopf NG, Hripcsak G, Swaminathan S, Weng C. Defining and measuring completeness of electronic health records for secondary use. J Biomed Inform. 2013;46(5):830-836.

76. Williams JG. Measuring the completeness and currency of codified clinical information. Methods Inf Med. 2003;42(4):482-488.

77. Hebert M. Telehealth success: evaluation framework development. Stud Health Technol Inform. 2001;84(Pt 2):1145-1149.

78. DeLone WH, McLean ER. Information Systems Success: The Quest for the Dependent Variable. Information Systems Research. 1992;3(1):60-95.

79. Hong JC, Fairchild AT, Tanksley JP, Palta M, Tenenbaum JD. Natural language processing for abstraction of cancer treatment toxicities: accuracy versus human experts. Jamia Open. doi:10.1093/jamiaopen/ooaa064

80. Devkota B, Salas J, Sayavong S, Scherrer JF. Use of an Online Patient Portal and Glucose Control in Primary Care Patients with Diabetes. Popul Health Manag. 2016;19(2):125-131.

81. Heron M. Deaths: Leading Causes for 2017. Natl Vital Stat Rep. 2019;68(6):1-77.

82. National Institutes of Health (NIH) Health Care Systems Research Collaboratory. Type 2 Diabetes Mellitus Phenotype Definitions.; 2014. https://dcricollab.dcri.duke.edu/sites/NIHKR/KR/Diabetes_Phenotype%20Definition%20Resources%20and%20Recommendations.pdf

83. Rubin DJ. Hospital readmission of patients with diabetes. Curr Diab Rep. 2015;15(4):17.

84. Zhuo X, Zhang P, Hoerger TJ. Lifetime direct medical costs of treating type 2 diabetes and diabetic complications. Am J Prev Med. 2013;45(3):253-261.

85. Grant RW, Wald JS, Poon EG, et al. Design and implementation of a web-based patient portal linked to an ambulatory care electronic health record: patient gateway for

diabetes collaborative care. Diabetes Technol Ther. 2006;8(5):576-586.

86. Darkins A, Ryan P, Kobb R, et al. Care Coordination/Home Telehealth: the systematic implementation of health informatics, home telehealth, and disease management to support the care of veteran patients with chronic conditions. Telemed J E Health. 2008;14(10):1118-1126.

87. Liu CW, Einstadter D, Cebul RD. Care fragmentation and emergency department use among complex patients with diabetes. Am J Manag Care. 2010;16(6):413-420.

88. Vo MT, Uratsu CS, Estacio KR, et al. Prompting Patients with Poorly Controlled Diabetes to Identify Visit Priorities Before Primary Care Visits: a Pragmatic Cluster Randomized Trial. J Gen Intern Med. 2019;34(6):831-838.

89. Faragli A, Abawi D, Quinn C, et al. The role of non-invasive devices for the telemonitoring of heart failure patients. Heart Fail Rev. Published online April 27, 2020. doi:10.1007/s10741-020-09963-7

90. The Cost of Diabetes | ADA. Accessed July 6, 2020. https://www.diabetes.org/resources/statistics/cost-diabetes

91. Klonoff DC. Using telemedicine to improve outcomes in diabetes--an emerging technology. J Diabetes Sci Technol. 2009;3(4):624-628.

92. Reed ME, Huang J, Brand RJ, et al. Patients with complex chronic conditions: Health care use and clinical events associated with access to a patient portal. PLoS One. 2019;14(6):e0217636.

93. Struijs JN, Baan CA, Schellevis FG, Westert GP, van den Bos GAM. Comorbidity in patients with diabetes mellitus: impact on medical health care utilization. BMC Health Serv Res. 2006;6:84.

94. Sun D, Zhou T, Li X, et al. Diabetes Comorbidity and All-Cause, Cardiovascular, and Cancer Mortality in U.S. Adults. Diabetes. 2018;67(Supplement 1):1580 - P. doi:10.2337/db18-1580-p

95. Katon WJ. The comorbidity of diabetes mellitus and depression. Am J Med. 2008;121(11 Suppl 2):S8-S15.

96. Price-Haywood EG, Luo Q. Primary Care Practice Reengineering and Associations With Patient Portal Use, Service Utilization, and Disease Control Among Patients With Hypertension and/or Diabetes. Ochsner J. 2017;17(1):103-111.

97. Zickmund SL, Hess R, Bryce CL, et al. Interest in the use of computerized patient portals: role of the provider-patient relationship. J Gen Intern Med. 2008;23 Suppl 1:20-26.

98. Hess R, Bryce CL, Paone S, et al. Exploring challenges and potentials of personal health records in diabetes self-management: implementation and initial assessment. Telemed J E Health. 2007;13(5):509-517.

99. Fonda SJ, McMahon GT, Gomes HE, Hickson S, Conlin PR. Changes in diabetes distress related to participation in an internet-based diabetes care management program and glycemic control. J Diabetes Sci Technol. 2009;3(1):117-124.

100. Hess R, Bryce CL, McTigue K, et al. The Diabetes Patient Portal: Patient Perspectives on Structure and Delivery. Diabetes Spectr. 2006;19(2):106-110.

