The Use of Predictive Analytics for Population Health Management,

Integrating Multilevel Data to Predict Colorectal Cancer Screening

by
Amanda Fleming Petrik

A dissertation submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy
in
Health Systems and Policy
Oregon Health Sciences University

Dissertation Committee:
Neal Wallace, Chair
Gloria Coronado
Sunny Lin
Eric S. Johnson
Lucy Savitz

Oral Defense: September 28, 2021

Abstract

The use of predictive analytics can help health systems target the right services to the right patients at the right time, while improving population health. Multilevel data, or data at the interpersonal, organizational, community and policy levels, is rarely sought after but may be used to improve risk prediction by providing information about a patient and the many groups to which they belong. Colorectal cancer screening promotion can be expensive and not all patients need it.

This study assessed the availability of multilevel data for use in a colorectal cancer screening risk prediction model in accordance with the Social-ecological Model (SEM) and assessed its ability to improve prediction over standard models based on individual level data. Model performance was evaluated overall and among critical subpopulations. This study found that while multilevel data is available, it did not improve the performance of the risk prediction models. Results were shared with health system leaders to assess their perceptions of the usefulness of the model in improving efforts to identify and target screening uptake efforts. The stakeholders found the individual and multilevel models to be potentially useful in efforts to increase colorectal cancer screening. While multilevel data was found to be interesting and potentially useful by stakeholders, they recognized that additional resources would be needed to fully use and maintain multilevel data.

Multilevel data is available and usable, but not consistently at all levels. The predictive models developed in this project were sufficient for predicting patient's likelihood of screening for colorectal cancer, but multilevel data but did not improve the performance of predictive models. Stakeholders found the individual and multilevel prediction models useful but reported potential barriers to implementation. Multilevel data should continue to be explored as potential predictors of health outcomes.

Dedication

To my children Charles Fleming Petrik and Cecelia Anne Petrik.

May you always have a love of learning, for life is a continually rewarding classroom.

Acknowledgments

I will be eternally grateful for a large group of mentors, friends, and family that have supported me throughout this process. As I started my journey at PSU hoping to gain expertise in Medical Sociology in 2008, I hardly believed it would end 13 years later in the OHSU/PSU School of Public Health in Health Systems and Policy.

First and foremost, I must thank my committee, especially Dr. Neal Wallace, who has endured countless hours of mentorship, reviewing, discussing risk prediction and multilevel modelling through both my exams and dissertation. Dr. Wallace continually guided the process and my personal growth. Dr. Gloria Coronado has not only been an amazing mentor, and committee member, but also friend over the past decade. Dr. Coronado has encouraged me to reach past my limits and had enough confidence in what I could do, that I ended up believing her. Dr. Eric Johnson has endured my learning curve for nearly two decades, but always amazed me in his endless mentorship, coaching and humor along the way. Dr. Sunny Lin has provided a fresh perspective, consistently pushing me to think deeper into concepts and arguments, convincing me that I am the expert. Finally, I'm very grateful that Dr. Lucy Savitz has been able to serve on this committee, as her extensive experience and expertise has provided insight that most students are not fortunate enough to have.

I also would like to acknowledge Dr. Melinda Davis, who talked me into going back to school during a poster presentation session at a conference. Her pep talks and introduction to Dr. Sherril Gelmon convinced me that I could thrive in this program. Dr.

Gelmon has not only acted as the program director, but as an advocate through the entire program. I am grateful for her guidance and friendship.

In the fall of 2017, I met a group of people who I knew would be incremental to my learning process. Our cohort leaned on and learned from each other over the next 4 years and provided support through a process that nobody else could understand. Soon to be Dr.'s. Lindsay Smith, Steven Fiala, and Sacha Walia have provided support, humor and advice when most needed. I would also like to add special thanks to the growing group of other PhD students in the program who through different stages of the process walked the same path providing companionship in a relatively lonely process.

Finally, from the bottom of my heart I would like to thank my family and friends. My extended family and friend group has endured four years of this program directing our lives and taking priority over weekends, events, holidays, and vacations. I especially thank my sister Samantha and my parents for feeding and loving me through the process, providing a Sunday dinner to my family when I was prioritizing school. I also want to thank my chosen family, the Ullan's, for endless love, support and understanding, and co-parenting of my children.

Lastly, the culture of my immediate family has been centered around studying and writing for the past four years. I was able to study for exams while on the bus driving throughout Europe and write while sitting on the banks of the Yellowstone River. To try to complete the program by the time Charles went to college required deep commitment from my partner Frank. Charles and Cecelia were 13 and 11 when I started

this program. Frank has been a loving and supportive partner who was able to pick up much of the parenting and household load to support my completion this program. Throughout this time, Frank has wiped many tears, reassured me when in doubt, encouraged my personal and professional growth, shifted schedules, and made excuses for my distraction and disconnection from our life together. My immediate family has listened to more discussions about colorectal cancer and risk prediction, policy, and health systems than most families ever imagine, or care to do. My family has grown with me, and I am truly grateful for their endless love and support.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Predictive analytics (PA) is the use of population-level data and statistics to predict future individual-level events. Prognostic or predictive risk models are often used to predict specific behaviors and/or risks. Several industries have relied on PA to optimize investments and reduce costs. For example, insurance companies may use credit scores to assess accident risk, online companies may use your purchase history for marketing, and airlines use models to estimate travel behaviors (Ravi B. Parikh et al., 2016; Parikh et al., 2019). Use of PA in healthcare has been steadily rising as health systems leaders are recognizing their value in targeting and optimizing care for patients at high risk of negative health outcomes (Miller, 2019). PA is a way to maximize the utility of healthcare expenses through precision delivery of care in our resource constrained healthcare climate (Bresnick, 2018). Understanding which patients most benefit can help care delivery systems to prioritize spending when resources are limited.

Current expectations for value-based health systems are to deliver accessible, patient-centered, affordable, and high-quality care. Health systems priorities informed by knowledge of costs and potential savings can drive leadership decisions on where and how resources can be used to identify health gaps and targeted care areas (Garber & Phelps, 1997). Identification of gaps in healthcare through PA reveals areas of inequity and allows programs to develop targeted or proportionate interventions appropriate for subpopulations (Benach et al., 2013). With the application of PA, and the successful

implementation of appropriate treatments or interventions, health systems may improve health equity and improve health care value (Braverman et al., 2017; NQF, 2008).

Risk prediction models can be used to identify populations at risk for disease or adverse events and have population health benefits (Benuzillo et al., 2019; Jeffery et al., 2019). The ecological fallacy clarifies that inferences are not about groups, rather about the whole individual and the many groups to which they belong (Kent et al., 2018). Risk prediction models can take into account the many groups to which an individual belongs, as well as their medical and other data, such as personal, family, clinic, community, and even policy level data which could influence health outcomes. However, models have not always been built using a broad patient population or multi-level data (Jeffery et al., 2019; Kent et al., 2018; Moons et al., 2019; Parikh et al., 2019). The use of PA in combination with multi-level data could be one way to recognize group membership and individual characteristics simultaneously. Advancements in the application of PA in addressing health needs may include increasing access to diverse sources of EHR and multilevel data.

Electronic Health Record (EHR) data is most commonly used in PA and is limited to information gathered during intake or during the clinical encounter. As patients are more likely to seek care when sick rather than well, these data can be biased and more frequently populated for patients with access to care. These data may include demographics, findings from the examination (like blood pressure and BMI), what types

of appointments or visits patients have, procedures, pharmacy prescriptions, diagnoses, laboratory results and some social determinants of health. However, the use of data outside of the EHR has become increasingly important as clinicians understand the importance of group membership or higher levels of influence on a person's health (Diez-Roux, 2000) and the need to address selection bias concerns. While group level data is often not readily available, some health systems see the importance of using external data such as Census data. Census data, as an example, has been used as a proxy for unavailable patient reported measures like income, education and health literacy, to measure community and neighborhood characteristics, or has been used as a proxy for community and social data such as a social deprivation (Knighton et al., 2016; Savitz, Bailey, et al., 2020). Identifying additional sources of data can give information to providers to improve health and drive the growth of big data for use in clinical practice or PA (Chambers et al., 2019).

Multi-level data is important to understanding health, as looking beyond individual level characteristics can tell you more about a person and the complex and interactive factors affecting care and outcomes (Diez-Roux, 1998; D. M. Rousseau, 1985). Levels of influence may include individual, interpersonal, provider, clinic, organizational, community or even policy level data (CDC, 2017). Understanding multi-level influences on outcomes may inform decisions about interventions (CDC, 2017; Kumar et al., 2012). However, multi-level data is not always available and often is not gathered in useful ways. Multi-level data is sometimes reported in ways that may

introduce measurement error (i.e. Census data) and affect predictive relations between the outcome and predictor (Woodhouse et al., 1996). Data quality sometimes limits the use of diverse datasets (Callahan et al., 2017).

The use of multilevel data may improve the accuracy of predictive modeling by including important non-clinical patient and community level characteristics. Many health systems-based studies to date have been limited to using available EHR and administrative data and fail to seek data sources outside of the EHR (Bhavsar et al., 2018; Goldstein et al., 2017). The Centers for Disease Control and Prevention (CDC) adapted the social-ecological framework (SEM) of health promotion for a multi-level approach to colorectal cancer (CRC) screening promotion (CDC, 2017). This framework addresses the need to consider individual, interpersonal, organizational, community and policy levels to understand screening patterns to optimize interventions (CDC, 2017). The addition of multilevel data may improve measures of prediction.

All data can vary in quality and availability. Data available in the EHR can vary by health system, and can be inconsistent by provider and over time (Savitz, Savitz, et al., 2020). Capture of data can depend on the structure of the EHR platform, data can be inaccurate or have errors, there may be missing data, and data can be subject to selective measurement (Savitz, Savitz, et al., 2020; Shah et al., 2018). Some important data, like family history or genetic data, is usually captured in the EHR but retained in notes and reports, so often unusable in analytics, although some systems are starting to integrated this data into the EHR (Ehrenstein V et al., 2019). Data like remote sensing

data from devices may be used in some systems, and projects, but is not often available. Claims data is sometimes used in analytics because claims tend to be more consistently recorded, yet they often lack the necessary detail needed, such as results of tests (Shah et al., 2018). Data sources beyond the EHR may include family history, genetic data, social, neighborhood or community level data, or data on social determinants of health (Bates, 2014). Yet these data sources may not contain all necessary data, and ideal data may be unavailable. While not all data in outside datasets are actionable, understanding characteristics of groups of patients could allow for assessment of appropriate interventions based on predictive characteristics.

Lack of access to multilevel data may limit the ability to create useful multi-level risk prediction models. The usefulness of multilevel models may be dependent on how easily accessible data points are, and while EHR data is easily accessible to a health system, multilevel data may be less so. Measuring attributes at different levels can be challenging and may be entangled in person level data, and there may be dependencies or confounding between levels (Diez Roux, 2008). Measurement of characteristics at the group level isn't always ideal, data quality and meaningfulness must be assessed (Diez Roux, 2008). While seeking to broaden data use to multiple levels, understanding the limitations of data and the potential for information bias, selection bias, measurement error and confounding at all levels is necessary. Understanding the importance of inclusion of multilevel data, attention must be paid considering the impact of adding in higher level data and the change in the effect of the individual characteristics. The ideal

data used in PA needs to be accurate, valid, and accessible. Further, the use of data should allow applicability across diverse sub-populations while understanding level interactions and factor relationships.

Generalizability is the ability to draw inferences on patient populations outside of a referent sample. In regards to predictive analytics, generalizability depends on the variables within a risk prediction model and on how generalizable or representative the population is in comparison to where the model will be applied (Steyerberg, 2019). Generalizable models may be used across settings and diverse populations. Models that are applicable to subpopulations provide promise in reducing bias and maintaining value to the health system. Use of multi-level data may improve generalizability, applicability, and accuracy of risk predictions models.

Generalizability allows for application of PA to other populations to improve health and outcomes. The internal and external validity of a risk prediction model indicates reproducibility and transportability to other systems (Austin et al., 2016). Samples used in PA can be assessed for heterogeneity to better understand the applicability of predictors effects on outside data (Debray et al., 2015; E. W. Steyerberg et al., 2019). As health systems seek information to allocate resources, use of models applicable to subpopulations or generalizable to outside populations may allow for streamlined implementation of effective interventions (Porter et al., 2006). Generalizability and applicability will increase the usefulness of a multi-level risk prediction model, although it may depend on the systems perspective, embedded

capacity to apply the use and results, available technology, and willingness to innovate (Cohen et al., 2014; Porter et al., 2006).

Given what we know about the potential for enhancing EHR-based PA models with multi-level data, the promise of addressing the needs of vulnerable populations warrants further study. For example, CRC is the 2nd leading cause of cancer deaths in the U.S. Nevertheless, CRC can be preventable if people are screened.

To date, predictive analytics have not been applied to predict a patient's likelihood of screening for colorectal cancer (CRC). However, there is potential that a multi-level risk prediction model could identify patients who are unlikely to screen for CRC. CRC screening can save lives because precancerous polyps or early stage cancers are detected and removed (Safayeva & Bayramov, 2019). Yet, only 69% of age eligible adults in the general population were screened for CRC in 2018 (CDC, 2018). This means that a staggering 21.7 million people have never been screened for CRC (CDC, 2018). Patient, provider, and system level characteristics have been found to contribute to the failure to screen for CRC (Muthukrishnan et al., 2019; Weiss et al., 2013; Yu et al., 2018). Multi-level factors influence CRC screening uptake and there is an opportunity in CRC to develop and investigate the use of multi-level PA as a tool to improve health outcomes for patients currently unlikely to screen.

Health systems or health plans can intervene to improve recommended screening adherence by targeting evidence-based interventions for vulnerable populations. These interventions could include multilevel interventions such as language

specific outreach, shared decision-making -, provider reminders, patient navigators, prompts and reminder calls, culturally appropriate mailings or letters, or centralized direct mailed screening programs, community media campaigns, or state policy recommending higher screening benchmarks (Coronado et al., 2018; Interventions, 2019). Patient navigation has been shown by multiple studies to be an effective intervention to increase CRC screening from 8-31% (Coronado et al., 2019; DeGroff et al., 2019; Horne et al., 2015; Rice et al., 2017).

Interventions like phone calls and patient navigation can be expensive, especially when applied to large groups of individuals such as the many patients unscreened for CRC. Health systems are more efficient when using analytic tools to help identify and target specific populations. Knowing who may be at risk for not adhering to screening recommendations could assist providers and clinics in identifying patients in need of early interventions, and how interventions should be tailored to be aimed at completing screening (Ravi B. Parikh et al., 2016). Critical evaluation before, during, and after implementation of PAs in clinical settings will ensure safe care, good outcomes, the elimination of waste, and more widespread uptake (Benuzillo et al., 2019).

Knowing which patients are likely to screen can allow clinicians and health systems in using scarce resources to optimally manage their at-risk populations. Known predictors of screening include individual level factors such as education and prior screening, system level factors like provider recommendations or a primary care relationship, and community level factors like healthcare access and rurality (Gimeno

García, 2012; Ni et al., 2020; Singal et al., 2017). People who are female, privately insured or live in urban areas are more likely to screen for CRC (Davis et al., 2017; de Moor et al., 2018). Multilevel data may improve models and predictors that may inform health systems of other effective interventions tailored for multilevel influences.

Clinics and systems face decisions in choosing and prioritizing interventions because it can be cost-prohibitive to intervene on all patients who have not completed screening, regardless of their likelihood of completing screening on their own. Stratifying patients in greatest need of interventions allows health systems to target resources to close screening gaps across populations. Improving risk prediction models by integrating readily available multi-level data could provide more benefit through improving model accuracy at very little cost. The actual benefit is determined by the systems leaders' perspectives on the value of the improvements. An improved and more accurate understanding of the groups unlikely to screen could lead to better matching with interventions. Precision delivery of interventions to those most likely to benefit might optimize patient outcomes, improve the targeting of interventions, and enhance opportunities to sustain successful interventions, especially in low-resource settings.

The broad adoption and use of PA can be influenced by systems leaders' willingness to invest and support PA. Organizational change is complex and dependent on knowledge of drivers of health gaps, the benefit of interventions addressing the gaps, and competing opportunities (Basu, 2011; McMillan, 2016). Further, the health system's capacity for change and technological limitations may impact a system's ability to use

multi-level PA (Scott, 2003). The use of multi-level information in PA promises to provide information that makes potential actions more effective. However, understanding health systems leaders' perspectives on the value and application of developed, multi-level models is critical to assessing their uptake and clinical applicability.

Problem Statement

PA can help direct healthcare resources in a way that improves care and reduces cost. PA has rarely been used to guide colorectal cancer screening promotion efforts, even though screening promotion is expensive, and some efforts are inefficiently directed. When PA has been used, it has only relied on EHR data.

- The EHR contains important individual information about a patient, but data external to the EHR is available and could be used to improve risk prediction. External data such as community and system level factors may improve model performance.
- Little is known about the value of external data sources in improving the predictive value or applicability of predictive models.
- For a PA model to be effectively used and sustained, health system leaders must invest and support adoption. Health systems leaders' perspectives are important and depend on the practical implementation of interventions, the cost to the system, and the evidence of value.

Research Questions and Aims

Research Questions

1) Can the inclusion of multi-level data improve the accuracy or applicability of a prognostic risk prediction model, in predicting patients' risk of failure to screen for colorectal cancer in order to target interventions?

2) How does the inclusion of multi-level data in a risk prediction model improve the usefulness to health system decision makers for managing population health?

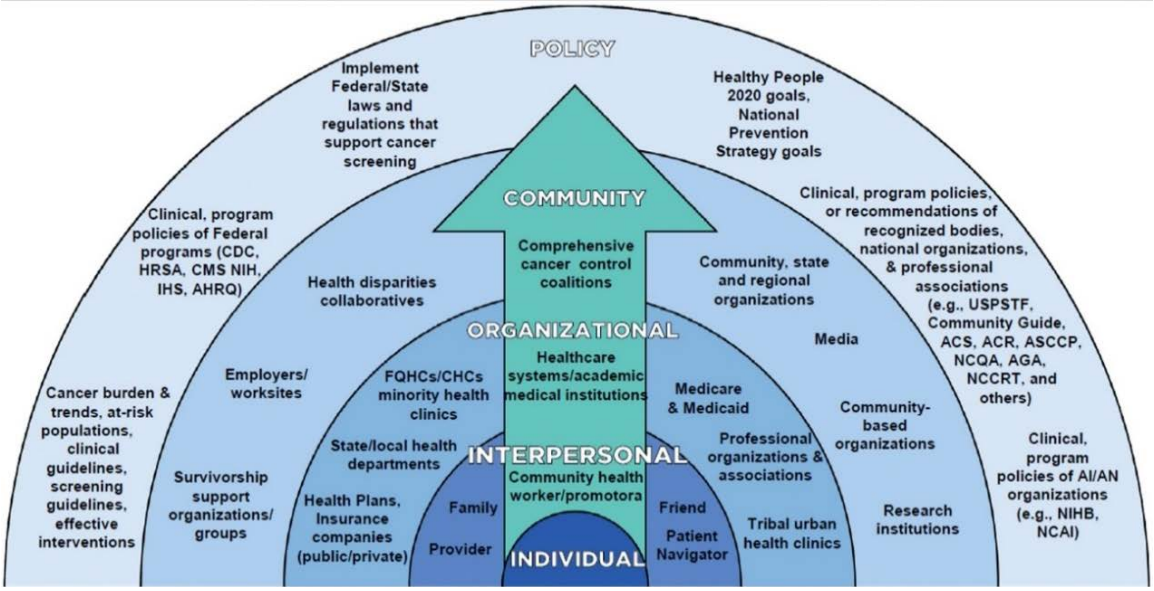
Aims

1) Assess data availability, quality, heterogeneity, and opportunities to broaden data availability at a health system based on the SEM framework. Categorize the data into levels according to the framework and ascertain availability and quality.
2) Develop risk prediction models using EHR and externally available multi-level data.
   a. First in a large dataset of patients due for CRC screening using standard individual level EHR data
   b. Assess the statistical improvements in models overall when adding multi-level data
   c. Assess applicability by applying the model to subpopulations, based on population characteristics such as race, ethnicity, or insurance type
3) Assess health system leaders' perceptions of the usefulness of the CRC screening model developed and the value of adding multi-level data, using qualitative research methods.

Theoretical Frameworks

The CDC's SEM Framework is an important tool for understanding the many levels of influence on our health (Figure 1.1) (CDC, 2017). Outcomes and health behaviors may be influenced by multiple levels, and understanding these levels gives a broader perspective on determinants and on designing interventions. Focusing on one level gives a narrow picture of factors impacting one's health behaviors and outcomes. The individual is at the center of the model, with surrounding levels of interpersonal, organizational, community and policy levels (CDC, 2017).

Figure 1.1 CDC's The Social Ecological Model



CDC, 2017

Use of the SEM model to drive a multi-leveled approach to PA may improve a model and add depth and dimension. The innermost individual band represents the individual, and characteristics which might affect their willingness to screen. Individual factors such as test preference, cultural norms and access to screening may influence the ability to complete screening. Interpersonal factors may include marital status, or the provider recommending screening. Organizational factors which could include access to colonoscopy clinics, clinic size, or waiting times. Community level factors may include colonoscopy capacity in gastroenterology clinics in an area, public awareness and media campaigns, or partnerships for cancer coalitions, population density, or unemployment rates. Policy level factors affecting screening may include insurance (HEDIS) or health organization (CCO, ACO) mandates for screening or participation in federally supported care delivery programs that require screening reporting (Uniform Data Systems (UDS)).

Following the guidance of the SEM data levels may improve subpopulation applicability and accuracy of risk predictions models. Integrating information from the various levels of influence augments EHR data with contextually rich external data, making the model more specific to outside influences. The SEM framework allows assessment in understanding the reciprocal interrelationships between the individual and population-level determinants of behavior and health (Moore et al., 2015).

Data from the different levels of the SEM framework may not always be available, ideal, or usable. Individual data is limited to data available in the EHR and may

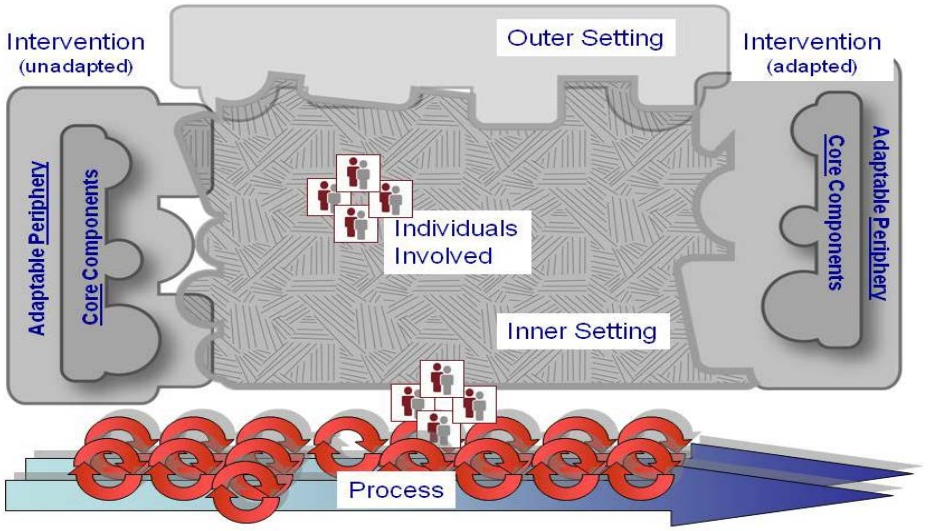
be lacking information about individual knowledge, attitudes, and beliefs toward healthcare or screening in general. While available person-level data may include provider characteristics or family history, ideal data like information about social networks or support systems may not be obtainable. Organizational data such as clinic size, location, and screening rates are available, but data on organizational rules or regulations is lacking. Community level data may include unemployment rates or population density but lacks data about relationships among organizations and partnerships. State and national policy data may be available, but local policies and institutional level data may be difficult to acquire. Regardless, assessing what data is available at all levels is essential because it can inform adequate interventions for specific patient populations. Perfect fidelity to data sources guided by the SEM, may not be feasible. Allowing flexibility and adaptability based on data availability may be necessary. This project will assess the availability and quality of data at the different SEM levels, the use of the data in risk prediction, and the value of the use of the data through qualitative interviews in Kaiser Permanente Northwest (KPNW).

The qualitative interviews will assess the functionality of the multi-level data from the health system's perspective and will be guided by relevant elements of the Consolidated Framework for Implementation Research (CFIR) (L. J. Damschroder et al., 2009). CFIR is a framework for seeking information for research to be translated into practice through a theory based on the various settings within and around a health system (L. J. Damschroder et al., 2009). The CFIR components are the intervention, inner

setting, outer setting, individuals and process (L. J. Damschroder et al., 2009). The questions for the qualitative interviews will be guided by the CFIR settings. This is the first step of understanding if and how a multi-level risk prediction model is useful from the perspectives of the decision makers.

The multilevel components of CFIR will guide the questions to understand how the implementation of the risk prediction model and use of multi-level data might occur (intervention) and if the findings are clinically meaningful from the health system perspective. The interviews will assess the characteristics of the health system that allow for adoption, such as priorities and goals (in the inner setting), the motivation to use multi-level data from the outer setting perspective, and the individual's perspective on the value of PA and if they find benefit in using multi-level PA including multi-level data.

Figure 1.2 Consolidated Framework for Implementation Research



(CFIR, 2020)

The use of risk prediction to identify patients unlikely to screen for colorectal cancer can allow health systems to systematically target interventions to the group needing it the most based on the components of the model. Systems can design interventions around the levels of influence in the multilevel risk models. The impact of the interventions can be maximized by considering all levels of influence. Behavioral interventions, health system interventions, or community interventions could be implemented based on knowledge gained from the integration of the SEM framework into analytics, and health system perspectives guided by CFIR.

For this study, understanding the various levels of influence and their effects on screening rates can provide insight into how to optimize screening and increase screening rates. CRC screening is appropriate for multivariate risk prediction because increasing timely CRC screening can prevent colorectal cancer.

Methods and Design Overview

In the existing literature, only one prior risk model has identified patients who are unlikely to show for a colonoscopy appointment, not CRC screening overall (Blumenthal et al., 2015). However, this model only took into account screening via colonoscopy (only one method of screening) and did not include multi-level data (Blumenthal et al., 2015).

Using data from KPNW, available EHR data together with external data across the multiple bands of influence framed by the SEM (Table 1.) will be aggregated. Guided by literature on predictors of screening, potential measures will be grouped by level of

influence in the SEM framework. Sources of data outside of the EHR will include data from the U.S. Census Bureau, American Community Survey, Department of Health, and locally available data.

Table 1. Sample Measures for Multivariate Risk Prediction Model

Level	Variable	Source
Individual	Perceptions of screening (test preference)	Clinical Visit (NLP*)
	EHR clinical data (demographics, vital data, diagnoses, comorbidity scores, encounters, prescriptions, procedures, labs)	EHR
	Social Determinants of Health	EHR
	Census data (neighborhood characteristics)	Census
	GINI Income Inequality Index	American Community Survey
Inter-personal	Family data (family history of disease, family size, marital status)	EHR (NLP)
	Number of address changes in past 2 years	EHR
	Provider panel screening rates	EHR
	Other provider characteristics (length of time in practice, panel size, panel characteristics)	EHR
Organizational	Clinic characteristics (size, location, funding, diversity of patients, provider FTE, availability of translation services?	Clinical data
	Clinic screening rates	Clinical data
	Colonoscopy characteristics (wait times for colonoscopy, distance to GI clinic, hospital/clinic partners)	Clinical data
Community	Community-level screening rates	Department of Health
	Public awareness (media campaigns)	Clinical knowledge/data
	GINI Income Inequality	American Community Survey
	Unemployment	American Community Survey
	Population density	American Community Survey
	Urban/Rural location	Urban/Rural Location
	Food access	Food Access
	Census data (median household income, unemployment rate, household poverty)	Census
	Area Deprivation Index	Neighborhood Atlas
Policy	Access to insurance (Medicaid Expansion, % uninsured)	State data
	Screening costs (Legislation covering follow-up colonoscopies)	State data
	Incentives for screening	State data, Department of Health

	CCO/ACO Coverage (care coordination by CCO/ACO)	State data, Department of Health, CMS for ACOs

*Natural Language Processing (NLP) may be used to identify information obtained in the clinical visit but not in a discrete field

Risk prediction models will then be created to predict the likelihood of screening for CRC. Models will be developed using individual level data and then with added multi-level data guided by the SEM framework. Model performance improvement will be assessed, through an improved internal validity and graphical calibration, when available SEM data is added to the model. Additionally, models will test the subpopulation applicability when applied to a subgroup population and non-Medicaid population separately.

The models will utilize a large retrospective sample of patients at KPNW. Members will be eligible for inclusion in the model if they are 50-75 years old in 2019 and were due for screening. The outcome of interest will be the completion of screening, based on HEDIS criteria, including fecal testing, FIT DNA, or colonoscopy (CDC, 2016).

After the risk models are complete, qualitative interviews with clinic leaders, clinicians, and decision makers will be conducted, based on relevant elements of the CFIR framework described above. The interviews will be centered around the value of a PA tool based on a multi-level data modeling practice. These interviews will be structured in-depth interviews of decision makers at KPNW, such as the Gastroenterology (GI) medical director, quality improvement manager, primary care project manager for CRC screening, or the director of quality improvement in both the

Northwest Permanente and Kaiser Foundation Hospitals organizations. Questions will be guided by relevant CFIR framework factors to understand the intervention, internal, external, and process context of making changes to use PA for CRC screening, and the feasibility of integrating multi-level data obtained outside of the EHR.

Questions framed by CFIR will specifically seek input regarding the clinical perspective of the usefulness of the PA tool in practice. For the intervention level, the multi-level risk prediction model, questions will be framed to understand the usefulness of the PA tool from the interviewee's perspective. The inner setting questions will focus on internal priorities, relative advantage, trialability, ease of use, and sustainability of acquiring multi-level data outside of the standard data sources. Outer setting questions will seek to gain insight on how the enhanced PA tool may be useful, concerning providing an advantage over competitors or motivation for improved resource allocation. Process questions will be framed to seek information on infrastructure and capacity issues that add to the usefulness of a PA tool. The interviews will additionally seek input on policy recommendations for using PA and expanding data use across the SEM levels of multi-level data.

Purpose and Significance of the Study

Predictive analytics has the potential to increase quality and saves wasted resources. When standardized and used appropriately, the use of multivariable predictive models can improve patient health. Some health systems are unable to use

PA to manage population health due to capacity or technology limitations. Other systems are expanding their use of PA to conduct better population health management. This study seeks to build enhanced PA models using multi-level data. Expanding data sources and the use of data across levels will be valuable in creating a more accurate and applicable model. Working with available data to most systems will make the model easily replicable. This study will assess the availability of multi-level data and the difference it makes in predictive model performance.

Health systems and health plans may recognize the benefit of using PA, and the positive impact on the way resources can be allocated to improve health through a better understanding of predictors. The performance of the models and improvement by adding multi-level data may influence adoption. The probability of adoption of multilevel risk prediction models is influenced by the reactions and perceived usefulness of such tools by decision makers. While health system leaders may adopt the use of predictive analytics as a tool to increase screening rates, it is only useful if health systems find it valuable and informative, and if the addition of multilevel data adds value. This study will understand the response to the use of multilevel data from a health system perspective. It will further ascertain implications for policy and practice and future research, including how PA tools are part of larger complex interventions aimed at improving targeted health outcomes (e.g., targeted CRC screening informed by enhanced PA and including patient reminders, mailings, improved ease of scheduling).

Organization of the Research

This research will be described in 5 chapters. Chapter 1 is the introduction and has provided an overview of the background, problem, proposed research questions and aims, theoretical framework guiding the research, and its purpose and significance. Chapter 2 will describe the state of the literature on the use of predictive analytics, how systems can benefit from predictive analytics, how predictive analytics can be used in population health, and the need for PA in CRC screening. Chapter 2 will also describe the literature on clinical data availability, usefulness, and multi-level data through a systems perspective. Chapter 3 will provide a thorough description of the methodology including the quantitative application of risk prediction models and evaluating improvements in the model, and the qualitative interviews to understand systems perspectives. Chapter 4 will describe the findings from the risk prediction model, the improvements through adding multi-level data, and the results of the qualitative interviews and key findings. Finally, Chapter 5 will include a discussion of the results outlining implications for policy makers in health systems, it will address the limitations of the study and make recommendations for future research.

Summary

Precision medicine and PA can be used to prioritize interventions and increase preventive screenings, which is especially useful when applied to identify populations in greatest need of costly interventions. Health systems are continually challenged to

reduce expenditures in a dynamically changing, resource constrained healthcare environment.

Multilevel data, when available, holds promise to improve the accuracy of models. Health systems may use multilevel data if it is available and increases the accuracy of risk prediction. Expanding data availability and usefulness is a step forward in improving clinically meaningful, targeted risk prediction models.

Increasing CRC screening is necessary to reduce cancer deaths, yet not enough people are screened, and unnecessary deaths occur. Using multilevel predictive analytics to identify people in need of CRC screening interventions may improve health outcomes among patients due for screening and the efficiency of interventions if used by health systems.

Definition of Terms

ACO	Accountable Care Organization
CCO	Coordinated Care Organization
CDC	Centers for Disease Control and Prevention
CFIR	Consolidated Framework for Implementation Research
CMS	Centers for Medicare and Medicaid Services
CRC	Colorectal Cancer
EHR	Electronic Health Record
FIT	Fecal Immunochemical Test

GI	Gastroenterology
HEDIS	Healthcare Effectiveness Data and Information Set
KPNW	Kaiser Permanente Northwest
NLP	Natural Language Processing
PA	Predictive Analytics
SDOH	Social Determinants of Health
SEM	Social Ecological Model
UDS	Uniform Data Systems
USPSTF	U.S. Preventive Services Task Force

Chapter 2: Background and Literature

Chapter Organization

This project seeks to understand if the inclusion of multilevel data will improve a prognostic risk prediction model for a patient's likelihood of colorectal cancer screening and if the health system finds the model useful for managing population health. This chapter provides a review and synthesis of the literature on the use of predictive analytics (PA), which serves as the basis for this study. This chapter also reviews the literature on the current use of multilevel data, the use of predictive analytics in health systems, preventative medicine, and colorectal cancer screening, and the interventions that increase screening. This chapter finally outlines the theories that will support and guide the methods of the project. Specifically, this chapter will:

1) Review the case for PA and how data is currently being used to improve health, the accuracy and applicability of enhanced models, and the importance of high quality and available data to reduce information bias.
2) Review the benefits of the use of multilevel data, access to multilevel data, and how multilevel data can reduce bias and improve applicability.
3) Describe how health systems are using predictive analytics, how systems can benefit from predictive analytics, and health systems' views on the use and usefulness of PA.
4) Discuss how predictive analytics can be used as part of complex interventions to improve adherence to recommended screenings.

5) Make the case for predictive analytics for CRC screening, assess why CRC screening is important, and describe the need for targeted interventions to increase CRC screening.
6) Provide an overview of the Social-ecological Model (SEM) and Consolidated Framework for Implementation Research (CFIR) frameworks.
7) Summarize barriers and limitations to the use of predictive analytics, including data limitations and availability.

This chapter provides the background as to how multilevel data can improve the accuracy of a prognostic risk prediction model in identifying patients at risk for failing to screen for colorectal cancer, and how health systems may apply a PA tool in practice.

The Current use and Purpose of Predictive Analytics

Precision delivery of medicine is the use of high value care that is cost effective. The use of analytics to deliver precision medicine is one way to maximize the utility of healthcare expenses, which is necessary now in our healthcare climate (Parikh et al., 2016). While other industries, such as insurance and advertising, have relied on predictive analytics to maximize investments, use in healthcare has been slow to adopt, sometimes due to limited clinical interest, health system capacity, or access to data (Parikh et al., 2016).

Precision medicine is the application of interventions, treatment, or specific care to individuals or groups of people based on personal, medical, lifestyle, social, clinical, or community characteristics. (R. B. Parikh et al., 2016) One method used in precision

medicine is predictive analytics (PA), where historical data is used to predict future events, outcomes, or behaviors. Models are typically developed or "trained" on one population, and then validated in an external population (Steyerberg & Harrell, 2016).

PA can improve the value of healthcare for some patients at high risk of negative health outcomes (Parikh, Kakad, & Bates, 2016). For example, PA has been used to predict specific diseases such as heart failure or end-stage renal disease prior to onset allowing for earlier treatment or prevention. (Johnson et al., 2008; N. Parikh et al., 2008); Bates and colleagues provide examples of opportunities to use big data and predictive analytics to reduce costs (Bates, 2014). These examples include using PA to address high-cost patients for actionable treatment, reduce emergency room readmissions, triage care for patients based on the risk of complications or decompensation, predict adverse events during a hospital stay and optimize treatment based on the prediction of the trajectory of disease (Bates, 2014). Bayati and colleagues studied the economic impact of the application of a readmissions prediction score and found that a post discharge plan based on the prediction and shared decision making for appropriate patients would have an 18.2% reduction in rehospitalizations, and save 3.8% of costs (Bayati et al., 2014).

PA can also be used in public health and population health management to predict outcomes and behaviors in groups of people to apply tailored interventions. Khoury and colleagues discuss the ability to deliver precision public health through the application of PA to determine the "right intervention, to the right population at the

right time" (Khoury et al., 2018). Finding these right populations and tailoring interventions may provide personalized care and system efficiency.

While the adoption of PA has been less expeditious than in other industries, the use of PA in health has spanned across the spectrum of health care. PA has been used in health care program management, reporting, and resource allocation. The National Quality Forum (NQF) standardized public reporting of national quality-based data to increase consistency in reporting by creating a model that predicts the risk of mortality in 30 days. (McNamara et al., 2015) The World Health Organization (WHO) used PA to identify geographical areas where mental health programs were underdeveloped and where mental health personnel would have the most impact. (Hudson, 2010) In Sub-Saharan Africa (data from 14 countries), PA was used to estimate human immunodeficiency virus (HIV) incidence infections for program planning. (Borquez et al., 2016) In Spain, chronic disease management groups have used PA to monitor pharmaceutical costs, reduce counteractive treatments, overdosing, and overprescribing. (Vivas-Consuelo et al., 2014) Similar strategies and use of PA could be applied to improve health equity through targeted interventions and address the social determinants of health (SDOH).

PA could also be utilized at the health system level to improve health, albeit integrated systems have a wider range of PA ability due to access to more data. PA has been used to reduce health care costs, increase health equity, and address SDOH. Socioeconomic factors are associated with high health care resource utilization. (Beck et

al., 2012) In Canada, PA was used to identify the top 5% of health care users in five years, and resources were provided to improve access to more appropriate health care. (Rosella et al., 2017) PA has identified individuals likely to readmit to the hospital, experienced complications based on the severity at presentation during hospital admission (triage status), or experience decompensation and adverse events during a hospital visit. (Bates, 2014) Health systems have identified interventions to improve health through PA such as recognizing people likely to need dialysis, likely to develop poor outcomes of heart failure or to fail to get a colonoscopy. (Blumenthal et al., 2015; Johnson et al., 2008; Smith et al., 2011)

Predictive models have been found in some cases to have a superior ability to predict clinical outcomes compared to clinical judgment. (Shah et al., 2018) Compared to human judgment precision medicine may have implications for the use and adoption of such models in clinical decision making, and for management policies that could improve health.

Generalizability

PA is more useful if models are generalizable to systems or populations outside of those where the models are developed and validated. Generalizability is achieved when the data used in the model is available in outside settings and when the model is applicable to heterogeneous populations (He et al., 2020; Ewout W. Steyerberg et al., 2019);. Data must be available for both the referent population and the external

population. While health systems do not consistently capture important socioeconomic and environmental data, the benefits of these data are widely known (Mahmoudi et al., 2020). Broadening the data to include publicly available data, or data from easily accessible data sources may help achieve generalizability.

Cohen and colleagues discuss the need to disseminate models more broadly, once they are built and validated (Cohen et al., 2014). Creating models that are easily validated and then used with data available at other systems allows the broad dissemination and ability to test and use, even if imperfect (Cohen et al., 2014). Generalizable models allow systems with limited analytic capacity to use PA, without the burden of customizing to their own system. In these cases, generalizability may be preferred over effectiveness based on specified data to a single system. McGraw and colleges go further to suggest that systems work cooperatively to combine data or distribute models (McGraw et al., 2012). Standardized models that are distributed across systems can inform research with more robust datasets, where models can be applied to large diverse populations (McGraw et al., 2012). The models could also be distributed for use by health systems (McGraw et al., 2012). For broad adoption or wide use of standardized models, systems must be open and transparent about the data and the PA models available.

The creation of generalizable models in the use of PA depends on having high quality, reliable, and available data. It also may depend on understanding the heterogeneity across subpopulations, testing models across health systems, and sharing

models that other systems may be able to use. Models can be more widely used when data inputs are readily available and not specific to a given health system.

Data and Predictive Analytics

Electronic health record (EHR) data, claims, and administrative data provide valuable information about a patient and their clinical encounters. Yet there are key limitations in this type of data including the lack of information about a patient, their family, or their community that could inform their health.

Having high quality, reliable, and available data is necessary to validate high performing PA models. PA has historically used administrative data, or data from the electronic health record (EHR) (Mahmoudi et al., 2020). These data fail to include family, interpersonal, organizational, community, or policy level data that may impact an individual's health.

Chambers and colleges adequately described the benefits and limitations of administrative and EHR data (Chambers et al., 2019). Administrative data includes data gathered from billing and administrative records or claims which includes diagnosis/CPT codes, inpatient and ambulatory encounters, services rendered such as tests and procedures, and provider services (Chambers et al., 2019). The purpose of these data is to optimize reimbursement. Administrative data is inexpensive to use, can offer large sample sizes with consistent data formatting, and provide little worry about uniformity; it is not complete (Chambers et al., 2019). Administrative data, can take time to be

available and documented, does not include important clinical information such as results or values of tests, and includes limited demographic data (Chambers et al., 2019).

While EHR data does include more nuanced information from the clinical visit, data is sometimes recorded in unstructured fields such as notes and reports (Chambers et al., 2019). The primary purpose of EHR data is to document treatment provided to an individual at a specific health system.

Because EHR data is often not standardized across systems (including in a given EHR product like EPIC)data is not necessarily stored or recorded in the same way (Chambers et al., 2019). Therefore, using data from different EHRs requires oversight, re-coding, and validation (Chambers et al., 2019). McVeigh and colleagues validated EHR data from a surveillance system and found that BMI is often not well measured, depression is often underdiagnosed, and there is insufficient capture of influenza vaccinations (McVeigh et al., 2016). One criticism of the use of EHR data in PA is that it is rarely validated prior to use (Mahmoudi et al., 2020). While clinical data is often missing or unavailable, the capture of social determinants of health and social data can be even more sparse.

Data from the EHR and administrative data are limited to patients who have interactions with healthcare systems; leaving out those who are unable or unwilling to interact with providers (e.g., people who don’t have insurance, or who are distrusting of providers). Building PA on these data without measures to evaluate bias could

perpetuate these inequities (Obermeyer et al., 2019). Using multilevel data may address this problem by expanding the usable data outside of the clinical encounter. However- multilevel data is not often readily available. Health systems do not consistently or systematically capture salient socioeconomic and environmental data about how and where a patient lives, although the benefits of these data are widely known (Mahmoudi et al., 2020). Gottlieb and colleagues noted that integrating social data can improve population health by improving panel management and expanding quality improvement beyond medical interventions (Gottlieb et al., 2016). Yet, social determinants and characteristics have yet to become a consistent part of the medical record (Gottlieb et al., 2016). Even when social determinants are sought, patients may be reluctant to respond because the questions can be sensitive (food insecurity and domestic abuse), and social determinants change frequently (transportation) (Andermann & Collaboration, 2016).

Data from the EHR are incompletely captured and have a high degree of missingness. Understanding the benefits and limitations of administrative, claims, and EHR data is important to improving predictive analytic models and adjusting the model to include the necessary data based on what we find. Selection or sampling bias can be found when omitting patients from a study sample based on missing data, or only including individuals based on a health event (Haneuse & Daniels, 2016; Kaplan et al., 2014). Training models on data where populations may have more missing data could provide a minority bias, or a training skew (Rajkomar et al., 2018). Finally, ascertaining

the missingness across groups is imperative to eliminating bias in models, and understanding where sensitivity analyses may be important (Kaplan et al., 2014). Finding ways to incorporate more data that can tell us even more about a specific patient type or sub-population can even further improve models. Understanding the benefits and limitations of administrative and EHR data is important to improving PA models and adjusting the model to include the necessary data based on what we find. Finding ways to incorporate more data that can tell us even more about a specific patient type or sub-population can even further improve models.

Multilevel Data

The individual, family, provider, system, and policy levels can affect a person's health, and obtaining data from multiple levels when available can improve prediction models. Using data from the EHR in collaboration with other sources provides a unique opportunity to identify patients at high risk of an event so that care can be optimally delivered and personalized. (Amarasingham, 2014; Heitmueller et al., 2014; Parikh et al., 2016). The importance of multiple levels of data for use in prediction models is being recognized globally (Heitmueller et al., 2014). For example, Hudson and colleagues discussed a multilevel model used and implemented by the World Health Organization to determine areas of high need for mental health services for prioritizing mental health staffing (Hudson, 2010). Here, data on the demographic, economic, political, social, cultural, environmental, and geographic conditions contributed to determining the

appropriateness of services (Hudson, 2010). One study in 14 African countries used multilevel predictive modeling to estimate HIV distribution using demographic and risk behaviors (Borquez et al., 2016). Another study in all African countries used multilevel predictive modeling to prioritize the treatment of malaria in children using estimates of fevers and risk of infection prevalence (Gething et al., 2010). Models are based on combinations of administrative, clinical, geographic, and policy data, and both studies suggest the models be used to optimize health care expenditures and clinical resources (Borquez et al., 2016; Gething et al., 2010). The use of data from multiple levels can improve the understanding of drivers of health and provide improved information for clinicians and decision makers (Chambers et al., 2019).

Recognizing the benefits and finding sources of multilevel data in determining care improvement and healthcare interventions is imperative. Using data to further understand multilevel influences is one way to integrate multilevel information into care. People, as patients, have distinct characteristics while simultaneously belonging to various groups. The literature has addressed the benefits of looking at micro and macro level data about a person and the groups to which they belong. The literature shows that the different levels are linked or interconnected and that levels can be synergistic (Denise M. Rousseau, 1985; Taplin et al., 2012; Weiner et al., 2012).

Zapka and colleagues specifically described the multilevel factors in cancer care that affect the quality of care and outcomes through two patient examples (Zapka et al., 2012). For Mrs. Smith, who opted for a catastrophic plan at retirement, factors that

impacted her care were her personal characteristics (she was active), missed opportunities at care, and lack of insurance to receive care when opportunities were found (Zapka et al., 2012). The outcomes of these multilevel factors of influence included the failure to seek preventative care, breast, colorectal, or cervical cancer screening for 25 years (Zapka et al., 2012). For Mrs. Adler, a young breast cancer survivor, her individual biology, family characteristics (genetics), and a fragmented healthcare system impacted the quality of care she received in that her physical and social needs were inadequately addressed (Zapka et al., 2012). Mrs. Adler's outcomes include the failure to receive follow-up care or management of conditions post treatment (Zapka et al., 2012).

Use of Predictive Analytics in Health Systems

Health systems face escalating costs while trying to accomplish more work with fewer resources (System, 2005). Health systems can gain an advantage from PA by improving the allocation of resources and providing more effective personalized care. This can be done by targeting resources through precise applications of interventions to patients that need them the most and by reducing unnecessary transactions. PA can be cost saving, by maximizing the investments for that organization, which will eliminate waste and reduce transaction costs in determining service needs. More directed resource allocation will improve healthcare quality and safety through a more precise and personalized application of care. Information technologies like PA can reduce

transaction costs, specifically resources spent on low value procedures or encounters by incorporating technology. Johnson and Rossow describe the use of health information technologies (HIT), such as PA, to reduce transaction costs, and improve resource allocation by improving processes. Process improvements may include improved or more applied communications with the patients, reduction of redundancy, or elimination of unnecessary procedures/tests within the system(Johnson, 2017). Improvements that have been strategically applied to increase the effectiveness of treatments to reduce costs are becoming more common in healthcare as a strategy for improved value.

The use of precision medicine might contribute to "strategic behavior" which would include employing and integrating resources for PA. PA is easier in health systems with analytic capabilities and available data, such as integrated health systems. Bates and colleagues outlined the strategic use of PA in identifying high cost patients, reducing hospital readmissions, identifying patients that may need in-patient triage, predicting decompensation in hospital visits, identifying adverse events to conditions, and predicting the trajectory of disease among multiple organ systems. (Bates, 2014) Among potentially high cost patients, systems were able to identify the patients most likely to benefit from actionable interventions such as care management. (Bates, 2014) The key in this example is not just to identify patients with high costs, but to identify those patients that would benefit from available interventions to coordinate services and navigate the system. (Bates, 2014) One such intervention might be to identify care

management or even an electronic device to monitor physiological attributes after a hospitalization to reduce readmission. (Bates, 2014)

The use of PA also can reduce redundancy, overuse of care, and treatment errors. Winters-Miner in their article discussed seven ways PA can improve healthcare. (Winters-Miner, 2014) Among the ways listed were the use of PA in preventive medicine and public health to improve health outcomes rather than treating "sick outcomes", targeting evidence-based medicine to populations to reduce wasteful treatments, and providing more accuracy to clinically significant predictions. (Winters-Miner, 2014) Patient outcomes may also improve by allowing patients to understand their own risks to disease and poor outcomes in an effort to engage them in improving their health. (Winters-Miner, 2014)

Precision medicine and PA can be an organizational asset for a health system if it improves resource allocation. (Williamson, 1981) Asset specificity, including technology specificity in the case of healthcare, transactions such as encounters, labs, and other testing can be reduced or improved to targeted populations. Asset specificity including technology specificity, such as PA, can reduce transaction costs. (Williamson, 1981) The more transaction cost minimizing PA can be, the more useful it will be to a health system. (Williamson, 2010) Treatments and screenings are specific assets that need to be applied to the right patients at the right time (Khoury et al., 2018).

Having asset specificity in PA programs targeted to predict healthcare outcomes may temporarily increase transactions through increasing preventive screening but

result in reducing more expensive chronic disease management while simultaneously improving quality of life. The prevention of disease, early detection of disease, or control of existing disease can all lead to reduced health system costs. (Woolf, 2010)

Information technologies and incorporating technology like PA reduce transaction costs such as time specificity. Organizational improvement through incorporating technology can use information processing to improve economics and the flow of information in an organization.

While PA can reduce transaction costs, improving the accuracy of risk prediction models may lower costs of interventions by more accurately identifying groups. Improved cost effectiveness is achieved by avoiding unnecessary transactions or by preventing poor outcomes. To improve accuracy and benefit from the use of predictive analytics, health systems will need robust data sources and multilevel data to enhance models and inform health systems of interventions for targeted populations. To effectively conduct PA, data systems will have to expand beyond the EHR and have more accessible data for system use. (Bates, 2014)

Health Systems Use of Multilevel Data

Health systems have access to data from multiple sources as mentioned above. These sources include administrative and operational data, data on the encounter (type and date), and other information from the medical record including diagnoses and procedures, claims data, and some social data from enrollment records. (Bates, 2014)

Health systems will benefit from exploring how to use multilevel data and data from other available sources. Using provider-level cancer screening rates or clinic-level screening rates will help systems know where resources may be best allocated. Interventions such as provider education or making systemic changes to increase screening could be more effective than patient level interventions.

Community- and policy level data may help address barriers to patient's health, such as food access or insurance stability issues. Pruitt and colleagues assessed the impacts of multiple levels on colorectal cancer (CRC) screening and found that screening varied the most across the neighborhood, physician, and clinician levels. (Pruitt et al., 2014) Further assessment of characteristics of successful screening levels could help systems understand how to best address levels of poor screening performance.

As more health systems access multilevel data, methods, and understanding about how to best do so will become established. Currently, health systems' access to data may be limited and inconsistent across systems. (Cohen et al., 2014) Integrated health systems may have more wide access to different types of data and normalizing the use of multilevel data may provide opportunities for it to become more widely available. Creating PA models using data that is easily available or disseminating methods for data acquisition will help produce improved adoption of these interventions.

Health Systems Approaches

An organization might be motivated to adopt PA because of its social values, organizational culture, or to reduce inequalities by directly providing improved healthcare to disadvantaged groups. (Cookson et al., 2017) The use of PA can improve equity of healthcare use among subpopulations by intentionally focusing on social determinants of health and the health of marginalized groups. Improving the overall efficiency with which health care is delivered can free up resources to be targeted to those in greatest need and reduce inequalities. Nonetheless, the organizational beliefs, values, and rules could drive this motivation. A health system is complex with organizational influences from individuals within the organization, the inner setting, and the outer setting. Health systems can also be racialized, may uphold racial inequities, or there may be an unequal distribution of resources (Ray, 2019). Understanding the influences of these contextual levels on organization performance will allow a better understanding of why they may see PA as useful.

Providers' and decision maker's beliefs about risk prediction may also drive the use of PA in health systems. Through qualitative interviews, Matthias and Imperiale found a prediction tool for identifying patients likely to have CRC was useful in that providers found that the model increased patients' likelihood to screen, increased efficiency of CRC screening, acceptability, and feasibility of implementation into clinical care (Matthias & Imperiale, 2020). In this case, the providers' beliefs about the PA tool increased the efficiency of use. For a PA tool to be effectively used and sustained, health system leaders

must buy into its value. Health systems leaders' perspectives are important and depend on the application of interventions and the cost to the system.

Health systems are often motivated to reduce escalating costs by efficiently using resources, saving time, and reducing redundancy. They are also motivated by regulatory requirements and published quality metrics. Health systems may strategically use technologies like PA to optimize the delivery of preventive medicine. An organization may see PA capacity as an organizational asset if it can improve the accuracy of care delivered. The organizational values and culture within a health system may be a catalyst for use of PA to improve efficiency in care.

Use of Predictive Analytics in Preventive Medicine

Predictive analytics can be useful in preventive medicine, especially when risk factors are modifiable. This section shows that risk prediction has been shown to be effective in the identification of opportunities for a more precise application of preventive medicine in the areas of predicting the likelihood of heart failure, diabetes and diabetes complications, chronic kidney disease (CKD), cancers, and a variety of other areas.

Cardiovascular Disease

The Framingham Heart Study Model was one of the first tools created to synthesize known risk factors into a prediction tool for cardiovascular disease and has

since been updated to predict other cardiovascular outcomes (Dawber et al., 1957; Mahmood et al., 2014). Conducted in Framingham, Massachusetts, it is perhaps the most widely known model of risk prediction (Mahmood et al., 2014). The Framingham Study, a longitudinal cohort study, followed patients aged 48-50 to track patient-level characteristics and outcomes found distinct characteristics of patients likely to develop cardiovascular disease. Since Framingham, thousands of articles have been written regarding interventions and prevention of cardiovascular disease (Mahmood et al., 2014). Of specific interest are modifiable behaviors like weight loss, dietary change, and pharmaceutical interventions that can change the course of disease if the risk is known early (N. I. Parikh et al., 2008).

Diabetes and Chronic Kidney Disease

Diabetes prevention also has benefitted from the use of PA. It has been found that early diabetes detection and treatment can reduce complications such as renal complications, stroke, and heart disease (Collins et al., 2011). Wilson and colleagues found that family history of diabetes and obesity, hypertension, low HDL cholesterol, elevated triglycerides, and impaired fasting glucose can identify patients likely to delay a diabetes diagnosis (Wilson et al., 2007). A variety of other studies have been conducted to create models that identify complications from diabetes, or the risk of other diseases due to diabetes such as cardiovascular disease or CKD (Zhao & Wong, 2018).

The early identification of patients who are likely to develop CKD can improve outcomes through early access to nephrology, and patient education and tracking(Nelson et al., 2019). Similar to diabetes, several studies have predicted adverse outcomes of CKD, including the need for renal replacement therapy, likelihood of developing hyperkalemia, or mortality (Johnson et al., 2010; Schroeder et al., 2017; Smith et al., 2010; Weiss et al., 2015). Early interventions, such as referral to nephrology can prevent adverse outcomes.

Cancer

Risk prediction has been widely used across the cancer spectrum. The most frequent uses have been in the most common cancers, lung, breast, prostate, and colorectal cancer.

Predictive models in lung cancer are prolific. In 2020, Toumazis and colleagues through a systematic review found 78 distinct models to predict the likelihood of developing lung cancer (Toumazis et al., 2020). The models included epidemiological models, and clinical models requiring clinical input (Toumazis et al., 2020). The use of risk prediction in lung cancer has decreased the number of screenings needed, increased detection of cancers, and reduced false positive tests (Tammemägi, 2015).

Breast cancer risk models look at potential risk over time at different time intervals such as in the next 5 and 10 years (Eriksson et al., 2017; Terry et al., 2019). Breast cancer models often include family history, age, race or ethnicity, age at and

status of menopause, BMI, and prior breast biopsies (Terry et al., 2019). Risk prediction is used to target patients who may need more frequent breast cancer screening intervals or may need to modify lifestyle risk factors such as weight control, exercise, and moderating alcohol intake (Howell et al., 2014).

Prostate cancers are primarily detected through a prostate specific antigen (PSA); however, risk prediction has been found to improve the accuracy of testing to detect prostate cancer (Louie et al., 2015). This is particularly important for prostate cancer patients, as the overdiagnosis of prostate cancers is common (Louie et al., 2015). The improvement in the accuracy of diagnosis can prevent harm to the patient caused by over-diagnosis.

Risk prediction models have also been used to identify patients with a likelihood of developing colon cancers through advanced proximal neoplasia, or polyps that may become cancerous (Cooper et al., 2018; Sung et al., 2018; Wong et al., 2016; Yeoh et al., 2011). This type of model can be useful in identifying which patients should go directly onto colonoscopy rather than to other methods of screening like fecal testing (FIT) or should be prioritized for colonoscopy following an abnormal FIT test. The Asia Pacific Colorectal Cancer Screening Score uses patient level characteristics like age, sex, smoking status, BMI, and family history to identify patients likely to have advanced neoplasia (Yeoh et al., 2011). Wong's model uses similar characteristics (age, gender, family history, BMI, and heart disease) to categorize patients at average and high risk of developing proximal neoplasia (Wong et al., 2016). Sekiguchi went further to identify

the cost effectiveness of this model compared to regular screening in finding neoplasia (Sekiguchi et al., 2018). They found that the risk prediction model was the most effective and cost effective of all other methods (Sekiguchi et al., 2018).

Daniel Blumenthal et. al, in 2015, created a risk model to predict a patient's non-adherence to colonoscopy (Blumenthal et al., 2015). This model found gender, mental health history, a non-adherence measure (missed appointments), number of prior missed appointments, education level, and wait times to the colonoscopy appointment predict the likelihood of adherence to completing the colonoscopy (Blumenthal et al., 2015). Notably, wait time introduced a clinic level characteristic, expanding beyond patient level characteristics (Blumenthal et al., 2015). To date, no risk prediction model has been created to predict the likelihood of overall screening for CRC to target interventions to increase screening rates.

Other Preventive Models

Other uses of PA in preventive medicine include identifying vulnerable patients in ambulatory surgery to adverse psychological vulnerability (Mijderwijk et al., 2016). Researchers in Canada have used PA to identify high health care users and to implement interventions to provide more appropriate forms of care (Rosella et al., 2017). The overprescribing practices of providers, concurrent use of counteractive pharmaceutical treatments, and potential overdosing have been monitored through PA (Vivas-Consuelo

et al., 2014). The opportunities for use of PA in population health management are endless.

Predictive Analytics in Colorectal Cancer Screening

The use of PA in preventive medicine could inform specific areas, such as colorectal cancer through increasing screening by targeting interventions. The Centers for Disease Control (CDC) and the American Cancer Society (ACS) continually update statistics regarding CRC prevalence and mortality. In 2019, they report that CRC was the 3rd most common cancer and 3rd leading cause of cancer related deaths in men and women in the United States (ACS, 2019; CDC, 2016). There are approximately 145,000 newly diagnosed cases of CRC each year, and more than 52,000 people died in 2019 from CRC (ACS, 2019).

The United States Preventive Services Task Force (USPSTF) recommends that all adults aged 50-75 years old should be screened regularly for CRC (US Preventive Services Task Force et al., 2016). Screening can prevent cancer by finding and removing pre-cancerous polyps. The CDC specifies that screening is considered complete if by a fecal immunochemical test (FIT) each year, a FIT DNA test every 3 years, by flex sigmoidoscopy, accompanied by a FIT every 5 years, a virtual colonoscopy every 5 years, or by a colonoscopy every 10 years (CDC, 2016). Screening intervals may vary by outcomes of prior screening or family history. In 2018, according to the CDC, only 68.8% of eligible people were screened for CRC in the US (CDC, 2018). Organizations like the

National Colorectal Cancer Roundtable (NCCRT) have identified disparities among populations. For example, in Federally Qualified Health Center's (FQHC's), which typically serve underprivileged people, the CRC screening rate in 2018 was 44.1% (National Colorectal Cancer Roundtable, 2018). Patients often served by FQHC's who are non-Hispanic black, American Indian or Alaska Native, Hispanic, rural, low income, or low education are less likely to complete screening (National Colorectal Cancer Roundtable, 2018). Failure to screen results in more advanced stages of cancer detection and decreased survival following diagnosis (ACS, 2019).

Patient reported barriers to completing screening in FQHC's were summarized by Muthukrishnan and colleagues and Stacy and colleagues. They found that these barriers may include lack of insurance, lack of regular care, lack of knowledge of the importance of or the need for screening, fear of results, or having competing health concerns (Muthukrishnan et al., 2019; Stacy et al., 2008). Anderson and Jetelina identified that completing colonoscopy can be particularly challenging because of the need to take time off from work, find someone to escort you to the appointment, and get necessary bowel preparation supplies (Anderson et al., 2011; Jetelina et al., 2019).

Provider and system-level barriers to completing CRC screening may include limited colonoscopy capacity, long wait times, failure to order a FIT test or refer the patient to the specialist, failure to schedule the procedure, failure to communicate expectations about the procedure or preparation for the procedure, and lack of adequate workflows to complete the referral (Weiss et al., 2013; Yu et al., 2018). These

barriers may inform interventions for patients at high risk of failing to complete screening.

Colorectal Cancer Screening Interventions

Interventions for increasing CRC screening can be applied at multiple levels. Patient, provider, organizational, community, or policy level interventions all may add to increased screening rates. Health systems need to determine the efficacy of interventions at the different levels in conjunction with the needs of the patient population to increase screening.

Patient level interventions

Common predictors of screening include both modifiable and nonmodifiable patient level attributes; modifiable include knowledge and awareness, risk perception, and attitudes. The most common and effective evidence-based interventions (EBI) at the patient level to increase CRC screening are centered around providing greater knowledge about the importance of screening (Gimeno Garcia et al., 2014). To do so, health systems may use intervention tactics like issuing information through reminders including phone, mail, or text reminders to the patient reminding them it is time to screen or return FIT tests. Djenaba Joseph and colleagues outlined other acceptable patient level EBI's which include patient facing videos, printed materials, brochures, and patient navigation (Joseph DA et al., 2016).

Provider level interventions

Joseph also outlined EBI's at the provider level, which include provider reminders, and provider assessment and feedback about their screening rates and unscreened patients, and patients needing follow-up testing (Joseph DA et al., 2016; Sarfaty & Wender, 2007). Other provider level EBI's include professional development and provider education to discuss the importance and ease of screening and the value of conversations with the patient (DeGroff et al., 2018). Direct patient and provider conversations have consistently been found to be effective yet finding time during the clinical visit can be challenging (Thompson et al., 2019).

Systems level interventions

Organizational level interventions to increase CRC screening may include identifying and reducing patient level barriers and burdens to screening (Gimeno Garcia et al., 2014; Joseph DA et al., 2016; Sarfaty & Wender, 2007). Barrier reduction may come through programs, such as through direct mailed FIT testing programs or implementation support on structured weekly, monthly, quarterly, semi-annual, or annual schedules (DeGroff et al., 2018). Clinical champions may also direct the CRC screening programs to allow for response to patient needs as learned through outreach programs (DeGroff et al., 2018). Patient navigators may be able to reduce barriers and provide patient education to increase knowledge and awareness (Gimeno Garcia et al.,

2014). Organizational level interventions offer a unique opportunity to impact screening outside of the clinical encounter.

Community level interventions

Community level interventions reach a wide net of people to raise awareness. These interventions could include small media or awareness campaigns (Joseph DA et al., 2016). These types of interventions are often accomplished through partnerships with organizations such as the American Cancer Society and may include awareness events that include group education, inflatable colons, parades, fundraising activities, and testimonials from locals impacted by CRC (ACS, 2019; Gimeno Garcia et al., 2014; Greenwald, 2006). Other community-specific efforts may include health fairs, local public service announcements on radio or tv, and messaging through community leaders (e.g., churches, barber/beauty shops) (Morrell, 2020).

Policy level interventions

Policy level interventions might include required reporting, the USPSTF recommendation, HEDIS, and other quality metrics that contribute to Medicare 5-start ratings. More local policy interventions include the implementation of incentivized metrics, like the CCO metrics implemented in Oregon or funding CRC implementation support as mentioned in the systems level interventions (DeGroff et al., 2018; Stock, 2017; Townsend et al., 2009). Essentially, policy that supports CRC screening programs

or outreach efforts provides the support needed for effective implementation (Townsend et al., 2009).

Targeting Interventions for CRC Screening

Interventions to increase screening should be targeted and customized to the patients that need it the most, and in ways that address modifiable risk factors (Gimeno Garcia et al., 2014). The application of interventions at different levels may increase effectiveness and address system level barriers to screening. Finally, targeted interventions can be cost effective to systems through eliminating blanket interventions delivered to patients likely to screen on their own, and reduction of disease by effectively reaching patients who were unlikely to screen.

Increasing CRC screening not only reduces the incidence and mortality of CRC, but is cost effective compared to cancer treatment, and lifesaving to patients (Lansdorp-Vogelaar et al., 2009). One study from CDC funded programs found that among FQHC's, adding interventions increased screening rates from 4.9-26.7 percentage points, with costs ranging from $18.76-$144.55 per patient (Subramanian et al., 2020). The ability to target interventions could prove to be more cost effective, but for preventive care, this is over the long run and recent health policy coverage of preventive care services by the ACA has been an important motivator in making the business case (Chait & Glied, 2018).

Understanding a health system's patient population, the specific predictors of screening for that population, and applying targeted interventions to address those

predictors will increase CRC screening. Predictors for patients unlikely to screen can be at multiple levels, and CRC screening multilevel interventions can also address multilevel risk predictors.

The Use of Predictive Analytics in CRC Screening

Predictive analytics in CRC screening can inform the use of complex interventions to increase CRC screening. There is no risk prediction model to predict the likelihood of overall screening for CRC. The use of predictive analytics to identify who is likely to screen gives a health system the ability to target interventions to increase screening rates.

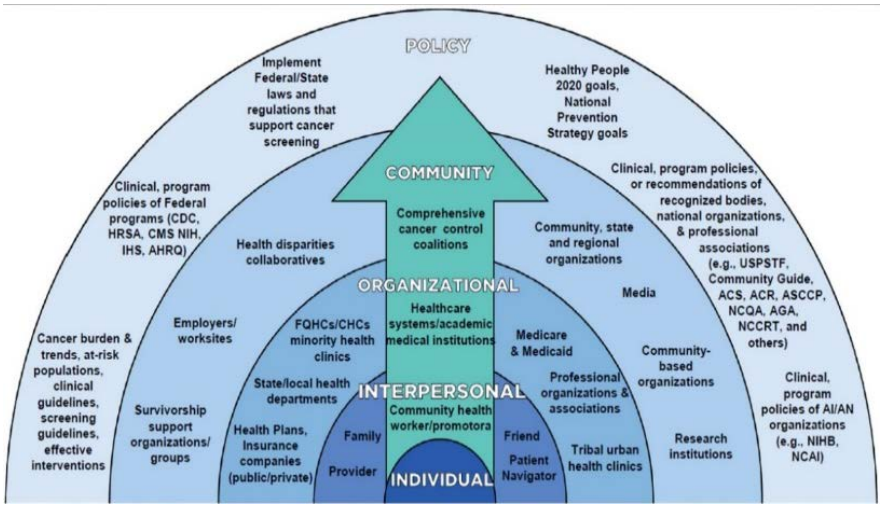
Predictive analytics have been used widely in preventive medicine. As many interventions have been found effective for CRC screening, PA could be used to find the right patients who need the right screening at the right time.

Theories and Frameworks

Methods to create risk prediction models will benefit from being based on a multilevel framework that considers different levels of influence on health such as the Social-ecological Model (SEM). The methods for conducting the qualitative interviews will benefit from being guided by relevant elements of the Consolidated Framework for

Implementation Research (CFIR) which seeks information from various settings within the health system (Laura J. Damschroder et al., 2009).

Social-ecological Model



The CDC's SEM Framework outlines that many levels influence health (CDC, 2017). Health behaviors have historically been found to be influenced by a variety of levels, but the SEM focuses on the personal, interpersonal, organizational, community, and policy levels (CDC, 2017). The innermost individual band of the SEM represents the individual and their characteristics which might affect their health behaviors. Regarding CRC screening, individual or personal preferences, cultural beliefs, and access to care may influence health outcomes. Interpersonal factors that may facilitate or minimize barriers to health seeking behavior may include provider recommending screening, or even marital status in that the spousal support may increase screening behavior. Organizational factors that influence health and screening include access to colonoscopy clinics, clinic size, and wait times. Community level factors may include colonoscopy capacity in gastroenterology clinics, or partnerships with cancer coalitions, population density, or unemployment

rates. Policy level factors affecting screening and health behaviors may include insurance or health plan coverage mandates for screening or participation in programs that require screening reporting (Uniform Data Systems (UDS)).

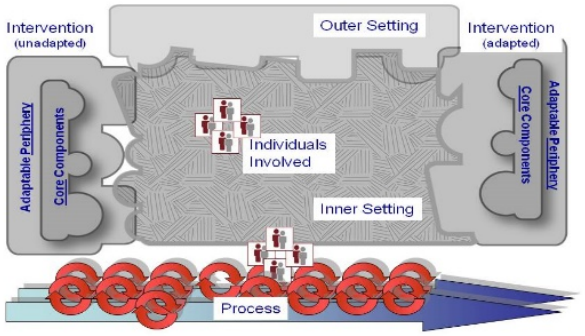
Zapka and colleagues, as mentioned above, recognized that multilevel factors that influence care are especially prevalent in the cancer care continuum (Zapka et al., 2012). They used the ecological model as it highlights the complex and interactive factors that affect the quality of care and outcomes at all levels (Zapka et al., 2012). When treating cancer patients, treatment should consider the individual, family, community, health provider, medical care system, community, state, and national and policy levels in attributing to care (Zapka et al., 2012). They used the two patient scenarios to exhibit this point, Mrs. Smith, and Ms. Adler were found to have very different issues with cancer care. Mrs. Smith lacked adequate insurance, and the health system failed to track her ongoing health issues (Zapka et al., 2012). Ms. Adler's breast cancer survivorship depended on her individual biology, genetics, stress, her provider's confusion about care continuity, a fragmented healthcare system, inadequate attention to psychological and social needs, the availability of community support groups, and the communication and coordination of the health system (Zapka et al., 2012). These two examples are not unusual for cancer patients. The interaction of the characteristics at different levels in care and health exhibits the need to use models like the SEM to frame exploration into predictors of screening.

Knowing what multilevel predictors affect screening could impact multilevel approaches to interventions. Multilevel interventions could mitigate factors that impact healthcare quality and facilitate addressing the causes of inadequate screening. Watson and colleges found that in lung cancer, multilevel approaches based on the SEM were able to address disparities in screening (Watson et al., 2019). Patient navigators were able to address individual, interpersonal, and some community level factors by applying motivational interviewing, addressing cultural norms, and bringing in cultural relevance to the importance of screening (Watson et al., 2019). Organizational level factors were addressed by creating systems improvements to the EHR, and community factors were addressed through facilitating partnerships (Watson et al., 2019).

The SEM is a guide for ensuring that predictors and interventions at all levels are considered. For this project, the SEM will guide the choice of multilevel predictors for use in the model. Having high quality, available data will make models more applicable and accurate. While PA has historically used administrative or EHR data, the SEM will guide the discovery of available multilevel data (Mahmoudi et al., 2020). At each level, potential predictors will be considered and determined if they are accessible through publicly available data or from within the health system data. Predictors at the SEM's multiple levels may include Census data tied to the individual's census block, provider level screening rates, clinic level screening rates, community data from the American Community Survey, or state level incentives for screening. The use of the SEM

framework in identifying relevant interventions will rely on the predictors in the model and the health systems' willingness to apply precision delivery of interventions.

Consolidated Framework for Implementation Research



The Consolidated Framework for Implementation Research (CFIR) will guide the qualitative interview protocol to learn about health system staff perceptions of the utility of the CRC screening model developed and the value of adding multilevel data. The CFIR components to be used include selected attributes of the intervention, inner setting, outer setting, individuals, and process (L. J. Damschroder et al., 2009).

Alison Cole and colleagues used the CFIR framework to guide qualitative interviews to understand facilitators and barriers to the implementation of a CRC screening program in an FQHC. They found that the CFIR constructs successfully drew out information regarding the implementation. Facilitators included previous quality initiative experience (intervention), engagement of leadership (inner setting), and champions for leading the intervention (process) (Cole et al., 2015). Barriers included limitations with communication with a diverse population (outer setting), and tensions for change (individuals) (Cole et al., 2015).

Safaeinili and colleagues adapted the CFIR framework for use in a learning health system to evaluate evidence based interventions (Safaeinili et al., 2020). Their goal was to expand CFIR to more accurately evaluate interventions to facilitate, design, and evaluate the use of evidence based interventions (Safaeinili et al., 2020). They expanded the outer setting to specifically address patient needs and resources, the inner setting to understand levels within that setting, and tailored constructs to specifically fit a primary care setting (Safaeinili et al., 2020). Here, the use of CFIR gave a general framework, but adaptations were made for the specifications needed for the learning health system.

The multilevel components of CFIR will guide the questions to understand how the implementation of the risk prediction tool that incorporates multilevel data might occur (intervention) and if the findings are clinically meaningful from the health system perspective. The interviews will assess the characteristics of the health system that allow for adoption, such as priorities and goals (in the inner setting), the motivation to use multilevel data from the outer setting perspective, and the individual's perspective on the value of PA and if they find benefit in using multilevel PA including multilevel data.

The proven frameworks of the SEM and CFIR guide this project for a systematic approach and associated methods. The SEM framework will guide the discovery of multilevel predictors. The CFIR framework will guide the interview questions with a purposive sample of health system staff to understand the components of the organization and the likelihood of implementing a multilevel risk prediction model.

Although adaptions may be necessary, the solid foundations will provide structure and direction.

Barriers and Limitations of Multilevel Predictive Analytics

While creating a multilevel risk prediction model could be useful to a health system, there are barriers and limitations that must be recognized. One prominent limitation is access to data both inside and outside of electronic databases. It has already been established that clinical data alone is not always sufficient in predicting health outcomes. Many EHRs exclude data outside of the information captured in discrete fields in a clinical encounter. Other characteristics of individuals that may play an important role in a person's health may simply be unavailable. Even when the use of data from an EHR and administrative databases is possible, data quality can restrict the use of this data in PA. (Parikh et al., 2019) Moreover, because EHRs are designed to capture transactional patient-level data, capture is often incomplete and subject to the care delivered. Therefore, EHRs will have missing data on those not accessing care that may be important (e.g., those who need but do not get care). Lack of access to long term data and robust populations in a clinical system will also limit the ability to adequately use PA.

For the proposed project, data from sources outside of the EHR will be used to help better predict health outcomes. These data may include social determinants of health, community level data, family characteristics, or social deprivation scales.

However, some data may be inaccessible. Social determinants of health are not routinely collected in a clinical encounter and are not available outside of the EHR (Hatef et al., 2019). Community level data may be limited to what is publicly available (Diez Roux, 2008). Family history data is ideal in understanding screening patterns yet is rarely collected in a meaningful way (Ginsburg et al., 2019). For this project, all levels of data for the model will be accessed, and limitations and barriers regarding data availability and quality will be documented.

For PA to be useful, clinicians must be equipped with robust tools and actionable outcomes that can be responsive to the model findings. The problem must be clear, and interventions should be specified prior to the deployment of prediction (Parikh et al., 2019; Steyerberg & Vergouwe, 2014). Not every decision needs a predictive risk model, therefore, assessing the outcomes and actions that might be taken after a predictive model is created is necessary before the execution or deployment of a model.

Those that use predictive analytics must also understand the risk of perpetuating bias. The purpose of prediction is to improve health, yet if groups of people fail to have adequate health information to contribute to prediction, treatment for that group may be averted only to groups with adequate information. For this reason, deeply understanding data missingness and data characteristics with groups where data is missing is imperative.

When conducting predictive analytics, a system must identify infrastructure limitations. (Parikh et al., 2019) In addition to data limitations, analytic capacity and

expertise are necessary. Advanced analytics can be limited to relatively large systems or systems with robust data structures. Analytic expertise can address model limitations. These limitations may include overfitting, calibration, or discrimination. (Shah et al., 2018; Steyerberg & Vergouwe, 2014) Lack of calibration would indicate a lack of agreement between the observed and predicted risk, and discrimination assesses the ability to distinguish between patients with and without the endpoint (Steyerberg & Vergouwe, 2014). A bootstrap corrected C-statistic should be used to determine concordance and calibration will be assessed through plots (Steyerberg & Harrell, 2016). Once a model is sufficient, it must continually be audited for applicability to changing populations within a clinic or system. (Parikh et al., 2019) The methods being developed in this area of PAs have been rapidly advancing our ability to address these problems.

Tools have been developed specifically to ensure that those who are conducting predictive analytics are following standards and guidelines, minimizing bias, and increasing applicability. (Moons et al., 2015; Moons et al., 2019; Wolff et al., 2019) PROBAST (Prediction Model Risk of Bias Assessment Tool) is the most recent and validated tool that signals questions regarding participants, predictors, outcomes, and analysis. (Moons et al., 2019) Using tools and checklists like PROBAST can improve the quality and reduce the bias of predictive analytical tools.

Predictive models are sometimes created for specific systems but could be more useful if interoperable and generalizable. (Parikh et al., 2019) Increasing data opportunities and the availability of data outside of the EHR may allow for more

generalizable models. Models will then be able to be translated from the statistical output and put into practice. Regression coefficients can be translated into a risk score or points-based system where a higher number of points mean a higher likelihood of screening. (Harrell, 2015) This will allow clinicians to translate a model to be put into practice without calculating the regression equation exactly. The code will assign one point to the smallest increment. The clinician could add up the points to determine the likelihood of screening.

Conclusion

This project seeks to understand if the inclusion of multilevel data can improve the accuracy or generalizability of a prognostic risk prediction model, in predicting patients' probability to screen for CRC. This project also seeks to understand the health systems perspective on how the inclusion of multilevel data in a risk prediction model improves the usefulness for managing population health. Specifically, the research questions are:

1) *Can the inclusion of multilevel data improve the accuracy or applicability of a prognostic risk prediction model, in predicting patients' risk of failure to screen for colorectal cancer in order to target interventions to the right patients at the right time?*

2) *How does the inclusion of multilevel data in a risk prediction model improve the usefulness to health system decision makers for managing population health?*

To answer these research questions, data availability, quality, heterogeneity, and opportunities to broaden data availability at a health system based on the SEM framework will be assessed. Data will be categorized into levels according to the framework and ascertained for availability. A risk prediction model will be developed using EHR and externally available, multilevel data. This will be done first in a large dataset of patients due for CRC screening using standard EHR data, then statistical improvements in models overall will be assessed when adding multilevel data, heterogeneity and transportability will be assessed Finally, this project seeks to understand health system decision makers' perceptions of the usefulness of the CRC screening model developed and the value of adding multilevel data, using qualitative research methods.

The literature suggests that PA can be used in systems to improve efficiency and increase the use of targeted interventions for appropriate preventive care. Although there are data limitations, expanding the use of multilevel data may improve the accuracy of models. Health systems will benefit from the use of predictive analytics in that it may reduce unnecessary costs and direct resources where they are most needed.

PA is already being used in a variety of ways to increase preventive screening. The use of PA in cancer prevention shows the impact and value of cost savings and

health promotion. CRC screening is important and can prevent cancer. Interventions targeting multiple levels from reminders to incentivized metrics are available to promote CRC screening. PA can reduce costs and increase the rates of CRC screening by identifying the populations that need interventions and by pinpointing which interventions would be most appropriate for the targeted population.

The SEM framework is appropriate to guide the creation of the risk prediction model by outlining levels of data that should be considered and reviewed for usability. The CFIR framework is appropriate to frame the semi-structured interviews to address all organizational levels that may impact the adoption of PA. These frameworks will direct the process of collecting and interpreting both qualitative and quantitative data.

Barriers to data acquisition and limitations of modeling need to be attempting to use PA. Limitations may include data availability and usability. Limitations to the methods can sometimes be overcome through analytic methods and following standards and guidelines. Broadening the data, increasing the usefulness of the data, and using multilevel data in risk prediction will improve the accuracy of risk prediction models in identifying which patients are at high risk of failing to screen for CRC. Interventions can then be targeted to this population. The decision maker's response to the inclusion of multilevel data will guide policy recommendations to increase the usefulness of PA.

Chapter 3: Methodology

Chapter Overview

This chapter provides an overview of the methodology to understand if the inclusion of multilevel data will improve a prognostic risk prediction model in predicting patients' likelihood of screening for colorectal cancer and if health system decision-makers find value in using a multilevel model in managing population health. The added value of multilevel data in improving prognostic risk prediction will be quantitatively determined by improved discrimination, calibration, net reclassification improvement, and applicability to subpopulations. The added value of the model in managing population health will be determined qualitatively through interviews with relevant health system decision makers to understand if the model is useful in practice. The mixed methodology is designed to test the value of adding multilevel data both in terms of improved performance in predicting the risk for failure to screen for colorectal cancer and its usefulness in guiding decision maker's efforts to improve screening rates. Specifically, this chapter will:

1) Review the research questions and aims of the overall project including a figure outlining the conceptual design of the project.
2) Provide an overview of the research design, including discussing the use of available data, the social ecological model (SEM) framework, the creation of the models, and conducting the qualitative interviews.
3) Discuss the justification of the research design.