101. Wade-Vuturo AE, Mayberry LS, Osborn CY. Secure messaging and diabetes management: experiences and perspectives of patient portal users. J Am Med Inform Assoc. 2013;20(3):519-525.

102. Robinson SA, Zocchi MS, Netherton D, et al. Secure Messaging, Diabetes Self-

management, and the Importance of Patient Autonomy: a Mixed Methods Study. J Gen Intern Med. Published online May 21, 2020. doi:10.1007/s11606-020-05834-x

103. Amante DJ, Hogan TP, Pagoto SL, English TM. A systematic review of electronic portal usage among patients with diabetes. Diabetes Technol Ther. 2014;16(11):784-793.

104. Jones JB, Weiner JP, Shah NR, Stewart WF. The wired patient: patterns of electronic patient portal use among patients with cardiac disease or diabetes. J Med Internet Res. 2015;17(2):e42.

105. Sutherland JM, Steinum O. Hospital factors associated with clinical data quality. Health Policy. 2009;91(3):321-326.

106. Sholle ET, Pinheiro LC, Adekkanattu P, et al. Underserved populations with missing race ethnicity data differ significantly from those with structured race/ethnicity documentation. J Am Med Inform Assoc. Published online April 26, 2019. doi:10.1093/jamia/ocz040

107. Koleck TA, Dreisbach C, Bourne PE, Bakken S. Natural language processing of symptoms documented in free-text narratives of electronic health records: a systematic review. J Am Med Inform Assoc. 2019;26(4):364-379.

108. American Medical Association. Special Coding Advice during COVID-19 Public Health Emergency. https://www.ama-assn.org/system/files/2020-05/covid-19-coding-advice.pdf

109. [Updated December 14] AHIMA and AHA FAQ: ICD-10-CM/PCS Coding for COVID-19. Published March 20, 2020. Accessed December 30, 2020. https://journal.ahima.org/ahima-and-aha-faq-on-icd-10-cm-coding-for-covid-19/

110. Huang J, Chen Y, Landis JR, Mahoney KB. Difference Between Users and Nonusers of a Patient Portal in Health Behaviors and Outcomes: Retrospective Cohort Study. J Med Internet Res. 2019;21(10):e13146.

111. Lyles CR, Sarkar U, Ralston JD, et al. Patient-provider communication and trust in relation to use of an online patient portal among diabetes patients: The Diabetes and Aging Study. J Am Med Inform Assoc. 2013;20(6):1128-1131.

112. Jackson SL, DesRoches CM, Frosch DL, Peacock S, Oster NV, Elmore JG. Will use of patient portals help to educate and communicate with patients with diabetes? Patient Educ Couns. 2018;101(5):956-959.

113. Sun R, Korytkowski MT, Sereika SM, Saul MI, Li D, Burke LE. Patient Portal Use in Diabetes Management: Literature Review. JMIR Diabetes. 2018;3(4):e11199.

114. Sun R, Burke LE, Saul MI, Korytkowski MT, Li D, Sereika SM. Use of a Patient Portal for Engaging Patients with Type 2 Diabetes: Patterns and Prediction. Diabetes Technol Ther. 2019;21(10):546-556.

115. Jensen RE, Chan KS, Weiner JP, Fowles JB, Neale SM. Implementing electronic health record-based quality measures for developmental screening. Pediatrics. 2009;124(4):e648-e654.

116. Chouvarda IG, Goulis DG, Lambrinoudaki I, Maglaveras N. Connected health and integrated care: Toward new models for chronic disease management. Maturitas. 2015;82(1):22-27.

117. Delone WH, Mclean ER. Measuring e-commerce success: Applying the DeLone & McLean information systems success model. International Journal of electronic commerce. 2004;9(1):31-47.

118. Bostrom RP, Heinen JS. MIS Problems and Failures: A Socio-Technical Perspective. Part I: The Causes. Miss Q. 1977;1(3):17-32.

119. Capobianco E. Data-driven clinical decision processes: it's time. J Transl Med. 2019;17(1):44.

120. Singh H, Meyer AND, Thomas EJ. The frequency of diagnostic errors in outpatient care: estimations from three large observational studies involving US adult populations. BMJ Qual Saf. 2014;23(9):727-731.

121. Marani H, Halperin IJ, Jamieson T, Mukerji G. Quality Gaps of Electronic Health Records in Diabetes Care. Can J Diabetes. 2020;44(4):350-355.

122. Neufeld JD, Doarn CR. Telemedicine Spending by Medicare: A Snapshot from 2012. Telemed J E Health. 2015;21(8):686-693.

123. Menachemi N, Burke DE, Ayers DJ. Factors affecting the adoption of telemedicine--a multiple adopter perspective. J Med Syst. 2004;28(6):617-632.