4) Describe the methods to accomplish the aims:

a. First, the methods to accomplish Aim 1 are described regarding the assessment of multilevel data sources and variables through the framework of the SEM including a description of concepts, variables, and sample measures for the models and reviewing data collection and participant selection criteria.

b. The methods to accomplish Aim 2 are described regarding the implementation of the multilevel PA model including the setting, model development, and the analytic methods applied.

c. Finally, the methods to accomplish Aim 3 are described regarding assessing health systems perceptions including the use of the Consolidated Framework for Implementation Research (CFIR), interview sampling, qualitative questions, data collection, and analysis of interviews.

5) Describe limitations and purpose and significance.

Research Questions and Aims

Research Questions

1) Can the inclusion of multilevel data improve the accuracy or applicability of a prognostic risk prediction model, in predicting patients' risk of failure to screen for colorectal cancer in order to target interventions to the right patients at the right time.

2) How does the inclusion of multilevel data in a risk prediction model improve the usefulness to health system decision makers for managing population health?

I will accomplish this in the 3 aims described below. The first 2 aims will address question 1, the third aim will address question 2.

Aim 1: Assess data sources and variables

The first aim is to assess the availability, quality, and opportunities to broaden data use external to electronically available data in an integrated health system for determining patients' probability of screening for colorectal cancer. Data will be assessed and acquired from a variety of sources. Data sought will be across multiple levels (SEM), publicly available, and will be retained at the lowest level when multiple levels are available (i.e., census tract vs. zip code). Data must have variance across the sample, be relevant to CRC screening, and not be colinear with other available data. The data will then be categorized into levels according to the SEM framework and availability and the use of the data will be determined. Datasets will be assembled with all available data at the SEM levels for use in Aim 2.

Aim 2: Develop risk prediction models

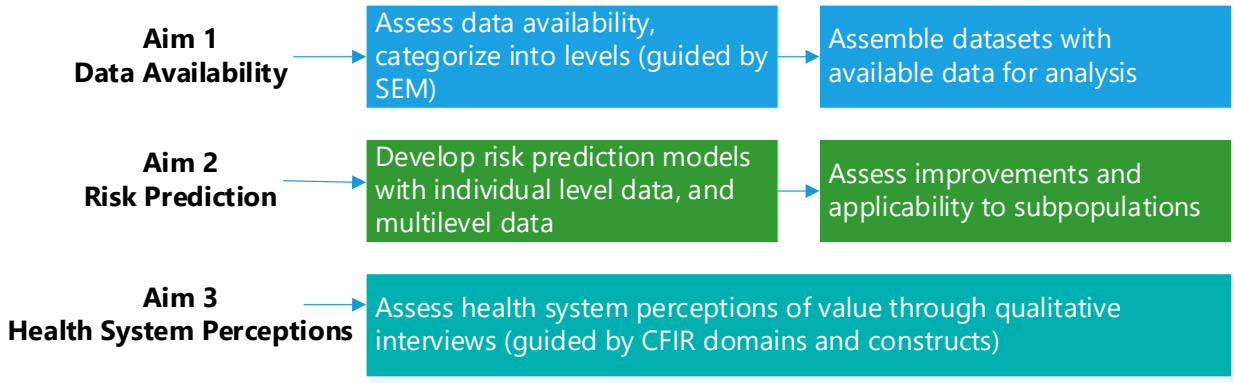
The second aim is to develop risk prediction models using individual level data from the EHR and available multilevel data as determined in Aim 1. The first model will be developed in a large dataset of patients who are due for CRC screening using individual level data. The statistical improvements in the model will be assessed when multilevel data is added. The Prediction Risk of Bias Assessment Tool (PROBAST) will be used to assess the risk of bias (ROB) (Wolff et al., 2019). This is done by assessing participants, predictors, outcomes, and analysis bias (Wolff et al., 2019). Finally,

subpopulation applicability based on the PROBAST assessment will be assessed by applying the model to subpopulations, based on population characteristics such as race, ethnicity, or insurance type.

Aim 3: Assess health system perceptions

Finally, the usefulness of the developed CRC screening model with added multilevel data will be assessed using qualitative research methods. Semi-structured interviews will be framed around the domains and constructs of the Consolidated Framework for Implementation Research (CFIR).

Figure 3.1 Conceptual Design



Research Design Overview

This is a mixed-methods project, first assessing data availability and external data sources, then developing risk prediction models, and finally assessing health systems perceptions related to the adoption of the innovated multilevel model informed by qualitative interviews.

In the existing literature, only one prior risk model has identified patients who are unlikely to show for a colonoscopy appointment, but no models to date predict CRC

screening overall (Blumenthal et al., 2015). The Blumenthal et al. colonoscopy model performed well, meaning the prediction is better than chance, (Area Under the Receiver Operating Curve (AUC) = 70.2%), but only took into account screening via colonoscopy (only one method of screening) and did not include data from multiple levels (Blumenthal et al., 2015).

Data exploration will be guided by the SEM and include data within and external to the EHR across the multiple bands of influence. For this project, available data will be identified from outside of the EHR and administrative databases and may include data from the U.S. Census Bureau, American Community Survey, Department of Health, and locally available data. Guided by literature on predictors of screening, potential measures will be grouped by level of influence in the SEM framework. All possible available multilevel data will be considered, and the final predictors will be determined in the model. Data that is sought after, but unavailable across all bands of influence, will be documented for reference for future models wishing to use multiple levels of data.

Risk prediction models, as described in more detail below, will then be created to predict the likelihood of screening for CRC Models will be developed using individual level data only and then with added external multilevel data guided by the SEM framework. Model performance improvement will be assessed by an improvement in internal validity, net reclassification improvement, and graphical calibration when available SEM data is added to the model. Additionally, the full sample initial and

multilevel models will test how applicable it is to the subpopulation when separately applied to subgroups.

The models will utilize a large retrospective sample of patients at Kaiser Permanente Northwest (KPNW). Subjects will be eligible for inclusion in the model if they are recommended for screening at KPNW, 45-75 years old (45+ for black patients, and 50+ for non-black patients) and were due for screening in January of 2019. The outcome of interest will be the completion of screening within a year, based on HEDIS criteria, including fecal testing, FIT DNA, colonoscopy, or CT colonography (CDC, 2016). The first risk model will use available EHR and administrative data, the second will incorporate data from outside sources including data from multiple levels. The final step will assess the applicability and model performance of either model one or model two (whichever produces greater performance characteristics) when applied to subpopulations.

After the risk models are complete, qualitative interviews will be conducted with clinic leaders, administrators, and decision makers within KPNW, based on the domains and constructs of the Consolidated Framework for Implementation Research (CFIR) framework. The interview guide will be designed to solicit reactions to how results from the model improve their ability to target at risk populations and increase screening rates. These structured, in-depth interviews will involve decision makers at KPNW, such as the Gastroenterology (GI) medical director, quality improvement manager, primary care project manager for CRC screening, or the director of quality improvement.

Questions will be guided by the CFIR framework to understand the influences of the risk prediction model, the internal context to KPNW, the external context to KPNW decisions and the process context of making changes to use PA for CRC screening, and the feasibility and perceived value of the use of the model to improve patient health.

Questions framed by the domains and constructs of CFIR will specifically seek input regarding the clinical perspective of the added value of a multilevel model in terms of improving population health. The intervention level within CFIR is the multilevel risk model, therefore qualitative questions will be framed to understand the current use of risk prediction. The inner setting of CFIR's framework guides questions focused on internal priorities and goals, barriers, and facilitators to the use of multilevel data, and sustainability of acquiring multilevel data outside of the standard (EHR) data sources. Outer setting level questions will seek to gain insight on how the multilevel model may be useful, concerning providing an advantage over outer setting influencers, such as other health plans or motivation for improving processes to save resources. Process level questions will ask about infrastructure or capacity issues and how the model may change processes to be more efficient or effective. The interviews will additionally seek insight into what external factors or internal policies could encourage the uptake on the use of PA tools and expanding data use across the SEM levels of data across multiple levels.

Justification of Research Approach

The research design includes a data exploration step, the creation of risk prediction models, and qualitative interviews. This design will provide a comprehensive understanding of the usefulness of risk prediction in an integrated health system.

The first aim, to assess data sources and variables, is integral to understanding the availability of data, the completeness and quality of the data, and the limitations to incorporating multilevel data into a risk prediction model. Data access, quality, and applicability may impede the use of some data elements. Thoroughly investigating what data is available and if the quality and completeness of the data are sufficient will inform other research as they seek external multilevel data.

The second aim, to develop the risk prediction models, uses the findings in Aim 1, and incorporates the external multilevel data into risk prediction models. This allows the understanding of the contribution of the addition of external multilevel data to the model created on EHR and administrative data. The second aim also allows for the understanding of the applicability and translation of the model to a subpopulation.

The third aim, to assess health system perceptions, allows for a sequential qualitative exploration to understanding system level limitations and perceived benefits of incorporating an improved predictive analytic tool for CRC screening in an integrated health system. Here I hope to explore what is found in the first two aims to understand the potential use of a multilevel risk prediction model.

Methods for Aim 1: Assess Data Sources and Variables

Conceptual Frameworks

The Social-ecological Model (SEM) will guide the data exploration to identifying components for the risk prediction model. This framework outlines the levels of influence in health (CDC, 2017). The use of the SEM to drive a multileveled approach to PA will add depth and dimension to the PA model. The innermost band, the individual level, contains the characteristics which might affect a person's willingness and ability to screen including insurance coverage, transportation, or taking time off from work. Individual factors to be included are the length of membership at KPNW, prior screening, prior test preference, individual level data from the EHR, and administrative data such as clinical data, medical history or demographics, and individual social determinants of health (SDOH) from the medical record. Interpersonal factors may include marital status, racial concordance (shared identity between a physician and patient), the provider's screening rate, or other provider characteristics, or neighborhood level factors like census level data from the individual's census block. Organizational factors will include clinic characteristics, screening rates, or distance to facilities that provide colonoscopies. Community level factors may include community level screening rates, health literacy, income inequality, unemployment rates, or other community characteristics from the American Community Survey. For this project, I define community at the level of census block group, zip, or county with a preference for the most specific level of data available. Policy level factors affecting screening may

include insurance or health plan (CCO) mandates for screening or incentives for CRC screening. Individual data may be limited to data available in the EHR or administrative data. Available interpersonal and organizational level data may come from administrative databases that contain provider level data. State and community policy data may be available from CCO's, but some data may be difficult to acquire, and few patients at Kaiser are CCO members. Regardless, assessing what data is available at all levels is essential for determining access to multilevel data. Allowing flexibility and adaptability based on data availability may be necessary.

Data Sources

This project will assess the availability of external data at the different SEM levels and the practicality of using the data in risk prediction. This data will be from publicly available databases and include the U.S. Census Bureau, American Community Survey, Department of Health, and locally available data from the Oregon Health Authority. Non-EHR and administrative data will be acquired separately and mapped to the patient level data by geographic level such as census block, zip code, county, or state. Data will specifically be sought that provides the most potential variation across the sample.

While concordance with providing data at all levels is desirable, the availability and access to data will be determined and recorded. First, data from the EHR and administrative data will be categorized across the SEM levels. Additional data from sources outside of the EHR and administrative data will be sought based on known predictors from prior literature. Relevant data will be sought from all known available

sources, each data element will have a credible evidence-based justification for inclusion. Data will be assessed for missingness and the ability to link to patients in the dataset. When multiple geographic levels are available, the lowest will be retained (i.e., census block vs. county). This project will strive to acquire as much data as possible from the multiple levels, and collinearity will be assessed, use in the model will be determined by the predictive model.

Gaps in data across the levels will prompt a search for data from that level from variable sources.

Concepts and Variables

Using data from KPNW, available data will be identified from the electronic databases (virtual data warehouse) and external sources across the multiple levels of influence framed by the SEM. Variables identified in the literature as predictors of screening will be sought for inclusion, such as age, insurance type, number of prior visits, health literacy, or prior preventive screenings (Petrik et al., 2018). Potential measures will be grouped by level of influence in the SEM framework.

Variables will be selected based on quality, completeness, availability, and the aspiration to fulfill diverse data from multiple levels. Data will be used if it is publicly available or available upon request. Additional variables may be sought based on availability and information gained in the process of variable identification.

Predictor characteristics will be measured at the time closest to but before their birthday in 2018 unless otherwise specified to assess baseline characteristics prior to eligibility for screening. Some variables to be included are shown in Table 1.

Table 3.1. Sample Measures for Multivariate Risk Prediction Model

Level	Variable	Source
Individual	Perceptions of screening (test preference)	Clinical Visit (NLP*)
	EHR clinical data (demographics, vital data, diagnoses, comorbidity scores, encounters, prescriptions, procedures, labs)	EHR
	Social Determinants of Health	EHR
	Length of membership	EHR
	GINI Income Inequality Index	American Community Survey
Inter-personal	Family data (family history of disease, family size, marital status)	EHR (NLP)
	Number of address changes in the past 2 years	EHR
	Provider panel screening rates	EHR
	Other provider characteristics (length of time in practice, panel size, panel characteristics)	EHR
Organizational	Clinic characteristics (size, location, funding, diversity of patients, provider FTE, availability of translation services?	Clinical data
	Clinic screening rates	Clinical data
	Colonoscopy characteristics (wait times for colonoscopy, distance to GI clinic, hospital/clinic partners)	Clinical data
Community/Neighborhood	Census data (neighborhood characteristics)	Census
	Community-level screening rates	Department of Health
	Public awareness (media campaigns)	Clinical knowledge/data
	GINI Income Inequality	American Community Survey
	Unemployment	American Community Survey
	Population density	American Community Survey
	Urban/Rural location	Urban/Rural Location
	Food access	Food Access
	Census data (median household income, unemployment rate, household poverty)	Census
	Area Deprivation Index	Neighborhood Atlas
	Health Literacy	Health Literacy Map
Policy	Access to insurance (Medicaid Expansion, % uninsured)	State data
	Incentives for screening	State data, Department of Health

	CCO/ACO Coverage (care coordination by CCO/ACO)	State data, Department of Health, CMS for ACOs

*Natural Language Processing (NLP) may be used to identify information obtained in the clinical visit but not in a discrete field

Data Collection

A request for a database to be created will be placed at the Center for Health Research at KPNW. Patients first will be selected based on the inclusion and exclusion criteria below. Retrospective data will be requested for as far back as 10 years to determine past colonoscopy screening, and for at least a year to determine CRC screening. Patient level data will then be pulled for use in the individual level model. Then multilevel data will be acquired for subsequent analysis with patient data.

Methods for Aim 2: Develop Risk Prediction Models

Selection of Participants

Subjects will be included if they are members of Kaiser Permanente for at least a year prior to their birthday in 2018 and if they are recommended for screening at KPNW. Black patients have a recommended earlier age for screening due to higher overall incidence rates and younger mean age at diagnosis for colorectal cancer (Rex et al., 2017). Specifically, patients will be eligible if they are 45-75 years old if black or 50-75 years old if non-black and are due for screening on their birthday in 2018. Patients are excluded if current for CRC screening, defined as fecal testing in the past 12 months, FIT DNA in the past 3 years, flexible sigmoidoscopy or virtual colonoscopy in the past 5 years, or colonoscopy in the past 10 years (Centers for Disease Control and Prevention,

2020b). Patients will also be excluded if they had co-morbid conditions that would make screening inappropriate, such as a history of CRC, colectomy, or who are receiving end-of-life care. (Centers for Disease Control and Prevention, 2020a)

The outcome of interest will be CRC screening completion in 2019, based on HEDIS criteria, including fecal testing, FIT DNA, or colonoscopy in the year (2019). (Centers for Disease Control and Prevention, 2020a)

Characteristics of the population of patients due for screening will be assessed for variability in demographics, such as race, ethnicity, or insurance type. The Prediction Risk of Bias Assessment Tool (PROBAST) will be used to assess the risk of bias (ROB) (Wolff et al., 2019). Participants, predictors, outcomes, and analysis bias will be assessed through approximately 20 questions to determine population characteristics disproportionate to the overall population, predictors with missingness in populations, and outcomes ratios based on the original population. Then observed screening rates for different groups identified in PROBAST will be evaluated for differences (in % screened) from the overall eligible population. This assessment and model characteristics may guide further analysis into subpopulation applicability.

Setting

This retrospective analysis will use the inception cohort from the population defined in Aim 1. First, this analysis will use available data at Kaiser Permanente Northwest, and then external sources identified in Aim 1.

Risk Prediction Data Analysis

Characteristics predicting the likelihood of screening will be evaluated using a multivariate logistic regression model in SAS® System Software. For the first model, only available individual level data from the EHR and administrative datasets will be used. Subsequent models will first incorporate the external multilevel data, and then test the applicability of the final model will be determined as the model is applied to a subpopulation.

For each model, a full model of patients with complete data for that model will be fit. Then, guided by Harrell's methods, a step-down method will be used to manually remove the weakest characteristics one covariate at a time to make the model more simple so that the final model will retain at least 95% of the variation explained in the full model. (Harrell Jr. et al., 1998; Heinze et al., 2018) This retention of the strongest characteristics will simplify the model so that the final model will retain at least 95% of the variation explained by the full model. (Harrell, 2015; Harrell Jr. et al., 1998)

For the final model, the observed and predicted risk of screening will be calculated and plotted in quintiles using risk predictiveness curves that show the distribution of observed and predicted risks of completing screening. (Lumley et al., 2002; Pepe et al., 2008) These graphs will show the mean predicted risk of obtaining screening plotted against the observed risk to understand the agreement between estimated and observed risk by quintile. Discrimination, or the ability to determine who will screen versus not, will be measured by a bootstrap corrected C-statistic. (Moons et al., 2019)

The explained variation will be measured with an R^2 statistic. The calibration will be measured by the Integrated Calibration Index (ICI), which assesses the difference between the model's calibration and perfect calibration. (Austin & Steyerberg, 2019)

Guidelines set forth by PROBAST will be followed, as discussed in Chapter 2. (Moons et al., 2019; Wolff et al., 2019) Specifically, the sources of data will be assessed for the risk of bias and the applicability to the question. Then, final predictors will be assessed for missingness for all eligible patients in the model, and screening completion determined prior to analysis. Data missingness may trigger imputation if appropriate depending on the variable in question. The variables excluded from the model, due to data missingness, will be evaluated. People eligible for screening who are excluded from the model due to membership or comorbidity exclusions will be examined and the missingness of specific predictors assessed.

The model will then be applied to a subpopulation. The performance of the model by age, race, ethnicity, insurance type will be reviewed. The model will be applied to the subpopulation, such as Medicaid patients, and model performance characteristics will be assessed. The poorer performance of the model will be determined by the model based C-statistic (*c-mbc).* Finally, Pepe plots of the overall population and subpopulation will be created for the assessment of the calibration of observed and predicted risk before and after adding the multilevel data (Pepe et al., 2008).

Methods for Aim 3: Assess Health Systems Perceptions

Aim 3 seeks to assess the health systems' perceptions of how the CRC screening model can be used in practice, using qualitative research methods. The functionality of the models including multilevel data will be assessed from the health system perspective.

Conceptual Framework

The domains and constructs of the Consolidated Framework for Implementation Research (CFIR) framework will be used to frame the qualitative interviews. (Laura J. Damschroder et al., 2009) This framework guides the collection of information about the various settings within the health plan, the context of making changes to CRC screening practices within the health system, and the feasibility and value of integrating multilevel data obtained outside of the EHR. The settings include the inner setting, outer settings, individual, process, and intervention. The questions will seek information from all 5 settings and a variety of topics. The interviews will focus on how the risk prediction models are used in care delivery and population health management, and what multilevel data add to the models.

Questions framed by CFIR will gather the perspective of clinicians and administrators of the usefulness of the model and the addition of multilevel data. For the intervention level, the multilevel risk prediction model, questions will be framed to understand the usefulness of the model in practice from the interviewee’s perspective.

The inner setting questions will focus on internal priorities and goals, barriers, and facilitators to using risk prediction tools in practice, and what the multilevel data adds to the use of the model. Outer setting questions will ask how the multilevel model may be useful, providing an advantage over competitors or motivation for improved resource allocation or improvement in performance measurement. Process questions will be framed to seek information on infrastructure and capacity issues that add to the usefulness of tools that offer an improved prediction. The interviews will additionally seek input on policy recommendations for using PA and expanding data use across the SEM levels of multilevel data.

Qualitative Interviews

Qualitative methods are effective in understanding complex relationships with innovations such as the use of PA in care delivery. Interviews can gain contextual information that is unavailable from other sources, such as perspectives and beliefs in a health system. Semi-structured interviews will be designed to gather the participants' perspectives and expertise in population health management, CRC screening, and the use of risk prediction.

For this project, interviews will be structured as in-depth interviews of decision makers at KPNW, such as the GI medical director, quality improvement manager, primary care project manager for CRC screening, or the director of quality improvement.

Design

Participants for the interviews will be selected if they are frontline caregivers or work in quality improvement at KPNW which includes Northwest Permanente and Kaiser Foundation Health Plan and Hospitals (Table 3.5.). Recommendations from Dr. Raj Mummadi, the gastroenterologist, and the Chief Quality and Population Health Officer at Northwest Permanente will be sought to inform key informant identification. Participants will be diverse in their roles and departments within KPNW and Northwest Permanente. Approximately eight key informants will be recruited through internal emails or phone messages. Participants will be employed by Kaiser Foundation Health Plan Hospitals.

Table 3.2 Participant Selection

Setting	Northwest Permanente	Kaiser Foundation Health Plan/ Hospitals
Front line caregiver	2	2
Quality Improvement	2	2

A structured guide will be developed and refined after an informal practice with a member of the committee. Approximately 12 key questions will be developed based on the performance and content of the final risk prediction model. A series of prompts for each question will also be outlined.

All interviews will be conducted by telephone. Interviews are expected to last 30-45 minutes and will be recorded. Participants will provide assent before the interview.

Data Collection

The interview guide will be comprised of approximately 12 questions exploring the CFIR domains relevant to this project. The guide will elicit questions about the participants' role in the organization and the process of making decisions in their department or work area.

For the inner setting, questions will ask about the types of data that are typically available for their population health management efforts, their knowledge, and view on the use of external and multilevel data, attitudes, and beliefs about the use of PA, and prior use of PA. Further information from the inner setting will be sought on an existing partnership with clinical care teams that may improve the implementation of new programs or strategies, and the potential sustainability of programs that uses PA and multilevel data. For the outer setting, the information will be elicited regarding how other health plans or organizations are using PA and multilevel data, and if this may provide a competitive advantage. Finally, for the process evaluation, the information will be sought on this person's perception of the organizational capacity for analytics and the use of multilevel data, and barriers and facilitators for the use of multilevel data and PA in the health system. All interviews will be audio-recorded and transcribed for content analysis.

Table 3.3 Sample Qualitative Questions

Setting	Question content	Topic
Person	Role in the organization?	Decision making ability

	Process of making decisions?	Decision making ability
Inner Setting	What types of data or evidence are typically used in your area?	Technological assets
	Will the use of multilevel data improve what you know about patients unlikely to screen? (Value of adding multilevel data)	Individual beliefs
	Attitudes and beliefs about predictive analytics?	Individual beliefs
	What mechanisms are used to evaluate the PA results for use in practice?	Capacity
	Prior partnerships with clinical care teams for intervention implementation?	Organizational structure
	Costs and sustainability of a program (and interventions) that uses predictive analytics and multilevel data?	Capacity, relative advantage, compatibility
	Attitudes and beliefs about the application of predictive analytics to subpopulations?	Individual beliefs
Outer Setting	Are your competitors or other organizations using predictive analytics? Multilevel data?	Motivation
	Would the use of predictive analytics and multilevel data provide a competitive advantage?	Motivation
	How has Covid impacted how predictive analytics may be used at this organization? Application to subpopulations?	Capacity
Process Evaluation	Capacity for analytics infrastructure, and analytic capacity, use of multilevel data?	Technological assets
	Human capacity within this organization to increase CRC screening?	Human assets
	Potential barriers and facilitators to utilizing multilevel data and predictive analytics?	Capacity

Analysis

All transcribed interviews will be recorded and analyzed using simple content analysis. Themes and summaries will be created for each question following the interviews, based on CFIR domains and constructs. A theme will be identified if mentioned by at least two individuals.

Limitations

There are barriers and limitations to this project. The main limitations are data availability, completeness, collinearity of multilevel data, and a limited context for qualitative interviews and data. EHR and administrative data are limited to the data available in analytic databases. External data is limited to data that is publicly available or available upon request. Some data may be inaccessible or not collected in the electronically available databases consistently such as social determinants of health, some community level data, and family history data. Multilevel data that is determined to be usable for this study may be colinear and very little predictive value to data is available in the EHR or administrative databases. Further, ideally, a model would be created using data from multiple systems with diverse geographical and patient populations, for a greater understanding of applicability to subpopulations and generalizability. Conducting this project only at KPNW limits the generalization of the mixed-methods findings to other settings.

Another limitation is that the qualitative interviews are from a single integrated delivery system. With unlimited time and resources, a wider range of participants would be selected from several health systems to be able to acquire the knowledge and opinions of decision makers from a variety of settings for greater generalizability.

Purpose and Significance

PA has the potential to increase patient safety and save resources. (R. B. Parikh et al., 2016) Multivariate risk prediction can improve population health when

standardized and used appropriately. (Cohen et al., 2014) PA is not always able to be used in population health management because of capacity or technology limitations. (Bates, 2014; Leininger, 2017) However, some health systems are expanding their use of PA to conduct better population health management. (Leininger, 2017) In the current setting, the COVID pandemic created a decrease in screening and has created a backlog of colonoscopies, and cancer screening guidelines are being expanded to include age 45-75. (Mazidimoradi et al., 2021; Tinmouth et al., 2021) This will create an increased need to know who will be able to complete screening independent of interventions. The application of this model could solve this contemporary problem of needing to know who is best served by targeted intervention. This study seeks to build more applicable models using multilevel data and to learn about data available from public sources across the SEM framework. Expanding data sources and the use of data across levels will be valuable in creating more accurate and applicable models. Finding what data is available to more systems will allow the use of multilevel data more common and make this model easily replicable. This study will assess the availability of multilevel data and the difference it makes in predictive model performance.

Health systems are more likely to use PA if they see a positive impact on resource use and patient health (R. B. Parikh et al., 2016); therefore, it is important to identify how PA can be integrated with clinical care in meaningful ways (R. B. Parikh et al., 2016). The adoption of multilevel risk prediction models is influenced by the perceived usefulness and reactions of the model by decision makers. Health system

leaders may adopt the use of PA as a tool to increase CRC screening rates, but it is only useful if health systems find it valuable and informative or useful in enhancing screening or performance rankings, and if the addition of multilevel data adds value. This study will understand the response to the use of multilevel data from a health system perspective and how models may inform practice and outreach. It will further ascertain implications for policy and practice and future research.

The mixed methods research design of this project was chosen to be able to answer the research questions:

1) Can the inclusion of multilevel data improve the accuracy or applicability of a prognostic risk prediction model, in predicting patients' risk of failure to screen for colorectal cancer in order to target interventions to the right patients at the right time?

2) How does the inclusion of multilevel data in a risk prediction model improve the usefulness to health system decision makers for managing population health?

The first aim seeks to assess data sources and variables through a comprehensive exploration of data within the electronic databases and at then at multiple levels in external data sources, guided by the SEM. The second aim uses quantitative methods and logistic regression to create a series of risk prediction models to assess the value of adding external multilevel data, and the applicability of the model to subpopulations. The final and third aim qualitatively assesses the health plan perceptions on PA and the use of multilevel data, guided by the domains and constructs of the CFIR.

Chapter 4: Presentation and Analysis of Data

This chapter presents the findings of each aim. The introduction summarizes each aim's results. Then the results of each aim are presented; specifically: 1) the results from the assessment and quality of data sources and variables; 2) results of the development of the risk prediction models and application to the non-White (minority) subpopulation; and 3) the results of the assessment of the health system perceptions through qualitative interviews.

The project was approved by the KPNW IRB (#00000405) on January 6, 2021. KPNW IRB agreed to accept review authority and continuing oversight from the Portland State University IRB under authorization agreement. A waiver of consent and HIPPA authorization was granted for Aims 1 and 2; a waiver of documentation of consent was granted for the interviews for Aim 3.

Introduction

Aim 1: Assess data sources and variables

The first aim was to assess data availability within an integrated health system and identify opportunities to include multilevel external data following the SEM framework. This data will be used in Aim 2 to develop predictive models to determine patients' probability of screening for colorectal cancer. Prediction models seek to identify predictors that are available, easy to obtain, and that can be measured with

reasonable precision. (Steyerberg, 2019) When data is unavailable there will be limits to how a model will perform and whether it will be adopted into practice.

Data were assessed and acquired from a variety of internal sources at the health system including the Kaiser Permanente Northwest (KPNW) Virtual Data Warehouse (VDW). The VDW includes electronic health record (EHR) data, administrative data, and external data resourced for research purposes (census data), among other data.

Data was also sought across the multiple levels (SEM) that were publicly available. Data was retained at the lowest level when multiple levels were available. For example, when a variable was available at the individual and community levels, the individual level was retained. Variables were selected that are relevant to CRC screening. Systematic reviews and key articles were assessed to ensure that variables that were previously identified as predictive of screening were included if available. Additional variables that have not been previously determined as predictors of screening were included if a correlation was suspected based on prior literature or similarities between other variables.

Feasible predictors have been categorized into levels according to the SEM framework including the individual, interpersonal, organizational, community, and policy levels. The lower level data (individual and interpersonal) was primarily guided by literature on predictors of cancer screening. The upper levels (organizational, community, and policy) were guided by data availability.

Aim 2: Develop risk prediction models

For Aim 2, prognostic risk prediction models were developed to identify patients' likelihood of screening for CRC. The models were developed to assess if risk prediction could identify patients' likelihood of screening for CRC. This retrospective cohort analysis identified patients due for screening on their birthday in 2018. Patients with missing data were removed from the model. The final individual model for the full population included 59,234 patients and the multilevel model 58,040 patients who were aged 50-75 and due for screening. While screening recommendations were recently reduced to age 45, the recommended age of screening was 50 during the period for which the data were analyzed. (US Preventive Services Task Force, 2021) Models were developed using logistic regression and a stepdown method for reducing the model, first using the individual level data, and then multilevel data based on the SEM framework. Both models were also applied to the subpopulation of *non-White* patients and performance was assessed. Subpopulation models included 12,676 and 12,184 patients for the individual and multilevel models respectively who were non-White.

The full population models of the individual and then multilevel data performed well with an R^2 of 0.1108 and 0.1119, and bootstrap corrected C- statistics of 0.7220 and 0.7218, respectively. These models may be useful in that risk prediction can be used to identify patients' likelihood of screening for colorectal cancer.

Aim 3: Assess health system perceptions

To assess the health system perceptions of predictive analytics, multilevel data, and the use of the risk prediction model in CRC screening, purposive sampling was used to recruit 5 current KPNW staff members working in predictive analytics or gastroenterology (CRC screening). Participants included frontline and quality improvement personnel from the Kaiser Foundation Hospitals group and Northwest Permanente. Semi-structured interviews consisted of 17 specific questions regarding their departments, CRC screening, data, predictive analytics, and usability of the model. Through thematic analysis, constructs within the interviews across the domains were identified.

The intervention guide was created under the constructs of the CFIR framework to assess the individual, inner setting, outer setting, process, and intervention components of their perceptions. Policy questions were asked to ascertain policies driving decisions at KPNW. The questions asked about the decision making process, CRC screening, use of data, use of predictive analytics, and model usability.

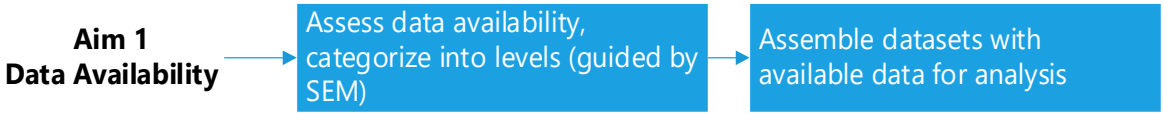
Inner setting findings include that team and group decisions are made, they all use PA to some extent, and the Quality Team facilitates the use of analytics in their departments. Outer setting questions revealed that many other departments and outside organizations are using PA and that this method of improvement is cost-efficient. The individual level revealed all individuals use data and IT tools in their daily work and have a favorable view on the use of PA. The process construct questions

revealed barriers including access to multilevel data and incorporating models into their general workflow and EHR. The Intervention construct revealed high usability of the model overall in the use of the information it provided, and that it could be used in a variety of ways to make their jobs easier. They identified using external data as difficult. The policy questions indicate no formal policies regarding the use of data or analytics, but rather an encouragement to use analytics to improve the quality of care for KPNW patients.

Aim 1: Assess data sources and variables

Aim 1 sought to assess data availability within electronically available data at a single integrated delivery system (KPNW), and opportunities to incorporate external data to evaluate predictors at all levels of the SEM framework. The potential predictors at the individual, interpersonal, organizational, community and policy levels were each assessed for data availability.

Figure 4.1 Conceptual Design: Aim 1



Known or previously examined predictors of screening were identified through prior literature, systematic reviews, and key articles. These predictors were categorized into SEM levels; 44 known predictors were identified (Table 1). Then, new variables

were identified for use if there was a suspected correlation with CRC screening based on content knowledge; this resulted in 49 new variables being identified. Combined, 93 potential predictors were retained, 47.3% are known predictors and 52.7% are new potential predictors.

KPNW databases were reviewed to determine what data was available. The KPNW VDW databases include data from the EHR, administrative data, claims, state records, legacy data systems, and linked data from the census. The data obtained from these databases includes utilization data, demographics, vital statistics, lab tests, treatments, enrollment, death, pharmacy, tumor registry, oncology, pathology, and patient-reported measures. Data from these sources go back as far as 1994 and most of the tables are updated daily or weekly.

Publicly available external databases were reviewed for content within the database and applicability to CRC screening and usability of the dataset regarding the ability to link it to patient level data. Databases were found on publicly available websites, and relevant variables were downloaded and categorized into the SEM levels.

Each of the variables was assessed for availability. Data were determined to be *common* or *emerging* based on the Agency for Healthcare Research and Quality (AHRQ) User's Guide for Registries for Evaluating Patient Outcomes. (Ehrenstein V et al., 2019) This resource identifies data as common if it is generally available in EHRs, and emerging if generally available but of interest and ability to be integrated into current data registries. (Ehrenstein V et al., 2019) Further, data were categorized as *calculated* if

programming is required to create the data point, *public* if it is publicly available, or *system* if availability is dependent on the health system in which the data resides. Data categories were mutually exclusive. Of the 93 total variables, 29 (31.2%) were determined to be commonly available, 9 (9.7%) emerging, 8 (8.8%) would need to be calculated to use, 13 (14.0%) are system specific variables, and 41 (45.1%) are publicly available.

Table 4.1 Data Characteristics

	SEM Level					
	Individual	Interpersonal	Organizational	Community	Policy	Total
Total variables (n)	32	13	6	35	7	93
Known predictors of CRC screening*, n(%)	27 (84.4%)	10 (76.9%)	2 (33.3%)	5 (14.3%)	0 (0.0%)	44 (47.3%)
New variables**, n(%)	5 (15.6%)	3 (23.1%)	4 (66.7%)	30 (85.7%)	7 (100.0%)	49 (52.7%)
Availability+, n(%)						
Common	25 (78.1%)	2 (15.4%)	2 (33.3%)			29 (31.2%)
Emerging	7 (21.9%)	2 (15.4%)				9 (9.7%)
System		9 (69.2%)	4 (66.7%)			13 (14.0%)
Public				35 (100.0%)	7 (100.0%)	41 (45.1%)
Calculated^, n(%)	1 (3.1%)	5 (41.7%)	2 (33.3%)			8 (8.8%)

*Known predictors of screening are variables identified through prior literature

**New variables are exploratory in prediction of CRC screening

+Common if generally available, emerging if not generally available but of interest, public if publicly available, system if system specific data

^ Calculated if programming required

Individual Level Data

The innermost individual band of the SEM represents the individual patient and their characteristics which might affect their health behaviors. Individual level data is the most accessible data at a health system, as most data is captured at the individual level for each patient. Typical patient level data includes demographics, clinical encounters, enrollment data, diagnoses, and procedures.

At the individual level, 32 variables were initially identified for potential use in the predictive models, 84.4% of which were known predictors of CRC screening (Table 4.1). Individual level data applicable to CRC screening are outlined in Table 4.2. Data applicable to CRC screening was determined by prior literature on known predictors of screening and by subject knowledge and suspected correlation. Known predictors include demographics such as age, education, ethnicity, race, language, sex, income, insurance, immigration status, and need for an interpreter. (Beydoun & Beydoun, 2008; Guessous et al., 2010; Gupta et al., 2014; Joseph DA et al., 2016; Linsky et al., 2011; Percac-Lima et al., 2009; Petrik et al., 2018; Vahabi et al., 2021; Weiss et al., 2013) Known predictors from encounter data include the use of other health services, the number of inpatient and outpatient visits, missed visits, or dental visits. (Beydoun & Beydoun, 2008; Liang et al., 2006; Petrik et al., 2018; Seeff et al., 2004; Vahabi et al., 2021; Young et al., 2007; Zimmerman et al., 2006) Enrollment predictors include membership length, patient portal enrollment, primary care physician assignment. (Beydoun & Beydoun, 2008; Guessous et al., 2010; Joseph DA et al., 2016; Vahabi et al., 2021) Diagnosis predictors include the presence of chronic disease, comorbidities that limit screening, comorbidity score, polysubstance abuse, and social determinants of health (determined by Z-codes in ICD-10). (Beydoun & Beydoun, 2008; Denberg et al., 2005; Gimeno Garcia, 2012; Ioannou et al., 2003; Weiss et al., 2013) Predictors from procedure data include prior CRC screening and test preference, flu shots, mammograms for women, and prostate screening for men. (Beydoun & Beydoun, 2008;

Gimeno Garcia, 2012; Murphy et al., 2014; Petrik et al., 2018; Zimmerman et al., 2006) Finally, Body Mass Index (BMI) is retained from vitals. (Beaber et al., 2019)

The 5 variables that are not known predictors of screening are believed to be correlated with the likelihood of CRC screening, based on subject specific knowledge. These include continuity of care (measured as the length of membership at the health plan), membership length <2 years (as an indicator of incomplete CRC screening history information in the patient record). Also, comorbidities that limit life expectancy, COPD, and dementia, were explored as new potential predictors of screening. Finally, health care utilization, including the number of hospital stays and the number of missed visits in the past year were included as new potential predictors.

Of the 32 variables identified, 78.1% are common, 21.9% are emerging and 3.1% (1 variable), the Charlson Comorbidity Score would have to be calculated. The Charlson score is commonly used to classify patient's comorbidities. (Charlson et al., 1987) Although common data, not all health systems will have access to specialty care data including dental visits and hospitalization data. The two unavailable variables are immigration status and the Charlson comorbidity score. The components of the Charlson score would be available in a typical EHR (diagnosis codes) but would require advanced programming to calculate the score. A patient's individual test preference is identified as a known predictor of screening but is not captured in retrospective data through an EHR, however, prior screening type (FIT vs. Colonoscopy) is an available proxy.

Table 4.2 Prospective Individual Level Data

Source	Variable	Description	If Known Predictor of CRC Screening, Source	Data Availability
Demographic Data				
EHR	Age	Age, >=45 years on patient's index_date (birthday 2018)	Beydoun, Guessous	Common
EHR	Education	Years of education completed	Beydoun, Guessous, Joseph	Emerging
EHR	Ethnicity	Hispanic	Beydoun, Guessous	Common
EHR	Immigrant	Recent immigrant	Vahabi, Gupta	Emerging
EHR	Income	Income	Beydoun	Emerging
EHR	Insurance	Insurance	Beydoun, Guessous	Common
EHR	Interpreter	Interpreter needed (Y/N)	Linsky, Percac-Lima	Common
EHR	Language	Preferred language	Petrik, Weiss	Common
EHR	Race	Race	Beydoun, Guessous, Joseph	Common
EHR	Sex	Gender	Beydoun, Guessous, Joseph	Common
EHR	BMI	Body Mass Index	Beaber (breast - PROSPER)	Common
Encounter Data				
EHR	Dental Visits	Seen a dentist in the prior year	Liang	Common
EHR	Out-patient Visits	Number of visits (OP) in prior year (0, 1, 2+)	Vahabi, Young, Petrik, Seeff, Zimmerman, Beydoun	Common
EHR	In-patient Visits	Hospital stays in prior year (y/n)	*New potential predictor*	Common
EHR	Missed Visits	Number of missed visits in prior year	*New potential predictor*	Common
Enrollment Data				
EHR	Usual source of care	Assigned primary care physician	Beydoun, Guessous, Vahabi, Joseph	Common
EHR	Continuity of Care	Membership > 5 years	*New potential predictor*	Common
System Administrative Data	Membership length	Membership >2 years	*New potential predictor*	Common
EHR	Patient portal	Enrollment in patient portal	*New potential predictor*	Common
Diagnoses Data				
EHR	Chronic disease	presence of chronic disease	Beydoun	Common
EHR	Comorbidity	Other health problems	Denberg	Common
EHR	Comorbidity Score	Charleson Comorbidity Score (calculated from diagnoses codes)	Weiss	Common/ Calculated
EHR	Polysubstance abuse	Smoking and other substance abuse	Beydoun, Ioannou, Gimeno Garcia	Common
EHR	SDOH Economic	Z59 - Problems related to housing and economic circumstances (low income housing, etc.)	Hammond, Calo, Ahmed	Emerging
EHR	SDOH Education	Z55 - Problems related to education and literacy; Less than a high school degree; High school diploma or GED	Hammond, Calo, Ahmed	Emerging
EHR	SDOH Social	Z60 - Problems related to social environment (stress, social enviornment)	Hammond, Calo, Ahmed	Emerging
EHR	Comorbidities that limit screening	Comorbidities that may limit screening (COPD, Dementia)	*New potential predictor*	Emerging
Proceure Data				
EHR	Prior CRC screening	Patient has had prior CRC screening	Young	Common
EHR	Flu shot	Patient has had prior flu vaccination (in prior year)	Petrik, Zimmerman	Common
EHR	Mammogram	Females have had prior mammogram	Petrik	Common
EHR	Prostate cancer screening		Gimeno Garcia	Common
EHR	Test preference	Perceptions of screening (test preference)	Beydoun, Murphy	Common

Interpersonal Level Data

The interpersonal level is the second level in the SEM. Interpersonal characteristics typically refer to the social influences and norms within a person's social networks. (Kumar et al., 2012) Interpersonal characteristics can impact individual behavior through norms, interpersonal connections, and relationships that influence health behaviors. This may include personal connections and an individual's relationship with the provider. Past literature provided a roadmap for finding interpersonal level data. The interpersonal level data was less widely available than the individual level data.

At this level, 13 variables were identified for use in the models, 76.9% of which were known predictors of screening (Table 4.1). Retained variables of interest are outlined in Table 4.3. These variables include marital status, physician recommendation of screening, provider screening rates, provider panel size, provider specialty, provider gender and provider and patient gender match, ethnicity match, and race match. (Beydoun & Beydoun, 2008; Guessous et al., 2010; Henderson & Weisman, 2001; Laiyemo et al., 2014; Seeff et al., 2004; Strumpf, 2011; Weiss et al., 2013; Zimmerman et al., 2006) Available interpersonal level data was found in provider datasets, procedures, and demographics. Some data had to be calculated from combinations of data, such as provider gender match and provider race match.

Three new variables were explored as potential new predictors of CRC screening. The new variables are provider screening rates, provider gender, and the patient's

number of address changes in the past 2 years. Provider screening rates could explain screening behaviors among the patient panel. The number of address changes could explain housing stability which may influence screening patterns.

There were three known predictors identified in the literature that were unavailable in the retrospective datasets at KPNW but may be available elsewhere. Unavailable data includes address changes, family history of colorectal cancer, family size, and provider practice time. (Beydoun & Beydoun, 2008; Katz et al., 2000; Weiss et al., 2013; Young et al., 2007)

Only 2 (16.7%) of the total variables in this level are common, and another 2 (16.7%) are emerging in health systems retrospective datasets. The only common variables are marital status and physician recommendation of screening. The emerging variables are family history and family size. A large proportion, 41.7% (5 variables) must be calculated, including provider panel size, screening rates and match with gender, race, and ethnicity of the patient. Another 9 variables (69.2%) are system specific variables such as patient address changes, provider specialty, and the number of years a provider has been practicing.

Table 4.3 Prospective Interpersonal Level Data

Source	Variable	Description	If Known Predictor of CRC Screening, Source	Data Availability
EHR	Marital Status	Marital Status	Beydoun, Guessous, Zimmerman	Common
EHR	Physician recommendation	Physician recommendation (prior order of CRC test)	Beydoun, Guessous, Laiyemo, Seeff	Common
System Administrative Data	Address changes	Number of address changes in past 2 years	*New potential predictor*	System
Provider Data	Provider panel size	Number of patients in the providers panel	Weiss	System/ Calculated
Provider Data	Provider specialty	Provider specialty (IM, FP, NP, PA)	Weiss	System
Provider	Panel screening rates	Provider panel screening rates	*New potential predictor*	System/ Calculated
Provider	Provider Gender	Provider gender	*New potential predictor*	System/ Calculated
Provider and EHR	Provider gender match	Does provider gender match patient's gender	Henderson	System/ Calculated
Provider and EHR	Provider ethnicity match	Does provider ethnicity match patient's ethnicity	Strumpf	System/ Calculated
Provider and EHR	Provider race match	Does provider race match patient's race	Strumpf	System/ Calculated
Unknown	Family history	Family history of colorectal cancer (1st degree)	Beydoun, Young	Emerging
Unknown	Family size	Family size	Katz	Emerging
Unknown	Provider practice time	Years in practice	Weiss	System

Organizational Level Data

Organizational or institutional level data represent characteristics of the organization that may impact its ability to screen patients. Characteristics of the system could provide barriers or facilitators for screening. KPNW is a single organization, however, the organization is made up of 34 distinct medical offices that have distinct characteristics. All organizational data are considered at the clinic level.

At this level, only 6 variables were identified for analysis (Table 4.1). The data at this level are outlined in Table 4.4. Two variables (33.3%) at the organizational level were identified as predictive of CRC through the literature. First, the clinic where a patient receives care has been identified as a predictor of screening (Petrik et al., 2018); patients will be linked to their primary clinic, where their primary provider is located.

Wait times or scheduling challenges have also been identified as predictors; (Denberg et al., 2005) wait time will be calculated at the clinic level from procedural data from order to scheduled colonoscopy dates. Some known organizational level predictors of CRC screening were determined not to be valuable to this project. Translational services are available to all patients, although variation in utilization among patients is present, variation across clinics is limited. Also, distance from a patient's home to the clinic is irrelevant for patients undergoing FIT testing, as the test generally is mailed to the patient's home and returned to the lab by mail.

Four new potential predictors (66.7% of total variables) at the organizational level are included. These variables are from administrative data and include clinic location, clinic size, clinic patient to provider ratio, and clinic level screening rates. Clinic level characteristics may contribute to screening patterns but have not been explored in past literature. However, clinic level screening rates are unavailable.

Of the six total variables, two variables (33.3%) were common, clinic assignment and wait time. Four of the variables (66.7%) are specific to the health system. These system specific variables include clinic size, clinic to provider ratio, clinic location, and clinic screening rates. Two variables (33.3% of total), wait time for a colonoscopy, and clinic level screening rates would have to be calculated.

Table 4.4 Prospective Organizational Level Data

Source	Variable	Description	If Known Predictor of CRC Screening, Source	Data Availability
Clinic data	Clinic	Clinic assignment	Petrik	Common
EHR	Wait time	Long wait times/scheduling challenges (time between order and scheduled colonoscopy)	Denberg	Common/ Calculated
System Administrative Data	Clinic size	Number of patients in patient's primary clinic	*New potential predictor*	System
System Administrative Data	Clinic size FTE	Clinic patient to provider ratio	*New potential predictor*	System
System Administrative Data	Location	Clinic location	*New potential predictor*	System
Unknown	Clinic screening rates	Clinic level screening rates	*New potential predictor*	System/ Calculated

Community Level Data

Community level data is the fourth level of the SEM framework. Data at this level may impact screening if community characteristics influence individual behavior. Characteristics of a community could impact screening likelihood by increasing awareness or by having decreased access to care. Patient data is linked by census block or zip code to community level data.

Community level data was widely available, as many publicly available data sets were found. A variety of resources offer county level data. AHRQ offers publicly available county level social determinants of health data drawn from different sources. This data includes over 337 variables on a wide variety of topics. The data are compiled by AHRQ from federal and other data sources over multiple years. The Robert Wood Johnson Foundation hosts a dataset called the “County Health Rankings and Roadmaps” for counties across the country that includes 77 measures of health. Data in this dataset is from multiple agencies including AHRQ, American Community Survey, Area Health Resource File from the American Medical Association, Behavioral Risk Factor Surveillance System, Bureau of Labor Statistics, Centers for Medicaid and Medicare

Services, Environmental Public Health Departments, National Assessment of Adult Literacy, National Center for Health Statistics, Office of Rural Health, and the Oregon Department of Education.

Additionally, Health Literacy Scores by census block are publicly available through the University of North Carolina at Chapel Hill as this data was created through funding by the National Institutes on Aging (University of North Carolina at Chapel Hill, 2014). Finally, the KPNW Utility for Care Department of Analytics (UCDA) links member level data to census data, which provides census block group level information on some community characteristics (e.g., area deprivation index).

Levels within the community level data include the census block and county. Data were retained if applicable to CRC screening. Community level data retained for use in the model are outlined in Table 4.5. After the reduction of the dataset to only include variables applicable to CRC screening, 35 community level variables were retained (Table 4.1). Only 5 (14.3%) of these variables are previously examined predictors of CRC screening. These known predictors were all at the census block level and include the GINI income inequality index, population density, health literacy, neighborhood deprivation index, education median household income, and neighborhood income. (Beydoun & Beydoun, 2008; Joseph DA et al., 2016; Petrik et al., 2018; Vahabi et al., 2021)

Most of the variables (n=30; 85.7%) are new to CRC screening literature because literature has rarely looked at community level variables. Two variables, health literacy,

and neighborhood deprivation index are at the census block level. The other data from the county level are all new potential predictors and include population characteristics like county size, median age, premature death, and cardiovascular disease and community level healthcare characteristics like health professional population and facilities.

In addition to the variables that did not apply to CRC screening, some data was not retained at the county level because it was available at the individual level. These variables include education, primary care access, preventive care, and available census data. Some preferred data was not available, such as the presence of media campaigns to increase screening awareness and community level screening rates.

All the data in this level, 35 variables (100%), are publicly available through the public datasets described above.

Table 4.5 Prospective Community Level Data

Source	Variable	Description	If Known Predictor of CRC Screening, Source	Data Availability
Census Block Level				
American Community Survey	GINI	GINI Income Inequality	Petrik	Public
American Community Survey	Population density	Population density by census block (Joseph found metropolitan protective)	Joseph	Public
VDW/Census	Education	Average education by census block	Beydoun	Public
VDW/Census	Median household income	Median household income by household block	Beydoun	Public
VDW/Census	Neighborhood income	Low income neighborhood (y/n) (% of households below poverty level)	Vahabi	Public
VDW/Census	Health Literacy	Health Literacy Estimate	*New potential predictor*	Public
VDW/Census	NDI	Neighborhood deprivation index (higer values mean more deprivation)	*New potential predictor*	Public
County Level				
Agency for Healthcare Research and Quality	Percent limited English	Percentage of households with limited English speaking	*New potential predictor*	Public
Agency for Healthcare Research and Quality	Pecent cititzens	Percentage of population who are not U.S. citizens	*New potential predictor*	Public
Agency for Healthcare Research and Quality	Total population	Total weighted population	*New potential predictor*	Public
Agency for Healthcare Research and Quality	Median age	Median age	*New potential predictor*	Public
Agency for Healthcare Research and Quality	Percent 45-64	Percentage of population age 45–64	*New potential predictor*	Public
Agency for Healthcare Research and Quality	Percent 65+	Percentage of population age 65 and over	*New potential predictor*	Public
Agency for Healthcare Research and Quality	Percent Hispanic	Percentage of population reporting Hispanic ethnicity	*New potential predictor*	Public
Agency for Healthcare Research and Quality	Percent white	Percentage of population reporting White race	*New potential predictor*	Public
Agency for Healthcare Research and Quality	Health professional shortage	HPSA code—shortage of primary care physicians (1=whole county, 2=part of county)	*New potential predictor*	Public
National Center for Health Statistics	Healthcare facilities	Number of facilities	*New potential predictor*	Public
Center for Disease Control (Atlas)	Cardiovascular death rate	Total cardiovascular disease death rate per 100,000 population	*New potential predictor*	Public
Community Health Rankings	Premature death rate	Premature deaths: age-adjusted deaths per 100,000 population aged 74 and under	*New potential predictor*	Public
American Community Survey, 5-year estimates	Income inequality	GINI index of income inequality	*New potential predictor*	Public
Behavioral Risk Factor Surveillance System	Quality of life	Percentage of adults reporting fair or poor health (age-adjusted).	*New potential predictor*	Public
Behavioral Risk Factor Surveillance System	Poor physical health days	Average number of physically unhealthy days reported in past 30 days (age-adjusted).	*New potential predictor*	Public
Behavioral Risk Factor Surveillance System	Poor mental health days	Average number of mentally unhealthy days reported in past 30 days (age-adjusted).	*New potential predictor*	Public
Behavioral Risk Factor Surveillance System	Excessive drinking	Percentage of adults reporting binge or heavy drinking.	*New potential predictor*	Public
Centers for Medicaid and Medicare Services	Primary Care Physicians	Rate of primary care physicians per 100,000 population	*New potential predictor*	Public
Environmental Public Health Tracking Network	Air pollution - particulate matter	Average daily density of fine particulate matter in micrograms per cubic meter (PM2.5).	*New potential predictor*	Public
Centers for Medicaid and Medicare Services	Preventable hospital stays	Discharges for Ambulatory Care Sensitive Conditions per 100,000 Medicare Enrollees	*New potential predictor*	Public
National Center for Health Statistics - Mortality Files	Premature death	Years of potential life lost before age 75 per 100,000 population (age-adjusted).	*New potential predictor*	Public
National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention	Sexually transmitted infections	Number of newly diagnosed chlamydia cases per 100,000 population.	*New potential predictor*	Public
Census Population Estimates	Rural	% of Rural Residents	Joseph, Ojinnaka	Public
Area Health Resources File (Rural RUCCA Code)	RUCCA	Rucca Code	Joseph, Ojinnaka	Public
Safe Drinking Water Information System	Drinking water violations	Indicator of the presence of health-related drinking water violations. 'Yes' indicates the presence of a violation, 'No' indicates no violation.	*New potential predictor*	Public
Uniform Crime Reporting - FBI	Violent crime	Number of reported violent crime offenses per 100,000 population.	*New potential predictor*	Public
United States Diabetes Surveillance System	Physical inactivity	Percentage of adults age 20 and over reporting no leisure-time physical activity.	*New potential predictor*	Public
USDA Food Environment Atlas, Map the Meal Gap from Feeding America	Food environment index	Index of factors that contribute to a healthy food environment, from 0 (worst) to 10 (best).	*New potential predictor*	Public

Policy Level Data

The outer band of the SEM is the policy level. Characteristics at this level may impact screening behaviors through policies that impact healthcare access, guidelines for when screening should occur, or screening awareness. Publicly available data at this level was limited in what might influence CRC screening.

Little policy level data was publicly available. Only 7 variables were identified at this level (Table 4.1). Policy level data retained for this project are outlined in Table 4.6. Policy is known to impact CRC screening rates. Policies such as HEDIS, recommendations by the USPSTF, CMS changes to recommendations, and the American Cancer Society's drop in age for recommended screening knowingly have impacted screening rates. The KPNW policy to screen African American's aged 45-49 came from updated national recommendations for screening and is an example of organizational policy driving screening practices. (Rex et al., 2017) However, policy level data is rarely used as a predictor of patient level screening rates. Therefore, there were no known policy level predictors of CRC screening from prior literature (0%). The exploration of this level of data is new to CRC screening literature, as little variation is expected across the population from a single health system in two states. Available data at this level includes data from Coordinated Care Organizations (CCOs), and state level policy. All data was found on the Oregon Health Authority website. New potential predictors retained for analysis include CCO screening rates, CCO incentive achievements, and screening programs. New potential predictors from state data retained for analysis include

insurance rates, incentivized metrics, and if CCO programs were in place. Some data, such as Medicaid expansion, will not be retained for analysis, as there is no variability across Washington and Oregon.

All 7 variables (100%) from this level are publicly available through public datasets for use in the analysis.

Table 4.6 Prospective Policy Level Data

Source	Variable	Description	If Known Predictor of CRC Screening, Source	Data Availability
CCO Level				
Oregon Health Authority	CCO Screening rate	2019 Colorectal cancer screening rate	*New potential predictor*	Public
Oregon Health Authority	CCO 2019 Incentive Rate	0=nothing; 1= achieved improvement; 2=achieved benchmark	*New potential predictor*	Public
Oregon Health Authority	CCO Screening programs	Colorectal cancer screening programs put into place	*New potential predictor*	Public
State Level				
State data	State Medicaid Expansion	Medicaid Expansion	*New potential predictor*	Public
State data	State Uninsured residents	Percent of residents uninsured	*New potential predictor*	Public
State data, Department of Health, CMS for ACOs	State CCO Program	CCO Program in Place for care coordination	*New potential predictor*	Public
State data	State CRC incentivized metric	Is there an incentivized metric for CRC screening?	*New potential predictor*	Public

Data Quality and Availability

Once the final list of variables was determined, each data element was pulled and examined among the eligible population. The eligible population is described in the Results section of Aim 2. The eligible population includes 60,220 patients identified as eligible for screening in 2018.

All variables were assessed for data quality for use in the risk prediction models. Each variable was assessed for redundancy with other variables, health system

availability according to the KPNW analyst (assigned to pull the data), missingness among eligible patients, and the extent of variation across eligible patients (Table 4.7). Redundancy was assessed because variables with the same meaning do not add to the models. Availability was assessed to eliminate unavailable variables. Missingness was assessed because regression analysis deletes subjects with any missing values, therefore requiring understanding the meaning of a missing value. The extent of variation across eligible patients was also assessed to ensure that variables contribute to the model.

Redundant variables were identified if there was obvious collinearity with another variable based on overlap and crosstabulations. Availability was determined by the KPNW analyst, if they were unable to acquire a particular variable or acquiring the variable was not feasible. Variables with greater than 5% missingness were assessed for possible imputation or removed from the potential predictor list. Finally, the distribution of the predictor is assessed for variance. Variables with less than 1% variance were also removed from the potential predictor list.

Of the 93 original variables identified as prospective measures, 7 (7.5%) variables were determined to be redundant with other measures, 10 (10.8%) variables were excluded due to availability, 6 (6.5%) were excluded due to missingness, and 9 (9.7%) were excluded for distribution reasons. The remaining 61 variables (64.5%) will be used in the EHR and multilevel prediction models.

Table 4.7 Data Quality for Use in Risk Prediction

	SEM Level					
	Individual	**Interpersonal**	**Organizational**	**Community**	**Policy**	**Total**
Total variables (n)	32	13	6	35	7	93
Variables removed from potential predictors, n(%)						
Not Available	2 (6.3%)	5 (38.5%)	1 (16.7%)	2 (5.7%)	0 (0.0%)	10 (10.8%)
Missing	1 (3.1%)	2 (15.4%)	0 (0.0%)	0 (0.0%)	3 (42.9%)	6 (6.5%)
Distribution too limited	5 (15.6%)	1 (7.7%)	0 (0.0%)	2 (5.7%)	1 (14.3%)	9 (9.7%)
Redundant with another measure	3 (9.4%)	0 (0.0%)	0 (0.0%)	2 (5.7%)	2 (28.6%)	7 (7.5%)
Variables remaining	21 (65.6%)	5 (38.5%)	5 (83.3%)	29 (82.9%)	1 (14.3%)	61 (65.6%)

Individual Level

At the individual level, three variables (9.4% of total variables) were redundant with other measures. The *presence of chronic disease* and *comorbidities that limit screening* were redundant with the Charlson comorbidity score. *Prior CRC screening* was redundant with *test preference*, as the latter variable's missing values indicated not having prior screening.

Two variables (6.3% of total variables) were not available at the individual level. The unavailable variables were *recent immigrant* and *income*. Both variables were categorized as "emerging" but were not in the KPNW databases.

One variable (3.1% of total variables), *years of education completed*, was determined to have too many missing values for use. Of all eligible patients, 95.9% of the values were missing for *years of education completed.* BMI had 9.0% of the values missing, but single imputation was performed as the data is presumed to be missing at random.

Limited distribution caused five variables (15.6% of the total variables) to be removed from the analysis. The variables included *assigned a primary care physician* (99.3% were yes), *other health problems that limit screening* (0.43% were yes), and the social determinant of health variables *SDOH problems related to housing and economic circumstances* (2.83% yes), p*roblems related to education and literacy* (0.05% yes), and *problems related to social environment* (0.08% yes). The remaining variables (n=21, 65.6% of total variables) are used for the primary analysis of the individual variables as well as the multilevel model.

Interpersonal Level

At the interpersonal level, there were 13 original variables identified for analysis. Of those, none were identified as redundant with other measures, and one variable was removed due to distribution. *Family history*, identified by diagnosis code, was found for <1% (0.82%) of the eligible population and therefore removed.

The five variables (38.5% of total variables) that were not available at the interpersonal level included *marital status, physician recommendation, panel screening rates, family size,* and *physician's years in practice.* The *physician recommendation*, determined by a prior order of a screening test, was found to be missing for most patients, as the ordering of FIT is commonly done through a centralized team. *Panel screening rates* were also unable to be calculated by the analyst, as the entire panel would have to be pulled and proved to be too cumbersome for this project.

Two variables (15.4%) were removed from the analysis due to missingness. These variables were: *does the provider race match the patient* and *does provider ethnicity match the patient.* The provider race and ethnicity variables in the VDW dataset were missing 100% of the time not allowing for a calculated match of race and ethnicity. The five remaining variables (38.5%) are used in the multilevel model.

Organizational Level

At the organizational level, six variables were identified as prospective data. Of this data, 1 variable (16.7%) was unavailable. The *Clinic screening rates* variable was unable to be calculated by the analyst, as the entire clinic population would have to be pulled, and this work proved to be too cumbersome for this project. No variables were redundant with other measures, no variables had too much missingness. The remaining five variables were retained for analysis.

Community Level

At the community level, 35 variables were originally identified as potential predictors. Two variables from the census block level (5.7% of total variables) were not available in the KPNW Census databases in the VDW. These variables are *GINI* and *Population density*.

Of the eligible population, only 691 patients (1.1%) did not have a zip code on file. All community level data at the county level are missing for these patients. Two

variables (5.7% of total variables) were found to be redundant with other variables. *Premature death* from the National Center for Health Statistics database was redundant with another premature death variable from the County Health Ranking database. *Poor physical health* was redundant with the *Quality of life* variable which is measured by the number of days in good health.

Two variables (5.7% of total), *Rural RUCCA* and *Health literacy* were found to have fewer than 1% variation, so eliminated because the distribution is too limited. In the *Health literacy* variable, <1% of the counties had "basic" literacy, 99% had "intermediate" literacy.

The remaining 29 variables (82.9% of total variables) will be used in the multilevel model.

Policy Level

There were seven original policy level variables identified as potential predictors. The distribution of each variable was assessed among the eligible population. The CCO level policy variables (n=3, 42.9% of total variables) were missing for all patients that do not live in Oregon (26.3%), so were therefore removed as potential predictors for the full model. These variables were: *CCO 2019 colorectal cancer screening rate, CCO 2019 incentive rate,* and *CCO screening programs.*

The state level policy variables were determined to be completely redundant with each other and colinear. *State, CRC incentivized metric* variable measures the

potential difference in the incentivized metric policy with 2 values one for each state (Oregon=yes and Washington=no). Two variables (28.6% of total variables) were removed due to redundancy. The *State Uninsured Residents* and *State CCO Program* variables have the exact population with the same variation by patient's state, Oregon, and Washington, so are colinear with the *State, CRC incentivized metric* variable. One variable (14.3% of total variables) was removed due to distribution, *State Medicaid Expansion,* which had no variation, as both Oregon and Washington are Medicaid Expansion states.

Therefore, only one policy level variable *State, CRC incentivized metric* was retained (14.3% of total variables). This variable will be retained for the multilevel model.

Summary of Aim 1 findings

Datasets were assembled containing 93 available variables at the SEM levels for use in Aim 2. Data at the individual and community levels were the most available (32 and 35 variables respectively) (Table 4.1). The individual level had the most known predictors (84.4%), and the policy level had the least known predictors (0%). Data at the interpersonal, organizational, and policy levels were limited (13, 6, and 7 variables respectively). Overall, 93 variables across all 5 levels were identified. Of those, 44 (47.3%) are known predictors of CRC screening, and 49 (52.7%) are new potential predictors. Of the 93 total variables, 31.2% were common variables from EHR and

administrative data, 9.7% emerging, 8.8% would have to be calculated, 14% are system specific, and 45.1% are publicly available.

Of the 93 originally identified variables 10.8% were excluded because they were unavailable, 6.5% were excluded because too many values were missing, 9.7% were excluded for distribution reasons, and 7.5% were excluded due to redundancy with other variables. Most of the variables excluded due to availability were at the policy level (85.7%). The policy level had the most variables excluded due to missingness (42.9%). The individual level had the most variables excluded for distribution reasons (15.6%).

In sum, interpersonal, organizational, and policy level data is less available than individual or community level data. Fewer known predictors are at the upper levels. About two-thirds of the proposed data at the multi-levels were determined usable for the models. More proposed data was retained at the organizational and community levels (83.3% and 82.9% respectively), and the least amount of data was retained at the policy and interpersonal levels (14.3% and 38.5% respectively).

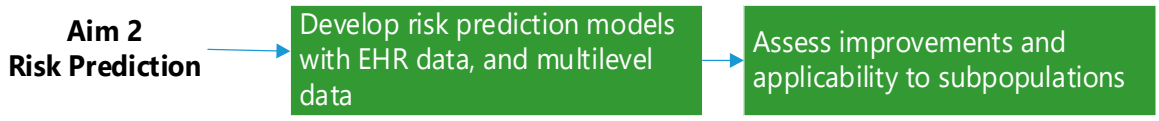
Aim 2: Develop risk prediction Models

The second aim is to develop risk prediction models using the KPNW data sources and available multilevel data as determined in Aim 1. This Aim includes developing a series of models to identify the risk of failing to complete CRC screening. These models are created to determine if a risk model can adequately identify patients unlikely to

screen for CRC. A retrospective cohort of patients due for screening was used for all models.

There is no external validation to these development models, it was not validated in an external dataset. Discrimination, or the ability to determine who will screen versus not, was measured by a bootstrap corrected C-statistic. (Moons et al., 2019) The discrimination and calibration was measured with the Neglekerke R^2 statistic. Calibration was assessed by the integrated calibration index (ICI). (Austin & Steyerberg, 2019) All analysis was initially conducted in STATA 17©, and then validated in SAS (bootstrapping, C and R^2) and R (ICI) by a KPNW analyst.

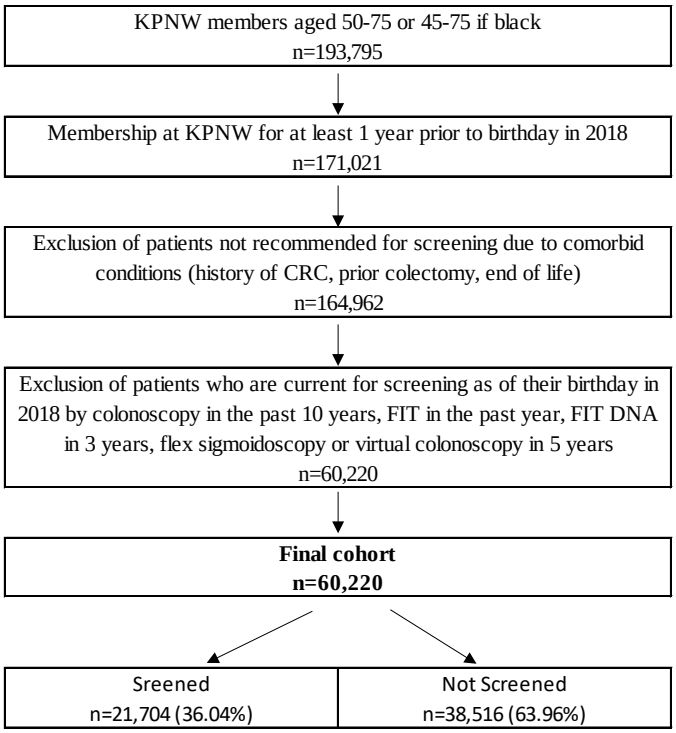
Figure 4.2 Conceptual Design: Aim 2



The first model was developed in a large full dataset of patients who are due for CRC screening using standard EHR individual level data. The multilevel model used the same individual level dataset and incorporated multilevel data from databases at KPNW, and publicly available data. The statistical improvements in the model were then assessed when multilevel data was added.

The Prediction Risk of Bias Assessment Tool (PROBAST) was used to assess the risk of bias (ROB) (Wolff et al., 2019). This is done by assessing participants, predictors, outcomes, and analysis bias (Wolff et al., 2019). Subpopulation applicability based on the PROBAST assessment was assessed by applying the models to a subpopulation, based on population characteristics. The Transparent Reporting of Multivariable Prediction Model for Individual or Prognosis and Diagnosis (TRIPOD) checklist was also used in reporting and clarification of the model development (Collins et al., 2015).

Figure 4.3 Eligible Patient Consort Diagram



Eligible Population

Eligible patients were identified through the EHR at KPNW. KPNW is an integrated health system that provides medical insurance coverage to about 606,000 members and dental insurance coverage to approximately 280,000 members in northwest Oregon and southwest Washington.

Patients were determined eligible if on their birthday in 2018 if they were due for screening, aged 50-75 and not Black or 45-75 and Black (n=193,795). Eligible patients had to have at least 1 year of membership prior to their birthday (n=171,021). Predictors were included from clinical encounters closest to but prior to their birthday in 2018. Outcomes were assessed for up to one year following their birthday.

Patients were excluded if they were not recommended due to comorbid conditions like a history of CRC, prior colectomy or if they were in end-of-life care (n=164,962 remaining eligible patients). Patients were also excluded if they were current for screening, by way of FIT in the past year, colonoscopy in the past 10 years, FIT DNA in the past 3 years, or flex sigmoidoscopy or virtual colonoscopy in 5 years (n=104,742 (63%) excluded). Of the remaining 60,220 eligible patients, 36% were found to have been screened, which is the primary outcome for analysis.

Outcomes and Predictors

The outcome is any CRC screening in the year following their birthday in 2018. The outcome includes any type of CRC testing including FIT, colonoscopy, FIT-DNA, virtual colonoscopy, or sigmoidoscopy. The outcome was identified by a CHR analyst, it was prespecified at the time of the data pull, and no predictors were excluded from the outcome definition. The CRC screening outcome was defined for all participants at the

Table 4.8 Modality of Screening

Screening Outcome	
	n (%)
Total	21704
FIT	19681 (90.7%)
Colonoscopy	3017 (13.9%)
Flexible Sigmoidoscopy	44 (0.2%)
FIT DNA (Cologuard)	7 (0.0%)

time of the data pull and was determined without prior knowledge of predictor information. Table 4.8 shows the distribution of modality of screening, 90.7% of patients were screened by FIT, and 13.9% were screened by colonoscopy. The categories are not mutually exclusive, patients may have been screened by more than one modality, for example, 1,011 patients were screened by FIT and Colonoscopy. Less than 1% of patients were screened by Flexible Sigmoidoscopy or FIT DNA.

The overall screening population (eligible for this analysis) is approximately 9.8% (60,220/617,073) of the overall population of KPNW. Compared to the overall population of KPNW, a greater proportion of the population identified as due for screening has more females, Black, White, and enrolled in Medicare (Table 4.9). There are fewer patients in the screening population who are Asian, Hawaiian or Pacific Islander, Native American or American Indian, patients that list "Unknown" race, Hispanic ethnicity, and have Medicaid insurance.

Table 4.9 Screening Population Comparison

		Total KPNW Population*	CRC Screening Population
	Total	617,073	60,220
Gender	Female	51.8%	52.3%
Race	Asian	6.1%	5.3%
	Black	3.4%	4.3%
	Hawaiian/ Pacific Islander	1.0%	0.1%
	Native American/ American Indian	0.8%	0.1%
	Other	0.6%	0.1%
	Unknown	22.7%	10.8%
	White	65.5%	77.3%
Ethnicity	Hispanic	8.4%	5.4%
Insurance	Medicare	18.4%	26.3%
	Medicaid	10.7%	4.3%

*Results as of 2/28/21 (closest date to data pull on 4/2/21)

The screening population is 52.3% female, 5.3% Asian, 4.3% Black, 0.1% Hawaiian or Pacific Islander, Native America or American Indian, or identify as “other” race, 10.8% have “unknown” race, 77.3% White, 5.4% Hispanic, and 26.3% are on Medicare, and 4.3% are on Medicaid.

A comparison analysis was conducted to assure that the model was not only applicable to the Northwest demographic population (insured and White patients). Subpopulations considered for a comparison analysis were the Medicaid and non-White groups. The Medicaid patient subgroup (n=2613) was determined to be too small to handle the multilevel analysis. Therefore, the *non-White* group (n=13,655) is the subpopulation for comparative analysis.

The *Data Quality* section in Aim 1 above describes the potential predictors considered in the analysis. As this is a retrospective inception cohort, all predictors were identified in clinical records closest to but prior to the patient’s birthday in 2018. Predictors were defined and assessed in the same way for all participants. Predictor assessments were made without knowledge of the outcome data. All predictors were available at the time of analysis and are expected to be available if the model is used in clinical practice.

The analyst at KPNW identified predictors for all patients without consideration of the outcomes. Predictors were identified at the individual, interpersonal, organizational, community and policy levels. Fewer known predictors are at the upper levels. About two-thirds of the proposed data at the multi-levels were determined

usable for the models. More proposed data was retained at the organizational and community levels (83.3% and 82.9% respectively), and the least amount of data was retained at the policy and interpersonal levels (14.3% and 38.5% respectively).

Table 4.7 describes 7 potential predictors which were omitted from analysis due to missingness. Imputation was performed when possible. In the individual characteristics, *BMI* was imputed for missing values using single imputation. Biological *sex* had 15 missing values (0.02%). Language is not an imputable variable, so patients missing *language* were grouped with the most common category, English. At the interpersonal level, provider was missing for 404 patients (0.67%). At the organizational level, clinic assignment was missing for 159 patients (0.26%). At the community level, county links were missing for 39 patients (0.06%). At the policy level, CCO information was missing for anyone who did not live in Oregon (26.3%), so those variables were removed from the analysis. Patients with missing data were removed from the analysis, as the models fit a population with no missing data.

Correlations of predictors were assessed analytically using correlation and cross tabulations. While clear correlations were assessed in the *Data Quality* assessment in Aim 1, the analytical correlation was assessed in STATA. I suspected that *Dental visits* and *insurance group* were colinear due to the unavailability of dental insurance to Medicare patients, yet 27.6% of Medicare patients had a dental visit. *Insurance group* was also suspected to be highly colinear with age, yet 8.7% of Medicare patients are less than 65 years old, 11.4% of patients aged 65-70 had Medicaid, and commercial

insurance was more evenly distributed across the age groups. The correlation of the *Hispanic* and *interpreter variables* was checked, and only 38% of Hispanic patients also used an interpreter. *Race* and *Hispanic* variables were checked, and 17.2% of Hispanic patients identify as White, and 75.4% as unknown race.

The o*utpatient visits* variable was assessed for correlation with the *missed visits*, *Charlson comorbidities*, and *influenza shots*. The *missed visits* variable was correlated with the *outpatient visits* variable, in that 50.9% of those with 2+ missed visits had 10+ outpatient visits, and 27.2% of the patients with no missed visits had no outpatient visits. Similarly, patients with no outpatient visits also had zero comorbidities in the Charlson score, and 45.1% of those with 2+ comorbidities also had 10+ visits. Having a prior influenza vaccination was also correlated in that 32.4% of those with no outpatient visits also had no flu vaccination. The correlation of *missed visits* and *Charlson comorbidities* was checked, and a significant correlation was not found; only 34.9% of the patients with 2+ comorbidities had 2+ missed appointments.

Individual Level Full Model Specification

For the individual and multilevel models, full models of patients with all potential predictors and complete data for the model were fit. The associations between the predictors and screening were assessed by using full logistic regression analysis and bootstrapping (random sampling). The analyses of the models, step-down methods, and performance of the model using bootstrapping and Integrated Calibration Index (ICI)

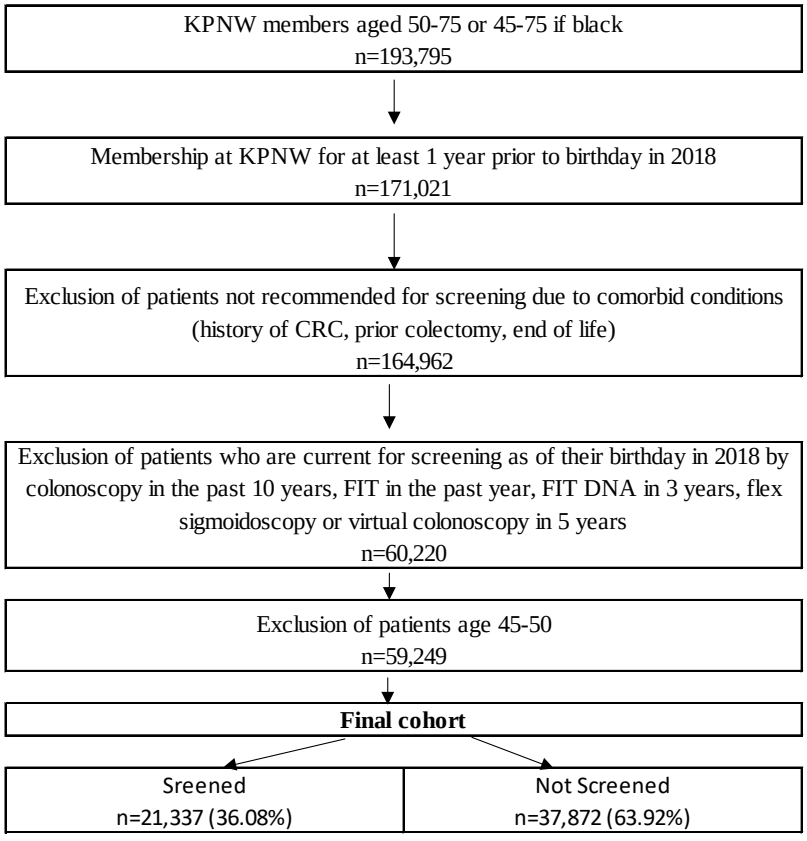
were validated using Harrell's code in SAS and R (Austin & Steyerberg, 2019; Steyerberg, 2019).

The dependent variable is the presence of screening (Table 4.8). Variables with more than two categories were modeled as dummy variables. Age was grouped into 5 categories with 5-year age bands (i.e., 45-49, 50-55). Patients who speak languages other than English or Spanish were grouped as "other". *Dental visits* were categorized into 3 levels if the patient did not have a dental plan if they had a plan but no visit, or if they had a dental visit. Outpatient visits were grouped linear, 0-9 visits and then 10 or more visits.

Individual Level Model (EHR data)

The full population model was fit for 99.98% of the population (n=60,205), 15 patients were missing sex, and therefore eliminated from the analysis. The first run of the full model identified a critical error in the patient eligibility criteria. The 45-49 age group is only patients who are black. The KPNW guidelines for screening were changed in 2018 to include black patients aged 45-49. The model interpretation was difficult because this age group was not representative of the remaining population of patients included in the other age groups, and the applicability of the model was compromised. The complicated nature of having an increased risk of CRC, with the group only being comprised of black patients aged 45-49 limited the model's interpretability. Therefore, due to the complexity and difference in characteristics of the age 45-49-year-old black patients, they were removed from

Figure 4.4 Eligible Patient Consort Diagram (45-50



analysis for this and all subsequent models. The full model was then rerun without the 45-49 age group. Figure 4.4 shows the eligibility of the population without the 45-49 age group. The model without the 45-49 aged patients (n=971) included 98.97% of the population (n=59,234) as 15 patients were missing *sex* and removed from the analysis. The screening outcome was completed for 36.1% of the population (n=21,337).

The variables in the full model include all 21 remaining individual characteristics (Table 4.10). Two variables, prostate screening, and mammogram screening were combined with sex for simplicity; women with mammogram screening and men with prostate screening were combined to eliminate coding for each of the screening variables as none for the gender where the screening does not apply. This leaves 19 remaining variables in the final full model. The performance of the full was adequate with a C-statistic of 0.7239 and an R^2 of 0.1116.

The step-down process was then used to simplify the model, based on Akaike's information criteria (AIC) and the change in R^2 (Steyerberg, 2019). First, the AIC was determined for each variable to understand the contribution to the model. The AIC was ranked from lowest to highest, and the least contributory variables were removed one by one in AIC order until the model R^2 dropped to no lower than 0.1105 (99%). *Insurance group, language group, interpreter, BMI, inpatient visits,* were all removed from the model while retaining 99% of the predictive value.

The 14 retained characteristics include *prior CRC screening and preference, age group, prior preventive screening and sex,* enrollment in KP.org (*patient portal*)*, dental*

membership and visits, number of outpatient visits, number of missed appointments, prior influenza vaccination, Charlson comorbidity score, race, Hispanic ethnicity, membership for 5 years or more, membership for less than 2 years, and *substance abuse.* Table 4.11 shows the results of the multivariate logistic regression analysis of predictors of CRC Screening in the data models. The final reduced model R^2 was 0.1108 and has 38 degrees of freedom, the reduced C-statistic was 0.7232, and the ICI was 0.0134 (Table 4.12).

The performance measures used for evaluation were bootstrapped C and R^2 and calibration. The model was validated internally using bootstrapping (500 bootstraps), which showed adequate performance with a bootstrap corrected C-statistic of 0.722 (Table 4.12). The calibration was also determined by plotting the observed and predicted risk of the reduced model (Figure 4.5), and by calculating the ICI, which shows excellent calibration (0.013). The calibration of the observed and predicted risk appears to be sufficient, with predictive accuracy at all levels.

Individual Level Subpopulation Model

The subgroup model was fit for 99.94% of the population (n=12,676), 8 patients were missing sex and eliminated from the model. Patients aged 45-49 years were removed from this model as they were in the full population. The screening outcome was completed for 31.0% of the 50-75 aged population (n=3,933).

The final variables included in the full population, reduced, individual level model were applied to the subpopulation. The subpopulation model is slightly improved from the full population with an R^2 of 0.1364, and the C-statistic of 0.7490 (Table 4.12).