124. Hollander JE, Carr BG. Virtually Perfect? Telemedicine for Covid-19. N Engl J Med. 2020;382(18):1679-1681.

125. Opportunities and Barriers for Telemedicine in the U.S. During the COVID-19 Emergency and Beyond. KFF. Published May 11, 2020. Accessed July 10, 2020. https://www.kff.org/womens-health-policy/issue-brief/opportunities-and-barriers-for-telemedicine-in-the-u-s-during-the-covid-19-emergency-and-beyond/

126. Gordon HS, Solanki P, Bokhour BG, Gopal RK. "I'm Not Feeling Like I'm Part of the Conversation" Patients' Perspectives on Communicating in Clinical Video Telehealth Visits. J Gen Intern Med. 2020;35(6):1751-1758.

127. Miller EA. The technical and interpersonal aspects of telemedicine: effects on

doctor-patient communication. J Telemed Telecare. 2003;9(1):1-7.

128. Uscher-Pines L, Mehrotra A. Analysis of Teladoc use seems to indicate expanded access to care for patients without prior connection to a provider. Health Aff . 2014;33(2):258-264.

129. Freckmann G, Pleus S, Grady M, Setford S, Levy B. Measures of Accuracy for Continuous Glucose Monitoring and Blood Glucose Monitoring Devices. J Diabetes Sci Technol. 2019;13(3):575-583.

130. Izquierdo RE, Knudson PE, Meyer S, Kearns J, Ploutz-Snyder R, Weinstock RS. A Comparison of Diabetes Education Administered Through Telemedicine Versus in Person. Diabetes Care. 2003;26(4):1002-1007.

131. Carter EL, Nunlee-Bland G, Callender C. A patient-centric, provider-assisted diabetes telehealth self-management intervention for urban minorities. Perspect Health Inf Manag. 2011;8:1b.

132. American Diabetes Association. Standards of medical care for patients with diabetes mellitus. Diabetes Care. 2003;26 Suppl 1:S33-S50.

133. Goudswaard AN, Lam K, Stolk RP, Rutten GEHM. Quality of recording of data from patients with type 2 diabetes is not a valid indicator of quality of care. A cross-sectional study. Fam Pract. 2003;20(2):173-177.

134. McDonald CJ, Overhage JM, Barnes M, et al. The Indiana network for patient care: a working local health information infrastructure. An example of a working infrastructure collaboration that links data from five health systems and hundreds of millions of entries. Health Aff . 2005;24(5):1214-1220.

135. Hills RA, Revere D, Altamore R, Abernethy NF, Lober WB. Timeliness and data

element completeness of immunization data in Washington State in 2010: a comparison of data exchange methods. AMIA Annu Symp Proc. 2012;2012:340-349.

136. O'Connor PJ. Organizing diabetes care: identify, monitor, prioritize, intensify. Diabetes Care. 2001;24(9):1515-1516.

137. Ross SE, Haverhals LM, Main DS, Bull SS, Pratte K, Lin CT. Adoption and use of an online patient portal for diabetes (Diabetes-STAR). AMIA Annu Symp Proc. Published online 2006:1080.

138. Genes N, Violante S, Cetrangol C, Rogers L, Schadt EE, Chan YFY. From smartphone to EHR: a case report on integrating patient-generated health data. NPJ Digit Med. 2018;1:23.

139. Bell SK, Delbanco T, Elmore JG, et al. Frequency and Types of Patient-Reported Errors in Electronic Health Record Ambulatory Care Notes. JAMA Netw Open. 2020;3(6):e205867-e205867.

140. Aziz F, Talhelm L, Keefer J, Krawiec C. Vascular surgery residents spend one fifth of their time on electronic health records after duty hours. J Vasc Surg. 2019;69(5):1574-1579.

141. Final Policy, Payment, and Quality Provisions Changes to the Medicare Physician Fee Schedule for Calendar Year 2021. Accessed November 6, 2021. https://www.cms.gov/newsroom/fact-sheets/final-policy-payment-and-quality-provisions-changes-medicare-physician-fee-schedule-calendar-year-1

142. Medicare Program; CY 2022 Payment Policies under the Physician Fee Schedule and Other Changes to Part B Payment Policies; Medicare Shared Savings Program Requirements; Provider Enrollment Regulation Updates; Provider and Supplier

Prepayment and Post-Payment Medical Review Requirements.; :1-1747. https://public-inspection.federalregister.gov/2021-14973.pdf

143. Foley & Lardner LLP. CMS Proposes New Remote Therapeutic Monitoring Codes: What You Need to Know. Published July 15, 2021. Accessed December 23, 2021. https://www.foley.com/en/insights/publications/2021/07/cms-new-remote-therapeutic-monitoring-codes

144. Redman TC. Measuring data accuracy: A framework and review. In: Information Quality. Routledge; 2014:33-48.

145. Michalakidis G, Kumarapeli P, Ring A, van Vlymen J, Krause P, de Lusignan S. A system for solution-orientated reporting of errors associated with the extraction of routinely collected clinical data for research and quality improvement. Stud Health Technol Inform. 2010;160(Pt 1):724-728.

146. Rodríguez JE, Campbell KM. Racial and Ethnic Disparities in Prevalence and Care of Patients With Type 2 Diabetes. Clin Diabetes. 2017;35(1):66-70.