Table 4.11 shows the results of the multivariate logistic regression analysis of predictors of CRC Screening in the individual level data model for the full population and the recalibration of the model for the *non-White* subgroup analysis. The table shows the observed odds ratio for each variable, the p-value, and 95% confidence interval. The model performance is also demonstrated in the calibration plot (Figure 4.6), where observed and predicted risks appear to agree, with wider variation at the 7^{th} and 10^{th} decile. The ICI is 0.0183 showing greater variance between the observed and predicted risk, and the calibration at the lower deciles is sufficient for use if seeking to identify patients unlikely to screen.

Table 4.10 Individual Level (EHR) Characteristics

Individual Characteristic	patients, no. (%) Without screening n=38,516		With screening n=21,704		Individual Characteristic	patients, no. (%) Without screening n=38,516		With screening n=21,704	
Age in years (mean (SE))					Dental visit in prior year				
45-50	***644***	***(1.7%)***	***327***	***(1.5%)***	No dental plan	24414	(63.4%)	12525	(57.7%)
50-55	10779	(28.0%)	6870	(31.7%)	Plan, no visit	7584	(19.7%)	3571	(16.5%)
55-60	9443	(24.5%)	4340	(20.0%)	Visit	6518	(16.9%)	5608	(25.8%)
60-65	7909	(20.5%)	4205	(19.4%)	Out-patient Visits				
65-70	5667	(14.7%)	3322	(15.3%)	0	10103	(26.2%)	3427	(15.8%)
70+	4074	(10.6%)	2640	(12.2%)	1	5632	(14.6%)	3124	(14.4%)
Ethnicity (Hispanic)	1989	(5.2%)	1261	(5.8%)	2	4428	(11.5%)	2691	(12.4%)
Insurance					3	3332	(8.7%)	2179	(10.0%)
Medicaid	1855	(4.8%)	758	(3.5%)	4	2525	(6.6%)	1612	(7.4%)
Medicare	9520	(24.7%)	5958	(27.5%)	5	2038	(5.3%)	1313	(6.0%)
Commercial	27141	(70.5%)	14988	(69.1%)	6	1534	(4.0%)	1074	(4.9%)
Interpreter needed	1835	(4.8%)	1053	(4.9%)	7	1287	(3.3%)	911	(4.2%)
Language (English)					8	1039	(2.7%)	757	(3.5%)
English	36358	(94.4%)	20553	(94.7%)	9	878	(2.3%)	584	(2.7%)
Spanish	915	(2.4%)	539	(2.5%)	10+	5720	(14.9%)	4032	(18.6%)
Other	1243	(3.2%)	612	(2.8%)	In-patient visits in prior year >0	1672	(4.3%)	792	(3.6%)
Race (White)					Missed a visit in past year				
White	29121	(75.6%)	17444	(80.4%)	0	30597	(79.4%)	17397	(80.2%)
Asian	1869	(4.9%)	1299	(6.0%)	1	4706	(12.2%)	2818	(13.0%)
Black	1664	(4.3%)	920	(4.2%)	2+	3213	(8.3%)	1489	(6.9%)
Hawaiian/Pacific Islander	320	(0.8%)	176	(0.8%)	Membership >5 years	19137	(49.7%)	13371	(61.6%)
American Indian	369	(1.0%)	198	(0.9%)	Membership <2 years	7742	(20.1%)	3133	(14.4%)
Other	234	(0.6%)	107	(0.5%)	Patient portal enrollment (kp.org)	26527	(68.9%)	17615	(81.2%)
Unknown	4939	(12.8%)	1560	(7.2%)	Comorbidity Score >0				
Sex					0	28586	(74.2%)	15741	(72.5%)
Male+ prostate screening	4622	(12.0%)	3069	(14.1%)	1	5143	(13.4%)	3324	(15.3%)
Male + no prostate screening	13998	(36.3%)	6994	(32.2%)	2+	4787	(12.4%)	2639	(12.2%)
Female + mammogram	14523	(37.7%)	10535	(48.5%)	Tobacco or other substance	12286	(31.9%)	7248	(33.4%)
Female no mammogram	5364	(13.9%)	1100	(5.1%)	Flu shot in prior year	17230	(44.7%)	13099	(60.4%)
BMI*					Prior Screening				
Underweight	292	(0.8%)	127	(0.6%)	No prior screening	26920	(69.9%)	8756	(40.3%)
Normal	6772	(17.6%)	4286	(19.7%)	Prior FIT	10570	(27.4%)	11886	(54.8%)
Overweight	15562	(40.4%)	7827	(36.1%)	Prior Colonoscopy	1026	(2.7%)	1062	(4.9%)
Obese	15890	(41.3%)	9464	(43.6%)					

Table 4.11. Multivariate Logistic Regression Analysis of Predictors of CRC Screening Individual Level

Characteristic	Full population (n=59,234) Standard EHR Data			Subpopulation (Non-white) (n=12,676) Standard EHR Data		
	Odds ratio	(95% CI)	*p* value	Odds ratio	(95% CI)	*p* value
Individual Level Characteristic						
Age in years (mean (SE))						
50-54	ref			ref		
55-59	0.471	(0.446 , 0.496)	0.000	0.469	(0.418 , 0.526)	0.000
60-64	0.531	(0.503 , 0.562)	0.000	0.563	(0.499 , 0.635)	0.000
65-69	0.534	(0.502 , 0.567)	0.000	0.567	(0.488 , 0.658)	0.000
70-75	0.524	(0.490 , 0.560)	0.000	0.513	(0.425 , 0.619)	0.000
Ethnicity (Hispanic)	1.431	(1.296 , 1.579)	0.000	1.538	(1.357 , 1.742)	0.000
Race						
White	ref					
Asian	1.202	(1.110 , 1.301)	0.000	ref		
Black	1.054	(0.943 , 1.178)	0.355	0.850	(0.740 , 0.976)	0.021
Hawaiian/Pacific Islander	0.919	(0.753 , 1.120)	0.401	0.746	(0.603 , 0.922)	0.007
American Indian	0.857	(0.712 , 1.030)	0.101	0.594	(0.568 , 0.850)	0.000
Other	0.774	(0.598 , 1.003)	0.053	0.627	(0.477 , 0.825)	0.001
Unknown	0.720	(0.662 , 0.782)	0.000	0.570	(0.505 , 0.643)	0.000
Sex						
Male+ prostate screening	ref			ref		
Male + no prostate screening	1.020	(0.961 , 1.083)	0.515	0.929	(0.803 , 1.074)	0.318
Female + mammogram	1.041	(0.984 , 1.101)	0.164	0.983	(0.853 , 1.132)	0.809
Female no mammogram	0.583	(0.534 , 0.636)	0.000	0.540	(0.444 , 0.656)	0.000
Dental visit in prior year						
No dental plan	ref			ref		
Plan, no visit	0.886	(0.844 , 0.931)	0.000	0.943	(0.845 , 1.053)	0.297
Visit	0.131	(1.251 , 1.372)	0.000	1.306	(1.169 , 1.458)	0.000
Out-patient Visits						
0	ref			ref		
1	1.371	(1.285 , 1.463)	0.000	1.523	(1.326 , 1.750)	0.000
2	1.448	(1.351 , 1.552)	0.000	1.647	(1.414 , 1.918)	0.000
3	1.541	(1.429 , 1.662)	0.000	1.970	(1.666 , 2.329)	0.000
4	1.523	(1.402 , 1.656)	0.000	1.804	(1.493 , 2.181)	0.000
5	1.532	(1.399 , 1.676)	0.000	1.923	(1.562 , 2.366)	0.000
6	1.682	(1.524 , 1.856)	0.000	1.970	(1.568 , 2.474)	0.000
7	1.688	(1.520 , 1.875)	0.000	1.888	(1.475 , 2.417)	0.000
8	1.738	(1.551 , 1.947)	0.000	2.043	(1.552 , 2.688)	0.000
9	1.567	(1.383 , 1.776)	0.000	1.769	(1.230 , 2.414)	0.000
10+	1.830	(1.702 , 1.968)	0.000	2.073	(1.741 , 2.468)	0.000
Missed a visit in past year						
0	ref			ref		
1	0.852	(0.806 , 0.902)	0.000	0.887	(0.779 , 1.009)	0.068
2+	0.639	(0.593 , 0.688)	0.000	0.613	(0.520 , 0.723)	0.000
Membership >5 years	1.181	(1.130 , 1.234)	0.000	1.032	(0.932 , 1.142)	0.549
Membership <2 years	0.792	(0.747 , 0.839)	0.000	0.792	(0.703 , 0.893)	0.000
Patient portal enrollment	1.402	(1.339 , 1.468)	0.000	1.212	(1.110 , 1.325)	0.000
Comorbidity Score						
0	ref			ref		
1	0.904	(0.856 , 0.955)	0.000	0.973	(0.857 , 1.105)	0.672
2+	0.701	(0.659 , 0.746)	0.000	0.691	(0.593 , 0.805)	0.000
Flu shot in prior year	1.296	(1.246 , 1.349)	0.000	1.315	(1.198 , 1.442)	0.000
Tobacco or other substance	0.879	(0.843 , 0.915)	0.000	0.985	(0.887 , 1.094)	0.778
Prior Screening						
No prior screening	ref			ref		
Prior FIT	3.815	(3.659 , 3.978)	0.000	3.640	(3.302 , 4.012)	0.000
Prior Colonoscopy	3.231	(2.935 , 3.555)	0.000	3.272	(2.476 , 4.324)	0.000

Table 4.12 Performance Statistics for Individual Level Prediction Models

	Individual Level	
Statistic	**Full Population**	**Subpopulation (non-White)**
Number of observations	59,234 (0.02% missing data, n=15)	12,676 (0.06% missing data, n=8)
C-statistic	0.7232	0.7501
Bootstrap-corrected C-statistic	0.7220	0.7457
R^2	0.1108	0.1364
Integrated calibration index (ICI)	0.0134	0.0183

Figure 4.5 Calibration of Full Population Individual Level Model

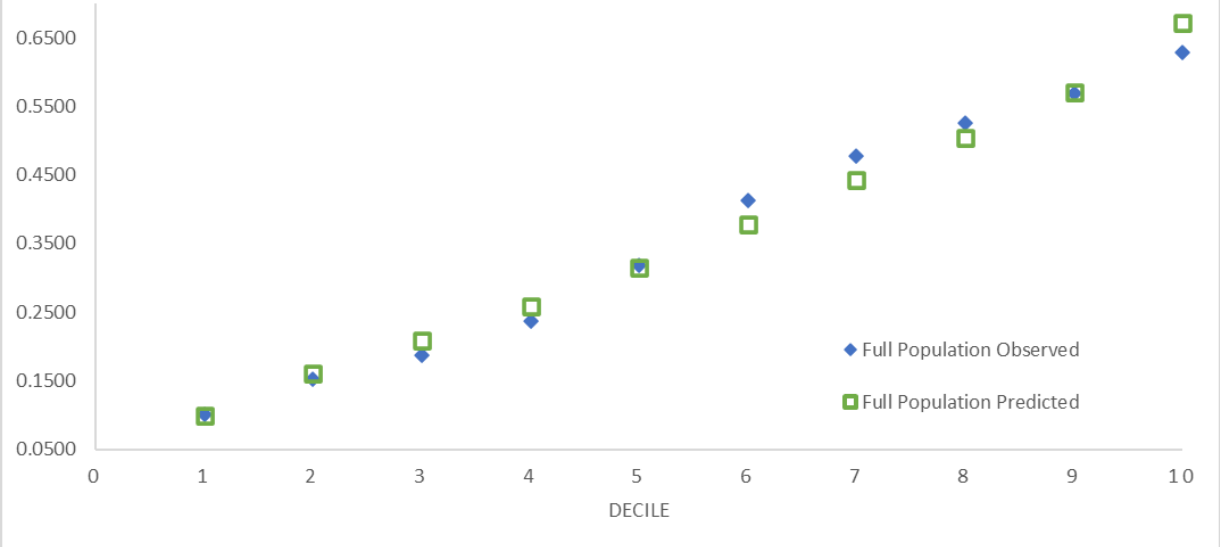
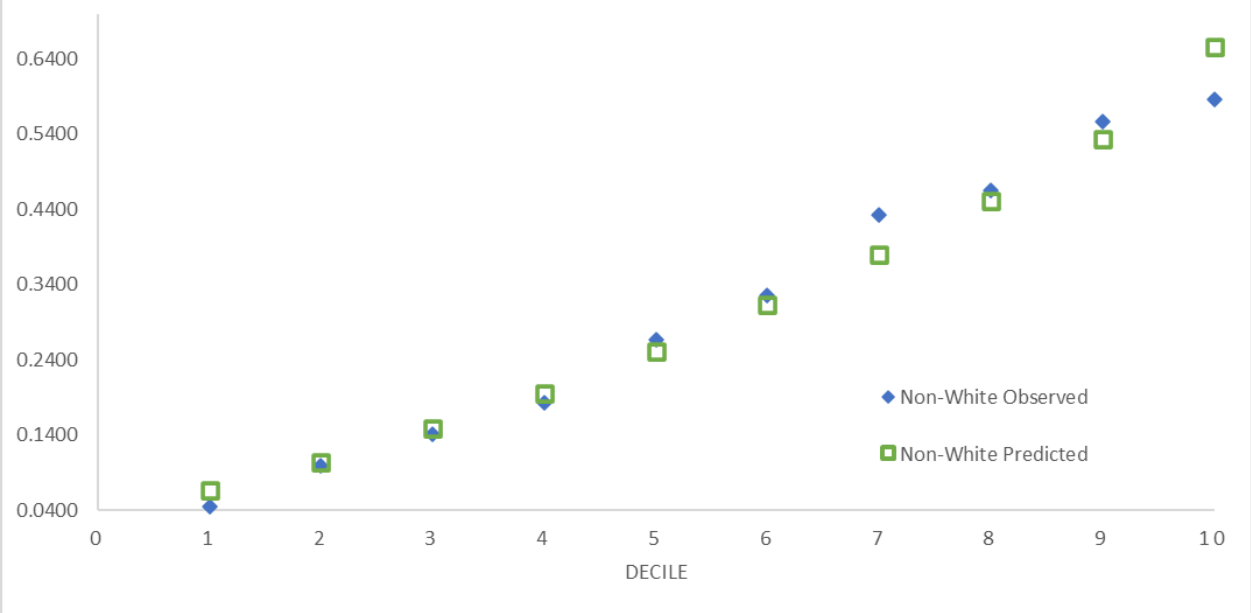


Figure 4.6 Calibration of the Nonwhite Population Individual Level Model



Multilevel Model

The multilevel model incorporates data outside of the individual level based on the SEM Framework. Tables 4.13-4.16 show the distribution of the interpersonal, organizational, community, and policy characteristics retained in the reduced models by the screened and non-screened populations.

Table 4.13 Interpersonal Level Characteristics of Patients

Multi-level Characteristic	Without screening n=38,516		With screening n=21,704	
Interpersonal Level				
Address changes				
0	28052	(72.8%)	16369	(75.4%)
1	6549	(17.0%)	3443	(15.9%)
2+	3885	(10.1%)	1888	(8.7%)
Provider panel size (large)				
<1000	4264	(11.1%)	1863	(8.6%)
1000-1499	11941	(31.0%)	6397	(29.5%)
1500-1999	17686	(45.9%)	10961	(50.5%)
2000+	4267	(11.1%)	2437	(11.2%)
Provider specialty				
Family Practice	22468	(58.3%)	12203	(56.2%)
Internal Medicine	13300	(34.5%)	7880	(36.3%)
Other (NP/PA)	2390	(6.2%)	1575	(7.3%)
Provider gender (male)	21337	(55.4%)	11076	(51.0%)
Provider gender match	25453	(66.1%)	14980	(69.0%)

Table 4.14 Organizational Level Characteristics of Patients

Multi-level Characteristic	patients, no. (%) Without screening		With screening	
Organizational Level				
Clinic assignment				
Battleground	421	(1.1%)	269	(1.2%)
Beaverton	2448	(6.4%)	1293	(6.0%)
Cascade	2137	(5.5%)	1322	(6.1%)
Eugene	72	(0.2%)	30	(0.1%)
Gateway	2149	(5.6%)	1140	(5.3%)
Hillsboro	583	(1.5%)	316	(1.5%)
Interstate	4192	(10.9%)	2325	(10.7%)
Keizer Station	872	(2.3%)	505	(2.3%)
Longview/Kelso	2388	(6.2%)	1448	(6.7%)
Mt. Scott	5155	(13.4%)	2985	(13.8%)
Murray Hill	925	(2.4%)	537	(2.5%)
North Lancaster	1664	(4.3%)	905	(4.2%)
Orchards	1894	(4.9%)	1103	(5.1%)
Rockwood	3218	(8.4%)	1732	(8.0%)
Salmon Creek	2638	(6.8%)	1596	(7.4%)
Skyline	1512	(3.9%)	831	(3.8%)
Sunset	2833	(7.4%)	1584	(7.3%)
Tualatin	2555	(6.6%)	1322	(6.1%)
West Salem	709	(1.8%)	453	(2.1%)
Wait time				
<=30 days	6974	(18.1%)	3887	(17.9%)
31-45 days	21765	(56.5%)	12034	(55.4%)
45+ days	9602	(24.9%)	5772	(26.6%)
Clinic size				
<10,000	11399	(29.6%)	6218	(28.6%)
30,000+	27117	(70.4%)	15486	(71.4%)

Table 4.15 Community Level Characteristics of Patients

Multi-level Characteristic	patients, no. (%) Without screening		With screening	
Community Level				
Education				
High school or less	2900	(7.5%)	1460	(6.7%)
Some college	24023	(62.4%)	13511	(62.3%)
Degree	11561	(30.0%)	6726	(31.0%)
Median household income				
<45K	1618	(4.2%)	789	(3.6%)
45-85K	20889	(54.2%)	11365	(52.4%)
85-140K	14297	(37.1%)	8467	(39.0%)
140K+	1654	(4.3%)	1063	(4.9%)
Neighborhood households below poverty (>10%)	10645	(27.6%)	5547	(25.6%)
Neighborhood deprivation index (>0 = deprivation)	11741	(30.5%)	6175	(28.5%)
Limited English proficiency (>5%)	3518	(9.1%)	1966	(9.1%)
Percent noncitizens (>5%)	21622	(56.1%)	11971	(55.2%)
Total population (>500K)	16706	(43.4%)	9123	(42.0%)
Median age (30's=0, 40+=1)	8858	(23.0%)	5016	(23.1%)
Age 45-64 (>25%)	23570	(61.2%)	13424	(61.9%)
Age 65+ (>15%)	10752	(27.9%)	6125	(28.2%)
Percent Hispanic (>15%)	10631	(27.6%)	5825	(26.8%)
Percent white (>85%)	18288	(47.5%)	10619	(48.9%)
Health professional shortage (whole county)	1013	(2.6%)	554	(2.6%)
Number of healthcare facilities (>15/100K)	10239	(26.6%)	5632	(25.9%)
Cardiovascular death rate (>200/100K)	4340	(11.3%)	2412	(11.1%)
Premature death rate (>300/100K)	18163	(47.2%)	10048	(46.3%)
Income inequality GINI (>0.45)	10378	(26.9%)	5694	(26.2%)
Quality of life (% w/ poor health >15%)	5987	(15.5%)	3338	(15.4%)
Poor mental health days (<4)	13557	(35.2%)	7808	(36.0%)
Excessive drinking (>20%)	15674	(40.7%)	8751	(40.3%)
Number of primary care physicians (>100/100K)	10325	(26.8%)	5675	(26.1%)
Air pollution (>9 average daily PM2.5)	8949	(23.2%)	5061	(23.3%)
Preventable hospital stays (>3000/100K)	24741	(64.2%)	14146	(65.2%)
Sexually transmitted infections (>500/100K)	10301	(26.7%)	5657	(26.1%)
Rural (rural residents >15%)	10780	(28.0%)	6129	(28.2%)
Drinking water violations (yes)	19917	(51.7%)	11019	(50.8%)
Violent crime (>400/100K)	10193	(26.5%)	5621	(25.9%)
Physical inactivity (>20%)	4428	(11.5%)	2464	(11.4%)
Food environment index (>8)	24017	(62.4%)	13758	(63.4%)

Table 4.16 Policy Level Characteristics of Patients

Multi-level Characteristic	patients, no. (%) Without screening		With screening	
Policy Level				
State				
Oregon	28557	(74.1%)	15865	(73.1%)
Washington	9488	(24.6%)	5667	(26.1%)
Other	471	(1.2%)	172	(0.8%)

The full population model was fit for the multilevel data in the same way it was fit for the individual level model described above. The full multilevel model was fit for the population removing the 45-49 year old patients due to the limited applicability of that unique population. Figure 4.4 shows the eligibility of the population without the 45-49 age group (n=59,249). The starting population was identical to the individual model, without the 45-49 aged patients (n=971 removed). Patients were also removed from the full model due to missing data (n=1,194). Missing data included patients who were missing *sex* (n=15), interpersonal level data (i.e., *address changes* n=186), provider level data (n=415), clinic level data (n=159), and miscellaneous information at the county level (i.e., *facilities* was missing for n=834 patients). The remaining population in the full model is 98.0% of the full population (n=58,040). The screening outcome was completed for 36.3% of the population in the model (n=21,068).

The 57 variables included in the full model include the individual characteristics described above (n=19, Table 4.10), interpersonal characteristics (n=5, Table 4.13), organizational characteristics (n=3, Table 4.14), community characteristics (n=29, Table 4.15), and policy level characteristic (n=1, Table 4.16). Collinearity was assessed among

the variables in the multilevel model. At the organizational level, the *patient to provider ratio* variable was colinear with clinic size and *clinic location (city)* was redundant with *primary care clinic;* the program removed these variables from the model.

The performance of the full model was adequate with a C-statistic of 0.7251 and an R^2 of 0.1129. The model used 110 degrees of freedom. The step-down process was then used to simplify the model, based on Akaike's information criteria (AIC) and the change in R^2 (Steyerberg, 2019). First, the AIC was determined for each variable to understand the contribution to the model. The AIC was ranked from lowest to highest, and the least contributory variables were removed one by one in AIC order until the model R^2 dropped to no lower than 0.1118 (99%).

Retained from the individual level were 23 variables, all the variables retained in the reduced individual level model and *inpatient visits,* and *BMI*. The complete individual variable list is *prior CRC screening and preference, age group, prior preventive screening and sex,* enrollment in KP.org (*patient portal*)*, dental membership and visits, number of outpatient visits, number of missed appointments, prior influenza vaccination, Charlson comorbidity score, race, Hispanic ethnicity, membership for 5 years or more, membership for less than 2 years, substance abuse, inpatient visits, Hispanic ethnicity,* and *BMI.*

Variables retained from the interpersonal level were *address changes* and *provider gender match.* From the organizational level, *wait time, clinic size,* and *primary clinic* were retained. From the community level, *facilities* (number of healthcare

facilities>15/100K residents) and *median family income* were retained. From the policy level, the only variable (*STATE*) was not retained.

Table 4.17 shows the results of the multivariate logistic regression analysis of predictors of CRC Screening in the multilevel data models. The final reduced model R^2 was 0.1119 and has 70 degrees of freedom, the reduced C-statistic was 0.7242, and the ICI is 0.0130 (Table 4.18).

Bootstrapping and calibration were used to evaluate the performance of the model. The model was validated internally using bootstrapping (500 bootstraps), which showed adequate performance with a bootstrap corrected C-statistic of 0.7218 (Table 4.18). The calibration was also determined by plotting the observed and predicted risk of the reduced model (Figure 4.19).

Multilevel Subpopulation Model

The subgroup model without 45-49 year old patients was fit for 96.05% of the population (n=12,184). Patients were missing data (n=500) as described above in the full multilevel model, including missing provider, clinic, and community level data. The screening outcome was completed for 31.5% of the population (n=3,838).

The final variables included in the full population, reduced, multilevel model were applied to the subpopulation of non-White patients. The subpopulation model is slightly improved from the full population with an R^2 of 0.1369, and the C-statistic of 0.7484, and the ICI is 1.0206 (Table 4.18).

Table 4.17 shows the results of the multivariate logistic regression analysis of predictors of CRC Screening in the multilevel data model for the full population and the *non-White* subgroup analysis. The table shows the observed odds ratio for each variable, the p-value, and 95% confidence interval. The model performance is also demonstrated in the calibration plot (Figure 4.20), where observed and predicted risks appear to agree.

Table 4.17 Logistic Regression Analysis of Predictors of CRC Screening Multilevel Model

Characteristic	Full Population + Multilevel Data			Subpopulation + Multilevel Data		
	Odds ratio	(95% CI)	*p* value	Odds ratio	(95% CI)	*p* value
Individual Level Characteristic						
Age in years (mean (SE))						
50-54	ref			ref		
55-59	0.466	(0.442 , 0.492)	0.000	0.466	(0.415 , 0.524)	0.000
60-64	0.532	(0.503 , 0.562)	0.000	0.578	(0.511 , 0.654)	0.000
65-69	0.533	(0.501 , 0.566)	0.000	0.577	(0.495 , 0.672)	0.000
70-75	0.518	(0.484 , 0.555)	0.000	0.524	(0.432 , 0.634)	0.000
Ethnicity (Hispanic)	1.396	(1.261 , 1.545)	0.000	1.418	(1.244 , 1.616)	0.000
Race (White)						
White	ref					
Asian	1.185	(1.092 , 1.287)	0.000	ref		
Black	1.093	(0.975 , 1.226)	0.128	0.842	(0.721 , 0.984)	0.030
Hawaiian/Pacific Islander	0.907	(0.742 , 1.109)	0.340	0.713	(0.572 , 0.888)	0.003
American Indian	0.849	(0.705 , 1.024)	0.086	0.670	(0.543 , 0.826)	0.000
Other	0.804	(0.619 , 1.045)	0.103	0.648	(0.490 , 0.856)	0.002
Unknown	0.735	(0.674 , 0.801)	0.000	0.572	(0.502 , 0.653)	0.000
Sex						
Male+ prostate screening	ref			ref		
Male + no prostate screening	1.021	(0.961 , 1.085)	0.495	0.932	(0.804 , 1.080)	0.350
Female + mammogram	1.042	(0.984 , 1.104)	0.159	0.980	(0.849 , 1.132)	0.783
Female no mammogram	0.591	(0.541 , 0.646)	0.000	0.543	(0.444 , 0.665)	0.000
BMI						
Underweight	ref			ref		
Normal	1.248	(0.991 , 1.571)	0.600	1.292	(0.760 , 2.197)	0.344
Overweight	1.217	(0.967 , 1.530)	0.940	1.295	(0.762 , 2.200)	0.340
Obese	1.151	(0.915 , 1.448)	0.229	1.412	(0.828 , 2.406)	0.205
Dental visit in prior year						
No dental plan	ref			ref		
Plan, no visit	0.897	(0.854 , 0.943)	0.000	0.953	(0.851 , 1.066)	0.396
Visit	1.308	(1.248 , 1.370)	0.000	1.302	(1.164 , 1.458)	0.000
Out-patient Visits						
0	ref			ref		
1	1.374	(1.287 , 1.462)	0.000	1.531	(1.329 , 1.763)	0.000
2	1.445	(1.347 , 1.550)	0.000	1.670	(1.429 , 1.951)	0.000
3	0.542	(1.429 , 1.665)	0.000	2.020	(1.703 , 2.397)	0.000
4	1.531	(1.407 , 1.666)	0.000	1.856	(1.528 , 2.254)	0.000
5	1.534	(1.400 , 1.680)	0.000	1.967	(1.591 , 2.431)	0.000
6	1.671	(1.513 , 1.846)	0.000	1.973	(1.564 , 2.489)	0.000
7	1.696	(1.525 , 1.886)	0.000	1.941	(1.508 , 2.497)	0.000
8	1.743	(1.553 , 1.956)	0.000	2.041	(1.538 , 2.709)	0.000
9	1.575	(1.388 , 1.786)	0.000	1.814	(1.319 , 2.496)	0.000
10+	1.880	(1.745 , 2.025)	0.000	2.188	(1.826 , 2.622)	0.000
In-patient visits in prior year >0	0.808	(0.731 , 0.893)	0.000	0.663	(0.492 , 0.894)	0.007
Missed a visit in past year						
0	ref			ref		
1	0.858	(0.812 , 0.908)	0.000	0.886	(0.777 , 1.010)	0.070
2+	0.665	(0.616 , 0.718)	0.000	0.627	(0.530 , 0.741)	0.000

Characteristic	Full Population + Multilevel Data			Subpopulation + Multilevel Data		
	Odds ratio	(95% CI)	*p* value	Odds ratio	(95% CI)	*p* value
Membership >5 years	1.174	(1.123 , 1.228)	0.000	1.019	(0.918 , 1.131)	0.724
Membership <2 years	0.793	(0.748 , 0.841)	0.000	0.803	(0.711 , 0.908)	0.000
Patient portal enrollment	1.401	(1.336 , 1.468)	0.000	1.182	(1.079 , 1.296)	0.000
Comorbidity Score						
0	ref			ref		
1	0.916	(0.867 , 0.968)	0.002	0.975	(0.856 , 1.110)	0.699
2+	0.737	(0.691 , 0.785)	0.000	0.707	(0.604 , 0.828)	0.000
Flu shot in prior year	1.297	(1.246 , 1.350)	0.000	1.323	(1.204 , 1.454)	0.000
Tobacco or other substance	0.883	(0.847 , 0.921)	0.000	0.985	(0.885 , 1.096)	0.777
Prior Screening						
No prior screening	ref			ref		
Prior FIT	3.823	(3.664 , 3.988)	0.000	3.603	(3.263 , 3.978)	0.000
Prior Colonoscopy	3.242	(2.943 , 3.571)	0.000	3.294	(2.486 , 4.364)	0.000
Multilevel Data						
Address changes						
0	ref			ref		
1	0.995	(0.946 , 1.047)	0.847	0.973	(0.868 , 1.090)	0.638
2+	0.897	(0.840 , 0.958)	0.001	0.942	(0.819 , 1.084)	0.407
Provider gender match	1.075	(1.033 , 1.118)	0.000	1.0619	(0.969 , 1.164)	0.201
Clinic assignment						
Battleground	ref			ref		
Beaverton	0.430	(0.273 , 0.679)	0.000	0.430	(0.273 , 0.679)	0.000
Cascade	0.499	(0.337 , 0.737)	0.000	0.499	(0.337 , 0.737)	0.000
Eugene	0.741	(0.350 , 1.568)	0.433	0.741	(0.350 , 0.157)	0.433
Gateway	0.834	(0.471 , 1.474)	0.532	0.834	(0.471 , 1.477)	0.532
Hillsboro	1.014	(0.568 , 1.812)	0.962	1.014	(0.568 , 1.813)	0.962
Interstate	0.434	(0.274 , 0.686)	0.000	0.434	(0.275 , 0.686)	0.000
Keizer Station	0.979	(0.552 , 1.735)	0.941	0.799	(0.552 , 1.736)	0.941
Longview/Kelso	0.503	(0.340 , 0.743)	0.001	0.503	(0.340 , 0.743)	0.001
Mt. Scott	0.483	(0.246 , 0.947)	0.034	0.483	(0.247 , 0.947)	0.034
Murray Hill	1.054	(0.594 , 1.869)	0.857	1.054	(0.594 , 1.869)	0.857
North Lancaster	1.015	(0.575 , 1.789)	0.960	1.015	(0.575 , 1.790)	0.960
Orchards	0.901	(0.749 , 1.084)	0.272	0.901	(0.749 , 1.085)	0.272
Rockwood	0.455	(0.288 , 0.721)	0.001	0.455	(0.288 , 0.721)	0.001
Salmon Creek	0.485	(0.328 , 0.715)	0.000	0.485	(0.328 , 0.715)	0.000
Skyline	0.925	(0.517 , 1.654)	0.792	0.925	(0.517 , 1.654)	0.792
Sunset	0.478	(0.304 , 0.754)	0.001	0.478	(0.304 , 0.754)	0.001
Tualatin	0.497	(0.319 , 0.776)	0.002	0.497	(0.319 , 0.776)	0.002
West Salem	1.089	(0.613 , 1.935)	0.770	1.089	(0.613 , 1.936)	0.770
Wait time						
<=30 days	ref			ref		
31-45 days	0.982	(0.594 , 1.621)	0.942	0.982	(0.594 , 1.621)	0.942
45+ days	1.038	(0.599 , 1.799)	0.893	1.038	(0.599 , 1.799)	0.893
Clinic size > 30,000	1.941	(1.374 , 2.742)	0.000	1.374	(1.374 , 2.742)	0.000
Median household income						
<45K	ref			ref		
45-85K	1.052	(0.954 , 1.160)	0.306	1.052	(0.954 , 1.160)	0.306
85-140K	1.223	(1.015 , 1.242)	0.025	1.123	(1.015 , 1.242)	0.025
140K+	1.138	(0.999 , 1.296)	0.052	1.138	(0.999 , 1.296)	0.052
Number of healthcare facilities (>15/100K)	1.088	(1.021 , 1.160)	0.009	1.088	(1.021 , 1.160)	0.009

Table 4.18 Performance Statistics for Multilevel Prediction Models

	Multilevel	
Statistic	**Full Population**	**Subpopulation (non-White)**
Number of observations	58,040 (2% missing data, n=1194)	12,184 (3.94% missing data, n=500)
C-statistic	0.7242	0.7505
Bootstrap-corrected C-statistic	0.7218	0.7384
R^2	0.1119	0.1369
Integrated calibration index (ICI)	0.0130	0.0206

Figure 4.7 Calibration of Full Population Multilevel Model

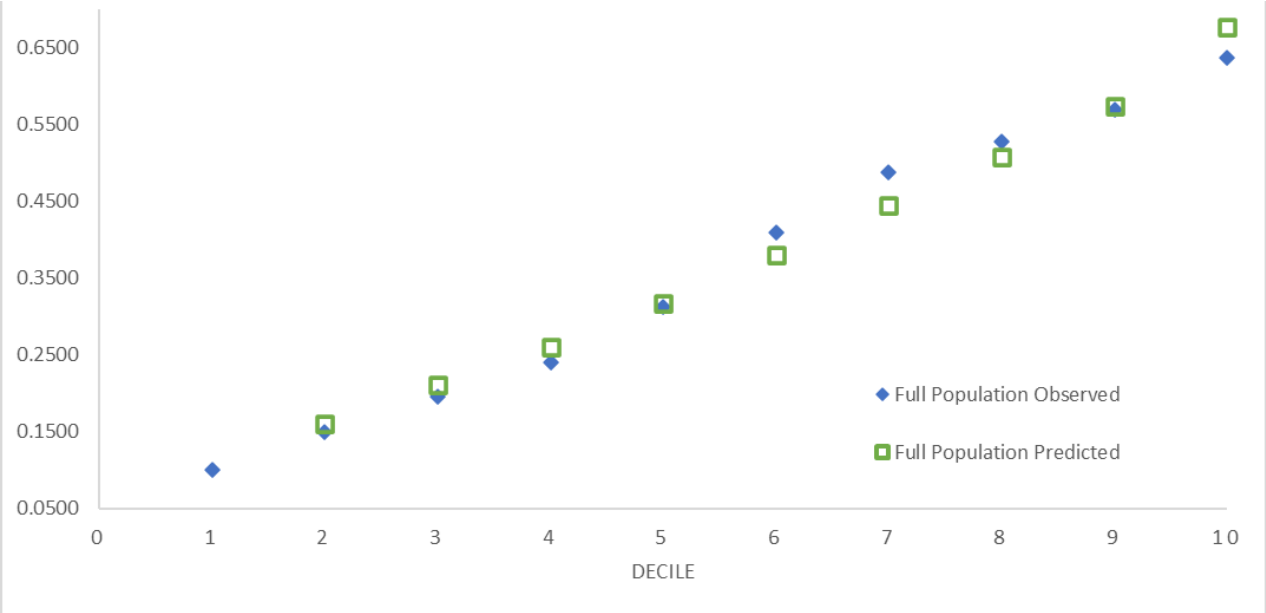
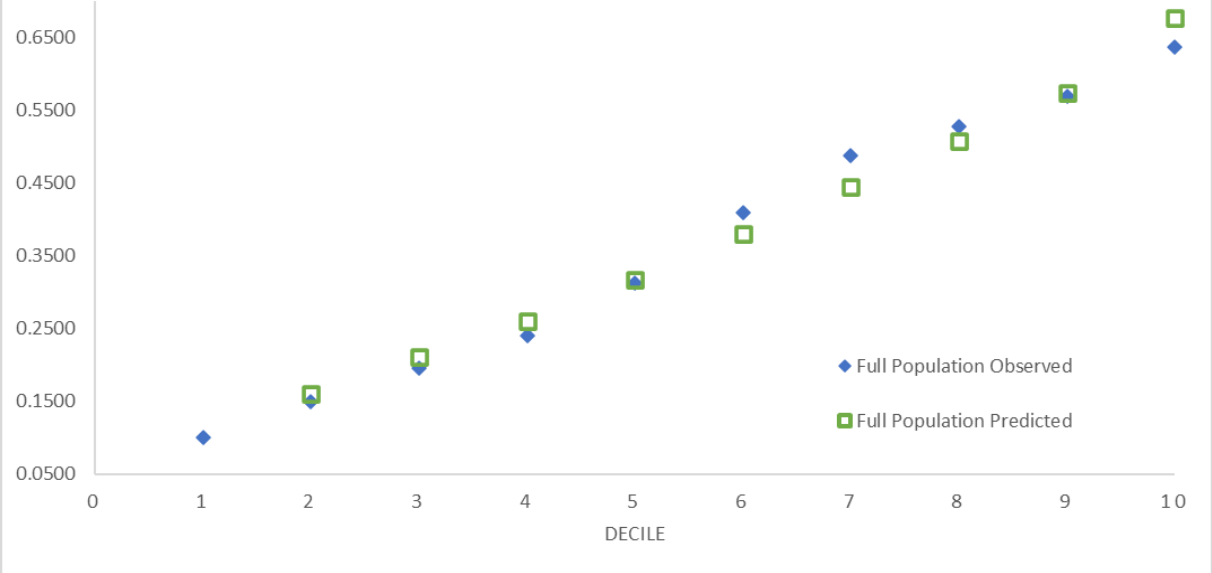


Figure 4.8 Calibration of Nonwhite Population Multilevel Model



Summary of Aim 2 Findings

The individual reduced model to predict CRC screening used 14 variables commonly available in EHR databases (Table 4.21). The multilevel reduced model added components at the individual, interpersonal, organizational, and community levels, but with limited improvement in the performance characteristics. The final reduced multilevel model used 23 variables, seven from the outer levels. Both models showed improved performance after application to a subpopulation (non-White patients), reassuring concerns of applicability when applied to less homogeneous populations. Table 4.22 shows the performance statistics for all models. The model performance is sufficient for clinical use. The performance of the model after adding multilevel data did not improve. The bootstrap-corrected c-statistics are identical; the R^2 statistic is only 1%

higher, and the ICI is 3% better on a relative scale for the multi-level model. The improvement in the calibration (accuracy) shows a minimal improvement, which was discussed with the stakeholders in the qualitative interviews. The reduced multilevel model incorporates the multilevel data, and the application of this model to a subpopulation proved the model to apply to non-White patients, but with reduced calibration (ICI). This risk prediction model can be used to identify patients unlikely to complete screening, for targeted application of interventions to increase CRC screening in a health system.

The performance characteristics and calibration plots show 2 sufficient models for identifying patients' likelihood of screening for colorectal cancer. The multilevel model applies multilevel data, and while some are retained in the final model, the impact of the inclusion of multilevel data is minimally beneficial. Both models perform well when applied to the non-White (minority) subpopulation.

One limitation of these models is that it was developed in an integrated health system in a single regional area. The geographical (community level) and organizational variation limit the impact of multilevel data, where it could be more valuable in a wider and more diverse population.

Table 4.19 Comparison of Individual and Multilevel Models

Individual Model	Multilevel Model	Level
Hispanic	✓	Individual
Substance Abuse	✓	
Member for 5 years	✓	
Member for less than 2 years	✓	
Race	✓	
Charleson comorbidity score	✓	
Missed appointments	✓	
Flu shot last year	✓	
Dental visit in past year	✓	
KP.org - Portal enrollment	✓	
Sex/preventive screeing	✓	
Outpatient Visits	✓	
Age group (5 year groups)	✓	
Test preference group	✓	
	Inpatient visits	
	BMI	
	Address changes	Interpersonal
	Provider gender match	
	Wait time for a colonoscopy	Organizational
	Clinic size	
	Primary care clinic	
	Number of healthcare facilities	Community
	Median household income	

Table 4.20 Performance Statistics for All Prediction Models

	Full Population		Subpopulation (NonWhite)	
Statistic	**Individual Level**	**+ Multilevel Data**	**Individual Level**	**+Multilevel Data**
Number of observations	59,234 (0.02% missing data, n=15)	58,040 (2% missing data, n=1194)	12,676 (0.06% missing data, n=8)	12,184 (3.94% missing data, n=500)
C-statistic	0.7232	0.7242	0.7501	0.7505
Bootstrap-corrected C-statistic	0.7220	0.7218	0.7457	0.7384
R^2	0.1108	0.1119	0.1364	0.1369
Integrated calibration index (ICI)	0.0134	0.0130	0.0183	0.0206

R^2 statistic, represents the model's discrimination and calibration

ICI statistic represents the model's weighted difference between observed and predicted probabilities

Aim 3: Assess Health System Perceptions

The final aim is to assess health system perceptions of predictive analytics, the use of multilevel data, and the model. The usefulness of the developed CRC screening model with added multilevel data was assessed using qualitative research methods, semi-structured interviews. The interview questions were framed around the domains and emerging themes were classified as constructs of the Consolidated Framework for Implementation Research (CFIR).

Figure 4.9 Conceptual Design: Aim 3

Purposive sampling was used to recruit 5 current KPNW staff members working in predictive analytics or gastroenterology (CRC screening). Semi-structured interviews consisted of 17 specific questions regarding their departments, CRC screening, data, predictive analytics, and usability of the model. Through thematic analysis, constructs within the interviews across the domains were identified. All data were collected between June 11 and June 30, 2021.

Qualitative data revealed 9 dominant constructs from the CFIR domains: compatibility, access to knowledge and information, peer pressure, needs and resources, knowledge and beliefs about the intervention, engagement, execution,

relative advantage, and complexity. The qualitative data also revealed that culture, not policy influenced the use of PA.

Inner setting findings include that team and group decisions are made, they all use PA to some extent, and the Quality Team facilitates the use of analytics in their departments. Outer setting questions revealed that many other departments and outside organizations are using PA and that this method of improvement is cost-efficient. The individual level revealed all individuals use data and IT tools in their daily work and have a favorable view on the use of PA. The process construct questions revealed barriers including access to multilevel data and incorporating models into their general workflow and EHR. The Intervention construct revealed high usability of the model overall in the use of the information it provided, and that it could be used in a variety of ways to make their jobs easier. They identified using external data as difficult. The policy questions indicate no formal policies regarding the use of data or analytics, but rather an encouragement to use analytics to improve the quality of care for KPNW patients.

Interview Guide

Table 4.21 outlines the characteristics of the interview guide. The interview guide contains a variety of questions framed by the CFIR framework within each CFIR domain: the inner setting, outer setting, individual, process, and intervention, and

policy. The questions seek to elicit views on the use of predictive analytics, multilevel data, and the model.

Of the questions in the guide, 8 questions addressed the *inner setting*. These questions elicited their views about decision making, increasing screening, and policies about the use of PA in their departments. Two questions addressed the *outer setting*, specifically asking about cost effectiveness and other organizations' use of predictive analytics. The two *process* questions asked about the acquisition of data and barriers and facilitators to the use of PA. The two *individual* questions ask about their personal use of data and IT tools, and their personal views on PA. The three *intervention* questions asked about the usability of the multilevel model. Policy questions were added to capture any relevant policies impacting the use of data or analytics. Across the domains, questions are about the decision making process, CRC screening, data access, predictive analytics, and model usability.

The individual and multilevel models, from Aim 2, were introduced to the participants to capture the usability of the model before Question #9. Reactions to the model were solicited in Questions #9 and #10.

Table 4.21 Interview Guide Characteristics

Interview Question	Domain	Theme	Sample Questions
1	Inner setting	Decision making process	How are decisions made in your department regarding strategies or information used to choose practices or care processes that improve the quality of care?
2	Inner setting	CRC screening	What if anything does your department currently do to increase the rates of colorectal cancer screening?
3	Individual	Data access	Do you use data or IT-driven tools to support increased CRC screening for your patients?
4	Process	Data access	What if the needed data or information came from a separate data source that you would have to connect with?
5	Inner setting	Predictive analytics	Are you aware of any tools your department currently uses that rely on predictive analytics?
6	Individual	Predictive analytics	How do you personally feel about using predictive analytics?
7	Process	Predictive analytics	Can you identify and describe barriers or facilitators to using predictive analytics that you have faced or anticipate facing?
8	Outer setting	Predictive analytics	Are you aware of other departments or organizations using predictive analytics to improve patient care?
9	Intervention	Model usability	Does the model provide information you would find useful in the development or choice of screening practices?
10	Intervention	Model usability	How would you use risk prediction models like the ones I have presented?
11	Outer setting	Predictive analytics	Do you see the use of predictive analytics for CRC as part of a care process as cost-efficient?
12	Intervention	CRC screening	Would the use of a CRC screening tool based on predictive analytics make your job easier?
13	Inner setting	Predictive analytics	In what ways does KPNW create potential barriers or facilitators to the use of predictive analytics in your line of work?
14	Inner setting	Data access	Are you able to ask for the types of data you need or want to do your job?
15	Policy	Predictive analytics	Does your department/KPNW have *policies about access and use of predictive analytics* ?
16	Policy	Data access	Does your department/KPNW have *policies about access to data* ?
17	Policy	CRC screening	Does your department/KPNW have policies about using data to understand trends in screening among subpopulations?

Interview Participants

Purposive sampling was used to identify interview participants. Interview participants were selected who work in colorectal cancer screening at KPNW and were recommended by the manager of the predictive analytics team that oversees quality and the practice director of Gastroenterology who also leads the Prediction Modeling Governance Committee. Participants were sought who worked in a variety of environments, including on the front line, and in the quality department. Further, because KPNW is comprised of two entities participants were sought from both entities,

Kaiser Foundation Hospitals, and Northwest Permanente, which are essentially the delivery system and the medical group respectively. A total of 6 participants were recommended because they are engaged in the day-to-day responsibilities of increasing CRC screening, quality of care improvement, are key opinion leaders, and had baseline knowledge of the use of analytics.

Recommended participants included the Program Manager in Clinical Quality and Population Health, a Gastroenterologist that also works in the Quality Department, the Practice Director of Gastroenterology, the Manager and Analyst of Clinical Information, and the Director of Clinical Quality Systems, and the Practice Director for Regional Surgical Access. Recommended participants were senior level staff, managers, and frontline workers who represented both the medical group and delivery system sides of the organization.

Participation

An email message outlining the purpose of the project and requesting participation was sent to each of 6 potential participants. Interview times were scheduled based on their availability. Once the participant agreed, the consent form and meeting time were sent. One participant was on family leave, the remaining 5 recruited participants agreed to be interviewed. Among the participants, 3 were female (60%), 2 were White (40%). The participants worked in four different departments, one participant worked both as a frontline caregiver and in Quality Improvement. Of the 5

total participants, 4 worked for Northwest Permanente (80%), 2 were frontline caregivers (40%), 2 worked in Quality Improvement (40%) and one worked in operations (20%) (Table 4.22).

Table 4.22 Participant Selection

Setting	Northwest Permanente	Kaiser Foundation Health Plan/ Hospitals
Front Line Caregiver*	2	
Quality Improvement*	2	1
Operations	1	

*One front line caregiver also worked in Quality Improvement

Data Collection

Semi-structured interviews were conducted through Microsoft Teams, recorded, and lasted between 46 and 60 minutes. An audio recording and transcript were generated through Microsoft Teams.

Interviews were conducted in a private office space. Using the semi-structured interview guide, participants were asked about their role and department and then were subsequently asked the remaining questions. Notes and observations about each interview (field notes) were documented during and after each interview, and by question.

Analysis

Audio recordings and field notes were analyzed using thematic analysis to identify which of the CFIR constructs emerged from the data. The CFIR provides a guide for assessing barriers and facilitators of the implementation of an innovation. (Safaeinili et al., 2020) CFIR is organized by five primary domains and 39 constructs within each domain. (Kirk et al., 2016) In this analysis, construct themes were determined by domain, guided by CFIR as they emerged from the interviews. Brief summaries were created for each of the 17 questions for each participant. Each question was accessed for constructs that are similar or redundant across questions within each domain. Quotes were collected within each domain and by question. As new constructs emerged within each domain, the prior interviews were revisited for reassessment.

Across the 5 CFIR domains (inner setting, outer setting, individual, process, and intervention), 9 constructs emerged from the qualitative interviews on the use and implementation of multilevel predictive analytics (Table 4.23). An additional domain, policy, was also assessed and one construct (culture not policy) emerged.

Table 4.23 Domains and Constructs

Interview Question	CFIR Domain	Emerged Constructs
1, 2, 5, 13, 14	Inner setting	Compatability, Access to knowlede and information
8, 11	Outer setting	Peer pressure, Needs and resources
3, 6,	Individual	Knowledge and beliefs about the intervention
4, 7	Process	Engagement, Executing
9, 10, 12	Intervention	Relative advantage, Complexity
15, 16, 17	Policy	Culture not policy

Results by Domain

Inner Setting

Inner Setting questions (1, 2, 5, 13, 14) included questions about the decision making process, CRC screening, predictive analytics, and data access. The major overarching constructs derived from responses to the inner setting questions are the compatibility of the use of analytics in the KPNW system, and the ease of access to knowledge and information.

Question 1 asked about how decisions are made in their departments; all participants said decisions were made by the team or group. Two participants went further to describe the quality department and managers as members of the group or teams. Question 2 asked about current efforts to increase CRC screening. All participants identified the use of endoscopy (colonoscopy) and FIT testing and identified

the quality department as executing the FIT mailing. Question 5 Asked about the departmental use of predictive analytics. All participants identified the current use of predictive analytics and separately listed the use of the National Cancer Institute tool, the Johns Hopkins tool, the use of analytics for tracking the numbers of screening, and prioritization. Question 13 asked about potential barriers or facilitators to the use of predictive analytics at KPNW. Two participants identified a facilitator as the "quality team" and the partnership between NW Permanente and analytics, encouragement to use the quality team to improve efforts.

"They (KPNW) have really embraced predictive analytics and predictive models, both in my department and in our organization. I've enjoyed observing the partnership between our health plan analytics colleagues to create that sort of community of practice around predictive analytics." – Participant #4

One participant identified a barrier to the use of analytics as not having enough resources to consistently staff the analytics department. Another participant identified barriers as the speed of implementation and not being able to study the implementation outcomes. Question 14 asked about the ability to access data to do their job. One participant said that the type of data they get is not always enough, and two participants identified wanting more "of this type of data" referencing the multilevel data in the model.

"I'd love to get this type of data breakdown more often; this model is more sophisticated and could guide our outreach efforts." – Participant #2.

The compatibility of the use of analytics in the KPNW system became evident as the participants consistently expressed support and coordination of analytics across departments. All participants mentioned collaboration and team decision making with the Quality department playing a major role in the use of analytics. There was no variation in CRC screening program descriptions. The consistent use and support of analytics seem to be built into the KPNW culture. The access to knowledge and information construct also consistently emerged when identifying the prioritization of use of analytics and the encouragement to partner with the quality department.

The available resources construct also emerged, in the desire to have more access to resources such as time and analytics. The success of the current use of analytics within KPNW appears to be drawing a need for more analytics, as the value is realized.

Outer Setting

Two questions (#8, #11) asked about the outer setting to Kaiser Permanente. The outer setting questions revealed the constructs of peer pressure and needs and resources for the use of analytics.

Question 8 asked if the participant was aware of other departments or organizations using predictive analytics to improve patient care. Participants #1 and #2 listed other internal departments using analytics for breast and cervical screening, and readmissions. Participant 3 identified the department (within Quality) that executes risk models for other departments. Participant 4 identified other outside organizations that use risk prediction, including KP Washington and Northern California. Participant 5 identified a common score used by similar departments in and out of the organization (surgical).

Question 11 asked if the use of PA for CRC improved cost efficiency. All 5 participants gave examples of where it has been used to improve costs. Examples included the use of analytics in scheduling, as used to identify caseloads for providers, and in targeting groups differently. One participant stated:

"…I think it has a lot of advantages…you know there's so many different components to consider and it's really complex, and I think pulling in all of those different factors together into a model that then gives you an output that's very clear and prioritizes, you know, who needs outreach or who is most at risk, or who is the most frail and needs interventions… higher utilization of the health system, higher costs, maybe more inpatient days, and it's like, we know what to do." – Participant #4

The peer pressure construct emerged as participants expressed consistent use of analytics in other departments and organizations. The needs and resources construct also emerged in that the participants expressed the use of analytics as cost saving and that targeting patients and groups was possible with analytics.

Individual

Questions #3 and #6 asked questions to ascertain the characteristics of individuals and their attitudes and beliefs about their use of data and IT tools and the intervention (PA). The overarching construct from the individual questions was a high level of knowledge and positive beliefs about the intervention (predictive analytics).

Question #3 specifically asks if they use data or IT-driven tools to support increased CRC screening for their patients. All respondents answered yes, and listed examples. Examples included reporting workbench databases, recall lists, the NCI risk model, specific platforms, IT-driven tools in SAS and SQL, and homegrown analytics.

Question #6 asked about their personal beliefs about using PA. All respondents answered favorably ("It's great"). One respondent specifically said:

"We use risk in clinical practice, whether we acknowledge it or not, but actually formalizing it in terms of clinical decision support would be helpful in prioritizing". – Participant #1

The high level of knowledge and positive beliefs about the intervention constructs were apparent as all 5 participants reported knowledge and current use of IT-driven tools and analytics, as well as positive personal feelings about the use of PA.

Process

The process questions (#4, #7) sought information on the degree to which PA and the use of multilevel data were available, and the ability to execute analytics at KPNW. The process domain of the CFIR framework is designed to understand the "improvement process" of executing an intervention. The questions for this domain were designed to understand difficulties in the execution of PA. The overarching constructs from the process questions are engagement (lack of engagement) in multilevel data and executing the intervention.

Question #4 asked, "*what if the needed data or information came from a separate data source*". All participants said access to outside data was limited or not available. Although Participant #5 listed the use of outside data such as published data, tracked data, and information gathered from interviews with staff. Participant #3 said they used data from state registries (immunization data) and claims.

Question #7 asked about the barriers and facilitators to using PA. The responses varied. Barriers were identified as slow moving to develop (n=2), a lag of awareness of the usability of PA, that PA is difficult to implement or integrate into workflows (n=2),

that the EHR is unable to support risk tools. All participants identified the "Quality Team" as a facilitator to using PA. Regarding the use of the EHR, one participant stated:

"*So, I think the biggest barrier is the existing EHR system we have which was developed as a transaction and billing record, it is not conducive (to PA), and I think the culture of the EHR companies is that they will not do anything to develop and incorporate data and link prediction models".* -Participant #1

The overall understanding of the potential for use of multilevel data was broadly dismissed by all participants. There are no structures at KPNW to explore the use of or access to external data, and the value of using such data was not exhibited through the process questions, prior to the introduction of the multilevel model. However, Question #14 (inner setting) was asked after the presentation of the multilevel model and elicited positive reactions to the use of multilevel data. While the quality department supports the use of analytics, there was an overarching theme of needing better coordination to more productively using predictive analytics.

Intervention

Questions #9, #10, and #12 seek the participant's perceptions of the intervention defined as the use of the individual and multilevel risk prediction model developed in Aim 2. These intervention characteristics questions seek these stakeholders'

perceptions of the quality and evidence of the models and perceived relative advantage. The overarching construct that emerged from the intervention characteristics questions is the relative advantage of implementation and complexity.

Question #9 asks if the multilevel model provides the information that they will find useful in the development of screening practices. The majority (4/5) of the participants thought it useful, with comments about the multiple uses of the model (administrative or QI), that it confirms the current strategies used and that knowing who the challenging population is to reach would be helpful.

"(would you find it useful?) Very much so, and again, it validates our current strategies of starting (interventions) with the positive FIT tests" - Participant #2

"Yes, it's not all quantitative, sometimes the model provides qualitative information that you can use." – Participant #1

The participant that questioned the multilevel model indicated that it would be difficult to implement and that getting external data is difficult:

"If you're showing that it doesn't have that big of an impact, I think the amount of effort and energy it would take to try to incorporate this, for such a small impact is just not a feasible thing." -Participant #3

Question #10 asked how the participants would use models like the one presented. The messages from the participants who answered were all positive. The participants separately stated that it could be used to validate data, close care gaps and those specific elements of the model are useful.

"Yes, we could use this to determine interventions and places to intervene, to deliver different interventions to different groups". -Participant #5

Question #12 asked if this tool could make their job easier, and responses were mixed. One participant stated that their use of analytics already makes their job easier and could lead to better health outcomes. Another participant identified that it would not make their work easier but could be a different method. One participant indicated that it would hypothetically be useful but could open more work to be done. Finally, one participant identified that it could help them "hone in" their screening practices.

Participant #3 also suggested that the multilevel data does not add anything, because what we are after is already captured in individual level data.

"You know patients that live in poor communities are not going to get screening as much, but it could be that we're just capturing that in the individual level data. Is it that

what we're trying to get from these other levels is already being reflected in the individual level data? So yeah, they live in lower income communities. Maybe they are more frequently missing appointments, or are a minority race, or don't come in for dental visits." – Participant #3

The relative advantage of implementation and complexity constructs were clear as the first two questions elicited the participant's general thoughts of the usefulness of implementing the intervention. Yet the perceived difficulty of implementation was evident from a few responses.

Policy

Questions #15, #16, and #17 asked about policies at KPNW regarding data and analytics. The overarching construct from the policy questions is that policy does not play a role in the use of data or analytics.

Question #15 asked if the department/KPNW has policies about access and use of PA, question #16 asked if there are policies about access to data, and question #16 asked about policies to understand trends in screening among subpopulations. All participants replied "no" to all policy questions. However, comments from other questions highlighted the expected use of analytics and a "culture" of the use of analytics at KPNW that may supersede the need for a formal policy mandating the use of analytics. The Quality and analytics departments seem to have created a local culture

making external policy unnecessary. In general, respondents noted that providers and staff are encouraged to use data and analytics and that policies generally address quality improvement and protection of PHI.

"*Some of the policies that come to mind are really around how we identify improvement efforts, based on regulatory need".* -Participant #4

While policies are evident in other areas, policy does not play a role in PA.

Summary of Aim 3 Findings

Qualitative interviews were conducted among 5 health system stakeholders. Participants were from various departments and in different roles. Common constructs were identified regarding the decision making process, CRC screening, use of data, use of predictive analytics, and model usability guided by the constructs of the CFIR framework.

In the inner setting, the stakeholders identified that team and group decisions are made, and they identified well-coordinated similar practices for increasing CRC screening. Stakeholders claimed that they all use PA to some extent and that the Quality Team facilitates the use of analytics in their departments. The interviews revealed the constructs of compatibility and access to knowledge and information in the inner setting.

The outer setting questions revealed that many other departments and outside organizations are using PA. They also identified that this method of improvement is cost efficient, indicating a potential external driving incentive. Major construct themes include peer pressure and needs and resources driving the use of PA.

The individual level is intended to identify the individual's personal knowledge and value they see in the intervention. All individuals use data and IT tools in their daily work. All individuals also indicated favorable knowledge and beliefs of PA, the individual CFIR construct.

Regarding the process of the potential use of PA and multilevel data, barriers and facilitators were identified. Barriers included access to multilevel data and incorporating models into their general workflow and EHR. However, participants identified the "Quality Team" as a facilitator to using data and PA. The overarching construct from these questions was decreased engagement (in multilevel data) and the difficulty in executing PA.

For the Intervention construct, feedback was sought on the usability and usefulness of the multilevel model. The participants found analytics and the multilevel model overall useful in the information it provided, and that it could be used in a variety of ways to make their jobs easier. The overall construct was a relative advantage of the use of PA, yet complexity of the implementing PA into their standard workflows.

All participants indicated that no internal policies impacted their use of data or analytics. It was clear that no relevant internal policies appear to play a role in this area

of their work, however, this level of employees is unlikely to know about governance and policy that may be driving decisions made that impact their work.

The interviews provide insight on the organizational perspective of the use of data and analytics to increase CRC screening. No major differences were found across participants from different departments. KPNW overall seems to have a culture of encouraging the use of data, analytics, and predictive analytics in daily work regarding increasing CRC screening. The use of multilevel data was not common, but after the presentation of the model, a few participants found value in the use of multilevel data, regardless of the lack of performance improvement.

Chapter 5: Discussion, Conclusion and Future Research

Predictive analytics (PA) and multilevel data are increasingly used in population health management and offer clinical decision support (CDS) at the point of care. The use of PA can allow systems to personalize care based on an individual's risk of certain events. (R. B. Parikh et al., 2016) The data most available for use in clinical settings are individual level data from the EHR. Multi-level data did not make a difference in this project but continues to have the potential to add to analytic models and to what we know about a patient and their risk of events.

The purpose of this study was to understand the value of the inclusion of multi-level data in a risk prediction model in predicting a patients' risk of failure to screen for colorectal cancer and to understand the health systems perspective of the use of multi-level data. The purpose was accomplished by assessing the data, developing risk prediction models, and conducting interviews with health system personnel.

This chapter will discuss the implications and meaning of the results across each of the three Aims. The Aims were to 1) assess data sources and variables, 2) develop risk prediction models, and 3) assess health system perceptions. The use of the data in the models and the application of the models to the subpopulations is reviewed, as is the reaction to the models and the performance.

Following the discussion are summary conclusions drawn from the dissertation, limitations of this research, the justification and significance of this research including implications for practice and policy, as well as future research directions.

Discussion

Aim 1: Assess the Data

The first aim was to assess data availability and identify opportunities to include multilevel data following the SEM framework. Known or previously examined predictors of screening and new variables were identified for use. Databases were reviewed for data that was applicable to CRC screening. Data were first assessed for availability and then assessed for quality when collected among the eligible population. The quality assessment included examining missingness, distribution, and collinearity. It was important to assess missingness as patients with missing data are excluded from the risk models. Distribution was assessed because variation is important to contribute to the models. Collinearity was important to include variables that measure different aspects of the patient.

Data Availability

Data was sought based on prior literature including systematic reviews, and key articles. Data was acquired from KPNW databases, and publicly available data sources. The data was assessed by determining what data was available, and what was usable by examining the characteristics of the data across the eligible population.

Through the assessment of the data sources and measures for Aim 1, this study found that individual and community level data was most readily accessed, while interpersonal, organizational, and policy level data was more difficult to find and integrate into research.

EHR and administrative data primarily captured at the individual level were common and available for use in health systems. A large integrated health system like KPNW has a robust data library and policies and practices in place where data are relatively easy to acquire and use. The system's integration of pharmacy, specialty, dental, and medical care provides a rich data source that may not be generalizable to systems where services are not integrated. These data sources provide broad individual data, the ability to find and create organizational and interpersonal measures and connect individual-level data with community level data. The EHR contained some data at the interpersonal and organizational levels. The data at the interpersonal level that was available through EHRs, or administrative databases often had to be calculated to create useful measures for use in the risk prediction models. For example, the interpersonal variables from this study such as *provider panel size, panel screening rates*, *provider gender match, provider race match,* and *provider ethnicity match* all had to undergo some level of programming; the variables were not sitting in the EHR ready for use. The ability to calculate variables assumes that a health system has some analytic ability.

Community level data was also relatively accessible through public data sources. This easily available data offered a unique opportunity to integrate community level data to better understand trends in health and health outcomes by geographical area, and how this level influences health. Census data was easily accessible and can be linked to patient address when available and could provide insights into neighborhood characteristics and their impacts on health outcomes. However, other community level data may lack validation and not be reliable for use.

There are limited data sources for the interpersonal, organizational, and policy levels. Organizational data can be available but like interpersonal data would take analytic effort and coordination to use in analytics. Organizational data like gap closure rates are difficult to acquire. Some organizational level data must be calculated, such as *clinic screening rates* and *wait time*. The calculation of these data would require analytic time and expertise.

Interpersonal level data is simply rarely collected. Marital status, family size, and family history data could be very valuable to a health system and a patient's treatment. Yet, some EHR systems and medical records, like those used at KPNW, are not widely able to collect and store this level of information. When collected it is not stored in discrete fields and not collected for population health management. Family history may be one of the most useful pieces of information in assessing risk for common diseases. (Ginsburg et al., 2019) Although *family history* was an emerging variable, only 0.82% of the patient population had a discrete record of family history of CRC. In a systematic

review, Henrikson and colleagues found that the prevalence of having a first degree relative with CRC is estimated between 3.1% and 10%. (Henrikson et al., 2015) The completeness of the data is unknown in the EHR, patients with no family history of CRC are not able to be differentiated from patients who have not been asked for their family history. This could mean that the information is not adequately recorded or asked in the clinical encounter or is in clinical notes as part of a patient's history. A greater commitment to creating EHRs able to store this data and maintaining this data could improve risk assessment.

Policy level data was difficult to apply to this project. There was little policy variation across the sample, variables had such little variation that they were not usable. The policy level data found for use was at the Coordinated Care Organization (CCO) and state levels. CCOs are county based organizations that are comprised of providers and organizations that provide care to Medicaid patients. (OregonHealthAuthority, 2017) CCO data is relevant because colorectal cancer screening was an incentivized metric for clinics within CCOs, clinics were paid to reach set performance or improvement targets. The CCO level data does not exist or apply for Washington patients (26.3%) and therefore was unusable, as it would become a state specific model with only the inclusion of the Oregon patients. At the state level, Washington, and Oregon both expanded coverage of the Medicaid population with ACA, and the measures were redundant with the *state* variable. Policy level data would be more applicable in a study with patients from a wider variety of states where policy data is more varied.

Variable Variation and Usability

The usefulness of data depends on the meaningfulness and applicability to the population being studied. Some variables had very little variation and were determined to be unusable. For example, the *health literacy* variable was selected for use, as literacy was suspected to be a predictor of screening. Yet, the distribution of the variable showed that patients lived in communities where 99% of patients are at the intermediate level of literacy, there was no variation across the patient population. More specific measures of literacy or measures with wider distribution could be more beneficial. The community variable used linked census block with literacy scores. A patient specific measure of health literacy could provide even more specific information, such as a patient's understanding of disease prevention. (Sørensen et al., 2012) This type of measure could act as the foundation for an intervention in itself to build greater literacy and help patients navigate health systems and self-health promotion.

The *RUCA* (rural code) variable also had little variation (<1% of the eligible patients lived in a rural area). This indicates a largely homogenous population that lives in a widely suburban or urban area. These variables may be valuable in a larger more diverse project but provided little value in a study conducted in a large integrated health system in a single metropolitan area, in 2 states. Other variables that were deemed unusable due to lack of variation include *Medicaid expansion* and *health literacy* (mentioned above). This suggests that data collected at larger geographic levels may not

be as useful for predictive algorithms as community level variables at a more granular level.

While a large health system like KPNW provides robust data because patients are covered by KP insurance, some data was simply unattainable for this project. Physician recommendation was not useful, in that less than 5% of screening orders were made by a physician. The health system uses an organizational intervention through a centralized process of mailing FIT kits. This intervention allows clinicians to simply discuss screening with the patient, but not actually place an order if they already have an order in the EHR. Also, *panel screening rates* and *clinic screening rates* were too difficult to acquire in that data for the entire patient population would have to be obtained, and screening rates calculated. For this project, this was too time intensive for the analyst to pull and calculate these data.

Variables that were identified in prior literature, known predictors, decreased as the conceptual levels increased (more in individual, fewer in policy). At the highest level, the policy level, we know that some groups or organizations such as UDS (Uniform Data System from Health Resource Services Administration), USPSTF (U.S. Preventive Services Task Force), and CMS (Centers for Medicare and Medicaid Services) have policies regarding CRC screening. UDS has reporting requirements for federally qualified clinics. (National Center for Quality Assurance (NCQA), 2021) The USPSTF recommends screening types and frequencies. (US Preventive Services Task Force et al., 2016) Yet, the direct impact of these policies on screening rates is difficult to ascertain because

they apply to clinics across the country, there is little variability in the impact of the policy on screening, especially in large healthcare systems. Policy level variables also may have minimal impact at KP where Medicaid enrollment rates are low, and where screening practices typically go beyond national recommendations.

Aim 2: Develop Risk Prediction Models

The second aim was to develop risk prediction models using individual and available multilevel data as determined in Aim 1. The first model was developed in a large dataset of patients who are due for CRC screening using individual level data. Statistical improvements in the model were assessed when multilevel data was added. The incorporation of the multilevel data allowed the understanding of the contribution of the addition of external multilevel data to the model created first on individual level EHR and administrative data. Subpopulation applicability, based on the PROBAST (Prediction model Risk of Bias Assessment Tool) was assessed through the application of the model to a non-White subpopulation. PROBAST is a tool designed specifically to assess the quality of prediction model studies and specifically the risk of perpetuating systematically distorted estimates of model performance (bias) among populations. (Wolff et al., 2019) This approach allowed for the understanding of the applicability and translation of the model to a subpopulation.

In the development of these models, it was immediately determined that I was unable to use the 45-49 age group based on active guidelines for this time period. The

45-49 age group during this time period included only Black/African American patients, as the screening recommendation was lower for Black patients until May of 2021(US Preventive Services Task Force, 2021). This group is unique and complicated the interpretation of the model findings in that it added bias by providing base estimates for a unique group that does not apply to the larger group and removed algorithmic fairness. The inability to separate the variation of this age group from the general population required the removal of this population from the analysis. The unique screening recommendations and risk of CRC screening among this age group provided estimates of the likelihood of screening that were not comparable to the remaining population. A separate model for Black patients could be redeveloped and include the 45-49 age group, but the relatively small numbers and there are so few Black patients that it would need to be a simple model with few predictors to avoid over-fitting.

Concordance was measured by a bootstrap corrected C-statistic. (Moons et al., 2019) The discrimination and calibration was measured with an R^2 statistic. Calibration was assessed by the integrated calibration index (ICI), where 0.0 indicates perfect agreement between the observed and predicted risk. (Austin & Steyerberg, 2019) The individual model performed well in the full population (bootstrap corrected C-statistic=0.722; R^2=0.1108, ICI=0.134). The full individual model alone would be sufficient for identifying populations unlikely to screen for CRC where interventions to screen could be spared (i.e., the highly likely to screen population may not need additional outreach or interventions).

The multilevel data was added, 40 variables in all, including 5 variables at the interpersonal level, 5 variables at the organizational level, 29 variables at the community level, and 1 variable at the policy level. Adding the multilevel data provided minimal improvements to the model (bootstrap corrected C-statistic=0.7218; R^2=0.1119, ICI=0.130). The bootstrap-corrected c-statistics are close to identical (-0.0002), the R^2 statistic is 1% higher on a relative scale for the multi-level model (+0.0011), and the ICI is 3% better on a relative scale for the multi-level model (-0.0004), but still close to zero indicating excellent agreement. The limited statistical improvements in the multilevel model may indicate that the added effort of acquiring the multilevel data may not be worth it in terms of model performance.

The reduction of the individual level model removed 5 variables out of 19 including *insurance group, language group, interpreter services, BMI,* and *inpatient visit*, as they contributed very little to the Akaike Information Criterion (AIC), reducing the R^2 by less than 1%. The variation in screening caused by these variables is likely being captured in other individual level variables in this model. The reduction of variables in the model removes variables with little effect, and models with many variables are less practical (Steyerberg, 2019). Simplification allowed me to determine the variables that contributed the most to a model.

The reduction of the multilevel model removed 34 of the 61 total variables (55.7%). The model was redeveloped essentially to include the multilevel variables; therefore, the contribution of the individual variables was reassessed in terms of the R-

squared variation. The reduced model retained all variables from the reduced individual level model plus two additional individual variables (*inpatient visits, BMI)*, as well as two interpersonal variables (*address changes* and *provider gender match*), three organizational variables (*wait time for colonoscopy, clinic size,* and *primary care clinic*), and two community variables (*number of healthcare facilities, median household income)*. The retention of few multilevel variables again begs to question the usefulness of the data and how much information is added by incorporating external data. This analysis did not find any variable at the higher levels that significantly improved the model.

Subgroup Validation

Applicability was assessed for a non-White population. The project was originally designed to test the application of models to an outside population (FQHC or community population). However, limitations in time and resources led to the design of testing the application to a subpopulation of KPNW patients. It was also important to apply the model to a subpopulation to test the applicability of the individual and multilevel models. Applicability was tested on the non-White population, which was considered more likely to have variation in multilevel data.

The subgroup application of the model is important to be able to show that the model could apply to a population outside of the primarily white population in the Northwest. The non-White population was chosen partly because they have variation in

screening rates. The screening outcome was present in 36.1% of the full population and only 31.2% of the non-White subpopulation. The difference in screening rates is one reason the subpopulation application was important. Rates of CRC (adenoma prevalence) have been found to be higher in Black patients but determined predominately due to lower rates of screening. (Rutter et al., 2021) The Medicaid (n=2,613) population was considered as a subpopulation but was not chosen because it was too small to handle the multilevel model and the many degrees of freedom required by the predictors. There would not have been enough patients in any standalone racial minority group to test the model. The non-White population (n=13,655) provided an adequate subgroup for the application of the large model. Both the individual and multilevel models performed better statistically in the non-White population than in the full population, as the R^2 and C-statistics in both models were both higher.

The application of the *individual* level model to the non-White population showed improved performance with a 3.3% relative increase in the bootstrap corrected C-statistic, a 23.1% relative increase in the $R^2$and a 36% relative increase in the ICI, showing an inferior calibration as there is a greater difference in the observed and predicted probabilities. The model separates patients more effectively but is less accurate in the non-White population. The application of the *multilevel* model to the non-White population showed improved performance with a 2.3% relative increase in the bootstrap corrected C-statistic, a 22.3% relative increase in the $R^2$and a 58% relative

increase in the ICI. The first two statistics (C and R^2) indicate improvements, but the increase in the ICI shows such inferior calibration it would be inadvisable to use the multilevel model in non-White populations. The improved variation and discrimination show that the models adequately translate to the non-White population, but the decrease in calibration is concerning for application to that population.

Again, the subpopulation application was conducted to test algorithmic bias because the NW population is predominately White, and the KPNW population is primarily commercially insured. The PROBAST tool assesses the risk of bias (ROB) through assessing the distribution of participants, the definition of predictors and outcomes, and analysis of the model (Wolff et al., 2019). In PROBAST, 20 signaling questions are asked to assess the risk among the above mentioned categories to ensure the applicability of the model to participants in the setting, the definition, assessment, and timing of predictors, and the application to the model question (Wolff et al., 2019). Failing to assess ROB could result in models that do not apply to minority populations and subgroups. Other types of bias may be present but undetectable using this tool, such as bias in how data is collected, or the care patterns of minority patients.

Although the PROBAST tool was used to guide the assessment of bias through addressing signaling questions to review the population in the model compared to the overall population, it was important to evaluate the model specifically in a more diverse population than the overall KPNW population. (Obermeyer et al., 2019) The application to the subpopulation discretely addresses the applicability to the minority population. It

was unexpected that the variables included in the models would better predict the likelihood of screening in a non-White population as the model was trained on a predominantly White population. The ICI increase suggests algorithmic bias because of worse performance and calibration in the non-White population, especially for the multi-level model. The absolute increase in the ICI for the individual level model may be sufficient, but it requires a closer inspection of the thresholds that would be used for interventions to improve screening. In this case, the model may warrant updating to determine the non-White probability of screening.

Our study suggests that a deeper look into the differences in screening rates among the white and non-White populations mentioned above (36.1% vs. 31.2%) is warranted. It has been determined that the unknown bias within the models is also applicable to the subpopulation, but the drivers of a decreased screening rate should be examined. Multilevel data was used in this project to try to capture non-individual level data that typically is not available in health system datasets, but the multilevel data did not improve the performance of the model. The data included in the individual level model captures differences seen in the White versus non-White patient populations in this sample.

Multilevel Data

Some drivers of decreased screening rates are not quantitatively available, even when using multilevel data. For example, known predictors (predictors identified by the

literature) like *medical mistrust, social networks,* and *social support* are simply not available as data points in this healthcare setting. (Adams et al., 2017; Alema-Mensah et al., 2017; Dominic et al., 2020) The inability to measure these types of drivers of screening must be considered as a limiting factor of risk prediction overall. The model has room for improvement in terms of the C-statistic, yet the multilevel data available for this project does not accomplish this.

Some measures within the individual level model are worth discussing. Patients with a 2+ Charlson score had decreased odds ratios in all models (0.701 in the individual full model) for screening, which may be indicative of patients who have a long history of failure to seek preventive care or who are unable to screen due to comorbid conditions. The largest contributor to the AIC was the *prior screening* variables, where patients with prior FIT or prior colonoscopy had ORs greater than 3 (OR=3.815 (3.659, 3.978 CI) and 3.231 (2.935, 3.556 CI) respectively). Patients with dental visits (regardless of if they had a plan) versus had a dental plan but no visit had an increased likelihood of screening (1.31 (1.251, 1.372 CI) vs. 0.886 (0.844, 0.931 CI)). Hispanic patients were more likely to screen than white, non-Hispanic? (OR=1.396 (1.261, 1.545 CI)).

Further, in the multilevel data relationships with screening were also interesting. Clinic size was determined as small if fewer than 30,000 patients and larger clinics have a greater likelihood of screening (OR=1.94 (1.374, 2.742 CI)). Geographical areas with a higher number of healthcare facilities in a community have a slightly elevated likelihood of screening (OR=1.088 (1.021, 1.160 CI)). These variables could be picking up

geographical variation in areas with a greater population density that the RUCCA codes did not. It is also possible that there was not enough variance in this population to determine screening differences by rurality. Smaller clinics are in rural areas, so clinic size could be competing as a predictor with correlated RUCCA. Further, the odds of screening among multilevel data were interesting. Clinic size was determined as small if it had fewer than 30,000 unique patients, and larger clinics have a greater likelihood of screening (OR=1.94 (1.374, 2.742 CI)). Geographical areas with higher numbers of healthcare facilities in a community have a slightly elevated likelihood of screening (OR=1.088 (1.021, 1.160 CI)). Health center density has been found to improve health outcomes as it reflects primary care access (Evans et al., 2015). These variables could be picking up geographical variation in areas with a greater population density that the RUCA codes did not. It is also possible that there was not enough variance in this population to determine screening differences generally seen in rural areas.

The use of PA in combination with multi-level data was one way to recognize group membership and individual characteristics simultaneously. While this study did not find that multilevel data improved prediction, evidence continues to emerge that multilevel data does play a role in health outcomes. Viramontes and colleagues conducted a study on CRC screening rates among a national diverse population (BRFFS data) and found lower screening rates among Hispanic patients (53.4% vs. 70.4%) overall, but also found variation by state and territory (Viramontes et al., 2020). Yet, Schuler and colleagues found that neighborhood level SES did not improve prediction in

mortality (Schuler et al., 2020). While mortality is different than screening, understanding community impact on a variety of outcomes is important. It could be that the KPNW population community level characteristics are not diverse enough to make a difference in predicting screening, or that the right data for this analysis was not accessed.

Multilevel data is important to understanding what influences health. It is also important to recognize that some important data at all levels is not available at all. For example, individual level data such as psychosocial issues like awareness and fear are important but not available in our current data (Brill, 2020). There are limitations to the interpretation of data at the upper levels of SEM, as it is capturing characteristics of groups of people. When using individual level data simultaneously as multilevel data, the contribution of the multilevel data can be difficult to interpret. In some cases, multilevel data is capturing higher level characteristics of communities such as access to health services. In other cases, higher level characteristics are acting as a proxy for individual characteristics such as income. This is where the interpretation of the upper level variables is critical to informing systems how to use such variables.

Use of the Models

No other models have been developed to test the risk of missed, indicated screening for CRC. However, risk prediction has been found to be useful in tailoring screening programs to inform patients of the risk of developing cancers (Saya et al.,

2020). Some models have used a few available multilevel variables, but no risk prediction models have focused on the performance improvement after the addition of multilevel data. The lack of information about adding multilevel data could be attributed to limitations in the availability of these data, and the lack of contributions to the model, like what was found in this project.

There is potential clinical use for either the individual or multilevel model. There is a pragmatic use of the reduced individual model that is the most simplistic way to identify patients' likelihood of screening. In this sense, the individual model could be used to target patients for interventions aimed at increasing indicated screening. In addition, the simplicity of the individual model reduces the effort and potential resources needed to acquire and maintain multilevel data as well as the generalizability outside KP or other closed systems of care (e.g., VA).

The developed risk models could impact patients and systems by providing a tool to identify patients' likelihood of screening, using limited resources to target those most vulnerable to missed, indicated screening. Both the individual model and multilevel models are sufficient in discrimination (C>0.6) (Steyerberg et al., 2010)., While the multilevel data adds little to the performance of the model, it could provide information that is important to the health system in determining ways to intervene on patients to increase screening., such as targeting specific regions or clinics. However, this same information could be attained by providing analytics to the decision makers outside of the risk prediction model. The health system perspective is important to

understand. Working in collaboration with the health plan when creating the models would provide information that is useful to the development of the model and the end user.

The multilevel data did not have the expected impact on the model. Although this was an exploratory effort in using multilevel data, it was anticipated that the multilevel variables would contribute more to the model than they did. The multilevel model did not contribute more to explaining the likelihood of screening because of lack of variation in data due to homogeneity in patient sample and multilevel data acting as a proxy for individual level characteristics.

As discussed in Chapter 2, the literature has addressed the benefits of looking at micro and macro level data about a person and the groups to which they belong. The literature shows that the different levels are linked or interconnected and that levels can be synergistic (Denise M. Rousseau, 1985; Taplin et al., 2012; Weiner et al., 2012) However, the multilevel variables' influence on screening may be captured by already included individual level data, or the KPNW patient population may be too homogenous to capture community and policy level differences. Further, the multilevel data may be less reliable than individual level data.

Aim 3: Assess Health System Perceptions

Aim 3 assessed the perceived usefulness of the developed CRC screening model, multilevel data, and predictive analytics in colorectal cancer screening using qualitative

research methods among health system stakeholders who might use a model like this. Semi-structured interviews were framed around the domains and constructs of the Consolidated Framework for Implementation Research (CFIR). Conducting the interviews was an opportunity to understand how PA could help achieve system level goals, gain knowledge about the usefulness of PA in general and the individual and multilevel models developed in this project.

The major observations from the interviews were that PA was widely supported at the organizational level; PA was great but difficult to implement within the EHR (EPIC); and that multilevel data could be useful information for designing interventions but is not helpful in what it added to the model. It was also observed that more support (in analytics) was needed and that the participants had minimal knowledge of externally available multilevel data.