147. Shrivastav M, Gibson W Jr, Shrivastav R, et al. Type 2 Diabetes Management in Primary Care: The Role of Retrospective, Professional Continuous Glucose Monitoring. Diabetes Spectr. 2018;31(3):279-287.

148. Bowen ME, Merchant Z, Abdullah K, Bhat D, Fish J, Halm EA. Patient, Provider, and System Factors Associated With Failure to Follow-Up Elevated Glucose Results in Patients Without Diagnosed Diabetes. *Health Serv Res Manag Epidemiol*. 2017;4:2333392817721647.

149. Klimek P, Kautzky-Willer A, Chmiel A, Schiller-Frühwirth I, Thurner S.

Quantification of diabetes comorbidity risks across life using nation-wide big claims data. PLoS Comput Biol. 2015;11(4):e1004125.

150. Panser LA, Naessens JM, Nobrega FT, Palumbo PJ, Ballard DJ. Utilization trends and risk factors for hospitalization in diabetes mellitus. Mayo Clin Proc. 1990;65(9):1171-1184.

151. Rosenthal MJ, Fajardo M, Gilmore S, Morley JE, Naliboff BD. Hospitalization and mortality of diabetes in older adults. A 3-year prospective study. Diabetes Care. 1998;21(2):231-235.

152. Liaw ST, Rahimi A, Ray P, et al. Towards an ontology for data quality in integrated chronic disease management: a realist review of the literature. Int J Med Inform. 2013;82(1):10-24.

153. Pippitt K, Li M, Gurgle HE. Diabetes Mellitus: Screening and Diagnosis. Published January 15, 2016. Accessed March 12, 2022. https://www.aafp.org/afp/2016/0115/p103.html

154. US Preventive Services Task Force, Davidson KW, Barry MJ, et al. Screening for Prediabetes and Type 2 Diabetes: US Preventive Services Task Force Recommendation Statement. JAMA. 2021;326(8):736-743.

155. Francis BH, Song X, Andrews LM, et al. Progression to type 2 diabetes, healthcare utilization, and cost among pre-diabetic patients with or without comorbid hypertension. Curr Med Res Opin. 2011;27(4):809-819.

156. Shahraz S, Pittas AG, Saadati M, Thomas CP, Lundquist CM, Kent DM. Change in Testing, Awareness of Hemoglobin A1c Result, and Glycemic Control in US Adults, 2007-2014. JAMA. 2017;318(18):1825-1827.

157. Comino EJ, Harris MF, Islam MDF, et al. Impact of diabetes on hospital admission and length of stay among a general population aged 45 year or more: a record linkage study. BMC Health Serv Res. 2015;15:12.

158. Samuels TA, Cohen D, Brancati FL, Coresh J, Kao WHL. Delayed diagnosis of incident type 2 diabetes mellitus in the ARIC study. Am J Manag Care. 2006;12(12):717-724.

159. Dusheiko M, Doran T, Gravelle H, Fullwood C, Roland M. Does higher quality of diabetes management in family practice reduce unplanned hospital admissions? Health Serv Res. 2011;46(1 Pt 1):27-46.

160. Risk Factors for Diabetes-Related Complications. Published January 12, 2022. Accessed March 13, 2022. https://www.cdc.gov/diabetes/data/statistics-report/risks-complications.html

161. Hospital Readmissions Reduction Program (HRRP). Accessed March 27, 2022. https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Readmissions-Reduction-Program

162. CDC. Type 2 Diabetes. Centers for Disease Control and Prevention. Published March 2, 2022. Accessed March 13, 2022. https://www.cdc.gov/diabetes/basics/type2.html

163. Bo S, Ciccone G, Grassi G, et al. Patients with type 2 diabetes had higher rates of hospitalization than the general population. J Clin Epidemiol. 2004;57(11):1196-1201.

164. American Diabetes Association. 3. Prevention or Delay of Type 2 Diabetes: Standards of Medical Care in Diabetes-2021. Diabetes Care. 2021;44(Suppl 1):S34-S39.

165. Nelson LA, Wallston KA, Kripalani S, LeStourgeon LM, Williamson SE,

Mayberry LS. Assessing barriers to diabetes medication adherence using the Information-Motivation-Behavioral skills model. *Diabetes Res Clin Pract*. 2018;142:374-384.

166. Patel V, McNamara L, Dullabh P, Sawchuk ME, Swain M. Variation in interoperability across clinical laboratories nationwide. *Int J Med Inform*. 2017;108:175-184

167. Gazmararian JA, Ziemer DC, Barnes C. Perception of barriers to self-care management among diabetic patients. *Diabetes Educ*. 2009;35(5):778-788.

168. Delaronde S. Barriers to A1C testing among a managed care population. *Diabetes Educ*. 2005;31(2):235-239.

169. Hueth KD, Jackson BR, Schmidt RL. An Audit of Repeat Testing at an Academic Medical Center: Consistency of Order Patterns With Recommendations and Potential Cost Savings. Am J Clin Pathol. 2018;150(1):27-33.

170. Dovey S, Fryer GE, Green L, Phillips RL. Toxic cascades: a comprehensive way to think about medical errors. Am Fam Physician. 2001;63(5):847.

171. Smith PC, Araya-Guerra R, Bublitz C, et al. Missing clinical information during primary care visits. JAMA. 2005;293(5):565-571.

172. Holup AA, Dobbs D, Meng H, Hyer K. Facility characteristics associated with the use of electronic health records in residential care facilities. J Am Med Inform Assoc. 2013;20(4):787-791.

173. Perzynski AT, Roach MJ, Shick S, et al. Patient portals and broadband internet inequality. J Am Med Inform Assoc. 2017;24(5):927-932.

174. Klinger EV, Carlini SV, Gonzalez I, et al. Accuracy of race, ethnicity, and language preference in an electronic health record. J Gen Intern Med. 2015;30(6):719-

723.

175. Walker DM, Hefner JL, Fareed N, Huerta TR, McAlearney AS. Exploring the Digital Divide: Age and Race Disparities in Use of an Inpatient Portal. Telemed J E Health. 2020;26(5):603-613.

176. Sarkar U, Karter AJ, Liu JY, et al. The literacy divide: health literacy and the use of an internet-based patient portal in an integrated health system-results from the diabetes study of northern California (DISTANCE). J Health Commun. 2010;15 Suppl 2:183-196.

177. Staroselsky M, Volk LA, Tsurikova R, et al. An effort to improve electronic health record medication list accuracy between visits: patients’ and physicians' response. Int J Med Inform. 2008;77(3):153-160.

178. Pillemer F, Price RA, Paone S, et al. Direct Release of Test Results to Patients Increases Patient Engagement and Utilization of Care. PLoS One. 2016;11(6):e0154743.

179. Wager KA, Schaffner MJ, Foulois B, Swanson Kazley A, Parker C, Walo H. Comparison of the quality and timeliness of vital signs data using three different data-entry devices. Comput Inform Nurs. 2010;28(4):205-212.

180. Allison PD. Using Panel Data to Estimate the Effects of Events. Sociol Methods Res. 1994;23(2):174-199.

181. Sorondo B, Allen A, Fathima S, Bayleran J, Sabbagh I. Patient Portal as a Tool for Enhancing Patient Experience and Improving Quality of Care in Primary Care Practices. EGEMS (Wash DC). 2016;4(1):1262.

182. Urowitz S, Wiljer D, Dupak K, et al. Improving diabetes management with a patient portal: a qualitative study of diabetes self-management portal. J Med Internet Res. 2012;14(6):e158.

183. Lyles CR, Nelson EC, Frampton S, Dykes PC, Cemballi AG, Sarkar U. Using Electronic Health Record Portals to Improve Patient Engagement: Research Priorities and Best Practices. Ann Intern Med. 2020;172(11 Suppl):S123-S129.

184. Coughlin SS, Prochaska JJ, Williams LB, et al. Patient web portals, disease management, and primary prevention. Risk Manag Healthc Policy. 2017;10:33-40.

185. Ancker JS, Barrón Y, Rockoff ML, et al. Use of an electronic patient portal among disadvantaged populations. J Gen Intern Med. 2011;26(10):1117-1123.

186. Goel MS, Brown TL, Williams A, Hasnain-Wynia R, Thompson JA, Baker DW. Disparities in enrollment and use of an electronic patient portal. J Gen Intern Med. 2011;26(10):1112-1116.

187. Baird A, Nowak S. Why primary care practices should become digital health information hubs for their patients. BMC Fam Pract. 2014;15:190.

188. Centers for Medicare and Medicaid Services (CMS). Promoting Interoperability Programs: MEDICARE PROMOTING INTEROPERABILITY PROGRAM ELIGIBLE HOSPITALS, CRITICAL ACCESS HOSPITALS, AND DUAL-ELIGIBLE HOSPITALS ATTESTING TO CMS OBJECTIVES AND MEASURES FOR 2020. https://www.cms.gov/files/document/medicare-eh-2020-provide-patients-electronic-access-their-health-information.pdf

189. Horth RZ, Wagstaff S, Jeppson T, et al. Use of electronic health records from a statewide health information exchange to support public health surveillance of diabetes and hypertension. *BMC Public Health*. 2019;19(1):1106.

190. Annis T, Pleasants S, Hultman G, et al. Rapid Implementation of a COVID-19 Remote Patient Monitoring Program. J Am Med Inform Assoc. Published online May 11,

2020. doi:10.1093/jamia/ocaa097

191. American Diabetes Association. Standards of Medical Care in Diabetes—2020 Abridged for Primary Care Providers. Clin Diabetes. 2020;38(1):10-38.

192. Phelan M, Bhavsar NA, Goldstein BA. Illustrating Informed Presence Bias in Electronic Health Records Data: How Patient Interactions with a Health System Can Impact Inference. EGEMS (Wash DC). 2017;5(1):22.

193. Goldstein BA, Phelan M, Pagidipati NJ, Peskoe SB. How and when informative visit processes can bias inference when using electronic health records data for clinical research. J Am Med Inform Assoc. 2019;26(12):1609-1617.

194. Predmore ZS, Roth E, Breslau J, Fischer SH, Uscher-Pines L. Assessment of Patient Preferences for Telehealth in Post–COVID-19 Pandemic Health Care. JAMA Netw Open. 2021;4(12):e2136405-e2136405.