All interviewees held similar knowledge and beliefs about multilevel data and PA. Related to the inner setting, there was a consistent belief that analytics was a priority of the organization and was compatible with their work, yet more resources for analytics were desired in the form of support and coordination of analytics, as well as improvement in the EHR to support the use of PA. One participant described EPIC as a "billing system", not a health tracking system, and indicated that it was unable to integrate risk prediction information for patient management. In this sense, the use of risk prediction would have to occur outside of the EHR. The ability to track risk scores in the EHR would be ideal. Participants also desired more time and resources devoted to

analytics. KPNW seems to have been able to develop appropriate structures for managing and generating data and analytics, but it leaves data users wanting more (Dash et al., 2019).

In the outer setting questions, it was clear that there is peer pressure to use PA, yet limitations in the technological and staff resources to successfully do so. There was consistent knowledge of other departments and outside organizations using analytics. The use of PA was consistently identified as cost saving, and it was perceived to increase the ability to target patients for interventions and streamline work processes. The cost savings of using analytics is dependent on the intervention and how it is used (Weiner et al., 2018). Participants also noted that more work could be required in the implementation of interventions, increasing time spent and costs of implementation. Participants identified limitations of the EHR in implementing risk models, a challenge that has been noted in prior research (Sharma et al., 2021). Sharma et al. call out the potential of the use of risk models dampened by the inability to successfully integrate into EHRs (Sharma et al., 2021). Workflows then must be outside of the EHR and can complicate the success of such programs.

All individuals expressed knowledge of and positive beliefs about the use of PA. However, when asked about their current use of multilevel data, no individual expressed the knowledge of using external multilevel data prior to being shown the multilevel model. The participants identified a wider use of multilevel or outside data, that was inconsistent with the type of data used in the multilevel model. Examples of

external data used included information from schedulers, published data, and data from other risk models. However, after introducing the models, interest emerged in discussing specific multilevel data components within the multilevel model. In particular, the organizational data (*clinics, clinic size*) was of interest as a way to tailor interventions. While pragmatic, the addition of the multilevel data did not improve the performance of the model, the interviews showed that when included, there was interest in knowing more about those multilevel variables and how screening varied across those measures. The stakeholders who work in increasing CRC screening rates clearly were interested in the multilevel data and indicated that it may be useful in different ways, outside of the model.

When asked about policy, participants consistently identified that there is no policy driving the use of data, analytics, or looking at subpopulations within KPNW. However, their answers generally pointed to a well-developed culture of using PA in their work. All participants were engaged in the process and use of PA.

In all, the interviews revealed a desire to use and tailor analytics for specific departmental use. While complex and sometimes difficult to implement, PA using extended information about the patient from the organization and community was perceived as helpful in closing care gaps and streamlining daily work.

Conclusions

This project is useful in ascertaining ways to expand the use of PA model inputs beyond data captured in the EHR and administrative databases, in developing risk prediction models to identify a patient's likelihood of screening for CRC, and understanding stakeholder perspectives of the use of multilevel data and risk prediction. The mixed methods research design was chosen to answer the research questions:

1) Can the inclusion of multilevel data improve the accuracy or applicability of a prognostic risk prediction model, in predicting patients' risk of failure to screen for colorectal cancer in order to target interventions to the right patients at the right time?

2) How does the inclusion of multilevel data in a risk prediction model improve the usefulness to health system decision makers for managing population health?

The first aim sought to assess data sources and variables through a comprehensive exploration of data within the electronic databases and at then at multiple levels in external data sources, guided by the SEM. While multilevel data is available, it is not consistently available and usable at all levels.

The second aim used quantitative methods and logistic regression to create a series of risk prediction models to assess the value of adding external multilevel data, and the applicability of the model to subpopulations. The multilevel data added little to

the performance of the models. The application to the subpopulations was adequate, performance of the models improved.

The final and third aim qualitatively assessed the perceptions of the use of multilevel data, PA, and this model created for CRC screening guided by the domains and constructs of the CFIR. Stakeholders who work in colorectal cancer screening described a strong culture of using analytics in their work, found the model useful, and identified barriers to implementation.

This project provides information about the use of external multilevel data and risk prediction in assessing the likelihood of a patient screening for colorectal cancer. While multilevel data did not add value to the predictive models, there may be applicability of multilevel data to other areas, like developing interventions for CRC screening. Health systems could broaden the types of data collected about a patient, their interpersonal relationships, their community, clinics, and the organization for use in additional exploration of multilevel data and how it may benefit their efforts to reduce screening. Risk models like this model are sustainable, in the sense that they can continually provide information about patients that are likely to need additional interventions to achieve screening.

Research conducted in more than one organization could particularly benefit from further exploration of the use of multilevel data in assessing impacts of community level variables on health outcomes like CRC screening. Research could also benefit from the use of data across multiple health systems, especially including community clinics

where there may be more variation in characteristics of communities and organizations to which the patients belong.

Limitations of this Research

There are several limitations to this project across all Aims. The main limitations include threats to internal, external, and measurement validity including data availability, completeness, collinearity of multilevel data, and a limited context for qualitative interviews and data.

EHR and administrative data are limited to variables in analytic databases. External data is limited to data that is publicly available or available upon request. Some data was inaccessible or not collected consistently in the electronically available databases; these variables included social determinants of health, some community level data, and family history data. Policy level data could also have included gap closure rates for HEDIS 5 star ratings but was not accessible. Multilevel data that was determined to be usable for this study was sometimes found to be colinear with individual level data and offered very little additional predictive value to the individual level data.

In the assessment of data availability, KPNW is likely better positioned to have access to the majority of health data, because it is a closed insurance system. The system's integration of pharmacy, specialty, dental, and medical care provides a rich data resource that may not be generalizable to other organizations, FQHCs, or

community clinics. Outside of a system like KP, people may go to different hospital systems or physicians, and there is usually no data sharing or interoperability in services and records. Some systems rely on claims data to track services utilized by patients by external providers. Even other integrated systems will have some variation in the availability of data, regardless of the data being common in EHR or administrative data. Community data may not be reliable or validated. It also may not be representative of the broader population.

Another limitation is that this work was done in one health system in the Pacific Northwest which could limit generalizability to the extent that practice patterns within KPNW affect the likelihood of screening. KPNW's mailed FIT testing program provides screening opportunities that may not be available in other settings, changing the likelihood of patient screening.

Further, ideally, a model would be created using data from multiple systems with diverse geographical and patient populations, for a greater understanding of applicability to subpopulations and generalizability. The use of data from hospital and specialty services could limit the usability of the risk prediction model if the model ends up predicting documented screening in non-interoperable EHRs. Additionally, other advanced analytics methods, like machine learning, could foster the use of large amounts of data to determine likelihood of screening.

There are factors that might have limited internal validity of the model. One variable, *membership less than 2 years*, was included to determine whether a patient

who was to the system, who may have obtained prior screening at a different facility. The variable was retained in the final models and showed a decreased likelihood of screening (OR=0.7923 (0.7033, 0.8926 CI) in the individual level model). This variable is an example of limitations of the interpretation of the model where prior screening history is inconsistently captured. Further, this model was also created with one single year of screening data from the patient's birthday in 2018. Updated screening practices, such as increased outreach or expansion to younger age groups, at KPNW could affect the current accuracy of the model.

Another limitation is that the qualitative interviews are from stakeholders in a single integrated delivery system, who work in CRC screening. With unlimited time and resources, a wider range of participants from various levels of the organization or from a variety of health care settings would produce greater generalizability. Conducting this project only at KPNW limits the generalization of the mixed-methods findings to other settings.

Efforts were made to minimize and adjust for limitations throughout the project. Applying the model to subpopulations was an effort to address the lack of diversity in the health system and patient population. Including variables, like the membership variables, was an attempt to better understand the validity of the models. Multilevel data was included to better understand community level impacts on a patient's screening.

Justification and Significance

PA has the potential to increase patient safety and save resources. (R. B. Parikh et al., 2016) Multivariate risk prediction can improve population health when standardized and used appropriately. (Cohen et al., 2014) The risk of bias among disparate groups of patients is important to consider when using PA; applying models to subpopulations is one way to understand bias. Staff capacity and technology resource limitations can hinder the use of PA for population health management. (Bates, 2014; Leininger, 2017) However, some health systems are expanding their use of PA to conduct better population health management. (Leininger, 2017) In the current setting, COVID has created a backlog of colonoscopies and cancer screening guidelines have been expanded to age 45 from 50 (US Preventive Services Task Force, 2021). This expansion will create an increased need to know who will be able to successfully complete screenings, even in the absence of tailored interventions. Application of this model could solve this contemporary problem of needing to know who is best served by targeted intervention using limited available resources. This study sought to build more applicable models using multilevel data and to learn about data available from public sources across the SEM framework. Expanding data sources and the use of data across levels is valuable in providing knowledge about patients likely to screen but did not improve the performance of the models. While multilevel data is available to systems, the use of the multilevel data is limited.

Health systems are more likely to use PA if they see a positive impact on resource use and patient health (R. B. Parikh et al., 2016); therefore, it is important to identify how PA can be integrated with clinical care in meaningful ways such as in clinical decision support and optimizing team based care (R. B. Parikh et al., 2016). The adoption of multilevel risk prediction models is influenced by the perceived usefulness and reactions of the model by decision makers. There is a clear distinction between the pragmatic use of the model in the ability to simply identify patient likelihood of screening, where multilevel data may not improve the performance of the model but may be relevant to intervention development. In this sense, working with the health system and key stakeholders when developing models, and providing information important to intervention development is imperative to success and usefulness. Health system leaders thought the models and the use of PA as a useful tool to increase CRC screening rates, however, the implementation of such a model could be difficult.

In practice, this research provides an example of usable multilevel data in analytics and a model for determining patient likelihood of screening for CRC. Multilevel data is available for use; however, some coordination is required for synthesizing multilevel and patient level data. The models created provide a tool for identifying patients likely to screen for CRC, and the model was perceived as useful by health system stakeholders. Either the individual or multilevel models are sufficient for use in general screening prediction.

There are several system and organizational implications from this work. First, health systems and organizations ideally would have analytics capabilities like KPNW where population health management is possible. While PA is meaningful, it is difficult without appropriate analytic capabilities.

PA is clearly already a useful tool used in the KPNW system. In speaking with the stakeholders, it was apparent that there would be value in feedback loops when developing models. The ability to provide feedback on the usability and meaningfulness of multilevel data components within the model helps optimize development and tailor interventions. As it has been noted, the multilevel data adds nothing to the models, but stakeholders were interested in what the multilevel data meant. Knowing this, it may be valuable to work with stakeholders during the development of models to better understand multilevel characteristics of patients within the models. Multilevel analytics could be used separately to inform interventions. With feedback loops, multilevel data could be more valuable in the information it provides to tailor interventions and understand populations in need of additional interventions.

There are also several policy implications from this work. First, the availability of data is critical to supporting analytics, yet individual and multilevel data is not always available to health systems, and the coordination of data can be challenging. Variables like health literacy and social determinates of health (SDOH's) for example may provide value in analytics yet are not widely available for a broad patient population and may not be validated and maintained in publicly available datasets. The Agency for

Healthcare Research and Quality (AHRQ) User's Guide for Registries for Evaluating Patient Outcomes identifies data generally available in EHRs and of interest to research. (Ehrenstein V et al., 2019) AHRQ could more strongly encourage standardization of variable values and measurement. Meaningful use policy could also encourage capture and standardization of multilevel data. This type of policy could improve data standardization and availability from EHR's that can be used in research. Finally, policy could require that publicly available multilevel data like the data used in this project be standardized and validated by the organizations that collect it.

Recommendations for Future Research

There are a variety of ways this project could expand to future research. Future research could include working to create and validate external data sources, externally validate the predictive risk models, or use the model in an implementation project and work with a health plan to optimize the use of the model.

The known predictors were most common at the lower levels. Integrating upper levels of data into research, despite the limitations can tell us more about the many levels of influence on our health. The exploration of the external multilevel data in other areas of health, in diverse geographical areas, or a variety of health care environments will determine if there is value in use. Literature on the use of external data could improve the validity and use of multilevel data in research.

A series of projects could externally validate the model. The model was created in a single integrated health system. Externally validating the model in community clinics, other integrated systems, other geographical areas, or in a combination of environments could determine the true utility and generalizability of the model.

This model could also be used to prioritize interventions to increase CRC screening. A randomized control trial (RCT) could test outreach in many forms including motivational interviewing, phone or automated reminders, patient navigation, or video based interventions. An RCT could look at the impact of such interventions on the rates of screening across the levels of risk determined by the model.

Finally, taking the information from the qualitative interviews, further work could be done to refine the model in collaboration with the delivery system to design interventions. Collaboratively working with the delivery system could examine risks across specific groups or subpopulations, within regions or clinics.

Funding

This project was not funded and is an independent student research study. However, analysis time was donated by the KPNW Center for Health Research to create a student dataset for analysis. Further, clinic personnel donated their time to participate in qualitative interviews. No organization outside of the committee had any role in the design, implementation, interpretation, and reporting of this project.

References

ACS. (2019). *Colorectal Cancer Facts & Figures 2017-2019*.

Adams, L. B., Richmond, J., Corbie-Smith, G., & Powell, W. (2017, Oct). Medical Mistrust and Colorectal Cancer Screening Among African Americans. *Journal of community health, 42*(5), 1044-1061. https://doi.org/10.1007/s10900-017-0339-2

Alema-Mensah, E., Smith, S. A., Claridy, M., Ede, V., Ansa, B., & Blumenthal, D. S. (2017, Winter). Social networks as predictors of colorectal cancer screening in African Americans. *J Ga Public Health Assoc, 6*(3), 369-372. https://doi.org/10.21633/jgpha.6.306

Andermann, A., & Collaboration, C. (2016). Taking action on the social determinants of health in clinical practice: a framework for health professionals. *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne, 188*(17-18), E474-E483. https://doi.org/10.1503/cmaj.160177

Anderson, J. C., Fortinsky, R. H., Kleppinger, A., Merz-Beyus, A. B., Huntington, C. G., 3rd, & Lagarde, S. (2011). Predictors of compliance with free endoscopic colorectal cancer screening in uninsured adults. *J Gen Intern Med, 26*(8), 875-880. https://doi.org/10.1007/s11606-011-1716-7

Austin, P. C., & Steyerberg, E. W. (2019). The Integrated Calibration Index (ICI) and related metrics for quantifying the calibration of logistic regression models. *Stat Med, 38*(21), 4051-4065. https://doi.org/10.1002/sim.8281

Austin, P. C., van Klaveren, D., Vergouwe, Y., Nieboer, D., Lee, D. S., & Steyerberg, E. W. (2016). Geographic and temporal validity of prediction models: different approaches were useful to examine model performance. *J Clin Epidemiol, 79*, 76-85. https://doi.org/10.1016/j.jclinepi.2016.05.007

Basu, A. (2011). Economics of individualization in comparative effectiveness research and a basis for a patient-centered health care. *Journal of health economics, 30*(3), 549-559. https://doi.org/10.1016/j.jhealeco.2011.03.004

Bates, D. (2014). Big Data In Health Care: Using Analytics To Identify And Manage High-Risk And High-Cost Patients. *Health Affairs, 33*(7), 1123-1131. https://doi.org/10.1377/hlthaff.2014.0041

Bayati, M., Braverman, M., Gillam, M., Mack, K. M., Ruiz, G., Smith, M. S., & Horvitz, E. (2014). Data-driven decisions for reducing readmissions for heart failure: general methodology and case study. *PLoS One, 9*(10), e109264-e109264. https://doi.org/10.1371/journal.pone.0109264

Beaber, E. F., Sprague, B. L., Tosteson, A. N. A., Haas, J. S., Onega, T., Schapira, M. M., McCarthy, A. M., Li, C. I., Herschorn, S. D., Lehman, C. D., Wernli, K. J., & Barlow, W. E. (2019, Aug). Multilevel Predictors of Continued Adherence to Breast Cancer Screening Among Women Ages 50-74 Years in a Screening Population. *J Womens Health (Larchmt), 28*(8), 1051-1059. https://doi.org/10.1089/jwh.2018.6997

Beck, A. F., Simmons, J. M., Huang, B., & Kahn, R. S. (2012). Geomedicine: area-based socioeconomic measures for assessing risk of hospital reutilization among children admitted for asthma. *Am J Public Health, 102*(12), 2308-2314. https://doi.org/10.2105/AJPH.2012.300806

Benach, J., Malmusi, D., Yasui, Y., & Martinez, J. M. (2013, Mar). A new typology of policies to tackle health inequalities and scenarios of impact based on Rose's population approach. *J Epidemiol Community Health, 67*(3), 286-291. https://doi.org/10.1136/jech-2011-200363

Benuzillo, J., Savitz, L. A., & Evans, S. (2019, Mar 25). Improving Health Care with Advanced Analytics: Practical Considerations. *EGEMS (Wash DC), 7*(1), 3. https://doi.org/10.5334/egems.276

Beydoun, H. A., & Beydoun, M. A. (2008, May). Predictors of colorectal cancer screening behaviors among average-risk older adults in the United States. *Cancer Causes Control, 19*(4), 339-359. https://doi.org/10.1007/s10552-007-9100-y

Bhavsar, N. A., Gao, A., Phelan, M., Pagidipati, N. J., & Goldstein, B. A. (2018). Value of Neighborhood Socioeconomic Status in Predicting Risk of Outcomes in Studies That Use Electronic Health Record Data. *JAMA network open, 1*(5), e182716-e182716. https://doi.org/10.1001/jamanetworkopen.2018.2716

Blumenthal, D. M., Singal, G., Mangla, S. S., Macklin, E. A., & Chung, D. C. (2015, Jun). Predicting Non-Adherence with Outpatient Colonoscopy Using a Novel Electronic Tool that Measures Prior Non-Adherence. *J Gen Intern Med, 30*(6), 724-731. https://doi.org/10.1007/s11606-014-3165-6

Borquez, A., Cori, A., Pufall, E. L., Kasule, J., Slaymaker, E., Price, A., Elmes, J., Zaba, B., Crampin, A. C., Kagaayi, J., Lutalo, T., Urassa, M., Gregson, S., & Hallett, T. B. (2016, Sep). The

Incidence Patterns Model to Estimate the Distribution of New HIV Infections in Sub-Saharan Africa: Development and Validation of a Mathematical Model. *PLoS Med, 13*(9), e1002121. https://doi.org/10.1371/journal.pmed.1002121

Braverman, P., Arkin, E., Orleans, T., Proctor, D., & Plough, A. (2017). *What is health equity: And what difference does a definition make?*

Bresnick, J. (2018, September 04, 2018). 10 High-Value Use Cases for Predictive Analytics in Healthcare. *Health IT Analytics*.

Brill, J. V. (2020, Nov). Screening for cancer: the economic, medical, and psychosocial issues. *Am J Manag Care, 26*(14 Suppl), S300-s306. https://doi.org/10.37765/ajmc.2020.88534

Callahan, T. J., Bauck, A. E., Bertoch, D., Brown, J., Khare, R., Ryan, P. B., Staab, J., Zozus, M. N., & Kahn, M. G. (2017). A Comparison of Data Quality Assessment Checks in Six Data Sharing Networks. *EGEMS (Wash DC), 5*(1), 8-8. https://doi.org/10.5334/egems.223

CDC. (2016). *Guidance for Measuring Colorectal Cancer (CRC) Screening Rates in Health System Clinics*. https://www.cdc.gov/cancer/crccp/pdf/Guidance_Measuring_CRC_Screening_Rates.pdf

CDC. (2017). *Social Ecological Model*.

CDC. (2018). *Colorectal Cancer Statistics*. Centers for Disease Control and Prevention.

Centers for Disease Control and Prevention. (2020a). *Colorectal (Colon) Cancer*. CDC.

Centers for Disease Control and Prevention. (2020b). *Colorectal Cancer Screening Saves Lives*. Centers for Disease Control and Prevention.

CFIR. (2020). *Consolidated Framework for Implementation Research (CFIR)*

Chait, N., & Glied, S. (2018, Apr 1). Promoting Prevention Under the Affordable Care Act. *Annu Rev Public Health, 39*, 507-524. https://doi.org/10.1146/annurev-publhealth-040617-013534

Chambers, D. A., Amir, E., Saleh, R. R., Rodin, D., Keating, N. L., Osterman, T. J., & Chen, J. L. (2019). The Impact of Big Data Research on Practice, Policy, and Cancer Care. *American Society of Clinical Oncology Educational Book*(39), e167-e175. https://doi.org/10.1200/edbk_238057

Charlson, M. E., Pompei, P., Ales, K. L., & MacKenzie, C. R. (1987). A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis, 40*(5), 373-383. https://doi.org/10.1016/0021-9681(87)90171-8

Cohen, I. G., Amarasingham, R., Shah, A., Xie, B., & Lo, B. (2014, Jul). The legal and ethical concerns that arise from using complex predictive analytics in health care. *Health Affairs, 33*(7), 1139-1147. https://doi.org/10.1377/hlthaff.2014.0048

Cole, A. M., Esplin, A., & Baldwin, L. M. (2015, Dec 3). Adaptation of an Evidence-Based Colorectal Cancer Screening Program Using the Consolidated Framework for Implementation Research. *Preventing chronic disease, 12*, E213. https://doi.org/10.5888/pcd12.150300

Collins, G. S., Mallett, S., Omar, O., & Yu, L. M. (2011, Sep 8). Developing risk prediction models for type 2 diabetes: a systematic review of methodology and reporting. *BMC Med, 9*, 103. https://doi.org/10.1186/1741-7015-9-103

Collins, G. S., Reitsma, J. B., Altman, D. G., & Moons, K. G. M. (2015, 2015/01/06). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. *BMC Medicine, 13*(1), 1. https://doi.org/10.1186/s12916-014-0241-z

Cookson, R., Mirelman, A. J., Griffin, S., Asaria, M., Dawkins, B., Norheim, O. F., Verguet, S., & J Culyer, A. (2017). Using Cost-Effectiveness Analysis to Address Health Equity Concerns. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research, 20*(2), 206-212. https://doi.org/10.1016/j.jval.2016.11.027

Cooper, J. A., Parsons, N., Stinton, C., Mathews, C., Smith, S., Halloran, S. P., Moss, S., & Taylor-Phillips, S. (2018, Jan). Risk-adjusted colorectal cancer screening using the FIT and routine screening data: development of a risk prediction model. *Br J Cancer, 118*(2), 285-293. https://doi.org/10.1038/bjc.2017.375

Coronado, G. D., Johnson, E. S., Leo, M. C., Schneider, J. L., Smith, D., Mummadi, R., Petrik, A. F., Thompson, J. H., & Jimenez, R. (2019, Dec 24). Patient randomized trial of a targeted navigation program to improve rates of follow-up colonoscopy in community health centers. *Contemp Clin Trials, 89*, 105920. https://doi.org/10.1016/j.cct.2019.105920

Coronado, G. D., Petrik, A. F., & Vollmer, W. M. (2018). Effectiveness of a mailed colorectal cancer screening outreach program in community health clinics: The stop crc cluster randomized clinical trial. *JAMA Internal Medicine*. https://doi.org/10.1001/jamainternmed.2018.3629

Damschroder, L. J., Aron, D. C., Keith, R. E., Kirsh, S. R., Alexander, J. A., & Lowery, J. C. (2009). Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. *Implement Sci, 4*, 50. https://doi.org/10.1186/1748-5908-4-50

Damschroder, L. J., Aron, D. C., Keith, R. E., Kirsh, S. R., Alexander, J. A., & Lowery, J. C. (2009, 2009/08/07). Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. *Implementation Science, 4*(1), 50. https://doi.org/10.1186/1748-5908-4-50

Dash, S., Shakyawar, S. K., Sharma, M., & Kaushik, S. (2019, 2019/06/19). Big data in healthcare: management, analysis and future prospects. *Journal of Big Data, 6*(1), 54. https://doi.org/10.1186/s40537-019-0217-0

Davis, M. M., Renfro, S., Pham, R., Hassmiller Lich, K., Shannon, J., Coronado, G. D., & Wheeler, S. B. (2017). Geographic and population-level disparities in colorectal cancer testing: A multilevel analysis of Medicaid and commercial claims data. *Preventive medicine, 101*, 44-52. https://doi.org/10.1016/j.ypmed.2017.05.001

Dawber, T. R., Moore, F. E., & Mann, G. V. (1957). Coronary heart disease in the Framingham study. *American journal of public health and the nation's health, 47*(4 Pt 2), 4-24. https://doi.org/10.2105/ajph.47.4_pt_2.4

de Moor, J. S., Cohen, R. A., Shapiro, J. A., Nadel, M. R., Sabatino, S. A., Robin Yabroff, K., Fedewa, S., Lee, R., Paul Doria-Rose, V., Altice, C., & Klabunde, C. N. (2018). Colorectal cancer screening in the United States: Trends from 2008 to 2015 and variation by health insurance coverage. *Preventive medicine, 112*, 199-206. https://doi.org/10.1016/j.ypmed.2018.05.001

Debray, T. P., Vergouwe, Y., Koffijberg, H., Nieboer, D., Steyerberg, E. W., & Moons, K. G. (2015, Mar). A new framework to enhance the interpretation of external validation studies of clinical prediction models. *J Clin Epidemiol, 68*(3), 279-289. https://doi.org/10.1016/j.jclinepi.2014.06.018

DeGroff, A., Gressard, L., Glover-Kudon, R., Rice, K., Tharpe, F. S., Escoffery, C., Gersten, J., & Butterly, L. (2019, Nov 6). Assessing the implementation of a patient navigation intervention for colonoscopy screening. *BMC Health Serv Res, 19*(1), 803. https://doi.org/10.1186/s12913-019-4601-4

DeGroff, A., Sharma, K., Satsangi, A., Kenney, K., Joseph, D., Ross, K., Leadbetter, S., Helsel, W., Kammerer, W., Firth, R., Rockwell, T., Short, W., Tangka, F., Wong, F., & Richardson, L. (2018). Increasing Colorectal Cancer Screening in Health Care Systems Using Evidence-Based Interventions. *Preventing chronic disease, 15*, E100-E100. https://doi.org/10.5888/pcd15.180029

Denberg, T. D., Melhado, T. V., Coombes, J. M., Beaty, B. L., Berman, K., Byers, T. E., Marcus, A. C., Steiner, J. F., & Ahnen, D. J. (2005, Nov). Predictors of nonadherence to screening colonoscopy. *J Gen Intern Med, 20*(11), 989-995. https://doi.org/10.1111/j.1525-1497.2005.00164.x

Diez-Roux, A. V. (1998, Feb). Bringing context back into epidemiology: variables and fallacies in multilevel analysis. *Am J Public Health., 88*(2), 216-222. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1508189/pdf/amjph00014-0044.pdf

Diez-Roux, A. V. (2000). Multilevel Analysis in Public Health Research. *Annual Review of Public Health, 21*(1), 171-192. https://doi.org/10.1146/annurev.publhealth.21.1.171

Diez Roux, A. V. (2008, Nov). Next steps in understanding the multilevel determinants of health. *J Epidemiol Community Health, 62*(11), 957-959. https://doi.org/10.1136/jech.2007.064311

Dominic, O. G., Chinchilli, V., Wasserman, E., Curry, W. J., Kambic, D. M., Caicedo, C. H., Ramirez, A. G., Ochoa, J. A., & Lengerich, E. J. (2020, Jun). Impact of Social Support on Colorectal Cancer Screening among Adult Hispanics/Latinos: A Randomized Community-based Study in Central Pennsylvania. *Cancer Prev Res (Phila), 13*(6), 531-542. https://doi.org/10.1158/1940-6207.Capr-19-0333

Ehrenstein V, Kharrazi H, Lehmann H, & Taylor, C. (2019). Chapter 4. Obtaining Data From Electronic Health Records. In Gliklich RE, Leavy MB, & D. NA (Eds.), *Tools and Technologies for Registry Interoperability, Registries for Evaluating Patient Outcomes: A User's Guide* (3rd ed.). Agency for Healthcare Research and Quality. https://www.ncbi.nlm.nih.gov/books/NBK551878/ (2019 Oct.)

Eriksson, M., Czene, K., Pawitan, Y., Leifland, K., Darabi, H., & Hall, P. (2017, Mar 14). A clinical model for identifying the short-term risk of breast cancer. *Breast Cancer Res, 19*(1), 29. https://doi.org/10.1186/s13058-017-0820-y

Evans, C. S., Smith, S., Kobayashi, L., & Chang, D. C. (2015, Aug). The Effect of Community Health Center (CHC) Density on Preventable Hospital Admissions in Medicaid and Uninsured Patients. *J Health Care Poor Underserved, 26*(3), 839-851. https://doi.org/10.1353/hpu.2015.0081

Garber, A. M., & Phelps, C. E. (1997, 1997/02/01/). Economic foundations of cost-effectiveness analysis. *Journal of health economics, 16*(1), 1-31. https://doi.org/https://doi.org/10.1016/S0167-6296(96)00506-1

Gething, P. W., Kirui, V. C., Alegana, V. A., Okiro, E. A., Noor, A. M., & Snow, R. W. (2010, Jul 6). Estimating the number of paediatric fevers associated with malaria infection presenting to Africa's public health sector in 2007. *PLoS Med, 7*(7), e1000301. https://doi.org/10.1371/journal.pmed.1000301

Gimeno Garcia, A. Z. (2012). Factors influencing colorectal cancer screening participation. *Gastroenterology research and practice, 2012*.

Gimeno García, A. Z. (2012). Factors influencing colorectal cancer screening participation. *Gastroenterology research and practice, 2012*, 483417-483417. https://doi.org/10.1155/2012/483417

Gimeno Garcia, A. Z., Hernandez Alvarez Buylla, N., Nicolas-Perez, D., & Quintero, E. (2014, 2014/03/05). Public Awareness of Colorectal Cancer Screening: Knowledge, Attitudes, and Interventions for Increasing Screening Uptake. *ISRN Oncology, 2014*, 425787. https://doi.org/10.1155/2014/425787

Ginsburg, G. S., Wu, R. R., & Orlando, L. A. (2019, Aug 17). Family health history: underused for actionable risk assessment. *Lancet, 394*(10198), 596-603. https://doi.org/10.1016/s0140-6736(19)31275-9

Goldstein, B. A., Navar, A. M., Pencina, M. J., & Ioannidis, J. P. (2017, Jan). Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. *J Am Med Inform Assoc, 24*(1), 198-208. https://doi.org/10.1093/jamia/ocw042

Gottlieb, L., Tobey, R., Cantor, J., Hessler, D., & Adler, N. E. (2016, 2016/11/01). Integrating Social And Medical Data To Improve Population Health: Opportunities And Barriers. *Health Affairs, 35*(11), 2116-2123. https://doi.org/10.1377/hlthaff.2016.0723

Greenwald, B. (2006, Mar-Apr). Promoting community awareness of the need for colorectal cancer screening: a pilot study. *Cancer Nurs, 29*(2), 134-141. https://doi.org/10.1097/00002820-200603000-00009

Guessous, I., Dash, C., Lapin, P., Doroshenk, M., Smith, R. A., & Klabunde, C. N. (2010, Jan-Feb). Colorectal cancer screening barriers and facilitators in older persons. *Preventive medicine, 50*(1-2), 3-10. https://doi.org/10.1016/j.ypmed.2009.12.005

Gupta, S., Sussman, D. A., Doubeni, C. A., Anderson, D. S., Day, L., Deshpande, A. R., Elmunzer, B. J., Laiyemo, A. O., Mendez, J., Somsouk, M., Allison, J., Bhuket, T., Geng, Z., Green, B. B., Itzkowitz, S. H., & Martinez, M. E. (2014). Challenges and possible solutions to colorectal cancer screening for the underserved. *Journal of the National Cancer Institute, 106*(4), dju032-dju032. https://doi.org/10.1093/jnci/dju032

Haneuse, S., & Daniels, M. (2016). A General Framework for Considering Selection Bias in EHR-Based Studies: What Data Are Observed and Why? *EGEMS (Wash DC), 4*(1), 1203-1203. https://doi.org/10.13063/2327-9214.1203

Harrell, F. (2015). *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*. Springer International Publishing.

Harrell Jr., F. E., Margolis, P. A., Gove, S., Mason, K. E., Mulholland, E. K., Lehmann, D., Muhe, L., Gatchalian, S., & Eichenwald, H. F. (1998). Development of a clinical prediction model for an ordinal outcome: the World Health Organization Multicentre Study of Clinical Signs and Etiological Agents of Pneumonia, Sepsis and Meningitis in Young Infants. *Stat Med, 17*(8), 909-944. https://doi.org/10.1002/(sici)1097-0258(19980430)17:8<909::Aid-sim753>3.0.Co;2-o

Hatef, E., Rouhizadeh, M., Tia, I., Lasser, E., Hill-Briggs, F., Marsteller, J., & Kharrazi, H. (2019, 2019/08/02). 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, 7*(3), e13802. https://doi.org/10.2196/13802

He, Z., Tang, X., Yang, X., Guo, Y., George, T. J., Charness, N., Quan Hem, K. B., Hogan, W., & Bian, J. (2020). Clinical Trial Generalizability Assessment in the Big Data Era: A Review. *Clin Transl Sci, 13*(4), 675-684. https://doi.org/10.1111/cts.12764

Heinze, G., Wallisch, C., & Dunkler, D. (2018). Variable selection – A review and recommendations for the practicing statistician. *Biometrical Journal, 60*(3), 431-449. https://doi.org/10.1002/bimj.201700067

Henderson, J. T., & Weisman, C. S. (2001). Physician Gender Effects on Preventive Screening and Counseling: An Analysis of Male and Female Patients' Health Care Experiences. *Med Care, 39*(12), 1281-1292. http://www.jstor.org/stable/3768101

Henrikson, N. B., Webber, E. M., Goddard, K. A., Scrol, A., Piper, M., Williams, M. S., Zallen, D. T., Calonge, N., Ganiats, T. G., Janssens, A. C. J. W., Zauber, A., Lansdorp-Vogelaar, I., van Ballegooijen, M., & Whitlock, E. P. (2015). Family history and the natural history of colorectal cancer: systematic review. *Genetics in medicine : official journal of the American College of Medical Genetics, 17*(9), 702-712. https://doi.org/10.1038/gim.2014.188

Horne, H. N., Phelan-Emrick, D. F., Pollack, C. E., Markakis, D., Wenzel, J., Ahmed, S., Garza, M. A., Shapiro, G. R., Bone, L. R., Johnson, L. B., & Ford, J. G. (2015, Feb). Effect of patient navigation on colorectal cancer screening in a community-based randomized controlled trial of urban African American adults. *Cancer Causes Control, 26*(2), 239-246. https://doi.org/10.1007/s10552-014-0505-0

Howell, A., Anderson, A. S., Clarke, R. B., Duffy, S. W., Evans, D. G., Garcia-Closas, M., Gescher, A. J., Key, T. J., Saxton, J. M., & Harvie, M. N. (2014, 2014/09/28). Risk determination and prevention of breast cancer. *Breast Cancer Research, 16*(5), 446. https://doi.org/10.1186/s13058-014-0446-2

Hudson, C. G. (2010, Dec). A predictive model of the development of national mental health systems. *Journal of Mental Health Policy and Economics, 13*(4), 175-187.

Interventions, I. W. G. o. t. E. o. C.-P. (2019). IARC Handbooks of Cancer Prevention. In *Colorectal cancer screening*. International Agency for Research on Cancer

(c) International Agency for Research on Cancer, 2019. For more information contact publications@iarc.fr.

Ioannou, G. N., Chapko, M. K., & Dominitz, J. A. (2003, 2003/09/01/). Predictors of colorectal cancer screening participation in the United States. *Am J Gastroenterol, 98*(9), 2082-2091. https://doi.org/https://doi.org/10.1016/S0002-9270(03)00423-4

Jeffery, A. D., Hewner, S., Pruinelli, L., Lekan, D., Lee, M., Gao, G., Holbrook, L., & Sylvia, M. (2019). Risk prediction and segmentation models used in the United States for assessing risk in whole populations: a critical literature review with implications for nurses' role in population health management. *JAMIA Open, 2*(1), 205-214. https://doi.org/10.1093/jamiaopen/ooy053

Jetelina, K. K., Yudkin, J. S., Miller, S., Berry, E., Lieberman, A., Gupta, S., & Balasubramanian, B. A. (2019, Jun 21). Patient-Reported Barriers to Completing a Diagnostic Colonoscopy Following Abnormal Fecal Immunochemical Test Among Uninsured Patients. *J Gen Intern Med*. https://doi.org/10.1007/s11606-019-05117-0

Johnson, E. S., Thorp, M. L., Platt, R. W., & Smith, D. H. (2008, 2008/10/01/). Predicting the Risk of Dialysis and Transplant Among Patients With CKD: A Retrospective Cohort Study. *American Journal of Kidney Diseases, 52*(4), 653-660. https://doi.org/https://doi.org/10.1053/j.ajkd.2008.04.026

Johnson, E. S., Weinstein, J. R., Thorp, M. L., Platt, R. W., Petrik, A. F., Yang, X., Anderson, S., & Smith, D. H. (2010, Mar). Predicting the risk of hyperkalemia in patients with chronic kidney disease starting lisinopril. *Pharmacoepidemiol Drug Saf, 19*(3), 266-272. https://doi.org/10.1002/pds.1923

Johnson, J. A. R., C.C. (2017). *Health Organizations*. Jones & Bartlett Learning. https://books.google.com/books?id=jKoyDwAAQBAJ

Joseph DA, Redwood D, DeGroff A, & Butler, E. (2016). Use of Evidence-Based Interventions to Address Disparities in Colorectal Cancer Screening. *Morbidity and Mortality Weekly Report*(65), 21-28. http://dx.doi.org/10.15585/mmwr.su6501a5external icon

Kaplan, R. M., Chambers, D. A., & Glasgow, R. E. (2014, Aug). Big data and large sample size: a cautionary note on the potential for bias. *Clin Transl Sci, 7*(4), 342-346. https://doi.org/10.1111/cts.12178

Katz, S. J., Zemencuk, J. K., & Hofer, T. P. (2000). Breast cancer screening in the United States and Canada, 1994: socioeconomic gradients persist. *American Journal of Public Health, 90*(5), 799-803. https://doi.org/10.2105/ajph.90.5.799

Kent, D. M., Steyerberg, E., & van Klaveren, D. (2018, Dec 10). Personalized evidence based medicine: predictive approaches to heterogeneous treatment effects. *BMJ, 363*, k4245. https://doi.org/10.1136/bmj.k4245

Khoury, M. J., Engelgau, M., Chambers, D. A., & Mensah, G. A. (2018). Beyond Public Health Genomics: Can Big Data and Predictive Analytics Deliver Precision Public Health? *Public Health Genomics, 21*(5-6), 244-250. https://doi.org/10.1159/000501465

Kirk, M. A., Kelley, C., Yankey, N., Birken, S. A., Abadie, B., & Damschroder, L. (2016, 2016/05/17). A systematic review of the use of the Consolidated Framework for Implementation Research. *Implementation Science, 11*(1), 72. https://doi.org/10.1186/s13012-016-0437-z

Knighton, A. J., Savitz, L., Belnap, T., Stephenson, B., & VanDerslice, J. (2016). Introduction of an Area Deprivation Index Measuring Patient Socioeconomic Status in an Integrated Health System: Implications for Population Health. *EGEMS (Wash DC), 4*(3), 1238. https://doi.org/10.13063/2327-9214.1238

Kumar, S., Quinn, S. C., Kim, K. H., Musa, D., Hilyard, K. M., & Freimuth, V. S. (2012). The social ecological model as a framework for determinants of 2009 H1N1 influenza vaccine uptake in the United States. *Health education & behavior : the official publication of the Society for Public Health Education, 39*(2), 229-243. https://doi.org/10.1177/1090198111415105

Laiyemo, A. O., Adebogun, A. O., Doubeni, C. A., Ricks-Santi, L., McDonald-Pinkett, S., Young, P. E., Cash, B. D., & Klabunde, C. N. (2014, Oct). Influence of provider discussion and specific recommendation on colorectal cancer screening uptake among U.S. adults. *Preventive medicine, 67*, 1-5. https://doi.org/10.1016/j.ypmed.2014.06.022

Lansdorp-Vogelaar, I., van Ballegooijen, M., Zauber, A. G., Habbema, J. D. F., & Kuipers, E. J. (2009). Effect of Rising Chemotherapy Costs on the Cost Savings of Colorectal Cancer Screening. *JNCI: Journal of the National Cancer Institute, 101*(20), 1412-1422. https://doi.org/10.1093/jnci/djp319

Leininger, L. D. T. (2017). Predictive Modeling for Population Health Management: A Practical Guide. *Mathematica Policy Research*(February 2017).

Liang, S. Y., Phillips, K. A., Nagamine, M., Ladabaum, U., & Haas, J. S. (2006, Oct). Rates and predictors of colorectal cancer screening. *Preventing chronic disease, 3*(4), A117.