195. Berkson J. Limitations of the application of fourfold table analysis to hospital data. Biometrics. 1946;2(3):47-53.

CURRICULUM VITAE

Kevin Keith Wiley, Jr.

EDUCATION

2022 **Ph.D. in Health Policy and Management**
Richard M. Fairbanks School of Public Health
Indiana University
Concentration: Public and Population Health Informatics
Dissertation: Electronic Health Record (EHR) Data Quality and Type II Diabetes Care

2012 **Master of Public Health**
Jiann-Ping Hsu College of Public Health
Georgia Southern University
Concentration: Health Policy and Management
Thesis: Associations between jurisdictional characteristics and provision of preventive services at local health departments (LHDs)

2008 **Bachelor of Arts**
College of Humanities
Coastal Carolina University
Major: Political Science
Minor: Biology

POSITIONS HELD

2021-Present **Phyllis Torda Health Care Quality and Equity Dissertation Fellow**
National Committee for Quality Assurance (NCQA)

2021-Present **Research Associate**
Center for Research on Inclusion and Social Policy (CRISP)
Public Policy Institute
Indiana University

2020-Present **Health Policy Advisor**
Supporting Pediatric Research on Outcomes and Utilization of Telehealth Collaborative - Clinical Translational Science Award
NIH National Center for Advancing Translational Sciences

2018-Present **Health Policy Research Scholar**
Robert Wood Johnson Foundation (RWJF)

2020-2021 **Community Engagement Associate (CEA) Scholar**

Clinic Coordinator
Research and Evaluation Clinic (pro bono)

Center for Research on Inclusion and Social Policy (CRISP)
Public Policy Institute
Indiana University

2017-2021 **T15 Biomedical Informatics and Data Science Trainee/Fellow**
National Library of Medicine/National Institutes of Health
Department of Health Policy and Management
Richard M. Fairbanks School of Public Health
Indiana University

2015-2017 **Telehealth Alliance and Outreach Coordinator**
Center for Telehealth
Medical University of South Carolina (MUSC)

2013-2015 **Program Coordinator**
Center for Global Health
Medical University of South Carolina (MUSC)

2012-2013 **Research Specialist**
Faculty of Environmental Engineering
Khon Kaen University, Thailand

2012-2012 **Fellow**
Office of Health Reform
United States Department of Health and Human Services

2010-2012 **Graduate Research Assistant**
Jiann-Ping Hsu College of Public Health
Georgia Southern University

2011-2012 **Research Associate**
World Health Organization (WHO)
Collaborating Center for Research and Training on Gender and Women's Health, Khon Kaen Thailand

2011-2012 **Fellow**
Minority Health International Research Training Program
National Institute on Minority Health and Health Disparities
National Institutes of Health (NIH)

2011-2012 **Contractor**
State Government Affairs
Healthcare Information and Management Systems Society (HIMSS)

TEACHING EXPERIENCE

2020-2021 **Teaching Assistant**
Bachelor of Science in Health Services Management program
Department of Health Policy and Management
Richard M. Fairbanks School of Public Health

Indiana University

WEB APPLICATIONS

Opioid Prescribing Rate Visualization Tool (RShiny)
Co-developed Interactive mapping visualization of CDC prescribing and death rates over time by county and state - Lead developer, Nate Apathy, PhD

COVID-19 Control Chart Visualization Tool (RShiny)
Interactive control charts showing 7-day moving averages of COVID-19 cases in Indiana

HONORS ANDS AWARDS

2021 **Finalist**
IUPUI's Sherry Queener Graduate Student Excellence Award nominee - nominated as the top Ph.D. student at the Richard M. Fairbanks School of Public Health, Indiana University

2020 **Awardee**
Health Policy Research Scholar (HPRS) Writing Retreat Scholarship

2020 **Health Scholar**
Aspen Ideas: Health

2012 **Inducted Member**
Delta Omega Honorary Society in Public Health

2011 **Inducted Member**
Delta Epsilon Iota National Academic Honor Society

2007 **Inducted Member**
Pi Sigma Alpha National Political Science Honor Society

2005 **Dean's List Recipient**
Coastal Carolina University

GRANTS AND FELLOWSHIPS

2021-Present **Phyllis Torda Health Care Quality and Equity Fellow**
National Committee for Quality Assurance (NCQA)

2021-Present **Health Policy Research Scholar (HPRS) Dissertation Grant Awardee**
Robert Wood Johnson Foundation, $10,000

2018-Present **Four year Pre-doctoral scholarship**
Health Policy Research Scholar (HPRS) program
Robert Wood Johnson Foundation (RWJF), $120,000

2017-2021 **T15 Biomedical Informatics and Data Science Trainee/Fellow**
National Library of Medicine (NLM) at the National Institutes of Health (NIH)

2016-2017 **Campus Life Enrichment Committee (CLEC) Grant Awardee**
Georgia Southern University (Delta Omega Honorary Society in Public Health), $984

2015-2016 **Campus Life Enrichment Committee (CLEC) Grant Awardee**
Georgia Southern University (Delta Omega Honorary Society in Public Health), $750

PEER-REVIEWED PUBLICATIONS

1. Bako, A.T., Taylor, H., **Wiley, K.K.**, Zheng, J., Vest, J.R. (2020). Using natural language processing to classify social work interventions. American Journal of Managed Care; 27(1):e24-e31. doi: 10.37765/ajmc.2021.88580. PMID: 33471465; PMCID: PMC8005360.

2. **Wiley, K.K.**, Hilts, K., Ancker, J., Unruh, M. Jung, A.R., Vest, J.R. (2020). Examining organizational characteristics and perceptions of clinical event notification services in health care settings. Journal of the American Medical Informatics Association - Open; 3(4):611-618. doi: 10.1093/jamiaopen/ooaa065. PMID: 33623895; PMCID: PMC7886547.

3. **Wiley, K.K.**, Dixon, B.E., Grannis, S., Menachemi, N. (2020). Underrepresented Racial Minorities in Biomedical Informatics Doctoral Programs: Graduation Trends and Academic Placement (2002-2017). Journal of the American Medical Informatics Association; 27(11):1641-1647. doi: 10.1093/jamia/ocaa206. PMID: 33053157; PMCID: PMC7671637.

4. Balio, C., **Wiley, K.K.**, Greene, M., & Vest, J.R. (2019). Opioid-Related Emergency Department Encounters: Patient, Encounter, and Community Characteristics associated with Repeat Encounters. Annals of Emergency Medicine; 75(5):568-575. doi: 10.1016/j.annemergmed.2019.12.005. Epub 2020 Jan 23. PMID: 31983498.

5. Vest, J.R., Jung, A.R., **Wiley, K.K**., Kooreman, H., Petit, L., Unruh, M. (2018). Adoption of health information technology among U.S. nursing facilities. Journal of Post-Acute and Long-Term Care Medicine; 20(8):995-1000.e4. doi: 10.1016/j.jamda.2018.11.002. Epub 2018 Dec 20. PMID: 30579920; PMCID: PMC6591108.

6. Welch, B. M., **Wiley, K. K**., Pflieger, L., Achiangi, A., Charle, D. K., Hughes, H. A. **(**2018). Review and comparison of electronic patient-facing family health history

tools. Journal of Genetic Counseling; 27(2):381-391. doi: 10.1007/s10897-018-0235-7. Epub 2018 Mar 6. PMID: 29512060; PMCID: PMC5861014.

RESEARCH IN PROGRESS

1. Armoudas, A., Arnaout, R., Chung, M., Hall, J., Knowles, J., Price, N., Rawat, D., Reigel, B., Bagdady, K.S., Wang, T., **Wiley, K.K.** Principles for Data Sharing, Privacy, and Security: An American Heart Association Policy Position Statement (In Preparation for Circulation)

2. **Wiley, K.K.,** Mendonca, E., Blackburn, J., De Groot, M., Menachemi, N., Vest, J.V. Measuring Data Quality in Telehealth and Office-based Diabetes Care (In Preparation for the Journal of the American Medical Informatics Association)

3. **Wiley, K.K.,** Mendonca, E., Blackburn, J., De Groot, M., Menachemi, N., Vest, J.V. Data Quality and Outpatient Visits in Diabetes Care (In Preparation for the American Journal of Managed Care)

4. **Wiley, K.K.,** Mendonca, E., Blackburn, J., De Groot, M., Menachemi, N., Vest, J.V. Patient Portal Use and EHR Data Quality in Diabetes Care (In Preparation for the Health Services Research)

5. **Wiley, K.K.,** Vest, J.R. Remote electronic device use and health care utilization. (In Preparation for the American Journal of Managed Care)

REPORTS, BOOKS, & OTHER PUBLICATIONS

1. **Wiley, K.K.** Bowling, E., Lawrence, R. (2021) Natural Helper Program Assessment: An Analysis of the Immigrant Welcome Center's Natural Helper Program. Center for Research on Inclusion and Social Policy, Public Policy Institute, Indiana University

2. **Wiley, K. K.** (2013). Neglected Tropical Diseases- A Local NGO's Challenges. In Johnson, J. A. and Musch, S. D. (Eds.). Multi-Sector Casebook in Health Administration, Leadership, and Management (pp.203-204). Stamford, CT: Delmar Cengage Learning Publishers.

3. **Wiley, K. K.** (2013). Case 22: Neglected Tropical Diseases- A Local NGO's Challenges. In Johnson, J. A. and Musch, S. D. (Eds.). Introduction to Public Health Management, Organization, and Policy (pp.384-385). Clifton Park, NY: Delmar Cengage Learning Publishers.