Linsky, A., McIntosh, N., Cabral, H., & Kazis, L. E. (2011, Feb). Patient-provider language concordance and colorectal cancer screening. *J Gen Intern Med, 26*(2), 142-147. https://doi.org/10.1007/s11606-010-1512-9

Louie, K. S., Seigneurin, A., Cathcart, P., & Sasieni, P. (2015, May). Do prostate cancer risk models improve the predictive accuracy of PSA screening? A meta-analysis. *Ann Oncol, 26*(5), 848-864. https://doi.org/10.1093/annonc/mdu525

Lumley, T., Kronmal, R. A., Cushman, M., Manolio, T. A., & Goldstein, S. (2002, Feb). A stroke prediction score in the elderly: validation and Web-based application. *J Clin Epidemiol, 55*(2), 129-136. https://doi.org/10.1016/s0895-4356(01)00434-6

Mahmood, S. S., Levy, D., Vasan, R. S., & Wang, T. J. (2014). The Framingham Heart Study and the epidemiology of cardiovascular disease: a historical perspective. *The Lancet, 383*(9921), 999-1008. https://doi.org/10.1016/S0140-6736(13)61752-3

Mahmoudi, E., Kamdar, N., Kim, N., Gonzales, G., Singh, K., & Waljee, A. K. (2020). Use of electronic medical records in development and validation of risk prediction models of hospital readmission: systematic review. *BMJ, 369*, m958. https://doi.org/10.1136/bmj.m958

Matthias, M. S., & Imperiale, T. F. (2020, Feb 26). A risk prediction tool for colorectal cancer screening: a qualitative study of patient and provider facilitators and barriers. *BMC Fam Pract, 21*(1), 43. https://doi.org/10.1186/s12875-020-01113-0

Mazidimoradi, A., Tiznobaik, A., & Salehiniya, H. (2021, Aug 18). Impact of the COVID-19 Pandemic on Colorectal Cancer Screening: a Systematic Review. *J Gastrointest Cancer*, 1-15. https://doi.org/10.1007/s12029-021-00679-x

McGraw, D., Rosati, K., & Evans, B. (2012, Jan). A policy framework for public health uses of electronic health data. *Pharmacoepidemiol Drug Saf, 21 Suppl 1*, 18-22. https://doi.org/10.1002/pds.2319

McMillan, K. (2016, Sep). Politics of change: the discourses that inform organizational change and their capacity to silence. *Nurs Inq, 23*(3), 223-231. https://doi.org/10.1111/nin.12133

McNamara, R. L., Wang, Y., Partovian, C., Montague, J., Mody, P., Eddy, E., Krumholz, H. M., & Bernheim, S. M. (2015, Sep). Development of a Hospital Outcome Measure Intended for Use With Electronic Health Records: 30-Day Risk-standardized Mortality After Acute Myocardial Infarction. *Med Care, 53*(9), 818-826. https://doi.org/10.1097/mlr.0000000000000402

McVeigh, K. H., Newton-Dame, R., Chan, P. Y., Thorpe, L. E., Schreibstein, L., Tatem, K. S., Chernov, C., Lurie-Moroni, E., & Perlman, S. E. (2016). Can Electronic Health Records Be Used for Population Health Surveillance? Validating Population Health Metrics Against Established Survey Data. *EGEMS (Wash DC), 4*(1), 1267-1267. https://doi.org/10.13063/2327-9214.1267

Mijderwijk, H., Stolker, R. J., Duivenvoorden, H. J., Klimek, M., & Steyerberg, E. W. (2016, Sep). Clinical prediction model to identify vulnerable patients in ambulatory surgery: towards optimal medical decision-making. *Can J Anaesth, 63*(9), 1022-1032. https://doi.org/10.1007/s12630-016-0673-3 (Un modele de prediction clinique permettant d'identifier les patients en chirurgie ambulatoire vulnerables: vers une prise de decision medicale optimale.)

Miller, J. A. (2019, October 8, 2019). How Predictive Analytics Is Impacting Patient Care. *HealthTech*, *Data Analytics*.

Moons, K. G., Altman, D. G., Reitsma, J. B., Ioannidis, J. P., Macaskill, P., Steyerberg, E. W., Vickers, A. J., Ransohoff, D. F., & Collins, G. S. (2015, Jan 6). Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med, 162*(1), W1-73. https://doi.org/10.7326/m14-0698

Moons, K. G. M., Wolff, R. F., Riley, R. D., Whiting, P. F., Westwood, M., Collins, G. S., Reitsma, J. B., Kleijnen, J., & Mallett, S. (2019, Jan 1). PROBAST: A Tool to Assess Risk of Bias and Applicability of Prediction Model Studies: Explanation and Elaboration. *Ann Intern Med, 170*(1), W1-w33. https://doi.org/10.7326/m18-1377

Moore, A. R., Buchanan, N. D., Fairley, T. L., & Lee Smith, J. (2015). Public Health Action Model for Cancer Survivorship. *American journal of preventive medicine, 49*(6 Suppl 5), S470-S476. https://doi.org/10.1016/j.amepre.2015.09.001

Morrell, R. (2020). *HAIR Program Provides Creative Platform to Increase Colon Cancer Awareness*. https://www.stopcoloncancernow.com/spread-awareness/news/hair-program-provides-creative-platform-to-increase-colon-cancer-awareness

Murphy, C. C., Vernon, S. W., Haddock, N. M., Anderson, M. L., Chubak, J., & Green, B. B. (2014, 2014/09/01/). Longitudinal predictors of colorectal cancer screening among participants in a randomized controlled trial. *Preventive medicine, 66*, 123-130. https://doi.org/https://doi.org/10.1016/j.ypmed.2014.06.013

Muthukrishnan, M., Arnold, L. D., & James, A. S. (2019). Patients' self-reported barriers to colon cancer screening in federally qualified health center settings. *Preventive medicine reports, 15*, 100896-100896. https://doi.org/10.1016/j.pmedr.2019.100896

National Center for Quality Assurance (NCQA). (2021). *Colorectal Cancer Screening (COL)*. https://www.ncqa.org/hedis/measures/colorectal-cancer-screening/

National Colorectal Cancer Roundtable. (2018). *The National Colorectal Cancer Roundtable*. Retrieved 2018 from http://nccrt.org/

Nelson, R. G., Grams, M. E., Ballew, S. H., Sang, Y., Azizi, F., Chadban, S. J., Chaker, L., Dunning, S. C., Fox, C., Hirakawa, Y., Iseki, K., Ix, J., Jafar, T. H., Köttgen, A., Naimark, D. M. J., Ohkubo, T., Prescott, G. J., Rebholz, C. M., Sabanayagam, C., Sairenchi, T., Schöttker, B., Shibagaki, Y., Tonelli, M., Zhang, L., Gansevoort, R. T., Matsushita, K., Woodward, M., Coresh, J., & Shalev, V. (2019, Nov 8). Development of Risk Prediction Equations for Incident Chronic Kidney Disease. *Jama, 322*(21), 2104-2114. https://doi.org/10.1001/jama.2019.17379

Ni, K., O'Connell, K., Anand, S., Yakoubovitch, S. C., Kwon, S. C., de Latour, R. A., Wallach, A. B., Sherman, S. E., Du, M., & Liang, P. S. (2020, Feb 3). Low colorectal cancer screening uptake and persistent disparities in an underserved urban population. *Cancer Prev Res (Phila)*. https://doi.org/10.1158/1940-6207.Capr-19-0440

[Record #350 is using a reference type undefined in this output style.]

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science, 366*(6464), 447. https://doi.org/10.1126/science.aax2342

OregonHealthAuthority. (2017). *Oregon Health System Transformation: CCO Metrics 2016 Final Report*.

Parikh, N., Pencina, M., Wang, T., Benjamin, E., Lanier, K., Levy, D., D'Agostino, R., Kannel, W., & Vasan, R. (2008, 01/15). A Risk Score for Predicting Near-Term Incidence of Hypertension: The Framingham Heart Study. *Ann Intern Med, 148*, 102-110. https://doi.org/10.7326/0003-4819-148-2-200801150-00005

Parikh, N. I., Pencina, M. J., Wang, T. J., Benjamin, E. J., Lanier, K. J., Levy, D., D'Agostino, R. B., Sr., Kannel, W. B., & Vasan, R. S. (2008, Jan 15). A risk score for predicting near-term incidence of hypertension: the Framingham Heart Study. *Ann Intern Med, 148*(2), 102-110. https://doi.org/10.7326/0003-4819-148-2-200801150-00005

Parikh, R. B., Kakad, M., & Bates, D. W. (2016, Feb 16). Integrating Predictive Analytics Into High-Value Care: The Dawn of Precision Delivery. *Jama, 315*(7), 651-652. https://doi.org/10.1001/jama.2015.19417

Parikh, R. B., Kakad, M., & Bates, D. W. (2016). Integrating Predictive Analytics Into High-Value Care: The Dawn of Precision Delivery. *Jama, 315*(7), 651-652. https://doi.org/10.1001/jama.2015.19417

Parikh, R. B., Obermeyer, Z., & Navathe, A. S. (2019, Feb 22). Regulation of predictive analytics in medicine. *Science, 363*(6429), 810-812. https://doi.org/10.1126/science.aaw0029

Pepe, M. S., Feng, Z., Huang, Y., Longton, G., Prentice, R., Thompson, I. M., & Zheng, Y. (2008, Feb 1). Integrating the predictiveness of a marker with its performance as a classifier. *Am J Epidemiol, 167*(3), 362-368. https://doi.org/10.1093/aje/kwm305

Percac-Lima, S., Grant, R. W., Green, A. R., Ashburner, J. M., Gamba, G., Oo, S., Richter, J. M., & Atlas, S. J. (2009). A culturally tailored navigator program for colorectal cancer screening in a community health center: a randomized, controlled trial. *J Gen Intern Med, 24*(2), 211-217. https://doi.org/10.1007/s11606-008-0864-x

Petrik, A. F., Le, T., Keast, E., Rivelli, J., Bigler, K., Green, B., Vollmer, W. M., & Coronado, G. (2018). Predictors of Colorectal Cancer Screening Prior to Implementation of a Large Pragmatic Trial in Federally Qualified Health Centers. *Journal of community health, 43*(1), 128-136. https://doi.org/10.1007/s10900-017-0395-7

Porter, M. E., Michael Porter, E. O. T., & Teisberg, E. O. (2006). *Redefining Health Care: Creating Value-based Competition on Results*. Harvard Business School Press. https://books.google.com/books?id=Kp5fCkAzzS8C

Pruitt, S. L., Leonard, T., Zhang, S., Schootman, M., Halm, E. A., & Gupta, S. (2014). Physicians, clinics, and neighborhoods: multiple levels of influence on colorectal cancer screening. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology, 23*(7), 1346-1355. https://doi.org/10.1158/1055-9965.EPI-13-1130

Rajkomar, A., Hardt, M., Howell, M. D., Corrado, G., & Chin, M. H. (2018). ENsuring fairness in machine learning to advance health equity. *Annals of Internal Medicine, 169*(12), 866-872. https://doi.org/10.7326/M18-1990

Ray, V. (2019, 2019/02/01). A Theory of Racialized Organizations. *American Sociological Review, 84*(1), 26-53. https://doi.org/10.1177/0003122418822335

Rex, D. K., Boland, C. R., Dominitz, J. A., Giardiello, F. M., Johnson, D. A., Kaltenbach, T., Levin, T. R., Lieberman, D., & Robertson, D. J. (2017, Jul). Colorectal Cancer Screening: Recommendations for Physicians and Patients from the U.S. Multi-Society Task Force on Colorectal Cancer. *Am J Gastroenterol, 112*(7), 1016-1030. https://doi.org/10.1038/ajg.2017.174

Rice, K., Gressard, L., DeGroff, A., Gersten, J., Robie, J., Leadbetter, S., Glover-Kudon, R., & Butterly, L. (2017, Sep 1). Increasing colonoscopy screening in disparate populations: Results from an evaluation of patient navigation in the New Hampshire Colorectal Cancer Screening Program. *Cancer, 123*(17), 3356-3366. https://doi.org/10.1002/cncr.30761

Rosella, L. C., Kornas, K., Yao, Z., Manuel, D. G., Bornbaum, C., Fransoo, R., & Stukel, T. (2017, Nov 17). Predicting High Health Care Resource Utilization in a Single-payer Public Health Care System: Development and Validation of the High Resource User Population Risk Tool (HRUPoRT). *Med Care*. https://doi.org/10.1097/mlr.0000000000000837

Rousseau, D. M. (1985). Issues of level in organizational research: Multi-level and cross-level perspectives. *Research in Organizational Behavior, 7*, 1-37.

Rousseau, D. M. (1985). Issues on level on organizational research: Multi-level and cross-level perspectives. *Research in Organizational Behavior, 7*(1), 1-37.

Rutter, C. M., Knudsen, A. B., Lin, J. S., & Bouskill, K. E. (2021, Jan). Black and White Differences in Colorectal Cancer Screening and Screening Outcomes: A Narrative Review. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology, 30*(1), 3-12. https://doi.org/10.1158/1055-9965.Epi-19-1537

Safaeinili, N., Brown-Johnson, C., Shaw, J. G., Mahoney, M., & Winget, M. (2020). CFIR simplified: Pragmatic application of and adaptations to the Consolidated Framework for Implementation Research (CFIR) for evaluation of a patient-centered care

transformation within a learning health system. *Learn Health Syst, 4*(1), e10201. https://doi.org/10.1002/lrh2.10201

Safayeva, A. K., & Bayramov, N. Y. (2019). The importance of colonoscopy in the treatment of colorectal polyps and colorectal cancer screening. *Ann Ital Chir, 90*, 311-317.

Sarfaty, M., & Wender, R. (2007). How to Increase Colorectal Cancer Screening Rates in Practice. *CA: A Cancer Journal for Clinicians, 57*(6), 354-366. https://doi.org/10.3322/ca.57.6.354

Savitz, S. T., Bailey, S. C., Dusetzina, S. B., Jones, W. S., Trogdon, J. G., & Stearns, S. C. (2020, Jan 28). Treatment selection and medication adherence for stable angina: The role of area-based health literacy. *J Eval Clin Pract*. https://doi.org/10.1111/jep.13341

Savitz, S. T., Savitz, L. A., Fleming, N. S., Shah, N. D., & Go, A. S. (2020, Sep). How much can we trust electronic health record data? *Healthc (Amst), 8*(3), 100444. https://doi.org/10.1016/j.hjdsi.2020.100444

Saya, S., Emery, J. D., Dowty, J. G., McIntosh, J. G., Winship, I. M., & Jenkins, M. A. (2020, Oct). The Impact of a Comprehensive Risk Prediction Model for Colorectal Cancer on a Population Screening Program. *JNCI Cancer Spectr, 4*(5), pkaa062. https://doi.org/10.1093/jncics/pkaa062

Schroeder, E. B., Yang, X., Thorp, M. L., Arnold, B. M., Tabano, D. C., Petrik, A. F., Smith, D. H., Platt, R. W., & Johnson, E. S. (2017, Jan 6). Predicting 5-Year Risk of RRT in Stage 3 or 4 CKD: Development and External Validation. *Clin J Am Soc Nephrol, 12*(1), 87-94. https://doi.org/10.2215/cjn.01290216

Scott, R. W. (2003). *Organizations: Rational, Natural and Open Systems* (5 ed.). Prentice Hall.

Seeff, L. C., Nadel, M. R., Klabunde, C. N., Thompson, T., Shapiro, J. A., Vernon, S. W., & Coates, R. J. (2004). Patterns and predictors of colorectal cancer test use in the adult U.S. population. *Cancer, 100*(10), 2093-2103. https://doi.org/https://doi.org/10.1002/cncr.20276

Sekiguchi, M., Kakugawa, Y., Matsumoto, M., & Matsuda, T. (2018, Oct). A scoring model for predicting advanced colorectal neoplasia in a screened population of asymptomatic Japanese individuals. *J Gastroenterol, 53*(10), 1109-1119. https://doi.org/10.1007/s00535-018-1433-7

Shah, N. D., Steyerberg, E. W., & Kent, D. M. (2018). Big data and predictive analytics: Recalibrating expectations. *Jama*. https://doi.org/10.1001/jama.2018.5602

Sharma, V., Ali, I., van der Veer, S., Martin, G., Ainsworth, J., & Augustine, T. (2021). Adoption of clinical risk prediction tools is limited by a lack of integration with electronic health records. *BMJ health & care informatics, 28*(1), e100253. https://doi.org/10.1136/bmjhci-2020-100253

Singal, A. G., Gupta, S., Skinner, C. S., Ahn, C., Santini, N. O., Agrawal, D., Mayorga, C. A., Murphy, C., Tiro, J. A., McCallister, K., Sanders, J. M., Bishop, W. P., Loewen, A. C., & Halm, E. A. (2017). Effect of Colonoscopy Outreach vs Fecal Immunochemical Test Outreach on Colorectal Cancer Screening Completion: A Randomized Clinical Trial. *Jama, 318*(9), 806-815. https://doi.org/10.1001/jama.2017.11389

Smith, D. H., Johnson, E. S., Thorp, M. L., Petrik, A., Yang, X., & Blough, D. K. (2010, Jan). Outcomes predicted by phosphorous in chronic kidney disease: a retrospective CKD-inception cohort study. *Nephrol Dial Transplant, 25*(1), 166-174. https://doi.org/10.1093/ndt/gfp387

Smith, D. H., Johnson, E. S., Thorp, M. L., Yang, X., Petrik, A., Platt, R. W., & Crispell, K. (2011, Fall). Predicting poor outcomes in heart failure. *Perm J, 15*(4), 4-11.

Sørensen, K., Van den Broucke, S., Fullam, J., Doyle, G., Pelikan, J., Slonska, Z., & Brand, H. (2012, Jan 25). Health literacy and public health: a systematic review and integration of definitions and models. *BMC Public Health, 12*, 80. https://doi.org/10.1186/1471-2458-12-80

Stacy, R., Torrence, W. A., & Mitchell, C. R. (2008, 2008/12/08). Perceptions of Knowledge, Beliefs, and Barriers to Colorectal Cancer Screening. *Journal of Cancer Education, 23*(4), 238-240. https://doi.org/10.1080/08858190802189030

Steyerberg, E. W. (2019). *Clinical Prediction Models* (2nd ed.). Springer.

Steyerberg, E. W., & Harrell, F. E., Jr. (2016, Jan). Prediction models need appropriate internal, internal-external, and external validation. *Journal of Clinical Epidemology, 69*, 245-247. https://doi.org/10.1016/j.jclinepi.2015.04.005

Steyerberg, E. W., Nieboer, D., Debray, T. P. A., & van Houwelingen, H. C. (2019). Assessment of heterogeneity in an individual participant data meta-analysis of prediction models: An

overview and illustration. *Stat Med, 38*(22), 4290-4309. https://doi.org/10.1002/sim.8296

Steyerberg, E. W., Nieboer, D., Debray, T. P. A., & van Houwelingen, H. C. (2019, Sep 30). Assessment of heterogeneity in an individual participant data meta-analysis of prediction models: An overview and illustration. *Stat Med, 38*(22), 4290-4309. https://doi.org/10.1002/sim.8296

Steyerberg, E. W., & Vergouwe, Y. (2014). Towards better clinical prediction models: seven steps for development and an ABCD for validation. *European Heart Journal, 35*(29), 1925-1931. https://doi.org/10.1093/eurheartj/ehu207

Steyerberg, E. W., Vickers, A. J., Cook, N. R., Gerds, T., Gonen, M., Obuchowski, N., Pencina, M. J., & Kattan, M. W. (2010). Assessing the Performance of Prediction Models: A Framework for Traditional and Novel Measures. *Epidemiology, 21*(1). https://journals.lww.com/epidem/Fulltext/2010/01000/Assessing_the_Performance_of_Prediction_Models__A.22.aspx

Stock, R., Goldberg, Bruce. (2017). *Health Reform Policy to Practice: Oregon's Path to a Sustainable Health System, A Study in Innovation*. Elsevier Inc.

Strumpf, E. C. (2011). Racial/Ethnic Disparities in Primary Care: The Role of Physician-Patient Concordance. *Med Care, 49*(5), 496-503. http://www.jstor.org/stable/23053808

Subramanian, S., Tangka, F. K. L., & Hoover, S. (2020). Role of an Implementation Economics Analysis in Providing the Evidence Base for Increasing Colorectal Cancer Screening. *Preventing chronic disease, 17*, E46-E46. https://doi.org/10.5888/pcd17.190407

Sung, J. J. Y., Wong, M. C. S., Lam, T. Y. T., Tsoi, K. K. F., Chan, V. C. W., Cheung, W., & Ching, J. Y. L. (2018, Jan). A modified colorectal screening score for prediction of advanced neoplasia: A prospective study of 5744 subjects. *J Gastroenterol Hepatol, 33*(1), 187-194. https://doi.org/10.1111/jgh.13835

System, N. A. o. E. U. a. I. o. M. U. C. o. E. a. t. H. C. (2005). Chapter 2: A Framework for a Systems Approach to Health Care Delivery. In R. PP, C. WD, & G. JH (Eds.), *Building a Better Delivery System: A New Engineering/Health Care Partnership*. National Academies Press (US) https://www.ncbi.nlm.nih.gov/books/NBK22878/?report=classic

Tammemägi, M. C. (2015, Mar). Application of risk prediction models to lung cancer screening: a review. *J Thorac Imaging, 30*(2), 88-100. https://doi.org/10.1097/rti.0000000000000142

Taplin, S. H., Anhang Price, R., Edwards, H. M., Foster, M. K., Breslau, E. S., Chollette, V., Prabhu Das, I., Clauser, S. B., Fennell, M. L., & Zapka, J. (2012). Introduction: Understanding and influencing multilevel factors across the cancer care continuum. *J Natl Cancer Inst Monogr, 2012*(44), 2-10. https://doi.org/10.1093/jncimonographs/lgs008

Terry, M. B., Liao, Y., Whittemore, A. S., Leoce, N., Buchsbaum, R., Zeinomar, N., Dite, G. S., Chung, W. K., Knight, J. A., Southey, M. C., Milne, R. L., Goldgar, D., Giles, G. G., McLachlan, S. A., Friedlander, M. L., Weideman, P. C., Glendon, G., Nesci, S., Andrulis, I. L., John, E. M., Phillips, K. A., Daly, M. B., Buys, S. S., Hopper, J. L., & MacInnis, R. J. (2019, Apr). 10-year performance of four models of breast cancer risk: a validation study. *Lancet Oncol, 20*(4), 504-517. https://doi.org/10.1016/s1470-2045(18)30902-1

Thompson, J. H., Schneider, J. L., Rivelli, J. S., Petrik, A. F., Vollmer, W. M., Fuoco, M. J., & Coronado, G. D. (2019, Jan-Dec). A Survey of Provider Attitudes, Beliefs, and Perceived Barriers Regarding a Centralized Direct-Mail Colorectal Cancer Screening Approach at Community Health Centers. *Journal of primary care & community health, 10*, 2150132719890950-2150132719890950. https://doi.org/10.1177/2150132719890950

Tinmouth, J., Dong, S., Stogios, C., Rabeneck, L., Rey, M., & Dubé, C. (2021, Mar). Estimating the Backlog of Colonoscopy due to Coronavirus Disease 2019 and Comparing Strategies to Recover in Ontario, Canada. *Gastroenterology, 160*(4), 1400-1402.e1401. https://doi.org/10.1053/j.gastro.2020.11.048

Toumazis, I., Bastani, M., Han, S., & Plevritis, S. (2020, Sep). Risk-Based lung cancer screening: A systematic review. *Lung Cancer, 147*, 154-186. https://doi.org/10.1016/j.lungcan.2020.07.007

Townsend, J. S., Richardson, L. C., Steele, C. B., & White, D. E. (2009). Evidence-Based Interventions and Screening Recommendations for Colorectal Cancer in Comprehensive Cancer Control Plans: A Content Analysis [Journal Article]. *Prev Chronic Dis. 6(4). 6*(4). https://stacks.cdc.gov/view/cdc/4164

University of North Carolina at Chapel Hill. (2014). *National Health Literacy Mapping to Inform Health Care Policy*. Retrieved February 22, 2021 from http://healthliteracymap.unc.edu/#

US Preventive Services Task Force. (2021). *Colorectal Cancer: Screening* https://www.uspreventiveservicestaskforce.org/uspstf/recommendation/colorectal-cancer-screening#fullrecommendationstart

US Preventive Services Task Force, Bibbins-Domingo, K., Grossman, D. C., Curry, S. J., Davidson, K. W., Epling, J. W., Jr., García, F. A. R., Gillman, M. W., Harper, D. M., Kemper, A. R., Krist, A. H., Kurth, A. E., Landefeld, C. S., Mangione, C. M., Owens, D. K., Phillips, W. R., Phipps, M. G., Pignone, M. P., & Siu, A. L. (2016, Jun 21). Screening for Colorectal Cancer: US Preventive Services Task Force Recommendation Statement. *Jama, 315*(23), 2564-2575. https://doi.org/10.1001/jama.2016.5989

Vahabi, M., Lofters, A. K., Kopp, A., & Glazier, R. H. (2021, Jan 4). Correlates of non-adherence to breast, cervical, and colorectal cancer screening among screen-eligible women: a population-based cohort study in Ontario, Canada. *Cancer Causes Control*. https://doi.org/10.1007/s10552-020-01369-y

Viramontes, O., Bastani, R., Yang, L., Glenn, B. A., Herrmann, A. K., & May, F. P. (2020, Sep). Colorectal cancer screening among Hispanics in the United States: Disparities, modalities, predictors, and regional variation. *Preventive medicine, 138*, 106146. https://doi.org/10.1016/j.ypmed.2020.106146

Vivas-Consuelo, D., Uso-Talamantes, R., Trillo-Mata, J. L., Caballer-Tarazona, M., Barrachina-Martinez, I., & Buigues-Pastor, L. (2014, Jun). Predictability of pharmaceutical spending in primary health services using Clinical Risk Groups. *Health Policy, 116*(2-3), 188-195. https://doi.org/10.1016/j.healthpol.2014.01.012

Watson, K. S., Hulbert, A., Henderson, V., Chukwudozie, I. B., Aponte-Soto, L., Lerner, L., Martinez, E., Kim, S., & Winn, R. A. (2019, 2019-March-12). Lung Cancer Screening and Epigenetics in African Americans: The Role of the Socioecological Framework [Hypothesis and Theory]. *Frontiers in Oncology, 9*(87). https://doi.org/10.3389/fonc.2019.00087

Weiner, B. J., Lewis, M. A., Clauser, S. B., & Stitzenberg, K. B. (2012). In Search of Synergy: Strategies for Combining Interventions at Multiple Levels. *JNCI Monographs, 2012*(44), 34-41. https://doi.org/10.1093/jncimonographs/lgs001

Weiner, M. G., Sheikh, W., & Lehmann, H. P. (2018). Interactive Cost-benefit Analysis: Providing Real-World Financial Context to Predictive Analytics. *AMIA Annu Symp Proc, 2018*, 1076-1083. https://pubmed.ncbi.nlm.nih.gov/30815149

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6371360/

Weiss, J. M., Smith, M. A., Pickhardt, P. J., Kraft, S. A., Flood, G. E., Kim, D. H., Strutz, E., & Pfau, P. R. (2013, Jul). Predictors of colorectal cancer screening variation among primary-care providers and clinics. *Am J Gastroenterol, 108*(7), 1159-1167. https://doi.org/10.1038/ajg.2013.127

Weiss, J. W., Platt, R. W., Thorp, M. L., Yang, X., Smith, D. H., Petrik, A., Eckstrom, E., Morris, C., O'Hare, A. M., & Johnson, E. S. (2015, Mar). Predicting mortality in older adults with kidney disease: a pragmatic prediction model. *J Am Geriatr Soc, 63*(3), 508-515. https://doi.org/10.1111/jgs.13257

Williamson, O. E. (1981, 1981/11/01). The Economics of Organization: The Transaction Cost Approach. *American Journal of Sociology, 87*(3), 548-577. https://doi.org/10.1086/227496

Williamson, O. E. (2010). Transaction Cost Economics: The Natural Progression. *American Economic Review, 100*(3), 673-690. https://doi.org/10.1257/aer.100.3.673

Wilson, P. W., Meigs, J. B., Sullivan, L., Fox, C. S., Nathan, D. M., & D'Agostino, R. B., Sr. (2007, May 28). Prediction of incident diabetes mellitus in middle-aged adults: the Framingham Offspring Study. *Arch Intern Med, 167*(10), 1068-1074. https://doi.org/10.1001/archinte.167.10.1068

Winters-Miner, L. A. (2014). Seven ways predictive analytics can improve healthcare. *Elsevier Connect*(October 6, 2014). https://www.elsevier.com/connect/seven-ways-predictive-analytics-can-improve-healthcare

Wolff, R. F., Moons, K. G. M., Riley, R. D., Whiting, P. F., Westwood, M., Collins, G. S., Reitsma, J. B., Kleijnen, J., & Mallett, S. (2019, Jan 1). PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Ann Intern Med, 170*(1), 51-58. https://doi.org/10.7326/m18-1376

Wong, M. C., Ching, J. Y., Chan, V. C., Tang, R. S., Luk, A. K., Lam, T. Y., Wong, S. S., Ng, S. C., Ng, S. S., Wu, J. C., Chan, F. K., & Sung, J. J. (2016, Feb 8). Validation of a risk prediction score for proximal neoplasia in colorectal cancer screening: a prospective colonoscopy study. *Sci Rep, 6*, 20396. https://doi.org/10.1038/srep20396

Woodhouse, G., Yang, M., Goldstein, H., & Rasbash, J. (1996). Adjusting for Measurement Error in Multilevel Analysis. *Journal of the Royal Statistical Society. Series A (Statistics in Society), 159*(2), 201-212. https://doi.org/10.2307/2983168

Woolf, S. (2010). Missed Prevention Opportunities: The Price Paid for Not Preventing Diseases. In P. L. Yong, R. S. Saunders, & L. Olsen (Eds.), *The Healthcare Imperative*. National Academies Press.

Yeoh, K. G., Ho, K. Y., Chiu, H. M., Zhu, F., Ching, J. Y., Wu, D. C., Matsuda, T., Byeon, J. S., Lee, S. K., Goh, K. L., Sollano, J., Rerknimitr, R., Leong, R., Tsoi, K., Lin, J. T., & Sung, J. J. (2011, Sep). The Asia-Pacific Colorectal Screening score: a validated tool that stratifies risk for colorectal advanced neoplasia in asymptomatic Asian subjects. *Gut, 60*(9), 1236-1241. https://doi.org/10.1136/gut.2010.221168

Young, W. F., McGloin, J., Zittleman, L., West, D. R., & Westfall, J. M. (2007). Predictors of colorectal screening in rural Colorado: testing to prevent colon cancer in the high plains research network. *The Journal of rural health, 23*(3), 238-245.

Yu, C., Skootsky, S., Grossman, M., Garner, O. B., Betlachin, A., Esrailian, E., Hommes, D. W., & May, F. P. (2018, Sep 4). A Multi-Level Fit-Based Quality Improvement Initiative to Improve Colorectal Cancer Screening in a Managed Care Population. *Clin Transl Gastroenterol, 9*(8), 177. https://doi.org/10.1038/s41424-018-0046-z

Zapka, J., Taplin, S. H., Ganz, P., Grunfeld, E., & Sterba, K. (2012, May). Multilevel factors affecting quality: examples from the cancer care continuum. *J Natl Cancer Inst Monogr, 2012*(44), 11-19. https://doi.org/10.1093/jncimonographs/lgs005

Zhao, Y., & Wong, N. D. (2018, Oct 11). The Evolving Cardiovascular Disease Risk Scores for Persons with Diabetes Mellitus. *Curr Cardiol Rep, 20*(12), 126. https://doi.org/10.1007/s11886-018-1069-5

Zimmerman, R. K., Nowalk, M. P., Tabbarah, M., & Grufferman, S. (2006, 2006/09/13). Predictors of colorectal cancer screening in diverse primary care practices. *BMC Health Serv Res, 6*(1), 116. https://doi.org/10.1186/1472-6963-6-116

Appendix A: Initial Email for Interview Participant Recruitment

Predictive Analytics Interview Recruitment Email: KPNW staff

To:

Subject line: Interview request for study on predictive analytics and colon cancer screening

Dear [CLINICIANS/STAFF],

For the requirements of my PhD program and dissertation, I have been exploring the use of multilevel external data and predictive analytics in colorectal cancer screening. The study, ***The Use of Multilevel External Data in the Development of a Multivariable Prediction Model for Population Health Management in an Integrated Health System to Predict Colorectal Cancer Screening***, tests the use of external data and predictive analytics as an approach to determine patients who are unlikely to screen for colorectal cancer.

We are inviting you to participate in a 30-45-minute phone interview to help us understand more about the use of data and analytics in preventive care. Your insight is important for helping us understand barriers of and facilitators to using external data and predictive analytics. If you agree to participate, we will work with you to find a convenient time to conduct the interview. Please let us know if you are interested in participating by replying to this email.

The interviews are voluntary, with no right or wrong answers to the questions being asked. Your comments will be kept in confidence. Interviews are audio recorded and content coded.

Thank you for your time and consideration. We look forward to the opportunity to speak with you about colorectal cancer screening and predictive analytics.

Warm regards,

Amanda Petrik,

PhD Student in Health Systems and Policy

Oregon Health Sciences University

Appendix B: Participant Consent Form

Kaiser Permanente Center for Health Research

OHSU/PSU School of Public Health Student Dissertation

Individual Interview Consent Form

Study Title: *The Use of Multilevel External Data in the Development of a Multivariable Prediction Model for Population Health Management in an Integrated Health System to Predict Colorectal Cancer Screening*

Principal Investigator: Amanda Petrik, Ph.D. (c)
Health Systems and Policy, School of Public Health
Oregon Health Sciences University

At the Center for Health Research (CHR), Amanda Petrik is conducting a study as part of the requirements for her PhD. The study seeks to understand if using multilevel data can improve a risk prediction model, in predicting patients' who are unlikely to complete their colorectal cancer screening. This may help target interventions to the right patients at the right time. It further seeks to understand how the inclusion of multilevel data in a risk prediction model improves the usefulness to health system decision makers for managing population health.

What will you be asked to do?

Participation in this interview is completely voluntary. There is no penalty if you decide not to take part. If you do decide to take part in this project, we will ask you to do two things:

1. **Participate in a telephone interview.**
   We are asking you to participate in a 30 to 45-minute telephone interview. During the discussion, you'll be asked questions that will help us understand how you might use multilevel data or a risk prediction model. We want to know more about if this model will help you in managing population health.

2. **Let us audio record the discussion.**

We can't write fast enough to record everything said in the interview, so we'd like your permission to audio record the discussion. The audio recording will be transcribed onto

paper, but the typed copy will not include your name or any other information that might let people know who you are. Only the study staff will see the typed transcripts. Quotes from the discussion may be used in presentations of study results, but they will not identify the speaker. The audio recording will be destroyed at the end of the study. If you do not want the discussion recorded, we will not be able to allow you to participate in the study.

Are there any risks?

There is a small risk of loss of confidentiality. You don't have to answer any questions you don't want to, and if you don't want to go on, you can stop at any time. All information gathered in this study will be kept confidential including your participation in this study.

Are there any benefits?

While you will not personally benefit from being in the interview, your participation will help researchers improve the new program so that it is helpful to others.

Do I have to be in the research study?

No, you do not have to join this research study and you can withdraw at any time. Your participation is completely voluntary and your decision to participate or not will not affect your medical care or health benefits in any way.

Compensation

You will not be paid for your participation.

Confidentiality

Kaiser Permanente is committed to protecting your confidentiality. State and federal laws also require Kaiser Permanente to maintain the privacy and security of your information in this study. If you agree to be in this study, researchers at Kaiser Permanente will be audio recording and transcribing the interview. Every reasonable effort will be made to keep your records confidential. To protect your confidentiality, we will ensure that any identifying information (example: *your name*) is removed from the transcripts by a CHR-approved transcriptionist who has signed a privacy agreement. Summary information from the interviews will be shared with members of the research team but will not include any identifiable information about you. All study-related materials will be stored in secured computers or locked files. The original audio recordings will be destroyed at the end of the study.

People and organizations involved in overseeing or auditing this study may also see or receive your information. These may include, for example, the Institutional Review Board (ethics review committee), research collaborators, and the Office of Human Research Protection (federal agency that oversees research).
We may publish the results of this research. However, we will not publish your name or any other identifying information.

Data from this study will not be used for future research.

Questions

If you have any questions about this study, please call Amanda Petrik, Ph.D(c). at the Center for Health Research. Amanda is available Monday through Friday and can be reached by calling 503-798-7271 or by email at amanda.f.petrik@kpchr.org.

If you have questions about your rights as a research subject, or to contact the Institutional Review Board (IRB), call Kaiser Permanente Northwest's Research Compliance Manager at 503-335-6725. The Institutional Review Board (IRB) is a committee of scientific, nonscientific, and community members who review research to protect the rights and welfare of participants.

If I agree, what does it mean?

This consent form contains important information. Agreeing to the interview means that:

You have read this form.

You are willing to take part in the study by:

1. Participating in a 30 to 45-minute interview over the telephone, and
2. Letting us audio record the discussion.

You do not have to take part in this study. Even if you agree to participate, you can change your mind at any time.
Your decision about taking part in, declining to take part in, or ending your participation will not affect you or your health care.

Do you have any questions? If we have your permission, we will now begin the interview.

Appendix C: Interview Guide

KPNW Staff Interview Questions

Introduction:

Hello, my name is Amanda Petrik and, while I work at the KPNW Center for Health Research, I'm reaching out to you as part of my dissertation research at the OHSU-PSU School of Public Health. My research focuses on how to support clinical care in using analytic tools to more efficiently and effectively identify patients in need of colorectal cancer screening. I want you to know that your participation is voluntary, and you may discontinue at any time. I am recording our conversation because it is part of the data I'll be using; all data will be aggregated and analyzed together from 8 interviews. I will not attribute any direct quotes from you personally without your permission, and you have provided verbal consent and understand the risks and benefits in participating in this interview.

Do you have any questions before I get started?

Well then, I will begin (start recording).

Thank you for being willing to speak with me today. I just want to confirm that you are a XXX in the department of X and your current job title is X. Is that correct?

(Front line caregiver, Quality improvement * Northwest Permanente, Kaiser Foundation Health Plan/Hospitals)

How many years have you worked as a X? and at KP?

INTERVENTIONS:

1. How are decisions made in your department regarding strategies or information used to choose practices or care processes that improve the quality of care?

[probe who is involved in this process]

2. What if anything does your department currently do to increase the rates of colorectal cancer screening?

a. Do other teams give any input on CRC screening practices or care processes used in your department?
b. Do you ever have input on how to reach patients to engage them in CRC screening?
c. Can you give me an example of a current CRC screening practice or care process your department is currently using?
1. Now thinking about that example, does your department have adequate resources to support CRC screening?
[probe for limitations in staffing];

[probe if there is enough staff to act on what is found in the model]

[probe could PA help focus limited resources]

DATA:

3. Do you use data or IT-driven tools to support increased CRC screening for your patients?

a. If YES,
Within your department, what types of information or data are typically used to design CRC screening improvement programs?

[probe for phone calls, letters, navigation]

b. Do you ever use information or data outside of EHR?
1. To the best of your knowledge, are other departments in KP using data outside of the EHR?

4. What if the needed data or information came from a separate data source that you would have to connect with? Would you be able to use this as part of the screening workflow?

a. How will you know that the source is reliable? (evidence-based)
b. What are potential barriers and facilitators to utilizing other sources of information coming from such an external source as part of your usual workflow?

ANALYTICS:

Predictive analytics is a type of analysis that uses data to predict which patients are likely or unlikely to receive care or to predict which patients at greatest risk for a disease or condition.

5. Are you aware of any tools your department currently uses that rely on predictive analytics?

a. Does your department have analytic support or is that a centralized function?

6. How do you personally feel about using predictive analytics?
a. Do you see predictive analytics as a useful tool in support of your screening activities?
b. Do you feel that predictive analytics captures variations in screening uptake among different subpopulations that you may serve?
[probe: how so?]

7. Can you identify and describe barriers or facilitators to using predictive analytics that you have faced or anticipate facing?

8. Are you aware of other departments or organizations using predictive analytics to improve patient care?

Description of the model and data:

As part of my research, I have developed a predictive analytic model to help identify patients to target for CRC screening. The tool would optimally be embedded in the EHR and a flag would appear for scheduled patients or a worklist would be generated for patient outreach. (insert details of the model after created)

What do you think about this model?

_____
_____
_____
_____

MODEL:

9. Does the model provide information you would find useful in the development or choice of screening practices?
[probe: perceptions of the multilevel data]

10. How would you use risk prediction models like the ones I have presented?
[probe: could a model like this could change your choice of intervention? How?]

[probe: does the model provide useful information to focus your screening efforts?]

ORGANIZATIONAL CLIMATE:

11. Do you see the use of predictive analytics for CRC as part of a care process as cost-efficient?

[probe: could PA make the screening outreach more efficient?

Will PA save your department money?

Will using PA be supported by the rest of KPNW (rewarded)?]

[probe: is this different in the long term (post-COVID-19) than the short term (in COVID-19)?]

12. Would the use of a CRC screening tool based on predictive analytics make your job easier?

[probe: is this different in the long term (post-COVID-19) than the short term (in COVID-19)?]

13. In what ways does KPNW create potential barriers or facilitators to the use of predictive analytics in your line of work?

[probe: is this different in the long term (post-COVID-19) than the short term (in COVID-19)?]

POLICIES

14. Are you able to ask for the types of data you need or want to do your job?

[probe: do you have the ability to get/ask for analytics or analytic support?

do you have decision making power?]

15. Does your department/KPNW have *policies about access and use of predictive analytics*?

[probe: KPNW policies?]

16. Does your department/KPNW have *policies about access to data*?

17. Does your department/KPNW have policies about using data to understand trends in screening among subpopulations?

a. Policies about tailoring outreach to subpopulations?

18. Is there anything I haven't asked you about that you think would be important for me to know about your work, policies, or practices used to provide preventive care screening or anything else that comes to mind?