4. **Wiley, K. K.,** & Vest, J.R. (2012). EHR Meaningful Use and State Health Information Technology Legislation Report. In States Will Transform Healthcare through Health IT and HIE Organizations. Healthcare Information and Management Services Society (HIMSS). [White Paper]. Available at: http://www.himss.org/files/HIMSSorg/policy/d/20120605StatesWillTransformHealthcare.pdf

PRESENTATIONS

1. Bear Don’t Walk, O. **Wiley, K.K.**, Walters-Threat, L., Rivera, R.L., Were, M.C., Bright, T.J. A Framework to Support Diversity, Equity, and Inclusion within AMIA Through Strengthened Pathways, Support, and Leadership. AMIA Annual Symposium 2021 (Submitted as Late Breaking abstract)

2. **Wiley, K.K.,** Lawrence, R. An Analysis of the Immigrant Welcome Center’s Natural Helper Program. Immigrant Welcome Center Board Meeting

3. **Wiley, K.K.** Measuring Data Quality in virtual care for patients with Type 2 Diabetes Mellitus. 2021 National Library of Medicine Informatics Training Conference

4. **Wiley, K.K.** Remote Electronic Device Use and Health Care Utilization. 2021AcademyHealth Annual Research Meeting (accepted poster)

5. **Wiley, K.K.** Introduction to Telehealth Policy. Indiana University-Purdue University, Indianapolis, IN, 2020*

6. **Wiley, K.K.** Bias in Health Care Artificial Intelligence. Indiana University-Purdue University, Indianapolis, IN, 2020*

7. **Wiley, K.K.** Big Data in HealthCare. City University of New York (CUNY) Public and Population Health Informatics program, 2020*

8. **Wiley, K.K.**, Hilts, K., Ancker, J., Unruh, M. Jung, H., Vest, J.R. Examining organizational characteristics and perceptions of clinical event notification services in health care settings. US National Library of Medicine Informatics Training Conference, 2020

9. **Wiley, K.K.**, Apathy, N. Analyzing Sentiment of Government Health Insurance Expansion on Twitter. US National Library of Medicine Informatics Training Conference, 2019

10. Danielson, E., **Wiley, K.K.**, Harle, C. Examining Factors Related to Patient Experiences of Missing Information and Repeat Testing during Healthcare Encounters. AcademyHealth Annual Research Meeting, 2019

11. Balio, C., **Wiley, K.K.** Opioid-Related Emergency Department Encounters: Patient, Encounter, and Community Characteristics associated with Repeat Encounters. AcademyHealth Annual Research Meeting, 2019

12. **Wiley, K.K.** Using Natural Language Processing to Identify Opioid Use Disorder in the Emergency Department. US National Library of Medicine Informatics Training Conference, 2018

13. Balio, C., Apathy, N., **Wiley, K.K.** Patient-Centered Medical Home Patients Have Fewer Hospitalizations: An Analysis Using MEPS Data. AcademyHealth Annual Research Meeting, 2018

14. **Wiley, K.K.** Telehealth: What it is and how it is changing care delivery. ECPI University of North Charleston, SC - Guest Lecture, 2016*

15. **Wiley, K.K.** Area Health Education Centers (AHEC) Bench to Bedside, 2016

16. **Wiley, K.K.** Telehealth: A Policy Primer on State and Federal Legislation and How it Affects Your Practice, Area Health Education Centers (AHEC), 2016*

17. **Wiley, K.K.** Telemedicine in Public Health. South Carolina Public Health Association, 2016

18. **Wiley, K.K.** Thinking Beyond the Degree. Gamma Theta Chapter of Delta Omega Honorary Society in Public Health, 2016

19. **Wiley, K.K.**, Shah, G., Markossian, T. Associations between Jurisdictional Characteristics and Provision of Preventive Services at Local Health Departments, Jiann-Ping Hsu College of Public Health, 2012

20. **Wiley, K.K.** Ban Khoknagnam Water-Analysis: Environmental and Agricultural Policy Implications. Ban Khoknagnam Municipality, 2011

21. **Wiley, K.K.** Gender Sensitivity and Water Management-Ban Khoknagnam Water Source Analysis, National Institute on Minority Health and Health Disparities-National Institutes of Health, 2011

22. **Wiley, K.K.** Cultural Competency and Research in Southeast Asia. Faculty of Nursing Colloquium, Khon Kaen University, Thailand, 2011

*Guest lecture

PROFESSIONAL ASSOCIATIONS & COMMITTEE MEMBERSHIP

Date	Position	Organization/Committee
2021-Present	Co-chair	American Medical Informatics Association (AMIA) Subcommittee for the Advancement of Health Equity and Antiracism in Health care
2021	Invited member	National Library of Medicine (NLM) Training Conference

2021-2022	Invited Member	American Medical Informatics Association (AMIA) DEI Task Force
2020	Reviewer	American Journal of Public Health
2019-2021	Fellow	IUPUI Graduate Student Emissaries
2019-2020	Reviewer	American Medical Informatics Association Annual Symposium
2019	Reviewer	MedInfo
2018-2019	President	IUPUI AcademyHealth Student Chapter
2017-Present	Member	American Medical Informatics Association (AMIA)
2015-2017	Member	American Telemedicine Association (ATA)
2015-2016	President	Delta Omega, Gamma Theta Chapter
2011-2015	Member	American Public Health Association (APHA)
2011-2012	President	Georgia Southern University AcademyHealth Student Chapter

STATISTICAL SOFTWARE PROFICIENCY

SAS, STATA, R, Python